



**DETRA NOTE**

2019-1



# DES TRIANGLES AGRÉGÉS AUX MÉTHODES INDIVIDUELLES

*Revue de méthodes de réservation en assurances  
dommages*

Par Gireg Willame, Michel Denuit et Julien Trufin

**Disclaimer**

Le contenu des Detra Notes est à usage exclusivement pédagogique. Les applications commerciales sont à ce point spécifiques qu'on ne pourrait leur appliquer une solution qui n'aurait pas été conçue pour le cas d'espèce. Detralytics n'assume donc aucune responsabilité en cas d'utilisation commerciale du contenu du présent document. Bien évidemment, toute l'équipe se tient à votre entière disposition si les techniques présentées dans cette note devaient retenir votre intérêt.

## Résumé

En assurances de dommages (IARD), lorsqu'un sinistre survient, une compagnie d'assurance doit estimer et réserver le coût total de ce dernier. Cette évaluation doit également être répétée tant que ce sinistre n'est pas liquidé. Ainsi, plusieurs techniques pour estimer ces coûts existent. Depuis plusieurs années, les actuaires sont habitués à évaluer les réserves à l'aide de méthodes collectives, n'incluant aucune information individuelle. Cependant, avec la disponibilité des données individuelles et l'accroissement de la puissance informatique, les actuaires sont de plus en plus tentés d'incorporer ces informations dans l'évaluation des réserves en vue d'améliorer leurs prédictions. Néanmoins, le coût algorithmique de telles approches individuelles reste relativement important. C'est pourquoi, depuis peu, des techniques mêlant des modélisations collectives en utilisant l'information au niveau individuel, ont également vu le jour. Dans cette note, nous nous contentons de présenter différentes approches et leurs aspects théoriques. Le but est d'avoir une idée suffisamment précise de chaque méthode, permettant ainsi de déterminer celle qui est la mieux adaptée aux besoins et aux données à disposition. Pour faciliter les comparaisons, un schéma explicatif commun est utilisé.

# Table des matières

<b>Table des matières</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Notations</b>	<b>3</b>
2.1 Notations pour les modèles agrégés . . . . .	4
2.2 Notations pour les modèles individuels . . . . .	4
<b>3 Description des méthodes existantes</b>	<b>7</b>
3.1 Méthodes agrégées . . . . .	7
3.1.1 Chain-Ladder (CL) . . . . .	7
3.1.2 Paid-Incurred Chain (PIC) . . . . .	9
3.1.3 Approche de Tweedie . . . . .	11
3.2 Méthodes individuelles pures . . . . .	15
3.2.1 ILR . . . . .	15
3.2.2 Approche de Larsen . . . . .	21
3.3 Méthodes collectives : CLR . . . . .	27
<b>4 Discussion finale</b>	<b>31</b>
<b>Bibliographie</b>	<b>37</b>
<b>5 Un mot sur la série et les auteurs...</b>	<b>39</b>
5.1 Les Detra Notes . . . . .	39
5.2 Biographies des auteurs . . . . .	39

# Chapitre 1

## Introduction

En assurances de dommages (IARD), les sinistres nécessitent parfois plusieurs années avant d'être liquidés. Durant ce temps, les assureurs doivent constituer des réserves représentant l'estimation de leurs obligations pour les sinistres qui sont arrivés à ou avant la date d'évaluation.

Afin d'évaluer ces réserves, beaucoup de techniques ont été proposées dans la littérature. Les premières sont les plus connues et consistent à uniquement utiliser les données agrégées, représentées dans les triangles de développement. Dans cette catégorie, nous pouvons trouver des méthodes d'évaluation des réserves telles que :

- Le modèle "Chain-Ladder" (CL) introduit par Mack [12];
- Le modèle "Paid-Incurred Chain" (PIC) présenté par Merz et Wütrich dans [14, 15];
- Le modèle de Tweedie proposé par Wütrich [22].

Grâce à la puissance informatique grandissante et l'envie d'utiliser le plus de données possibles, de nouvelles méthodes ont vu le jour. Dans un premier temps, des méthodes *purement individuelles* sont apparues, visant à développer l'évolution de chacun des sinistres, depuis leur occurrence jusqu'à leur liquidation. Le défaut majeur de ces méthodes est inhérent à leur grand avantage ; la simulation de tout l'historique des paiements pour chacun des sinistres provoque un temps de calcul relativement élevé.

Le but étant d'être proche du monde assurantiel, nous décidons de ne pas traiter les modèles de développement individuel des sinistres en temps continu mais nous renvoyons les lecteurs intéressés par un tel sujet vers Antonio et Plat [2], Haastrup et Arjas [8] et les références qui y sont citées. Nous nous intéressons aux trois méthodes de réservation individuelle pure suivantes :

- Modèle "Individual Loss Reserving" qui nécessite principalement le triangle des paiements, tel qu'introduit par Pigeon *et al.* [18]. Dans la suite du texte, ce modèle sera noté ILR. Une extension, également présentée par Pigeon *et al.* [19], mêlant les triangles de paiements et des charges sera également abordée ;
- Modèle de "Larsen", portant le nom de son créateur [11].

Récemment, des méthodes de réservation utilisant les données individuelles dans une optique de modélisation collective ont fait leur apparition. Contrairement aux modèles individuels purs, ces approches ne permettent pas de suivre l'historique des paiements pour un sinistre particulier mais

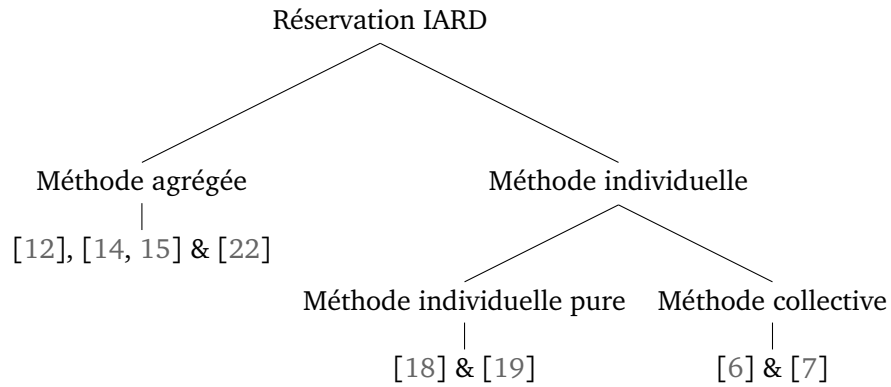


FIGURE 1.1: Structure de réservation IARD et méthodes considérées.

s'intéressent plutôt à la globalité des cash-flows. Cependant, la réduction de la complexité algorithmique constitue un atout majeur. Dans ce contexte, nous nous intéressons principalement au modèle introduit par Denuit et Trufin [6], que l'on désignera par CLR. Une extension de ce dernier, présentée par ces même auteurs [7], sera brièvement abordée.

Bien évidemment, la liste des modèles considérés n'est pas exhaustive mais nous considérons qu'elle fournit une base solide de comparaison. Le schéma de réservation décrit ainsi que les méthodes traitées sont visibles à la Figure 1.1 ci-dessus.

La note s'articule comme suit. Dans un premier temps, nous introduisons les notations nécessaires à cette dernière. Nous comparons ensuite les différents modèles énoncés, présents dans chacune des catégories introduites. Nous tâcherons de garder un schéma de comparaison identique afin de faciliter la lecture. Nous clôturons la note par une discussion ainsi qu'un tableau synthétisant les différentes méthodes proposées.

## Chapitre 2

### Notations

Nous supposons avoir à disposition  $n$  années d'observations. Les *années d'accident* varient de  $i = 1, \dots, n$  alors que les *années de développement* vont de  $j = 1, \dots, n$ . Ces données remplissent un triangle : Pour chaque année d'accident  $i$ , les années de développement varient de  $j = 1$  (c'est-à-dire l'année d'accident elle-même) jusqu'à  $n - i + 1$ , correspondant à la dernière *année calendrier*  $n$  (située sur la diagonale du triangle, donc). Notons  $\mathcal{D}_n$  l'ensemble des réalisations observées jusqu'à l'année calendrier  $n$ . Nous observons à la Figure 2.1 un graphique représentant la situation. Le but est donc de remplir la partie inférieure du triangle, ce pourquoi plusieurs méthodes existent.

Notons également  $\omega$  le *temps maximal* nécessaire à la liquidation d'un sinistre. Cela signifie que, pour un sinistre survenu l'année d'accident  $i$ , il sera clôturé au plus tard l'année calendrier  $i + \omega - 1$ . Remarquons qu'en cas de sinistre à développement long, il se peut que  $\omega > n$ . Cependant, la gestion de ce cas particulier dépend de la méthode de réservation utilisée. Dès lors, afin de ne pas surcharger les explications, nous supposons pour la suite de cette note que  $\omega = n$ . Par conséquent, la première année d'accident ( $i = 1$ ) est entièrement observée. L'éventualité où  $\omega > n$  sera néanmoins brièvement évoquée dans la discussion finale (voir Chapitre 4).

Nous attirons également l'attention du lecteur sur l'effet *calendrier*, qui consiste à comparer une année calendrier à une autre. Cet effet permet notamment d'expliquer l'inflation ou des changements de législation.

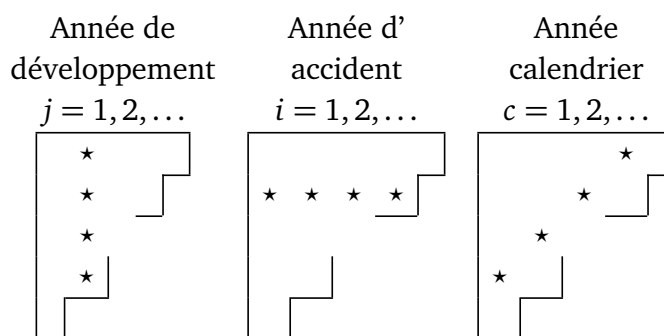


TABLE 2.1: Les trois dimensions d'un triangle de développement.

Commençons par les notations qui nous seront utiles au niveau agrégé.

## 2.1 Notations pour les modèles agrégés

Dans ces modèles particuliers, nous nous intéressons principalement aux paiements incrémentaux (comprenez annuels) ou cumulés effectués dans chacune des cellules  $(i, j)$  du triangle et ce pour la totalité des sinistres survenus durant l'année considérée. Par exemple, supposons que nous souhaitons connaître le paiement annuel effectué pour l'ensemble des sinistres survenus l'année d'accident  $i$  lors du  $j^{\text{ème}}$  développement. Clairement, si  $j \leq n - i + 1$ , cette valeur est observée. Il s'agit donc de la déterminer lorsque  $j > n - i + 1$ . Dans certains cas, nous pouvons avoir besoin du triangle constitué des *charges*, comprenez l'addition des paiements cumulés avec une estimation de ce qu'il reste à payer, au niveau global.

Par conséquent, nous notons :

- $P_{i,j}$  pour désigner le paiement annuel effectué pour la totalité des sinistres survenus l'année d'accident  $i$  à l'année de développement  $j$  ;
- $C_{i,j}$  pour désigner le paiement cumulé effectué pour la totalité des sinistres survenus l'année d'accident  $i$  à l'année de développement  $j$ . Mathématiquement, cela revient à écrire

$$C_{i,j} = \sum_{k=1}^j P_{i,k};$$

- $I_{i,j}$  pour désigner les charges pour la totalité des sinistres survenus l'année d'accident  $i$  à l'année de développement  $j$ . Il s'agit donc de l'addition de  $C_{i,j}$  avec une estimation de ce qu'il reste à payer pour les sinistres survenus l'année d'accident  $i$ .

## 2.2 Notations pour les modèles individuels

Durant la vie d'un sinistre, plusieurs événements marquants sont à relever. Premièrement, il s'agit de l'occurrence de ce dernier. Dès lors, notons  $K_i$  pour désigner le nombre de sinistres survenus l'année accident  $i$ . Soit  $k$  représentant le  $k^{\text{ème}}$  sinistre survenu l'année d'accident  $i$ . Une fois que le sinistre s'est produit, il existe un certain délai entre cet instant et la déclaration à l'assureur. Ce *délai de déclaration* pour le  $k^{\text{ème}}$  sinistre est noté  $T_{i,k}$ . Une fois déclaré, plusieurs paiements peuvent avoir lieu ainsi que de nombreux ajustements des charges. En fonction de la méthode, les paiements considérés peuvent être annuels ou cumulés. Nous notons :

- $P_{i,j,k}$  pour désigner le paiement annuel effectué dans la cellule  $(i, j)$  pour le  $k^{\text{ème}}$  sinistre ;
- $C_{i,j,k}$  pour désigner les paiements cumulés relatifs au  $k^{\text{ème}}$  sinistre de l'année d'accident  $i$  jusqu'à l'année de développement  $j$  ;
- $I_{i,j,k}$  pour désigner les charges relatives au  $k^{\text{ème}}$  sinistre de l'année d'accident  $i$  évalué à l'année de développement  $j$ .

Dans certains cas, il est préférable de travailler uniquement avec les paiements et/ou ajustements des charges *réellement* effectués au lieu de regarder l'évolution pour toutes les années de développement. Dès lors, quand nous le préciserons, l'indice  $j$  fera référence au  $j^{\text{ème}}$  paiement, paiement cumulé ou



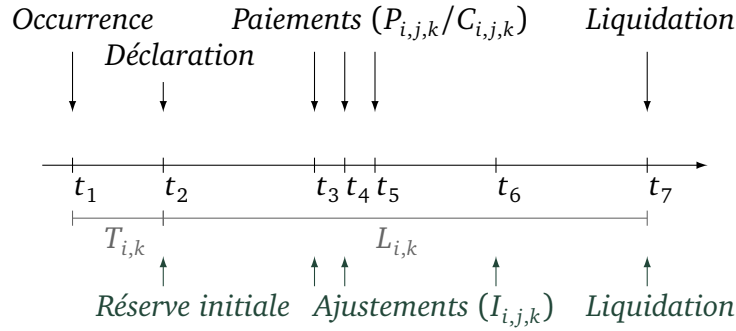


FIGURE 2.1: Ligne du temps d'un sinistre.

ajustement des charges, et non plus à l'année de développement. En d'autres termes, cela revient, d'une part, à supposer que

$$P_{i,j,k} > 0 \text{ pour } j \geq 1,$$

impliquant ainsi  $C_{i,j,k} - C_{i,j-1,k} > 0$ . Et d'autre part, que

$$I_{i,j,k} - I_{i,j-1,k} \neq 0 \text{ pour } j > 1.$$

Finalement, lorsque plus aucun paiement/ajustement n'est nécessaire, le sinistre est liquidé. Le *décalé de liquidation*, noté  $L_{i,k}$ , représente la durée écoulée entre la déclaration et la liquidation.

Nous désignons également par  $U_{i,k}^{(C)}$  (resp.  $U_{i,k}^{(I)}$ ) le nombre de variations effectives, c'est-à-dire différentes de 0, des paiements cumulés (resp. des charges) observées pour ce sinistre, *après le premier paiement* (resp. évaluation des charges). Avec cette notation, nous avons, au total,  $U_{i,k}^{(C)} + 1$  paiements et  $U_{i,k}^{(I)} + 1$  évaluations des charges. Nous utilisons une notation différente car les instants de paiements ainsi que leurs nombres peuvent être différents de ceux relatifs aux ajustements des charges.

Selon les méthodes, différentes durées, telles que, par exemple, le délai entre deux paiements, seront nécessaires. De même, différentes valeurs de comptage pourraient être requises.

Une représentation graphique est visible à la Figure 2.1. En noir, nous remarquons les mouvements relatifs au triangle des paiements alors qu'en vert, nous observons ceux relatifs aux charges. Les délais de déclaration et de liquidation sont également représentés en gris. De plus, pour cet exemple, le premier paiement (resp. évaluation du case estimate) a lieu en  $t_3$  (resp. en  $t_2$ ) et il en reste encore deux (resp. trois ajustements) à venir. Nous obtenons donc  $U_{i,k}^{(C)} = 2$  et  $U_{i,k}^{(I)} = 3$ .

Pour clôturer cette section, nous rappelons qu'il existe trois grandes catégories de sinistres. La première est formée par les *Incurred But Not Reported* (IBNR), qui concerne l'ensemble des sinistres qui ont eu lieu mais qui n'ont pas encore été notifiés à l'assureur. Ensuite, nous avons la classe formée par les *Reported But Not Paid* (RBNP), constituée des sinistres qui ont été notifiés à l'assureur mais pour lesquels aucun paiement n'a encore été effectué. De plus, ces sinistres ne sont pas encore clôturés. Finalement, nous avons les *Reported But Not Settled* (RBNS), représentant les sinistres déclarés à l'assureur, pour lesquels un premier paiement a au moins eu lieu mais qui ne sont pas encore clôturés. Ainsi, durant son développement, un sinistre va passer dans chacune des catégories citées ci-dessus. La ligne du temps d'un sinistre avec cette catégorisation est visible à la Figure 2.2. Comme nous le verrons pour les méthodes individuelles pures, cette classification aboutit à des besoins différents.

Une variable reviendra souvent pour modéliser la fréquence. Celle-ci permet d'exprimer le nombre de sinistres en fonction de l'année d'occurrence et du délai de déclaration. Nous notons alors  $K_{i,t}$  pour désigner le nombre de sinistres survenus l'année d'occurrence  $i$  et déclarés lors de la  $t^{\text{ème}}$  année de développement. Il peut également être important de distinguer la fréquence en fonction de la classification IBNR/RBNP/RBNS, nous notons donc :

- $K_{i,j}^{IBNR}$  le nombre de sinistres de type IBNR dans la cellule  $(i, j)$ , c'est-à-dire survenus l'année d'occurrence  $i$  et déclarés strictement après l'année de développement  $j$ . Mathématiquement, cela revient à écrire

$$K_{i,j}^{IBNR} = \sum_{t>j} K_{i,t}.$$

- $K_{i,j}^{RBNP}$  le nombre de sinistres de type RBNP dans la cellule  $(i, j)$ , c'est-à-dire survenus l'année d'occurrence  $i$ , déclarés avant l'année de développement  $j$  mais pour lesquels aucun paiement n'a été effectué avant l'année calendrier  $i + j - 1$  ;
- $K_{i,j}^{RBNS}$  le nombre de sinistres de type RBNS dans la cellule  $(i, j)$ , c'est-à-dire survenus l'année d'occurrence  $i$ , déclarés avant l'année de développement  $j$  pour lesquels un ou plusieurs paiements ont déjà eu lieu avant l'année calendrier  $i + j - 1$ .

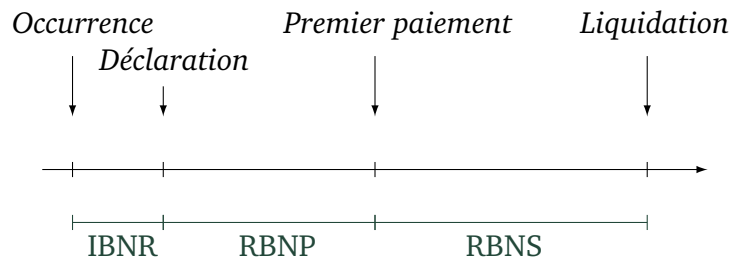


FIGURE 2.2: Ligne du temps d'un sinistre - Classification IBNR/RBNP/RBNS

## Chapitre 3

# Description des méthodes existantes

Le but de ce chapitre est de comparer différentes méthodes de réservation. Pour chacun des cas considérés, nous analysons les points suivants :

- Données nécessaires à la calibration ;
- Hypothèses relatives aux distributions ;
- Possibilité de traiter la dépendance/corrélation entre variables ;
- Estimation des différents paramètres ;
- Possibilité d'obtenir un déroulé de cash-flows par sinistre ;
- Limitations pour l'évaluation des réserves et du risque sur un horizon d'un an. Ce dernier est mieux connu sous la dénomination anglaise *One-year risk* ;
- Remarques diverses.

### 3.1 Méthodes agrégées

Commençons par analyser les méthodes les plus répandues, qui utilisent les données au niveau agrégé. Celles-ci sont présentes dans la branche gauche de l'arbre représenté à la Figure 1.1.

#### 3.1.1 Chain-Ladder (CL)

Nous présentons la méthode de Chain-Ladder telle qu'introduite par Mack [12].

• Données nécessaires à la calibration : Les seules données requises par cette méthode sont les paiements au niveau agrégé. Notons qu'elle peut également s'appliquer aux charges. Nous travaillons ensuite avec les paiements cumulés.

• Hypothèses sur les distributions : La méthode de Mack ne requiert pas la spécification d'une loi de probabilité particulière afin de décrire la dynamique des paiements. Cependant, celle-ci doit satisfaire deux critères :

- La moyenne de  $C_{i,j+1}$  sachant l'historique de paiements satisfait

$$\mathbb{E}(C_{i,j+1} | C_{i,j}, \dots, C_{i,1}) = C_{i,j} \lambda_j.$$

- La variance conditionnelle est de la forme

$$\text{Var}(C_{i,j+1}|C_{i,j}, \dots, C_{i,1}) = C_{i,j}\sigma_j^2, \sigma_j^2 > 0.$$

A travers ces hypothèses, nous constatons que l'évolution des paiements cumulés pour une année d'accident  $i$  fixée est uniquement expliquée par un effet année de développement. Dès lors, la modélisation, par exemple, d'un effet calendrier, est impossible.

• Dépendance entre variables : Nous supposons que les montants payés pour les sinistres provenant d'années d'accident différentes sont indépendants, ce qui signifie

$$\{C_{i_1,1}, \dots, C_{i_1,n}\} \perp \{C_{i_2,1}, \dots, C_{i_2,n}\},$$

pour  $i_1 \neq i_2$  deux années d'accident. Néanmoins, aucune hypothèse n'est faite au niveau individuel, excepté que deux sinistres d'année d'origine différentes sont indépendants (cela découle de l'assertion ci-dessus). Dès lors, une dépendance entre deux sinistres de la même année d'accident n'est pas exclue.

A partir des hypothèses sur la loi des paiements, nous pouvons affirmer qu'il existe une relation de dépendance entre deux paiements successifs. Cependant, cette dépendance n'est pas très effective. En effet, l'observation du dernier paiement se répercute uniquement dans la moyenne et dans la variance du paiement à prédire, ce qui limite la structure de dépendance.

• Estimation des paramètres : Grâce aux hypothèses sur les paiements  $(C_{i,j})_{i,j}$ , Mack a démontré que l'estimateur obtenu via moindres carrés pondérés pour  $\lambda_j$  est donné par

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}, j = 1, 2, \dots, n-1. \quad (3.1)$$

• Déroulé individuel de cash-flows : Nous travaillons avec une méthode agrégée. Un déroulé de cash-flows, sinistre par sinistre, n'est pas prévu.

• Limitations pour l'évaluation des réserves et du *One-year risk* : Rappelons que la méthode ne spécifie aucune distribution, nous avons uniquement accès à la moyenne et à la variance des paiements cumulés. Dès lors, aucune simulation des réserves ne sera obtenue. De même, le *One-year risk* ne sera pas calculable.

Cependant, le best-estimate des réserves est accessible. Grâce à la loi des espérances itérées, le best-estimate des réserves pour les sinistres survenus l'année d'accident  $i > 1$ , évaluées en fin d'année calendrier  $n$ , est donné par

$$\mathbb{E}(C_{i,n}|C_{i,1}, \dots, C_{i,n+1-i}) - C_{i,n+1-i} = C_{i,n+1-i} \times (\lambda_{n+1-i} \times \dots \times \lambda_{n-1} - 1).$$

Dès lors, en sommant sur  $i$ , nous pouvons estimer le montant total qui doit être réservé.

• Remarques diverses : Notons que les réserves obtenues à l'aide du Chain-Ladder, telle qu'introduite par Mack [12], sont également accessibles via le *modèle de Poisson surdispersé*, noté ODP, défini comme

$$\mathbb{E}(P_{i,j}) = \mu_i \gamma_j \text{ et } \text{Var}(P_{i,j}) = \phi \times (\mu_i \gamma_j), \quad (3.2)$$

pour  $\mu_1, \dots, \mu_n, \gamma_1, \dots, \gamma_n$  et  $\phi$  des constantes positives, avec la contrainte d'identification des paramètres suivante

$$\sum_{j=1}^n \gamma_j = 1.$$

Notons que nous travaillons dans ce cas avec les paiements sous la forme incrémentale. Dès lors, grâce à Merz et Wütrich [15, Théorème 3.4, p. 94], nous pouvons affirmer que l'estimation des réserves pour les sinistres survenus l'année d'accident  $i > 1$ , évaluées à la fin de l'année calendrier  $n$ , satisfait

$$\begin{aligned} \widehat{\mathbb{E}}(C_{i,n}|C_{i,1}, \dots, C_{i,n+1-i}) - C_{i,n+1-i} &= C_{i,n+1-i} \times (\widehat{\lambda}_{n+1-i} \times \dots \times \widehat{\lambda}_{n-1} - 1) \\ &= \sum_{j=n-i+2}^n \hat{\mu}_i \hat{\gamma}_j, \end{aligned}$$

où les facteurs de lien  $\widehat{\lambda}_j$  sont définis à l'équation (3.1). Nous avons donc deux spécifications différentes (Mack et ODP), qui nous donnent les mêmes estimations des réserves. Soulignons qu'une troisième méthode, nous donnant les mêmes estimations, existe et est connue sous la dénomination anglaise *The non-informative prior Gamma-Gamma Bayesian CL model*. Ces trois modèles sont longuement détaillés dans Merz et Wütrich [15].

### 3.1.2 Paid-Incurred Chain (PIC)

On s'intéresse maintenant au modèle (PIC). La conception de ce dernier est similaire à celle expliquée pour Chain-Ladder. La différence principale trouve sa source dans la volonté d'utiliser les charges. Nous repartons du même schéma afin d'expliquer cette méthode.

- Données nécessaires à la calibration : Dans ce contexte, les paiements et les charges, au niveau agrégé, sont nécessaires. Nous travaillons également avec les paiements sous forme cumulative.
- Hypothèses sur les distributions : Soit  $i$  une année d'accident fixée, cette méthode se base principalement sur quatre hypothèses :

- Premièrement, nous supposons que la charge terminale correspond au paiement total. Nous limitons ici notre analyse à  $n$  années, ce qui revient à écrire

$$C_{i,n} = I_{i,n} \text{ presque sûrement.}$$

- La seconde hypothèse est du même ordre que celle formulée dans Chain-Ladder. Nous supposons que

$$C_{i,j} = C_{i,j-1} \lambda_{j-1}, \forall j = 2, \dots, n. \quad (3.3)$$

Soulignons que le premier paiement cumulé  $C_{i,1}$  doit également être modélisé afin de déterminer l'évolution totale.

- Dans la même lignée, nous lions deux charges successives de la manière suivante

$$I_{i,j-1} = I_{i,j} \gamma_{j-1}, \forall j = 2, \dots, n. \quad (3.4)$$

- Il reste maintenant à fixer une loi pour les différentes valeurs. Grâce aux écritures ci-dessus, il est suffisant de préciser une distribution pour les variables  $C_{i,1}$ ,  $\lambda_j$  et  $\gamma_j$  pour  $i = 1, \dots, n$  et  $j = 1, \dots, n-1$ . Par convenance, nous spécifions une loi pour le logarithme de ces valeurs. De plus, afin d'inclure une structure de dépendance entre ces variables, nous modélisons leur loi jointe. Cette idée permet de créer une relation entre les paiements et les charges. Nous définissons alors le vecteur aléatoire suivant

$$\Lambda = \left( (\ln C_{i,1})_i, (\ln \lambda_j)_j, (\ln \gamma_j)_j \right)^t. \quad (3.5)$$

Nous devons maintenant spécifier la loi de  $\Lambda$ . D'une part, nous supposons que la moyenne des  $\ln C_{i,1}$  est de la forme

$$\mathbb{E}(\ln C_{i,1}) = \nu_i + \alpha_1,$$

permettant ainsi d'insérer les effets relatifs aux années d'occurrence et de développement. D'autre part, nous supposons que les moyennes des  $\ln \lambda_j$  et  $\ln \gamma_j$  sont de la forme

$$\mathbb{E}(\ln \lambda_j) = g_j \text{ et } \mathbb{E}(\ln \gamma_j) = h_j, 1 \leq j \leq n-1,$$

où seule l'année de développement a un impact. Dès lors, la moyenne de  $\Lambda$  peut s'écrire comme

$$\mathbb{E}(\Lambda) = \left( (\mathbb{E}(\ln C_{i,1}))_i, (\mathbb{E}(\ln \lambda_j))_j, (\mathbb{E}(\ln \gamma_j))_j \right)^t = A\Theta,$$

avec  $\Theta$  le vecteur reprenant les paramètres relatifs aux moyennes, c'est-à-dire

$$\Theta = (\nu_1, \dots, \nu_n, \alpha_1, g_1, \dots, g_{n-1}, h_1, \dots, h_{n-1})^t,$$

et  $A$  une matrice déterminée et uniquement constituée de 0 et de 1.

Enfin, sachant  $\Theta$ , nous admettons que le vecteur  $\Lambda$  suit une *loi normale multivariée* de matrice de covariance  $\Sigma$  et de moyenne  $A\Theta$ . Dès lors, en passant au logarithme aux équations (3.3) et (3.4), nous pouvons affirmer que  $\ln(C_{i,j}/C_{i,j-1})$  et  $\ln(I_{i,j-1}/I_{i,j})$  suivent également une loi normale.

Nous avons ici suivi l'approche proposée par Merz et Wütrich [14] et avons donc supposé que  $\Lambda$  suit une loi normale. Ce choix est uniquement effectué par convenance et trouve principalement sa source dans la faculté d'obtenir des formes analytiques déterminées. Cependant, plusieurs autres possibilités sont ouvertes pour définir cette loi jointe. En effet, nous aurions pu supposer que les lois marginales, c'est-à-dire celles de  $(\ln C_{i,1})_i$ ,  $(\ln \lambda_j)_j$  et  $(\ln \gamma_j)_j$ , sont normales avec une structure différente pour la loi jointe (e.g. une loi normale multivariée, une loi de Student multivariée, une loi elliptique ou encore une loi normale avec asymétrie).

Dans ce contexte, nous commençons avec le premier paiement  $C_{i,1}$  et nous construisons les paiements cumulés jusqu'à l'instant  $n$ , soit  $C_{i,n}$ . Nous construisons ensuite, à reculons, à partir de  $I_{i,n} = C_{i,n}$ , le pattern des charges.

Notons, qu'une nouvelle fois, la modélisation de l'effet calendrier n'est pas introduite dans le modèle.

- Dépendance entre variables : Sous l'hypothèse multinormale, les différents paiements ainsi que les charges peuvent être corrélés. Aucune hypothèse quant à l'indépendance est formulée.
- Estimation des paramètres : Dans ce cadre, nous devons estimer le vecteur moyenne et la matrice de covariance de  $\Lambda$ .

En ce qui concerne la moyenne, vu que la matrice  $A$  est connue, il suffit de déterminer les paramètres de  $\Theta$ . Deux possibilités s'offrent à nous

- Soit ces paramètres sont estimés *a priori*, ce qui facilite alors l'exposé ;
- Soit nous nous plaçons dans un cadre bayésien, et tout comme Merz et Wütrich [14], nous supposons que

$$\Theta \sim \text{Nor}(v, \Xi),$$

avec des estimations *a priori* de la moyenne  $v$  et de la matrice de covariance  $\Xi$ . A titre informatif, la matrice de covariance de  $\Theta$  peut être estimée à l'aide d'un processus auto-régressif d'ordre 1. Nous renvoyons le lecteur intéressé vers [15, p. 150].

Dans ce contexte, un résultat important (voir [15, p. 242]) découle de nos hypothèses

$$\Lambda \sim \text{Nor}(Av, \Sigma + A\Xi A^t).$$

Ce corollaire nous indique que le seul impact de  $\Theta$  sur la distribution de  $\Lambda$  provient de sa moyenne  $v$  et de sa matrice de covariance  $\Xi$ , fixées *a priori*. Cela permet d'éviter les simulations de  $\Theta$  en amont afin d'obtenir la distribution de  $\Lambda$ . On conçoit donc l'importance de ce corollaire dans la réduction de la complexité algorithmique et la mise en place d'un tel modèle.

Notons que la différence entre ces deux modélisations de  $\Theta$  se répercute principalement au niveau de la variance de  $\Lambda$  et symbolise la possible volatilité induite par les valeurs de  $\Theta$ .

Finalement, intéressons-nous à l'estimation de la matrice de covariance  $\Sigma$ . Pour ce faire, nous procédons à une estimation empirique de cette dernière. Une discussion sur la calibration de  $\Sigma$  peut être consultée dans [15, p. 246].

- Déroulé individuel de cash-flows : Nous travaillons avec une méthode agrégée. Un déroulé de cash-flows, sinistre par sinistre, n'est pas prévu.
- Limitations pour l'évaluation des réserves et du One-year risk : Supposons que les triangles supérieurs des paiements et des charges soient remplis jusqu'à l'année calendrier  $n$ . Une des propriétés remarquable de la loi gaussienne est la suivante

$$\Lambda = (\Lambda_1, \Lambda_2) \sim \text{Normale} \Rightarrow (\Lambda_1 | \Lambda_2 = x) \sim \text{Normale}. \quad (3.6)$$

Cette propriété permet donc d'affirmer que la loi du vecteur des log-ratios futurs sachant l'information disponible jusqu'à l'année calendrier  $n$ , suit une loi normale. Notez que le vecteur moyenne et la matrice de covariance devront être ajustés en fonction de l'information disponible.

Vu qu'une combinaison linéaire d'un vecteur gaussien reste gaussienne, nous pouvons facilement simuler les paiements futurs et ainsi les réserves pour chacune des années d'accident considérée. De plus, il est aisé de déterminer le best-estimate de ces dernières.

De la même manière, le *One-year risk* s'obtient assez facilement. En effet, grâce à cette approche paramétrique, nous pouvons simuler les réserves lors de l'année calendrier  $n$  et donc les paiements et les charges inhérents à l'année calendrier  $n + 1$ . En fonction des simulations réalisées entre les années calendrier  $n$  et  $n + 1$ , nous pouvons définir les paramètres de la loi normale des paiements futurs (compte tenu de l'information disponible jusqu'à l'année calendrier  $n + 1$ , donc) et effectuer des simulations de la réserve. Dès lors, la *Value-at-Risk* de la différence entre les réserves évaluées aux années calendrier  $n$  et  $n + 1$ , le tout diminué des paiements effectués entre ces deux années, définit les pertes sur l'année.

- Remarques diverses : Grâce à l'hypothèse de normalité, des formes plus ou moins aisées de l'erreur de prédiction moyenne sont disponibles. Pour plus de détails, nous renvoyons le lecteur vers Merz et Wütrich [15, p. 245].

### 3.1.3 Approche de Tweedie

Dans cette section, nous abordons un modèle basé sur les distributions de Tweedie, proposé par Merz et Wütrich [15] et Wütrich [22].

Avant de commencer la comparaison, il est préférable d'introduire quelques concepts. Commençons par définir les lois de Tweedie et les notions adjacentes.

Dist.	$\theta$	$a(\theta)$	$\phi$	$\mathbb{E}(Y)$	Var( $Y$ )
Bin( $1, \mu$ )	$\ln \frac{\mu}{1-\mu}$	$\ln(1 + \exp(\theta))$	1	$\mu$	$\phi \mu(1 - \mu) = \mu(1 - \mu)$
Nor( $\mu, \sigma^2$ )	$\mu$	$\frac{\theta^2}{2}$	$\sigma^2$	$\mu$	$\phi = \sigma^2$
Poi( $\mu$ )	$\ln \mu$	$\exp(\theta)$	1	$\mu$	$\phi \mu = \mu$
Gam( $\mu, \alpha$ )	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\frac{1}{\alpha}$	$\mu$	$\phi \mu^2 = \frac{\mu^2}{\alpha}$
IGau( $\mu, \sigma^2$ )	$-\frac{1}{2\mu^2}$	$-\sqrt{-2\theta}$	$\sigma^2$	$\mu$	$\phi \mu^3 = \sigma^2 \mu^3$

TABLE 3.1: Quelques membres de la famille exponentielle linéaire ( $w=1$ ).

**Définition 3.1.1 (Loi de Tweedie)** Les lois de Tweedie forment un sous-ensemble de la famille exponentielle linéaire (angl. Exponential Dispersion Family, dénommée dans la suite par ED) qui satisfont la propriété suivante

$$\text{Var}(Y) = \phi \mathbb{E}(Y)^p, \quad Y \in ED, \phi, p \geq 0.$$

**Définition 3.1.2 (ED family)** Une variable aléatoire  $Y$  appartient à la famille exponentielle linéaire si sa fonction de masse de probabilité/de densité est de la forme

$$\left. \begin{array}{l} \mathbb{P}(Y = y) \\ f_Y(y) \end{array} \right\} = \exp\left(\frac{y\theta - a(\theta)}{\phi/w}\right) c(y, \phi/w), \quad (3.7)$$

pour tout  $y$  appartenant au support de  $Y$ . Ici,  $\theta$  représente le paramètre canonique,  $\phi$  le paramètre de dispersion et  $w$  le poids.

La volonté principale derrière l'utilisation de la famille exponentielle linéaire est que celle-ci permet d'utiliser les modèles linéaires généralisés (GLM). En effet, une des conditions pour effectuer un GLM est que la fonction de densité ou de probabilité de la variable réponse s'écrive sous la forme (3.7).

Une propriété importante de la famille exponentielle linéaire permet d'affirmer que la moyenne arithmétique de  $n$  variables aléatoires i.i.d., présentes dans cette famille, reste au sein de celle-ci avec les mêmes paramètres excepté le poids  $w$ , qui évolue vers  $nw$ . Ce qui vient d'être énoncé sera notamment utile lors du passage au niveau individuel. Cette proposition ainsi que d'autres peuvent être consultées dans Kaas *et al.* [10] ainsi que dans les cours Denuit [4], Trufin [21].

Nous présentons au Tableau 3.1 quelques exemples de membres de la famille exponentielle linéaire, pour un poids  $w$  fixé à 1. Si l'on regarde la dernière colonne, relative à la variance, nous pouvons déduire certaines lois de Tweedie (e.g. lois normale, de Poisson, Gamma et inverse gaussienne).

Nous avons maintenant assez d'éléments à disposition pour entamer le schéma de comparaison des méthodes.

- Données nécessaires à la calibration : Dans ce contexte, seuls les paiements au niveau agrégé sont nécessaires. En particulier, nous travaillons ici avec les paiements incrémentaux.
- Hypothèse sur les distributions : A l'origine, une seule hypothèse relative à la loi des paiements incrémentaux  $P_{i,j}$  est nécessaire à cette méthode.

Soit  $p \in [1, 2]$  fixé. Supposons qu'il existe des constantes positives  $\mu_1, \dots, \mu_n, \gamma_1, \dots, \gamma_n$  et  $\phi$  telles que les deux égalités suivantes soient satisfaites

$$\mathbb{E}(P_{i,j}) = \mu_i \gamma_j \text{ et } \text{Var}(P_{i,j}) = \phi \times (\mu_i \gamma_j)^p, \quad (3.8)$$



pour  $i = 1, \dots, n$  et  $j = 1, \dots, n$ . Afin de déterminer ces paramètres de manière unique, nous imposons une des contraintes d'identification suivantes

$$\mu_1 = 1, \gamma_1 = 1 \text{ ou encore } \sum_{j=1}^n \gamma_j = 1.$$

En particulier, cette hypothèse est satisfaite par les lois de Tweedie (voir Définition 3.1.1) et nous choisissons de nous placer dans ce contexte pour la suite de la présentation. Notons que ce choix est uniquement effectué par convenance et n'est pas une demande explicite du modèle original. Cela dit, plusieurs facteurs nous poussent dans cette direction :

- Les lois de Tweedie appartiennent à la famille exponentielle linéaire. Par conséquent, une utilisation des GLM est envisageable. Cela permet donc à l'actuaire d'utiliser des outils connus et souvent à disposition ;
- Sous l'hypothèse de Tweedie, lorsque  $p \in ]1, 2[$ , cela revient à supposer que  $P_{i,j}$  suit une loi Poisson-Gamma composée. Autrement dit, nous retrouvons une approche collective, bien connue de tous, où nous modélisons le nombre de paiements par une loi de Poisson alors que le coût annuel de ceux-ci est donné par une loi Gamma. Une telle affirmation est notamment prouvée dans Smyth et Jorgensen [20].

De plus, si  $p = 1$ , nous travaillons avec un modèle de Poisson surdispersé, nous donnant ainsi, comme nous l'avons vu précédemment, la même estimation des réserves que Chain-Ladder, alors que si  $p = 2$ , les paiements  $P_{i,j}$  suivent une loi Gamma, conformément au Tableau 3.1.

De nouveau, notez que la modélisation d'un effet calendrier n'est pas prise en considération par le modèle.

- Dépendance entre variables : Nous supposons ici que les paiements incrémentaux  $P_{i,j}$  sont indépendants.
- Estimation des paramètres : Dans un premier temps, supposons que le paramètre  $p \in [1, 2]$  est connu. Un avantage de travailler avec un membre de la famille exponentielle linéaire est que sa fonction de vraisemblance, compte tenu des observations réalisées jusqu'à l'année calendrier  $n$  ( $\mathcal{D}_n$ , donc), est connue. Par conséquent, en maximisant cette dernière, nous pouvons estimer les valeurs  $\mu_i$  et  $\gamma_j$ . Le système à résoudre est disponible dans Merz et Wütrich [15, p. 87]. Dès lors, le paramètre  $\phi$  peut, par exemple, être estimé à l'aide de la formule des résidus de Pearson telle que présentée dans Merz et Wütrich [15, Eq. (3.12), p. 91].

Supposons que les paramètres mentionnés ci-dessus ont été estimés. Si nous souhaitons évaluer la loi des  $P_{i,j}$ , en fonction de  $p \in [1, 2]$ , nous distinguons les cas suivants :

- Si  $p = 1$  (resp.  $p = 2$ ), il s'agit d'une loi de Poisson surdispersée (resp. Gamma) et la connaissance des estimations de la moyenne et de la variance, modélisées à l'équation (3.8), est suffisante pour caractériser la loi des  $P_{i,j}$  ;
- Si  $p \in ]1, 2[$ , il s'agit cette fois d'une loi Poisson-Gamma composée. Dès lors, grâce à Zhang [23, Eq. (2.2)] et Kaas [9, p. 8], la fréquence est modélisée par une loi de Poisson de moyenne

$$\frac{(\widehat{\mu}_i \widehat{\gamma}_j)^{2-p}}{\widehat{\phi}(2-p)},$$

alors que les montants suivent la loi suivante

$$\text{Gam} \left( (2-p) \widehat{\phi} (\widehat{\mu}_i \widehat{\gamma}_j)^{p-1}, \frac{2-p}{p-1} \right),$$

où les paramètres symbolisent la moyenne et la *forme* de la distribution Gamma.

Généralement, le paramètre  $p$  n'est pas déterminé avant de débiter les estimations. Ce problème de sélection du modèle est technique et loin d'être évident, nous n'en donnons donc qu'une brève explication et renvoyons les lecteurs intéressés par plus de détails vers Merz et Wütrich [15] et Wütrich [22] ainsi que les références citées dans ces écrits. Une approche bayésienne est également proposée dans Peters *et al.* [17].

Grâce à de bonnes propriétés mathématiques, une estimation *a posteriori* de  $p$  est rendue possible. Dès lors, une manière de faire, consiste à estimer  $\mu_i$  et  $\gamma_j$  comme si la valeur de  $p$  était déterminée et à ajuster au fur et à mesure. Nous donnons ici les premières étapes de cette approche algorithmique :

1. Fixons  $p^{(0)}$  et estimons  $\mu_i^{(1)}$  et  $\gamma_j^{(1)}$  comme énoncé précédemment ;
2. Nous pouvons maintenant estimer  $\phi^{(1)}$  ainsi que  $p^{(1)}$  en maximisant la fonction de vraisemblance. Pour être tout à fait précis, il s'agit de la profile vraisemblance. Soulignons que cette maximisation peut être réalisée via le package ChainLadder disponible dans R ;
3. On recommence ces étapes jusqu'à une certaine convergence des valeurs estimées ci-dessus.

• Déroulé individuel de cash-flows : Nous travaillons avec une méthode agrégée. Un déroulé de cash-flows, sinistre par sinistre, n'est pas prévu.

• Limitations pour l'évaluation des réserves et du One-year risk : Supposons que le paramètre  $p$  est fixé.

Si nous sommes uniquement intéressés dans la détermination du best-estimate des réserves, seules les valeurs  $\mu_i$  et  $\gamma_j$  sont intéressantes et nous obtenons, pour les années d'accident  $i > 1$  et calendrier  $n$ , le résultat suivant

$$\mathbb{E} \left( \sum_{j \geq n-i+2} P_{i,j} \mid \mathcal{D}_n \right) = \sum_{j \geq n-i+2} \mu_i \gamma_j. \quad (3.9)$$

En sommant sur  $i$ , nous évaluons finalement la totalité des réserves pour l'année calendrier  $n$  considérée.

D'autre part, si nous le souhaitons, nous pouvons déterminer la loi des réserves. Une fois tous les paramètres estimés, nous connaissons la distribution des paiements incrémentaux dans chaque cellule  $(i, j)$ . Dès lors, en rappelant que les réserves pour l'année d'accident  $i > 1$ , évaluées à l'année calendrier  $n$ , sont définies par

$$\sum_{j \geq n-i+2} P_{i,j},$$

et que les réserves totales sont définies par

$$\sum_{i=2}^n \sum_{j \geq n-i+2} P_{i,j}, \quad (3.10)$$

nous pouvons déterminer (e.g. via simulations) la loi de ces dernières. Vu que pour  $p \in ]1, 2[$ , les paiements  $P_{i,j}$  suivent une distribution de type Poisson-Gamma composée et sont indépendants, la loi des réserves peut également être calibrée à l'aide de l'algorithme de Panjer.

Dans le même ordre d'idée, le *One-year risk* s'obtient assez facilement en procédant comme suit :

- Simuler, de manière indépendante,  $k$  fois les réserves totales, telles que définies par l'équation (3.10), compte tenu de l'information disponible jusqu'à l'année calendrier  $n$ .

Pour réaliser cette étape, nous avons dû simuler  $k$  fois les paiements  $P_{i,j}$  pour l'année calendrier  $n + 1$ , c'est-à-dire  $i + j - 1 = n + 1$  (la diagonale suivante, donc). Notons  $\mathcal{D}_{n+1}^{(k)}$  l'ensemble des observations réalisées jusqu'à l'année calendrier  $n + 1$  obtenu grâce à la  $k^{\text{ème}}$  simulation de ces paiements, c'est-à-dire

$$\mathcal{D}_{n+1}^{(k)} = \mathcal{D}_n \cup \{p_{i,j}^{(k)} : i + j - 1 = n + 1\};$$

- Conditionnellement à chaque ensemble  $\mathcal{D}_{n+1}^{(k)}$ , nous pouvons simuler, également  $k$  fois et de manière indépendante, les réserves totales en considérant cette information additionnelle;
- En prenant la différence entre les réserves calculées lors de l'année calendrier  $n$  et  $n + 1$  et en retranchant les paiements effectués lors de l'année calendrier  $n + 1$ , nous pouvons évaluer la *Value-at-Risk* à un an.

• Remarques diverses : Si  $p \in ]1, 2[$ , nous modélisons les paiements via un modèle Poisson-Gamma composé. Cependant, la loi Gamma utilisée pour représenter les paiements ne possède pas une queue épaisse. Dès lors, notre approche ne nous permet pas une très bonne approximation des frais élevés. Ce problème a d'ailleurs été relevé et résolu par Denuit et Trufin [6], comme nous le verrons plus tard.

D'autre part, à titre informatif, une approximation de l'erreur-carré moyenne de la prédiction est expliquée par Merz et Wütrich [15, p. 90].

Finalement, nous avons supposé  $\phi$  constant à travers les périodes de développement. Cependant, cette hypothèse peut être très restrictive et pourrait être adaptée en fonction de la base de données considérée. Nous renvoyons le lecteur intéressé par ce sujet vers Boucher et Davidov [3].

## 3.2 Méthodes individuelles pures

Nous nous intéressons maintenant aux méthodes purement individuelles. Comme énoncé précédemment, celles-ci permettent d'obtenir la projection de cash-flows pour chacun des sinistres. Cependant, cette faculté est contre-balançée par une complexité algorithmique importante. Nous commençons cette section par le modèle ILR, proposé par Pigeon *et al.* [18].

### 3.2.1 ILR

L'idée de cette méthode est similaire à celle du Chain-Ladder mais nous travaillons cette fois au niveau individuel. De plus, nous modélisons uniquement les paiements effectifs, on ne s'intéresse donc pas aux paiements nuls. Cette modélisation change quelque peu l'approche. En effet, pour chaque sinistre, nous travaillons maintenant par paiement et non plus par année de développement. Nous utilisons les mêmes notations que précédemment excepté que l'indice  $j$  ne fait plus référence au paiement effectué lors de l'année de développement  $j$  mais au  $j^{\text{ème}}$  paiement. Mathématiquement, le  $j^{\text{ème}}$  paiement cumulé pour le  $k^{\text{ème}}$  sinistre de l'année d'accident  $i$ , obéit à la relation suivante

$$C_{i,j,k} = C_{i,j-1,k} \lambda_{j-1}^{(k)} \text{ avec } \lambda_{j-1}^{(k)} \neq 1. \quad (3.11)$$

Le vecteur de développement pour le  $k^{\text{ème}}$  sinistre survenu durant l'année d'occurrence  $i$ , sachant qu'il y a, au total,  $u_{i,k}^{(C)} + 1$  paiements effectifs, est donné par

$$\Lambda_i^{(k)} := \left( \ln C_{i,1,k}, \ln \lambda_1^{(k)}, \dots, \ln \lambda_{u_{i,k}^{(C)}}^{(k)} \right). \quad (3.12)$$

Notez les similitudes entre ce vecteur et, en faisant abstraction des charges, celui introduit à l'Equation (3.5) lors de la présentation du PIC. Dès lors, si le timing des paiements ne nous intéresse pas, grâce à (3.11), la connaissance de ce vecteur est suffisante pour caractériser le sinistre. Au contraire, si la connaissance parfaite de la trajectoire est requise, nous devons également nous intéresser aux différents délais. Dans un premier temps, faisons abstraction du timing des cash-flows. Nous expliquerons brièvement, dans le schéma de comparaison, comment une telle possibilité peut être prise en considération.

La classification IBNR/RBNP/RBNS des sinistres, présentée à la Figure 2.2, joue un rôle important dans la projection individuelle des cash-flows. En effet, pour chaque catégorie, les besoins en simulations sont totalement différents. Soient  $i$  une année d'occurrence et  $n$  une année calendrier fixées, pour la classe IBNR, nous devons simuler :

- Le nombre total de sinistres qui n'ont pas encore été déclarés à l'année calendrier  $n$ , c'est-à-dire ceux déclarés strictement après l'année de développement  $n - i + 1$ ,  $K_{i,n-i+1}^{IBNR}$  donc.
- Pour chaque sinistre, le nombre de paiements qui sont effectués. Nous rappelons que le nombre total de paiements effectifs pour le  $k^{\text{ème}}$  sinistre survenu l'année  $i$  est modélisé par la variable aléatoire  $U_{i,k}^{(C)} + 1$  ;
- Une fois le nombre de paiements à effectuer connu pour chaque sinistre, nous devons spécifier leurs montants respectifs. Cela équivaut à identifier le vecteur (3.12).

Pour la classe des RBNP, le nombre total de sinistres  $K_{i,n-i+1}^{RBNP}$  survenus l'année  $i$  qui ont été déclarés avant l'année calendrier  $n$ , pour lesquels aucun paiement n'a été effectué, est connu. Il reste donc à déterminer le nombre de paiements  $U_{i,k}^{(C)} + 1$  pour chaque sinistre et le vecteur de développement (3.12). Finalement, pour les RBNS, nous avons également à disposition le nombre de sinistres  $K_{i,n-i+1}^{RBNS}$  survenus l'année  $i$ , déclarés avant l'année calendrier  $n$ , pour lesquels un ou plusieurs paiements ont été effectués et un historique de paiements existe. Nous connaissons donc une partie du vecteur de développement (3.12). Il reste à préciser la fin de la trajectoire des sinistres, et donc des paiements, en tenant compte des observations.

Grâce à cette classification, nous avons maintenant une vision claire des besoins du modèle proposé. Il reste à spécifier :

- Une loi pour le nombre de sinistres IBNR, c'est-à-dire  $K_{i,n-i+1}^{IBNR}$  ;
- Une loi pour le nombre de paiements effectifs, les  $U_{i,k}^{(C)}$  donc ;
- Une loi pour le vecteur de développement (3.12) lorsque le premier paiement n'a pas encore été effectué ;
- Lorsqu'un historique de paiements existe, une loi *conditionnelle* pour le nombre total de paiements sachant qu'il y en a déjà eu une certaine quantité (cela permet donc de déterminer le nombre de paiements restants et la longueur totale du vecteur de développement) et pour la fin du vecteur (3.12).

Maintenant que l'idée principale est présentée, nous pouvons nous concentrer sur le schéma de comparaison usuel.

• Données nécessaires à la calibration : Pour cette méthode, nous avons besoin :

- Des paiements individuels. Notons que ces derniers sont agrégés sur la période considérée. Ainsi, si la période de calcul est l'année, tous les paiements individuels effectués la même année sont sommés. Par facilité, nous travaillons ici par année ;
- D'une mesure d'exposition pour chaque année d'occurrence  $i$ , notée  $w(i)$ , ainsi que des occurrences et des délais de déclaration individuels  $T_{i,k}$  ;
- Si l'on souhaite simuler le timing des différents cash-flows, des délais de liquidation  $L_{i,k}$  et des instants de paiements, pour chaque sinistre.

• Hypothèses sur les distributions : Grâce à l'introduction, nous connaissons les lois qu'il reste à spécifier. Pour ce faire, nous suivons les hypothèses prises dans [18] et [19].

Premièrement, précisons la loi de  $K_{i,n-i+1}^{IBNR}$ , correspondant au nombre de sinistres de type IBNR pour les années d'occurrence  $i$  et calendrier  $n$ , avec  $1 \leq i \leq n$ . Cependant, la seule information à disposition dans la base de données est le nombre de sinistres déclarés. Dès lors, une manière de faire consiste à d'abord spécifier la loi relative aux délais de déclaration  $T_{i,k}$ . Celle-ci servira ensuite pour ajuster le nombre d'occurrences.

Comme énoncé dans [2, 18], nous pouvons supposer que  $T_{i,k}$  suit une loi géométrique avec quelques composantes dégénérées. Cela revient à définir sa fonction de densité de la manière suivante :

$$f_{T_{i,k}}(t; \nu) = \sum_{l=0}^m I_l(t) \nu_l + \left( 1 - \sum_{l=0}^m \nu_l \right) f_{T_{i,k}|T_{i,k}>m}(t; \nu_{m+1}), \quad (3.13)$$

où nous définissons :

- $I_l(t)$  comme étant égal à 1 si l'instant  $t$  se situe dans la  $l^{\text{ème}}$  année après l'occurrence et 0 sinon ;
- La fonction  $f_{T_{i,k}|T_{i,k}>m}(t; \nu_{m+1})$  comme la densité d'une loi géométrique de paramètre  $\nu_{m+1}$ , par exemple. Cependant, une loi de Poisson ou Binomiale Négative peut également être calibrée si elle correspond mieux aux données.

Sur les premières années, nous avons beaucoup de déclarations et il est donc préférable d'estimer les probabilités par les fréquences observées. Pour modéliser les délais de déclaration plus lointains, nous supposons que la queue de la distribution évolue selon une loi géométrique. La valeur charnière  $m$  et les différents paramètres du vecteur  $\nu$  doivent finalement être estimés. Soulignons que ceux-ci sont indépendants de l'année d'occurrence  $i$  et de l'accident  $k$ , ce qui implique que l'ensemble des délais de déclaration suivent la même loi. Dans la suite, nous notons  $F_{T_{i,k}}(t; \nu)$  pour désigner la fonction de répartition associée.

Notons que, si nous souhaitons connaître les différents délais, nous pouvons modéliser les nombres de périodes entre deux paiements/mouvements consécutifs comme proposé à l'Equation (3.13). Ainsi, nous modélisons :

- Le nombre de périodes entre la déclaration et le premier paiement ;
- Le nombre de périodes entre le premier et le deuxième paiement ;
- ...

- Vu que le dernier paiement peut être effectué avant liquidation, le nombre de périodes entre celui-ci et la liquidation.

D'autre part, la modélisation de  $K_{i,n-i+1}^{IBNR}$  se résume à celle d'une variable de comptage. Dès lors, une loi de Poisson, par exemple, semble appropriée. Finalement, il reste à ajuster par le délai de déclaration et nous supposons donc que  $K_{i,n-i+1}^{IBNR}$  suit une loi de Poisson de moyenne

$$\theta w(i) \left(1 - F_{T_{i,k}}(n - i; \nu)\right).$$

Ainsi, en fonction du délai  $n - i$  entre l'année d'occurrence et la date d'évaluation considérées, le nombre moyen de sinistres IBNR sera plus ou moins important. En effet, si la différence  $n - i$  est élevée, le terme entre parenthèses va tendre vers 0, impliquant un nombre moyen de sinistres non-déclarés, à la date d'évaluation, faible. Cette constatation est bien en ligne avec la réalité.

Nous précisons maintenant la loi de  $U_{i,k}^{(C)}$ . En suivant [18, 19], nous pouvons supposer que cette variable de comptage suit également une loi géométrique avec quelques composantes dégénérées telle que spécifiée à l'Equation (3.13). Néanmoins, en fonction de la base de données, d'autres modélisations peuvent être proposées si elles s'avèrent plus appropriées. Dès lors, grâce à la loi des probabilités totales, il est assez aisé de déterminer le nombre total de paiements sachant qu'un certain nombre d'opérations ont déjà été effectuées, ce qui est nécessaire dans le cadre des sinistres de type RBNS.

Nous insistons sur le fait que, à l'exception du nombre de sinistres qui dépend de l'année d'occurrence  $i$ , les autres lois qui viennent d'être présentées sont identiques pour toutes les années d'accident et pour tous les sinistres. Si nous possédons suffisamment de données et si une disparité est observée, l'exercice peut être effectuée à un niveau de granularité plus important.

Sachant  $U_{i,k}^{(C)} = u_{i,k}^{(C)}$ , le vecteur de développement est donc le suivant

$$\Lambda_i^{(k)} = \left( \ln C_{i,1,k}, \ln \lambda_1^{(k)}, \dots, \ln \lambda_{u_{i,k}^{(C)}}^{(k)} \right)^t.$$

Une première approche, assez usuelle, consiste à supposer que ce vecteur suit une loi normale multivariée avec moyenne et matrice de covariance définies par

$$\mathbb{E}(\Lambda_i^{(k)}) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_{u_{i,k}^{(C)}+1} \end{pmatrix} \text{ et } \text{Var}(\Lambda_i^{(k)}) = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,u_{i,k}^{(C)}+1} \\ \vdots & & & \vdots \\ \sigma_{u_{i,k}^{(C)}+1,1} & \sigma_{u_{i,k}^{(C)}+1,2} & \dots & \sigma_{u_{i,k}^{(C)}+1,u_{i,k}^{(C)}+1} \end{pmatrix}. \quad (3.14)$$

Dès lors, seule la dimension du vecteur  $\Lambda_i^{(k)}$  définit sa loi. En effet, si deux sinistres ont un vecteur de développement de même taille, nous supposons ici qu'ils possèdent *a priori* la même distribution.

D'une part, grâce à l'écriture (3.11), nous introduisons une corrélation entre les différents paiements d'un même sinistre.

D'autre part, nous pouvons imaginer que la corrélation au sein du vecteur de développement va varier avec le nombre de paiements, sa taille donc. En effet, pour un sinistre important, on pourrait s'attendre à une structure de corrélation entre les paiements différente de celle d'un sinistre peu coûteux. Cependant, avec cette approche, la matrice de corrélation va à chaque fois être construite sur la même base. Par exemple, si nous regardons les matrices de corrélations pour des vecteurs de

développement de taille 5 et 6, la matrice  $5 \times 5$  située dans le coin supérieur gauche de la matrice de corrélation pour le vecteur de développement de taille 6 est identique à la matrice de corrélation pour le vecteur de développement de longueur 5. Par conséquent, connaître la loi d'un vecteur de développement nous donne celle des vecteurs de longueurs inférieures.

De plus, cette hypothèse facilite les calculs et permet de définir directement la loi des vecteurs de développement pour les sinistres RBNS. En effet, conditionnellement à une partie du vecteur, l'hypothèse de normalité reste de mise, comme nous l'avons déjà cité à l'équation (3.6). Dès lors, il reste à déterminer les paramètres de cette loi normale, compte tenu de la trajectoire passée (voir Pigeon *et al.* [18]). Notons qu'ici, la trajectoire passée aura un impact dans l'évaluation des paramètres. Par conséquent, la loi ne dépend plus uniquement de la longueur du vecteur de développement.

Cependant, cette hypothèse de normalité a une limite importante. Elle ne permet pas de capter une possible asymétrie présente dans les données. Dès lors, comme proposée dans les articles mentionnés ci-dessus, une loi normale multivariée avec asymétrie, notée MSN, peut être utilisée pour caractériser la loi de  $\Lambda_i^{(k)}$ . Excepté l'asymétrie, les avantages et inconvénients exposés dans le cadre d'une loi normale restent également valables pour cette distribution. Nous renvoyons le lecteur intéressé par les propriétés d'une telle distribution vers [18, 19] et les références qui y sont citées.

• Dépendance entre variables : Comme nous l'avons vu, nous pouvons introduire une corrélation entre les différents facteurs de développement, et donc entre les paiements, pour un même sinistre. La forme de la matrice de covariance peut être spécifiée en fonction des données. Ainsi, plusieurs possibilités sont ouvertes :

- Si on suppose que cette matrice est diagonale, cela revient à admettre que les facteurs de développement sont indépendants, impliquant l'indépendance entre les paiements ;
- On pourrait uniquement s'intéresser à la corrélation entre le premier paiement et les facteurs de développement, sans dépendance intra-facteurs ;
- etc...

Finalement, nous admettons que les sinistres sont indépendants entre-eux et indépendants du nombre d'occurrences.

• Estimation des paramètres : Premièrement, pour le délai de déclaration  $T_{i,k}$ , nous faisons varier la valeur de  $m$  et estimons les différents paramètres par maximum de vraisemblance. On sélectionne ensuite le modèle le plus adéquat en accord avec le critère d'information bayésien (BIC), par exemple. Si la distribution n'est pas spécifiée, nous faisons également varier les différentes lois admissibles. Nous connaissons maintenant la fonction de répartition  $F_{T_{i,k}}(t; \hat{\nu})$ . En procédant de la même manière, nous pouvons calibrer les paramètres relatifs au nombre de paiements  $U_{i,k}^{(C)}$  et aux nombres de périodes entre deux paiements consécutifs, si requis.

Deuxièmement, pour le nombre de sinistres IBNR,  $K_{i,n-i+1}^{IBNR}$ , il reste à estimer le paramètre  $\theta$ . Cela se fait assez facilement via maximum de vraisemblance également.

Nous devons maintenant préciser les paramètres des lois normales multivariées avec asymétrie. Par ce qui a été dit avant, si l'on détermine la loi du plus grand vecteur de développement, nous connaissons celle de tous les autres. En effet, pour les paramètres relatifs aux autres sinistres, il suffira de prendre des sous-vecteurs et sous-matrices de ceux estimés. Dès lors, sur base d'une étude descriptive, nous pouvons extraire la longueur maximale observée pour les vecteurs de développement. Nous pouvons ensuite proposer plusieurs formes pour la matrice de corrélation et calibrer les différents paramètres.

Cette étape est loin d'être évidente et se base principalement sur une approche itérative du *maximum product of spacings method*. Nous renvoyons le lecteur intéressé par une introduction vers Pigeon *et al.* [19] et vers Akdemir [1] pour un développement plus complet. Finalement, nous retenons le modèle le plus adapté sur base du BIC, par exemple.

• Déroulé individuel de cash-flows : Supposons que les différents paramètres ont été estimés. Cette méthode nous permet d'obtenir un déroulé individuel des cash-flows. Pour ce faire, nous reprenons la classification IBNR/RBNP/RBNS. Pour une année d'occurrence  $i$  et une année calendrier  $n$ , pour les IBNR, nous devons :

- Simuler le nombre de sinistres  $K_{i,n-i+1}^{IBNR}$  ;
- Pour chaque sinistre, simuler le nombre total de paiements  $U_{i,k}^{(C)} + 1$  ;
- Pour chaque sinistre, sachant la réalisation de  $U_{i,k}^{(C)}$ , simuler une réalisation du vecteur de développement  $\Lambda_i^{(k)}$  ;
- Pour chaque sinistre, si le timing des paiements est requis, nous pouvons également le simuler.

Pour les sinistres de type RBNP, nous connaissons le nombre de sinistres mais ne savons rien à propos de leurs trajectoires. Dès lors, nous reprenons à partir de l'étape 2 de l'énumération ci-dessus.

Pour les RBNS, nous en connaissons le nombre et une partie du vecteur de développement. Pour un certain sinistre, supposons qu'il y ait déjà eu  $u^* + 1$  paiements, nous devons alors simuler :

- Le nombre total de paiements  $U_{i,k}^{(C)} + 1$  sachant  $U_{i,k}^{(C)} \geq u^*$  ;
- Supposons que le nombre total de paiements soit égal à  $u_{i,k}^{(C)} + 1 \geq u^* + 1$ , il nous reste à simuler la fin du vecteur de développement, c'est-à-dire

$$\left( \lambda_{u^*+1}^{(k)}, \dots, \lambda_{u_{i,k}^{(C)}}^{(k)} \right) \left| \left( C_{i,1,k}, \lambda_1^{(k)}, \dots, \lambda_{u^*}^{(k)} \right) .\right.$$

Par ce qui a été dit, ce dernier suivra une loi normale multivariée avec asymétrie ;

- Si le timing des paiements est requis, simuler le nombre de périodes entre deux paiements consécutifs.

Pour l'année calendrier  $n$ , nous pouvons ainsi simuler les trajectoires pour l'ensemble des sinistres survenus l'année d'occurrence  $i$ .

• Limitations pour l'évaluation des réserves et du One-year risk : Une évaluation de type best-estimate des réserves est envisageable sans une réelle complexité algorithmique. En effet, sachant le nombre de paiements (restants), Pigeon *et al.* [19] nous donne une forme analytique du paiement cumulé moyen au terme. Par hypothèse, nous savons que les sinistres sont indépendants entre-eux ainsi que du nombre d'occurrences. De plus, les sinistres possédant un vecteur de développement de même longueur sont identiquement distribués. Par conséquent, en utilisant la loi des espérances itérées, pour les sinistres IBNR, la valeur attendue pour les paiements totaux à l'année calendrier  $n$  est don-



née par

$$\begin{aligned}
\sum_{i=2}^n \mathbb{E} \left( \sum_{k=1}^{K_{i,n-i+1}^{IBNR}} C_{i,U_{i,k}^{(C)}+1,k} \right) &= \sum_{i=2}^n \mathbb{E}_{U_{i,k}^{(C)}} \left[ \mathbb{E} \left( \sum_{k=1}^{K_{i,n-i+1}^{IBNR}} C_{i,U_{i,k}^{(C)}+1,k} \mid U_{i,k}^{(C)} = u_{i,k}^{(C)} \right) \right] \\
&\stackrel{1+i.d.}{=} \sum_{i=2}^n \mathbb{E}_{U_{i,1}^{(C)}} \left[ \mathbb{E} \left( K_{i,n-i+1}^{IBNR} \mid U_{i,1}^{(C)} = u_{i,1}^{(C)} \right) \mathbb{E} \left( C_{i,U_{i,1}^{(C)}+1,1} \mid U_{i,1}^{(C)} = u_{i,1}^{(C)} \right) \right] \\
&\stackrel{1+i.d.}{=} \left( \sum_{i=2}^n \mathbb{E} \left( K_{i,n-i+1}^{IBNR} \right) \right) \mathbb{E}_{U_{i,1}^{(C)}} \left[ \mathbb{E} \left( C_{1,U_{1,1}^{(C)}+1,1} \mid U_{1,1}^{(C)} = u_{1,1}^{(C)} \right) \right],
\end{aligned}$$

où une forme analytique de l'espérance entre crochets de la dernière égalité est donnée par [19, Prop. 1].

Pour les sinistres de type RBNP, le nombre d'occurrences est observé. Il suffit donc de remplacer l'espérance sur la variable aléatoire  $K_{i,n-i+1}^{IBNR}$  ci-dessus par l'observation  $K_{i,n-i+1}^{RBNP}$ .

Finalement, pour les sinistres de type RBNS, nous devons tenir compte de l'historique de chacun d'entre-eux jusqu'à l'instant d'évaluation, notons-le  $\mathcal{D}_{n,k}$ . Rappelons que le nombre de sinistres est ici observé et nous avons maintenant

$$\sum_{i=2}^n \sum_{k=1}^{K_{i,n-i+1}^{RBNS}} \mathbb{E} \left( C_{i,U_{i,k}^{(C)}+1,k} \mid \mathcal{D}_{n,k} \right) = \sum_{i=2}^n \sum_{k=1}^{K_{i,n-i+1}^{RBNS}} \mathbb{E}_{U_{i,k}^{(C)}} \left[ \mathbb{E} \left( C_{i,U_{i,k}^{(C)}+1,k} \mid \mathcal{D}_{n,k}, U_{i,k}^{(C)} = u_{i,k}^{(C)} \right) \right], \quad (3.15)$$

où une forme analytique de l'espérance entre crochets est donnée par [19, Prop. 2]. Par conséquent, le best-estimate des réserves RBNS, peut être estimé en retranchant à la valeur calculée (3.15) le dernier paiement cumulé observé.

Si l'on souhaite estimer la loi des réserves, vu la dépendance entre les paiements pour un même sinistre, nous simulons un grand nombre de trajectoires pour tous les sinistres, comme nous l'avons expliqué ci-dessus, selon la classification IBNR/RBNP/RBNS. Dès lors, la complexité algorithmique devient très importante.

Par le même argument, l'évaluation du *One-year risk* est encore plus coûteuse en temps de calcul. En effet, nous simulons toutes les trajectoires pour chaque sinistre. Afin de savoir s'il y a eu un paiement durant la première année après l'année calendrier considérée, nous devons également simuler le délais entre le dernier paiement observé et le premier paiement futur, qui vient d'être simulé. Ensuite, nous pouvons, à nouveau, simuler les réserves avec cette information supplémentaire.

### 3.2.2 Approche de Larsen

Nous terminons les méthodes individuelles pures en présentant le modèle introduit par Larsen [11].

La première partie de l'article fait appel à la théorie des *Marked Poisson Process* (MPP), telle que définie par Norberg [16]. De plus, celle-ci est abstraite et requiert très peu d'hypothèses. En effet, le modèle général qui est introduit ne définit, par exemple, aucune loi et ne permet donc pas d'effectuer des simulations. Des hypothèses plus précises doivent être stipulées pour modéliser les processus et nous renvoyons le lecteur intéressé par le modèle au sens strict vers [11, Sections 1-3]. Par conséquent, nous nous focalisons sur une certaine spécification de ce modèle général, également introduite

par Larsen [11]. Cette dernière est principalement basée sur des GLM, permettant ainsi d'incorporer l'information individuelle durant la modélisation des processus.

Nous soulignons que la méthode proposée dans Larsen [11] est appliquée aux charges incrémentales mais, comme suggéré dans cet article, elle peut s'appliquer aux paiements incrémentaux, ce que nous faisons dans cette note. De plus, nous modélisons ici tous les paiements incrémentaux depuis l'occurrence jusqu'à l'année de développement  $n$ . Nous devons donc prendre en compte les années où aucun paiement n'est effectué. Par conséquent, une masse de probabilité en 0 devra être insérée. En outre, l'approche proposée par Larsen [11] permet de modéliser différemment les coûts en fonction des données observées. Ainsi, en fonction des données disponibles à la date d'évaluation, un paiement non-nul aura plus ou moins de chance d'être calibré à l'aide d'une loi à queue fine ou d'une loi à queue épaisse. Cette faculté permet notamment de mieux modéliser les frais importants, pour lesquels une loi à queue fine n'est pas adaptée.

Dans ce contexte, nous regardons l'évolution des sinistres cellule après cellule et nous analysons la fréquence pour chacune des années d'occurrence et de déclaration. Supposons que nous sommes à l'année calendrier  $n$ , sur base de la classification IBNR/RBNS/RBNP, nous avons besoin, pour les IBNR, de connaître :

- Le nombre de sinistres survenus l'année d'accident  $i$  et reportés après  $t$  années, c'est-à-dire  $K_{i,t}$  avec  $n \geq t > n - i + 1$ . Une nouvelle fois, soulignons que la condition  $t > n - i + 1$  correspond aux années de développement qui n'ont pas encore été observées pour l'année d'accident  $i$  ;
- Vu que nous analysons l'ensemble des paiements pour tous les sinistres, pour chaque année de développement à partir de leurs déclarations jusqu'à l'année  $n$ , les valeurs  $P_{i,t+l,k}$  pour chaque sinistre  $k$ , survenu l'année d'occurrence  $i$ , déclaré l'année  $t$  et évalué à l'année de développement  $t + l$ , avec  $i = 1, \dots, n$ ,  $t > n - i + 1$ ,  $k = 1, \dots, K_{i,t}$  et  $(n - t) \geq l \geq 0$ . Notez qu'ici  $l$  compte les années de développement *après la déclaration*.

Contrairement à l'ILR, introduit dans la Section 3.2.1, nous ne modélisons plus le nombre total de sinistres de type IBNR lors de l'évaluation mais les fréquences pour toutes les cellules futures. Cela augmente donc légèrement la complexité algorithmique lors des simulations. Cependant, cette approche est importante car nous modélisons ici les paiements incrémentaux dans chacune des cellules depuis la déclaration jusqu'à l'année de développement  $n$ . Dès lors, la connaissance exacte de l'instant de notification est primordiale.

Pour les sinistres de type RBNP, les fréquences sont observées mais aucun historique de paiements existe. En effet, au contraire des IBNR, nous connaissons les instants de déclaration  $t \leq n - i + 1$ , avec  $i = 1, \dots, n$  et nous savons qu'aucun paiement n'a été effectué. Nous devons uniquement déterminer les paiements ultérieurs à la dernière diagonale observée (correspondant à l'année calendrier  $n$ , donc), c'est-à-dire  $P_{i,t+l,k}$  pour  $n \geq t + l > n - i + 1$ ,  $i = 1, \dots, n$ .

Pour les sinistres de type RBNS, nous connaissons également les instants de déclaration  $t \leq n - i + 1$ , avec  $i = 1, \dots, n$  et un historique de paiements est disponible. Il reste à déterminer les paiements ultérieurs à la dernière diagonale observée, c'est-à-dire  $P_{i,t+l,k}$  pour  $n \geq t + l > n - i + 1$ ,  $i = 1, \dots, n$  compte tenu des paiements déjà effectués.

Notez que l'information ci-dessus ne nous spécifie pas quand le sinistre est liquidé. En effet, nous pouvons observer une succession de paiements nuls sans pour autant connaître l'instant où le sinistre est clôturé. Nous pouvons, si nous le souhaitons, requérir ce délai pour chaque sinistre. Pour ce faire, l'idée introduite par Larsen [11] consiste à modéliser la liquidation à chaque période, depuis la déclai-

ration jusqu'à l'année de développement  $n$ , à l'aide d'une variable binaire. Nous ne nous épanchons pas plus sur cette extension et nous renvoyons le lecteur intéressé vers l'article en question.

Finalement, comme nous le verrons, cette méthode permet d'introduire une structure de dépendance entre les paiements d'un même sinistre. Nous débutons maintenant le schéma de comparaison usuel.

- Données nécessaires à la calibration : Pour cette méthode, les instants ainsi que les montants des paiements individuels sont nécessaires. De plus, les occurrences et les délais de déclaration sont également requis.

Vu que le nombre maximum de développements est fixé à  $n$  et que nous modélisons tous les paiements, nous ne demandons aucune information à propos des délais de liquidation. Cependant, si l'actuaire est intéressé par la trajectoire exacte du sinistre, cette dernière est également souhaitée.

Comme mentionné, nous travaillons ici avec les paiements individuels sous forme incrémentale.

- Hypothèses sur les distributions : Soient  $i$  une année d'accident,  $k$  un sinistre survenu durant cette année et déclaré après  $t$  années. La première hypothèse formulée se fait au niveau de la dépendance entre les paiements pour un même sinistre. Larsen [11] suppose que le paiement effectué lors de  $(t + l)^{\text{ème}}$  année de développement, avec  $l \geq 0$ , ne dépend des paiements passés qu'à travers une fonction de ceux-ci, c'est-à-dire

$$P_{i,t+l,k} | P_{i,t+l-1,k}, \dots, P_{i,t,k} \stackrel{\text{Loi}}{=} P_{i,t+l,k} | h(P_{i,t+l-1,k}, \dots, P_{i,t,k}). \quad (3.16)$$

Il s'agit de l'hypothèse la plus forte formulée dans l'article (voir [11, Eq. (3.7)]). Celle-ci permettra néanmoins une inclusion plus aisée de l'effet des paiements passés dans les GLM. Notons que la fonction  $h$  reste à être spécifiée. Une approche intuitive consiste à supposer que le prochain paiement dépend du paiement total déjà effectué, ce qui revient à définir

$$S_{i,t+l-1,k} := h(P_{i,t+l-1,k}, \dots, P_{i,t,k}) = P_{i,t+l-1,k} + \dots + P_{i,t,k}. \quad (3.17)$$

Dans un premier temps, intéressons-nous à la modélisation du nombre de sinistres survenus l'année d'occurrence  $i$  et notifiés après  $t$  années,  $K_{i,t}$ . En suivant l'article, nous supposons que ces variables suivent une loi de Poisson de moyenne de la forme suivante

$$w(i, t) c \alpha_i \beta_t, \quad (3.18)$$

où  $w(i, t)$  est une mesure d'exposition pour ce type de sinistres et  $c$  est une constante indépendante de l'année d'occurrence  $i$  et du délai de déclaration  $t$ . Notez qu'une autre loi de comptage, telle que la Binomiale Négative, par exemple, peut également être envisagée.

Contrairement à la méthode ILR présentée, nous ne modélisons pas directement le vecteur des paiements pour un sinistre mais nous modélisons, de manière *indépendante*, les lois conditionnelles, tout en imposant une structure de dépendance au sein de ce vecteur. En effet, par le théorème des probabilités totales, nous savons que

$$\begin{aligned} \mathbb{P}(P_{i,t,k}, \dots, P_{i,n,k}) &= \mathbb{P}(P_{i,n,k} | P_{i,n-1,k}, \dots, P_{i,t,k}) \mathbb{P}(P_{i,n-1,k}, \dots, P_{i,t,k}) \\ &= \mathbb{P}(P_{i,n,k} | P_{i,n-1,k}, \dots, P_{i,t,k}) \mathbb{P}(P_{i,n-1,k} | P_{i,n-2,k}, \dots, P_{i,t,k}) \dots \mathbb{P}(P_{i,t,k}) \\ &\stackrel{(3.16)+(3.17)}{=} \mathbb{P}(P_{i,n,k} | S_{i,t+(n-t-1),k}) \mathbb{P}(P_{i,n-1,k} | S_{i,t+(n-t-2),k}) \dots \mathbb{P}(P_{i,t,k}). \end{aligned} \quad (3.19)$$

Dès lors, au niveau des paiements, en repensant à la classification IBNR/RBNP/RBNS, il est suffisant de fixer une loi pour :

- Le premier paiement  $P_{i,t,k}$ , effectué à la déclaration ;
- Le deuxième paiement connaissant le premier,  $P_{i,t+1,k}|P_{i,t,k}$
- Le troisième paiement connaissant la somme des deux premiers,  $P_{i,t+2,k}|S_{i,t+1,k}$  grâce aux hypothèses (3.16) et (3.17)
- ...
- Le dernier paiement connaissant la somme de tout l'historique,  $P_{i,t+(n-t),k}|S_{i,t+(n-t-1),k}$  grâce aux hypothèses (3.16) et (3.17).

Pour la modélisation du premier paiement  $P_{i,t,k}$ , nous utilisons une distribution mixte avec trois composantes, définie comme suit :

- Supposons que la séparation pour la distribution mixte se fasse au seuil  $\kappa > 0$ , nous fixons

$$\pi_{i,t} = \mathbb{P}(P_{i,t,k} > 0) \text{ et } \tilde{\pi}_{i,t} = \mathbb{P}(P_{i,t,k} > \kappa | P_{i,t,k} > 0).$$

Premièrement, nous obtenons une masse de probabilité en 0 définie par  $1 - \pi_{i,t}$ . Nous supposons que  $\pi_{i,t}$  (resp.  $\tilde{\pi}_{i,t}$ ) est modélisée de la manière suivante

$$\frac{\pi_{i,t}}{1 - \pi_{i,t}} = c\alpha_i\beta_t;$$

- Une composante à queue fine avec probabilité  $\mathbb{P}(0 < P_{i,t,k} \leq \kappa) = (1 - \tilde{\pi}_{i,t})\pi_{i,t}$  (e.g. Gamma, Inverse Gaussienne, ...). En suivant l'article, nous supposons que la queue fine est modélisée par une loi Gamma de moyenne

$$c\alpha_i\beta_t;$$

- Une composante à queue épaisse avec probabilité  $\mathbb{P}(P_{i,t,k} > \kappa) = \pi_{i,t}\tilde{\pi}_{i,t}$  (e.g. Pareto type 2, GPD, ...). Cette fois, nous supposons que la queue épaisse est modélisée par une GPD de seuil  $\kappa > 0$  et indépendamment des variables explicatives.

Cela revient donc à différencier la modélisation du premier paiement en fonction qu'il soit nul, compris entre 0 et  $\kappa$  ou supérieur à  $\kappa$ .

Nous souhaitons maintenant modéliser le  $(l + 1)^{\text{ème}}$  paiement,  $P_{i,t+l,k}$ , sachant  $S_{i,t+l-1,k}$ , avec  $l = 1, \dots, n - t$ . Deux grandes familles peuvent directement être distinguées. D'une part, si aucun paiement n'a été effectué jusqu'à présent, c'est-à-dire  $S_{i,t+l-1,k} = 0$ , nous pouvons déterminer la loi de  $P_{i,t+l,k}|S_{i,t+l-1,k}$  comme nous l'avons fait pour le premier paiement. Si nous le souhaitons, nous pouvons également incorporer une covariable relative à l'année de développement après la déclaration. D'autre part, si un ou plusieurs paiements ont été effectués, c'est-à-dire  $S_{i,t+l-1,k} > 0$ , afin d'utiliser cette variable résumant l'information passée, nous utilisons une discrétisation. Soient  $0 = x_0 < x_1 < \dots < x_h$  une discrétisation de  $\mathbb{R}$  pour une valeur  $x_h$  sélectionnée *a priori* et suffisamment grande. Nous notons

$$SG_{i,t+l-1,k} = \min \{x_s : S_{i,t+l-1,k} \leq x_s\}, \quad (3.20)$$

c'est-à-dire la borne supérieure de l'intervalle auquel  $S_{i,t+l-1,k}$  appartient. L'idée est également similaire à celle introduite pour la modélisation du premier paiement, en considérant cette fois l'information supplémentaire apportée, d'une part, par le nombre d'années de développement après la

notification et, d'autre part, par la variable  $SG$  symbolisant les paiements passés. Ces deux effets seront donc aussi utilisés pour la modélisation des probabilités et pour celle de la composante à queue fine. Notons que le seuil  $\kappa$  de la composante à queue épaisse reste commun à toutes les modélisations.

Nous insistons sur le fait que cette approche permet de mieux traiter les coûts relatifs aux différents sinistres. Ainsi, pour chaque sinistre, en fonction de nombreux critères, tels que l'année de développement après la notification ou l'historique de paiements, nous donnons entre autre plus ou moins de poids aux possibilités d'avoir un paiement nul, faible ou important.

• Dépendance entre variables : Compte tenu des spécifications proposées pour les lois, deux éléments importants ressortent :

- L'effet du passé impacte explicitement la loi du paiement futur, à travers la discrétisation de  $S_{i,t+l-1,k}$ . Ainsi, en fonction des paiements passés, nous pouvons
  - Définir les probabilités d'avoir un paiement nul, modéré ou important ;
  - Si le paiement est modéré, modéliser la moyenne de la loi Gamma.

Néanmoins, cette dépendance entre les paiements peut s'avérer plus faible que celle introduite dans l'ILR, présenté à la Section 3.2.1.

- Nous modélisons les moyennes (ou probabilités) comme un produit d'effets individuels. On ne considère donc aucune dépendance entre les variables. Cependant, une telle faculté pourrait être nécessaire pour introduire, par exemple, une relation entre l'année d'occurrence et le délai de déclaration, ou encore, entre le délai de déclaration et l'année de développement après celui-ci. Néanmoins, s'ils sont souhaités, de tels besoins peuvent facilement être inclus.

Nous supposons également que les fréquences  $K_{i,t}$  sont indépendantes entre-elles et indépendantes des coûts.

• Estimation des paramètres : Nous distinguons les différentes modélisations abordées ci-dessus et expliquons la démarche proposée par Larsen [11]. D'une part, pour les fréquences  $K_{i,t}$ , il s'agit d'une régression de Poisson, qui peut être effectuée via un GLM. D'autre part, pour les probabilités  $\pi$  et  $\tilde{\pi}$ , nous effectuons des régressions logistiques. Notons qu'une approche via régression logistique *multinomiale* aurait également pu être utilisée. Ensuite, pour la composante à queue fine, l'utilisation d'un GLM semble être indiquée. Par conséquent, les paramètres de ces spécifications peuvent être estimés via maximum de vraisemblance.

Finalement, pour la loi de Pareto, il reste à déterminer le seuil commun  $\kappa$  ainsi qu'une estimation des paramètres. Pour fixer ce seuil, une manière de faire repose sur le fait que la *mean excess function* d'une loi Pareto généralisée est approximativement linéaire à partir du seuil recherché. Ce dernier aurait également pu être déterminé via les estimateurs de Hill. Les autres paramètres de cette loi peuvent, quant à eux, être estimés via maximum de vraisemblance. Pour plus de détails techniques, nous renvoyons le lecteur intéressé vers McNeil *et al.* [13].

• Déroulé individuel de cash-flows : Supposons que les différents paramètres ont été estimés. Cette méthode nous permet également d'obtenir un déroulé individuel de cash-flows. Nous reprenons la classification IBNR/RBNP/RBNS introduite précédemment. Pour une année d'occurrence  $i > 1$  et une année calendrier  $n$  fixées, pour les sinistres de type IBNR, pour chaque  $n \geq t > n - i + 1$ , nous devons :

- Simuler le nombre de sinistres qui n'ont pas encore été déclarés à l'année de développement  $t$ , c'est-à-dire  $K_{i,t}$  ;

- Pour chaque sinistre  $k$ , simuler le premier paiement,  $P_{i,t,k}$  ;
- Pour chaque sinistre  $k$ , simuler les paiements futurs compte tenu des simulations passées, c'est-à-dire  $P_{i,t+l,k} | S_{i,t+l-1,k} = s_{i,t+l-1,k}$  avec  $n-t \geq l \geq 1$  et où la valeur  $s_{i,t+l-1,k}$  résulte des valeurs précédemment simulées.

Pour les RBNS, nous connaissons leurs délais de déclarations (et donc le nombre total) ainsi qu'un certain historique des paiements. Dès lors, pour chaque sinistre  $k$ , déclaré à l'année  $1 \leq t \leq n-i+1$ , notons la somme observée des paiements passés  $s_{i,n-i+1,k}$ , nous devons :

- Simuler le prochain paiement, compte tenu des observations effectuées, ce qui revient à simuler  $P_{i,t+(n-i+2-t),k} | S_{i,t+(n-i+1-t),k} = s_{i,n-i+1,k}$  ;
- Simuler les paiements futurs compte tenu des observations et simulations passées, c'est-à-dire  $P_{i,t+l,k} | S_{i,t+l-1,k} = s_{i,t+l-1,k}$  avec  $n \geq t+l > n-i+2$  et où la valeur  $s_{i,t+l-1,k}$  est une réalisation de  $S_{i,t+l-1,k} = s_{i,n-i+1,k} + P_{i,n-i+2,k} + \dots + P_{i,t+l-1,k}$ .

Finalement, pour les sinistres de type RBNP, nous connaissons leurs délais de déclaration  $1 \leq t \leq n-i+1$  (et donc le nombre total) lors de l'année calendrier  $n$ . Pour chaque sinistre  $k$ , déclaré l'année  $t \leq n-i+1$ , nous savons également qu'aucun paiement n'a été effectué. Nous pouvons alors reprendre le schéma introduit pour les RBNS, en fixant  $s_{i,n-i+1,k}$  à 0.

• Limitations pour l'évaluation des réserves et du *One-year risk* : Nous formulons maintenant les réserves et reprenons la classification IBNR/RBNP/RBNS usuelle.

D'une part, la valeur des réserves IBNR lors de l'année calendrier  $n$ , en supposant que tous les sinistres soient liquidés avant l'année de développement  $n$  comme mentionné précédemment, est donnée par

$$IBNR = \sum_{i=2}^n \sum_{t=n-i+2}^n \sum_{k=1}^{K_{i,t}} \sum_{l=0}^{n-t} P_{i,t+l,k}. \quad (3.21)$$

En d'autres termes, pour toutes les années d'occurrence, nous calculons le nombre de sinistres qui n'ont pas encore été déclarés et nous sommes leurs paiements, depuis leur déclaration jusqu'à l'année de développement  $n$ .

D'autre part, la valeur des réserves RBNP/RBNS à la fin de l'année calendrier  $n$ , est donnée par

$$RBNP + RBNS = \sum_{i=2}^n \sum_{t=1}^{n-i+1} \sum_{k=1}^{K_{i,t}} \sum_{l=n-i+2-t}^{n-t} P_{i,t+l,k}, \quad (3.22)$$

où le nombre de sinistres survenus l'année  $i$  et reportés après  $t$  années  $K_{i,t}$  est cette fois observé. Nous sommes, jusqu'à l'année de développement  $n$ , l'ensemble des paiements restants pour tous les sinistres qui ont été déclarés avant l'année calendrier  $n$ .

Pour obtenir une réalisation de la réserve totale, nous devons donc simuler l'entièreté des paiements ultérieurs pour chaque sinistre, impliquant une complexité algorithmique élevée. Comme énoncé dans la section réservée à l'ILR 3.2.1, l'évaluation du *One-year risk* est encore plus coûteuse en temps de calcul. Finalement, grâce au théorème des espérances itérées, une forme analytique, que nous ne détaillons pas ici, pourrait être dégagée en cas d'évaluation du best-estimate des réserves.

### 3.3 Méthodes collectives : CLR

On regarde maintenant les méthodes collectives. Nous nous trouvons donc dans la partie située à l'extrême droite de l'arbre représenté à la Figure 1.1.

Ces méthodes ne nous permettent pas de suivre la trajectoire d'un sinistre particulier. Dans ce cadre, nous allons modéliser l'ensemble des paiements effectués dans la cellule  $(i, j)$ , sans faire de distinction entre les différents dossiers ouverts. On conçoit donc bien l'aspect collectif du modèle. Cependant, des données individuelles sont utilisées afin de spécifier au mieux les différentes lois de probabilité.

Comme nous l'avons cité dans la section 3.1.3, l'approche de Tweedie pour  $p \in ]1, 2[$  équivaut à supposer que la somme des paiements dans chacune des cellules suit une loi de Poisson-Gamma composée. Cela revient à dire que le nombre de paiements suit une loi de Poisson, que les paiements individuels suivent une loi Gamma et que toutes les variables aléatoires sont indépendantes.

D'une part, admettre que la distribution des paiements suit une loi Gamma est contraignant. En effet, la loi Gamma possède une queue fine et ne permet pas de modéliser correctement les paiements élevés, ce qui pourrait, par exemple, être nécessaire lors de dommages corporels.

D'autre part, cette approche implique que l'impact des variables explicatives sur la fréquence et sur la sévérité évolue dans le même sens, ce qui, en pratique, n'est pas toujours le cas. Par exemple, nous observons qu'un assuré résidant dans un milieu urbain aura une fréquence plus élevée mais une sévérité plus faible (*a contrario* pour un assuré habitant dans une région rurale). Dès lors, les effets sont contradictoires, ce qui n'est pas en accord avec le modèle de Tweedie.

Afin de résoudre ces problèmes, une approche proposée dans Denuit et Trufin [6] consiste à utiliser des *lois mixtes* pour modéliser les paiements individuels. Dès lors, un paiement est expliqué par une loi à queue fine (e.g. Gamma) avec une certaine probabilité  $\pi$  ou par une loi à queue épaisse (e.g. Pareto) avec probabilité  $1 - \pi$ . Dans ce contexte, une information au niveau agrégé pour les paiements ne suffit plus. En effet, nous nécessitons les données individuelles afin d'évaluer les différentes composantes de cette loi mixte.

Maintenant que l'idée principale du modèle est introduite, nous effectuons le schéma de comparaison basé sur l'article Denuit et Trufin [6].

- Données nécessaires à la calibration : Pour modéliser les coûts, les paiements individuels sont indispensables. On peut déduire de cette information le nombre de paiements effectifs (les paiements  $> 0$ , donc) dans chacune des cellules  $(i, j)$ .

Nous travaillons ici avec les paiements effectifs sous la forme incrémentale.

- Hypothèses sur les distributions : Premièrement, nous supposons que les paiements incrémentaux individuels, dans chaque cellule  $(i, j)$ , sont identiquement distribués. Ce qui signifie que

$$P_{i,j,1}, P_{i,j,2}, P_{i,j,3}, \dots, \text{ pour } i = 1, \dots, n \text{ et } j = 1, \dots, n,$$

suivent la même loi.

Ensuite, conformément à l'explication précédente, nous supposons que la totalité des paiements pour les sinistres d'année d'occurrence  $i$  à l'année de développement  $j$ , c'est-à-dire dans la cellule  $(i, j)$ ,

est modélisée de manière collective. Cela revient à écrire

$$P_{i,j} = \sum_{k=1}^{N_{i,j}} P_{i,j,k}, \quad (3.23)$$

où  $N_{i,j}$  est le nombre de paiements effectifs dans la cellule  $(i, j)$ . Contrairement à l'ILR, présenté à la section 3.2.1, où l'indice  $j$  correspondait au  $j^{\text{ème}}$  paiement, ce dernier fait ici référence à l'année de développement. Nous analysons cette fois les paiements totaux, *cellule après cellule*. Il reste donc à spécifier et ajuster les lois des fréquences et des paiements incrémentaux individuels.

Pour la fréquence  $N_{i,j}$ , afin de travailler avec des lois de type Poisson-composée, nous admettons que  $N_{i,j}$  suit une loi de Poisson de moyenne

$$\mathbb{E}(N_{i,j}) = \alpha_i \beta_j, \quad (3.24)$$

avec la contrainte d'identification des paramètres usuelle

$$\sum_{j=1}^n \beta_j = 1.$$

Nous supposons donc que la fréquence moyenne des paiements est expliquée par l'année d'occurrence  $i$  et l'année de développement  $j$ . Notons qu'une autre loi aurait pu être choisie, on pense notamment à la loi Binomiale Négative.

Nous devons maintenant préciser la loi des paiements. Comme énoncé précédemment, supposer que les paiements individuels suivent une loi Gamma nous renvoie vers le modèle de Tweedie présenté à la section 3.1.3 et ne nous permet pas, d'une part, de modéliser les frais engendrés par certains sinistres et, d'autre part, de prendre en compte de potentiels effets contraires des variables explicatives sur la fréquence et la sévérité. De plus, contrairement à l'approche de Larsen (voir Section 3.2.2), vu que nous travaillons avec des paiements effectifs, nous ne devons pas modéliser de masse de probabilité en 0. Nous utilisons donc une distribution mixte avec deux composantes, définie comme suit

- Une composante à queue fine avec probabilité  $1 - \pi_j$  (e.g. Gamma, Inverse Gaussienne, ...) de moyenne

$$\delta_{1,j}(1 + g_1)^{i+j-2}. \quad (3.25)$$

- Une composante à queue épaisse avec probabilité  $\pi_j$  (e.g. Pareto type 2, GPD, ...) de moyenne

$$\delta_{2,j}(1 + g_2)^{i+j-2}. \quad (3.26)$$

Les variables  $g_1$  et  $g_2$  représentent les taux d'inflation, supposés constants, alors que  $\delta_{1,j}$  et  $\delta_{2,j}$  modélisent l'effet développement. Nous effectuons plusieurs remarques sur les deux lois continues :

- Grâce à cette modélisation, l'effet inflation est directement inclus dans le modèle ;
- En spécifiant l'inflation de la sorte, nous supposons, d'une part, que son impact est différent en fonction du type de paiement et, d'autre part, que l'année de référence utilisée pour cette valeur est la première année calendrier. En effet, pour cette première année,  $i = j = 1$  et l'effet inflation se résume à 1. On analyse donc, pour chacune des années calendriers, les déviations de cette valeur unitaire.



Notons que, dans ce contexte, la probabilité  $\pi_j$  dépend uniquement de l'année de développement. Une observation intéressante, notée par Denuit et Trufin [6], relève que le poids attribué à la composante à queue fine décroît avec l'année de développement  $j$  et donc, la probabilité  $\pi_j$  d'une composante à queue épaisse croît avec le lag  $j$ , jusqu'à atteindre 1. Nous pouvons expliquer cet effet par le fait que, lorsque peu de développements sont nécessaires, il s'agit principalement de petits sinistres, avec des paiements modérés. Au contraire, plus le sinistre va durer dans le temps, plus il aura de chance d'être important, avec des paiements relativement élevés, ce qui nécessite une loi à queue épaisse. Bien que l'article se limite à un effet développement, des informations individuelles pourraient également être incorporées dans la modélisation des  $\pi_j$ .

- Dépendance entre variables : Nous supposons ici que toutes les variables aléatoires sont indépendantes. Par conséquent, les paiements totaux obtenus dans chaque cellule  $P_{i,j}$  sont également indépendants. Notez que cette hypothèse est similaire à celle formulée dans l'approche de Tweedie présentée à la section 3.1.3 pour  $p \in ]1, 2[$ .

- Estimation des paramètres : En ce qui concerne la fréquence  $N_{i,j}$ , nous nous situons dans le domaine des GLM. Dès lors, une estimation via maximum de vraisemblance des différents paramètres est réalisable. Pour le modèle considéré, nous pouvons facilement estimer les valeurs  $\alpha_i$  et  $\beta_j$ .

La loi des paiements individuels dans la cellule  $(i, j)$  ne fait pas partie de la famille exponentielle linéaire et elle ne remplit donc pas les conditions pour effectuer un GLM. Pour ajuster une loi mixte avec composante Gamma et Pareto, par exemple, l'utilisation du package GAMLSS de R semble appropriée. Les GAMLSS sont des modèles de régression où plusieurs paramètres de la distribution supposée de la variable réponse (la loi de  $P_{i,j,k}$  dans notre cas) peuvent être modélisés comme des fonctions additives des variables explicatives. Cela contraste avec les GLM où uniquement la moyenne est modélisée comme une fonction additive des variables explicatives. De plus, ces modèles ne s'appliquent pas uniquement aux membres de la famille exponentielle linéaire. Certaines contraintes de dérivabilité sont néanmoins à souligner. Pour plus d'information sur cette méthode, nous renvoyons le lecteur intéressé vers Denuit et Trufin [6] et les références qui y sont citées. Grâce à cette approche, nous sommes capables d'estimer l'ensemble des paramètres relatifs à la distribution des paiements pour chaque cellule  $(i, j)$ , avec  $1 \leq i, j \leq n$ .

- Déroulé individuel de cash-flows : Dans ce contexte, nous ne suivons pas la trajectoire d'un sinistre particulier. Nous nous intéressons uniquement aux paiements totaux effectués dans chaque cellule  $(i, j)$ , sans identifier les paiements individuels.

- Limitations pour l'évaluation des réserves et du One-year risk : Supposons que les différents paramètres ont été estimés.

Si l'on s'intéresse uniquement au best-estimate des réserves pour les années d'occurrence  $i > 1$  et calendrier  $n$ , grâce à l'hypothèse d'indépendance entre les paiements et la fréquence et au fait que les paiements dans une même cellule sont identiquement distribués, en utilisant la définition de  $P_{i,j}$  introduite à l'Equation (3.23), nous avons

$$\mathbb{E} \left( \sum_{j=n+i-2}^n P_{i,j} \right) = \sum_{j=n+i-2}^n \mathbb{E} \left( \sum_{k=1}^{N_{i,j}} P_{i,j,k} \right) = \sum_{j=n+i-2}^n \mathbb{E}(N_{i,j}) \mathbb{E}(P_{i,j,1}).$$

Dès lors, grâce aux spécifications (3.24), (3.25) et (3.26), nous obtenons

$$\mathbb{E} \left( \sum_{j=n+i-2}^n P_{i,j} \right) = \sum_{j=n+i-2}^n \alpha_i \beta_j \left( (1 - \pi_j) \delta_{1,j} (1 + g_1)^{i+j-2} + \pi_j \delta_{2,j} (1 + g_2)^{i+j-2} \right).$$

Par conséquent, si les différents paramètres ont été préalablement calibrés, l'estimation du best-estimate des réserves est immédiat.

D'autre part, si l'on s'intéresse à la loi des réserves, une fois les paramètres estimés, nous pouvons procéder de deux manières :

- Chaque cellule suit une loi de Poisson-composée. De plus, ces dernières sont toutes indépendantes. Par conséquent, la somme des futurs paiements totaux, c'est-à-dire

$$\sum_{i=2}^n \sum_{j=n+i-2}^n P_{i,j}, \quad (3.27)$$

suit une loi de Poisson-composée et l'algorithme de Panjer peut être utilisé pour la déterminer ;

- Soit nous procédons par simulation. Pour ce faire, pour chacune des cellules à remplir, nous simulons le nombre de paiements  $N_{i,j}$ . Ensuite, pour chaque paiement et de manière indépendante, nous simulons son coût, à l'aide de la loi mixte. Ainsi, nous obtenons une réalisation du paiement total dans chacune des cellules  $(i, j)$  avec  $i + j - 2 \geq n$ . Nous pouvons finalement sommer les futurs paiements pour obtenir une réalisation de (3.27).

Finalement, dans le même ordre d'idée que pour les modèles collectifs, le *One-year risk* s'obtient assez facilement.

## Chapitre 4

# Discussion finale

Ce dernier chapitre a pour but de présenter certaines améliorations que l'on pourrait apporter aux différentes méthodes présentées. Les idées sont similaires à celles introduites, il s'agit principalement d'extensions utilisant plus de données, ou encore, des séparations entre les différents sinistres.

D'une manière générale, nous n'avons pas tenu compte d'un potentiel facteur d'actualisation lors de l'évaluation des réserves. A l'exception de l'ILR, introduit à la section 3.2.1, cet effet du temps peut facilement être incorporé. En effet, pour toutes ces méthodes, les instants relatifs aux flux qui nous intéressent sont déterminés. Néanmoins, soulignons que pour la méthode de Larsen, même si les délais de liquidation ne sont pas connus, l'actualisation ne pose aucun problème ; nous risquons juste d'actualiser des paiements nuls, obtenus une fois les sinistres clôturés. Au contraire, pour l'ILR, nous devons prendre en compte le timing des différents mouvements afin de pouvoir actualiser, impliquant de nombreuses simulations supplémentaires. De plus, les formules de type best-estimate des réserves proposées doivent également être revues.

Excepté pour l'approche collective, où nous avons suivi Denuit et Trufin [6], l'introduction de l'inflation n'a pas été abordée. Une manière de faire consiste à corriger les données par cet effet avant d'appliquer l'une ou l'autre méthode, évitant ainsi des modélisations impossibles ou trop complexes. Néanmoins, selon les modèles, cet effet peut facilement être incorporé. D'une part, soulignons que pour l'approche de Larsen, celui-ci peut directement être modélisé à travers les régressions. D'autre part, la méthode Chain-Ladder introduite par Mack [12] et présenté à la section 3.1.1, ne permet pas la modélisation d'un effet calendrier et donc celle de l'inflation. Cependant, en reprenant le modèle de Poisson surdispersé (voir (3.2)), qui, rappelons-le, fournit les mêmes estimations des réserves que le modèle introduit par Mack [12], nous sommes capables de prendre en considération une inflation à taux constant, noté  $g$ . Nous pouvons écrire

$$\mathbb{E}(P_{i,j}) = \alpha_i \beta_j (1+g)^{i+j-1} = \alpha_i (1+g)^i \beta_j (1+g)^{j-1} = \tilde{\alpha}_i \tilde{\beta}_j.$$

Dès lors, sans la spécifier explicitement, l'inflation sera automatiquement captée par les réserves obtenues via la méthode Chain-Ladder. Nous soulignons que :

- L'année de référence pour l'inflation correspond à l'année calendrier 1, c'est-à-dire  $i = j = 1$ . Dès lors, nous regardons ici les variations, apportées par l'inflation, par rapport à la première année calendrier ;
- Le taux d'inflation constant  $g$  n'est pas identifiable ;

- Si le taux d'inflation n'est plus constant (il pourrait, par exemple, varier avec l'année d'occurrence), la décomposition présentée n'est plus réalisable.

Certains sinistres nécessitent plus de  $n$  années de développement, c'est-à-dire  $\omega > n$ . Dès lors, l'introduction d'un *tail-factor* peut être envisagée. D'une part, pour le CLR et le modèle de Tweedie, de tels facteurs peuvent être introduits afin d'estimer les paramètres relatifs à l'effet développement pour  $\omega \geq j > n$ . Dès lors, grâce aux spécifications propres à chacune de ces méthodes, les différentes variables peuvent être calibrées. D'autre part, pour la méthode Chain-Ladder, le modèle PIC et l'approche de Larsen, l'étude est limitée à  $n$  années de développement et la prise en considération d'un tel mécanisme peut améliorer les prédictions. Finalement, notez que l'ILR modélise directement l'instant de liquidation, permettant ainsi, moyennant une éventuelle adaptation de la loi du délai de liquidation, de ne pas recourir à ce type d'extension.

Jusqu'à présent, nous n'avons introduit que les données minimales nécessaires. Cependant, il peut être intéressant de requérir plus d'informations au niveau individuel, permettant ainsi de traiter différemment les sinistres en fonction de leurs caractéristiques (e.g. Dommages matériels ou corporels, canal de distribution, ...). L'introduction de ces effets permet d'obtenir des nuances dans les calibrations des distributions, impliquant une meilleure prédiction. D'une part, pour la méthode ILR, une dichotomie *a priori* peut être effectuée et la méthode est alors répétée sur les différents sous-groupes, totalement distincts, ainsi obtenus. D'autre part, pour la méthode de Larsen ainsi que pour le CLR, de tels effets peuvent facilement être incorporés dans les régressions. En effet, en introduisant des covariables relatives aux caractéristiques des sinistres, nous pouvons directement modéliser leurs impacts dans les GLM. En outre, soulignons que les méthodes individuelles pures permettent de tenir compte des montants passés, soit de manière implicite (à l'aide d'une matrice de variance-covariance et de lois conditionnelles), comme nous l'avons vu pour l'ILR à la section 3.2.1, soit de manière plus explicite, où ils interviennent dans les régressions (voir section 3.2.2).

Une classification chère aux actuaires est la séparation des sinistres en fonction de leur coût total, à la liquidation. Généralement, deux classes sont créées. D'une part, nous avons les *petits* (ou *attritionnels*) sinistres et, d'autre part, nous avons les *grands* (ou *larges*) sinistres. Pour les méthodes agrégées, ce split peut être utilisé, créant ainsi deux triangles distincts. On peut alors appliquer les méthodes proposées sur chaque triangle ainsi obtenu. Pour les méthodes individuelles, nous pouvons modéliser ce split comme une caractéristique du sinistre, impliquant ainsi, pour le modèle ILR, une séparation *a priori* et pour les deux autres méthodes présentées dans ce volet, l'introduction de cette variable dans les régressions.

Basée sur la même idée de séparation des sinistres en fonction de leur coût total, une extension au modèle CLR, se base sur une remarque intéressante, faite par Denuit et Trufin [7], qui consiste à séparer *a priori* les petits des grands sinistres et à mêler *différentes approches* et méthodes de calibration. Les détails techniques étant similaires à ce qui a été présenté dans les sections 3.1.3 et 3.3, nous renvoyons le lecteur intéressé vers [7] et nous expliquons uniquement l'idée générale ci-dessous.

La dichotomie annoncée peut se réaliser de différentes manières :

- Selon le délai de notification et de liquidation. En effet, une corrélation importante entre les différents délais et l'ampleur du sinistre est observée. Les sinistres peu coûteux sont généralement notifiés et liquidés rapidement. Au contraire, les grands sinistres nécessitent plus de développement avant d'être liquidés.

Dès lors, on qualifiera de petit sinistre, un sinistre notifié et liquidé avant  $\omega_1$  années de déve-

loppement, où  $\omega_1$  doit être fixé sur base d'une étude descriptive de la base de données. L'autre partie des sinistres, c'est-à-dire ceux qui ne sont pas reportés ou liquidés avant  $\omega_1$ , forment l'ensemble des grands sinistres.

Cette dichotomie est d'ailleurs utilisée dans Denuit et Trufin [7]. Soulignons que, si nous n'observons pas  $\omega_1$  années de développement pour une année d'accident  $i$  fixée, nous ne pouvons pas décider si un sinistre encore ouvert est important ou non. La calibration se base donc sur les années d'occurrence  $i$  telles que  $i \leq n - \omega_1 + 1$ . Dès lors, si cette valeur pivot  $\omega_1$  est trop importante, trop de données risquent d'être exclues. Dans ce cas, un autre critère de segmentation devrait être envisagé ;

- Cette séparation peut également se faire en fonction de la première estimation du coût total du sinistre, comprenez le *case estimate*, qui est généralement la première évaluation des charges individuelles. Le montant à la base de la séparation devra alors être déterminé ;
- En fonction du type de sinistre : Dommages corporels ou non, décision du tribunal, ...

Rappelons que, pour la séparation petits-grands sinistres usuelle, un seuil est utilisé. Dès lors, pour les années d'occurrence plus lointaines, un grand nombre de sinistres importants a été observé. Alors que, pour celles plus récentes, un pourcentage élevé de grands sinistres n'ont pas encore dépassés le seuil. La classification est donc loin d'être évidente. Au contraire, ce type de dichotomie a le mérite de nous fournir *facilement* une information *rapide*. De plus, elle nous définit une approximation efficace de la séparation petits-grands sinistres.

Nous obtenons ainsi deux ensembles distincts pouvant être traités de manières différentes. D'une part, les petits sinistres sont généralement bien plus nombreux et une modélisation des paiements totaux dans chaque cellule via une loi Poisson-Gamma composée, telle que proposée dans l'approche de Tweedie (section 3.1.3), semble être appropriée. D'autre part, pour les grands sinistres, dont l'effectif est généralement plus faible, l'utilisation d'information individuelle supplémentaire est préférable avant d'agréger. Notons que ce n'est pas parce qu'un sinistre est considéré, à la liquidation, comme coûteux, que tous les paiements effectués sont très élevés. Certains paiements réalisés durant la durée de vie du sinistre peuvent être plutôt faibles mais la masse totale payée est, elle, importante. Pour les grands sinistres, afin de modéliser les paiements totaux dans chacune des cellules, on conçoit donc l'importance d'utiliser une loi mixte pour calibrer les paiements individuels. Cet ensemble sera donc modélisé comme expliqué à la section 3.3. Comme noté dans Denuit et Trufin [7], pour cette catégorie, il peut être judicieux d'intégrer le délai de déclaration individuel dans le paiement moyen. En effet, pour une cellule observée, certains sinistres importants peuvent être déclarés depuis une année de développement, d'autre depuis deux, ... impliquant des structures de coûts différentes. L'introduction de ce facteur permet donc de mieux capter la variation des paiements incrémentaux dans le temps. Dès lors, durant l'évaluation, un conditionnement par rapport à ce délai est nécessaire. Finalement, afin d'obtenir le paiement total effectué dans chaque cellule, il suffira de sommer les paiements pour les petits et grands sinistres.

Remarquons que le modèle proposé par Denuit et Trufin [7] ne considère plus le nombre de paiements effectifs mais s'intéresse plutôt au nombre de sinistres ouverts, dans chacune des cellules. Dès lors, une masse de probabilité en 0, afin de calibrer les paiements individuels, doit être prise en considération. Cependant, pour les petits sinistres, cette approche, comparée à l'utilisation des paiements effectifs, n'a pas de réel avantage. En adéquation avec les données disponibles, l'actuaire peut alors choisir la meilleure option.

Une telle idée pourrait également être utilisée pour les méthodes individuelles pures, permettant ainsi de limiter la complexité algorithmique.

Notez que, jusqu'à présent, aucune méthode individuelle n'utilise à la fois les paiements et les charges, comme cela a été présenté pour le PIC. Nous introduisons maintenant une extension à l'ILR vu à la section 3.2.1. Celle-ci a également été proposée par Pigeon *et al.* [19] et permet d'incorporer les données observées au niveau des charges individuelles.

Au lieu d'être basée sur Chain-Ladder, l'idée est cette fois similaire au modèle PIC excepté que nous travaillons de manière individuelle. Soient  $i$  une année d'occurrence fixée et  $k$  un sinistre survenu au cours de cette année. Dans ce contexte, l'indice  $j$  fait référence à la  $j^{\text{ème}}$  variation effective des paiements cumulés ou des charges. Sachant qu'il y a, au total,  $U_{i,k}^{(C)} + 1 = u_{i,k}^{(C)} + 1$  paiements effectifs et  $U_{i,k}^{(I)} + 1 = u_{i,k}^{(I)} + 1$  évaluations effectives des charges, les relations suivantes sont à la base du modèle

$$\begin{cases} C_{i,j,k} = \lambda_{j-1}^{(k)} C_{i,j-1,k} \text{ pour } j = 2, \dots, u_{i,k}^{(C)} + 1 \\ I_{i,j-1,k} = \gamma_{j-1}^{(k)} I_{i,j,k} \text{ pour } j = 2, \dots, u_{i,k}^{(I)} + 1 \end{cases} \quad (4.1)$$

avec la condition terminale,

$$C_{i,u_{i,k}^{(C)}+1,k} = I_{i,u_{i,k}^{(I)}+1,k}.$$

Avec cette approche, nous construisons, pas à pas, les paiements cumulés jusqu'au dernier et nous repartons ensuite, à reculons, pour déterminer les charges, c'est-à-dire

$$C_{i,1,k}, C_{i,2,k}, \dots, C_{i,u_{i,k}^{(C)}+1,k} = I_{i,u_{i,k}^{(I)}+1,k}, \dots, I_{i,2,k}, I_{i,1,k}.$$

Dès lors, grâce aux hypothèses (4.1), connaître  $C_{i,1,k}$  et l'ensemble des facteurs de développement  $\lambda$  et  $\gamma$  permet de déterminer les paiements et les charges pour ce sinistre. Nous définissons donc le vecteur de développement

$$\Lambda_i^{(k)} := \left( \ln C_{i,1,k}, \ln \lambda_1^{(k)}, \dots, \ln \lambda_{u_{i,k}^{(C)}}^{(k)}, \ln \gamma_{u_{i,k}^{(I)}}^{(k)}, \dots, \ln \gamma_1^{(k)} \right). \quad (4.2)$$

Comparé à l'équation (3.12), nous ajoutons l'information relatives aux charges. Dès lors, si nous faisons abstraction de cette information, nous retombons dans le cas présenté précédemment.

La classification IBNR/RBNP/RBNS définit également nos besoins. Ceux-ci sont quasiment similaires à ceux exposés à la section 3.2.1, en incorporant cette fois l'effet des données supplémentaires. Pour une année d'accident  $i$  et une année calendrier  $n$ , pour les IBNR, nous requérons :

- Le nombre de sinistre  $K_{i,n-i+1}^{IBNR}$  ;
- Pour chaque sinistre, le nombre de paiements effectifs et de variations dans les charges, soient  $U_{i,k}^{(C)} + 1$  et  $U_{i,k}^{(I)} + 1$  ;
- Pour chaque sinistre, le vecteur de développement défini à l'équation (4.2) ;
- Si nécessaire, pour chaque sinistre, les délais entre deux paiements consécutifs ou entre deux ajustements consécutifs des charges.

Pour les RBNP, nous connaissons le nombre de sinistre et nous reprenons à l'étape 2 citée ci-dessus. Finalement, pour les RBNS, nous connaissons une partie du vecteur de développement (4.2), il reste à déterminer le nombre (et leurs timings, si requis) de paiements et d'ajustements des charges restants ainsi que leurs montants.

Afin d'effectuer les différentes projections individuelles, il reste à spécifier et calibrer les distributions pour les variables nécessaires. D'une part,  $K_{i,n-i+1}^{IBNR}$  et  $U_{i,k}^{(C)}$  peuvent être déterminés comme expliqué précédemment. D'autre part, un choix naturel consiste à spécifier  $U_{i,k}^{(I)}$  de la même manière que pour le nombre de paiements (voir (3.13)) et d'en évaluer les paramètres. On supposera que les deux variables de comptage  $U_{i,k}^{(C)}$  et  $U_{i,k}^{(I)}$  sont indépendantes. Dès lors, pour les sinistres RBNS, par la loi des probabilités totales, la distribution du nombre total d'ajustements des charges compte tenu de ceux déjà observés, est facilement obtenue. Ensuite, nous pouvons supposer que le nouveau vecteur de développement suit également une loi normale multivariée (avec asymétrie, si nécessaire). Dès lors, en fonction des données disponibles, l'introduction d'une corrélation entre les paiements et les charges est envisageable. De plus, dans le cadre de sinistres RBNS, nous pouvons facilement simuler le reste de la trajectoire grâce à la normalité (avec asymétrie, si nécessaire) conditionnelle. Finalement, si nous souhaitons connaître les différents délais d'ajustements des charges, ceux-ci peuvent, comme nous le faisons pour les paiements, être spécifiés à l'aide d'une loi géométrique (ou autre) avec quelques composantes dégénérées (voir (3.13)).

Afin de conclure cette note, nous pensons qu'un tableau synthétisant ce qui vient d'être abordé est plus efficace. De plus, en présentant les avantages et inconvénients de chacune des méthodes, nous obtenons également une base de comparaison générale. Cette table est visible à la Figure 4.1 ci-dessous. Les détails étant disponibles dans le texte, nous ne faisons pas plus de commentaires.

Nous nous sommes contentés de présenter certaines méthodes, sans les appliquer sur une base de données. Cette application aurait néanmoins pu servir de benchmark afin de déterminer quelle approche nous donne les résultats les plus justes *pour la base de données considérée*. Cela dit, *pourvu que nous ayons les données nécessaires*, intuitivement, vu que le niveau de granularité est plus important, nous attendons des performances plus élevées des méthodes de réservation individuelles. Dès lors, deux cas concrets s'offrent à nous :

- Soit nous sommes intéressés par les flux de chaque sinistre, auquel cas les méthodes individuelles pures sont appropriées ;
- Soit les cash-flows individuels ne sont d'aucun intérêt et une méthode de réservation collective est tout indiquée.

En pratique, seul le paiement total nous importe et la connaissance de tous les cash-flows n'est pas nécessairement utile pour effectuer la réservation au sens strict du terme. De plus, la complexité algorithmique engendrée par les méthodes individuelles pures n'est pas anodine. C'est pourquoi nous jugeons que les méthodes collectives peuvent définir le meilleur équilibre possible. Cependant, nous insistons sur le fait qu'il n'y a pas de bonne ou mauvaise méthode, tout dépend principalement des données disponibles et des besoins de l'assureur.

Critères \ Méthodes	CL	PIC	Tweedie	ILR	Larsen	CLR
Niveau des données	Collectif	Collectif	Collectif	Individuel	Individuel	Individuel
Dépendance paiements pour un même sinistre	Via moyenne	Via matrice de covariance	Indépendance	Via matrice de covariance	Via probabilités/moyenne	Indépendance
Distribution paiements/vecteurs de paiements <sup>1</sup>	Aucune	Normale multidim.	Poisson-Gamma composée ( $p \in [1, 2[$ )	MSN	Split Gamma/GPD	Split Gamma/Pareto-2
Méthode de calibration des paiements	Non	Cadre bayésien	Découpe des propriétés du GLM	Approche itérative pour calibrer la MSN	GLM & Fit Pareto	GAMLSS
Modélisation de la fréquence <sup>2</sup>	Non	Non	Non	#Sinistres/AA Poisson ajustée	#Sinistres/AA+AD Poisson	#Paieiments/cellule Poisson
Méthode de calibration de la fréquence	Non	Non	Non	Sélection modèle (BIC) + Max. Vraisemblance	GLM	GLM
Régression avec covariables individuelles	Non	Non	Non	Non	Oui	Oui
Projections individuelles	Non	Non	Non	Oui	Oui	Oui
Complexité algorithmique importante	Non	Non	Non	Oui	Oui	Non
Difficulté principale	/	Estimation des paramètres	Estimation des paramètres	Estimation des paramètres	Split paiements + nombreux GLM	Stabilité du GAMLSS

TABLE 4.1: Tableau récapitulatif des méthodes.

1. Il s'agit ici des lois présentées mais, comme nous l'avons vu, d'autres spécifications sont possibles.

2. AA = Année Accident ; AD = Année Déclaration.



# Bibliographie

- [1] D. AKDEMIR : *A class of multivariate skew distributions : Properties and inferential issues*. Thèse de doctorat, Bowling Green State University, Ohio, 2009.
- [2] K. ANTONIO et R. PLAT : Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014(7):649–669, 2013.
- [3] J.P. BOUCHER et D. DAVIDOV : On the importance of dispersion modeling for claims reserving : An application with the Tweedie distribution. *Variance*, 5:158–172, 2011.
- [4] M. DENUIT : *Assurances de dommages 2 (actu2200)*. Matériel de cours, 2017.
- [5] M. DENUIT et A. CHARPENTIER : *Mathématiques de l'Assurance Non-Vie. Tome II : Tarification et Provisionnement*. Economica, Paris, 2005.
- [6] M. DENUIT et J. TRUFIN : Beyond the Tweedie reserving model : the collective approach to loss development. *North American Actuarial Journal*, 21(4):611–619, 2017.
- [7] M. DENUIT et J. TRUFIN : Collective loss reserving with two types of claims in motor third party liability insurance. *Journal of Computational and Applied Mathematics*, 335:168–184, 2018.
- [8] S. HAASTRUP et E. ARJAS : Claims reserving in continuous time : A non-parametric bayesian approach. *ASTIN Bulletin*, 26(2):139–164, 1996.
- [9] R. KAAS : Compound Poisson distributions and glm's : Tweedie's distribution. [https://lstat.kuleuven.be/research/seminars\\_events/files/3afmd/Kaas.PDF](https://lstat.kuleuven.be/research/seminars_events/files/3afmd/Kaas.PDF), 2005. Online ; accessed 13 April 2018.
- [10] R. KAAS, M.J. GOOVAERTS et M. DENUIT : *Modern Actuarial Risk Theory Using R*. Springer, New-York, 2008.
- [11] Ch. LARSEN : An individual claims reserving model. *ASTIN Bulletin*, 37(1):113–132, 2007.
- [12] T. MACK : Distribution-free calculation of the standard error of chain-ladder reserve estimates. *ASTIN Bulletin*, 23(2):213–225, 1993.
- [13] A. J. MCNEIL, R. FREY et P. EMBRECHTS : *Quantitative Risk Management : Concepts, Techniques and Tools*. Princeton University Press, 2005.
- [14] M. MERZ et M. V. WÜTRICH : Paid-Incurred Chain Claims Reserving Method. *Insurance : Mathematics and Economics*, 46(3):568–579, 6 2010.

- [15] M. MERZ et M. V. WÜTRICH : *Stochastic claims reserving manual : Advances in dynamic modelling*, 2015.
- [16] R. NORBERG : Prediction of outstanding claims 2 : Model variations and extensions. *ASTIN Bulletin*, 29(1):5–25, 1999.
- [17] G. W. PETERS, P. V. SHEVCHENKO et M. V. WÜTRICH : Model uncertainty in claims reserving within Tweedie’s compound Poisson models. *ASTIN Bulletin*, 39(1):1–33, 2009.
- [18] M. PIGEON, K. ANTONIO et M. DENUIT : Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin*, 43:399–428, 2013.
- [19] M. PIGEON, K. ANTONIO et M. DENUIT : Individual loss reserving using paid-incurred data. *Insurance : Mathematics and Economics*, 58:121–131, 2014.
- [20] G.K. SMYTH et B. JORGENSEN : Fitting Tweedie’s compound Poisson model to insurance claims data : dispersion modelling. *ASTIN Bulletin*, 32:143–157, 2002.
- [21] J. TRUFIN : Assurance non-vie 2 (actu-f506). Matériel de cours, 2017.
- [22] M. V. WÜTRICH : Claims reserving using Tweedie’s compound Poisson model. *ASTIN Bulletin*, 33:331–346, 2003.
- [23] Y. ZHANG : Likelihood-based and bayesian methods for Tweedie compound Poisson linear mixed models. *Statistics and Computing*, 23:743–757, 2013.

## Chapitre 5

# Un mot sur la série et les auteurs...

### 5.1 Les Detra Notes

Les Detra Notes sont une série de notes techniques à la fois pédagogiques et pertinentes pour les professionnels de l'actuariat. Ces notes, rédigées dans un langage simple et accessible, sont publiées par les membres de l'équipe Detralytics, un cabinet de conseil qui fournit des services d'expertise et innovation aux entreprises nécessitant un soutien, une formation ou de la R&D en actuariat, data science et gestion des risques. L'expertise Detralytics combine connaissance du marché et recherche académique. Detralytics a été fondée dans le but de soutenir les entreprises dans l'avancement des sciences actuarielles et de répondre aux futurs défis de la profession. C'est dans le cadre de cette mission que nous rendons accessibles les Detra Notes.

### 5.2 Biographies des auteurs

#### **Gireg Willame**

Gireg Willame est un consultant participant au « Talent Accelerator Program » (TAP) de Detralytics, un programme destiné à booster la carrière des actuaires récemment diplômés en alliant expériences de consultance, formation, et coaching. Gireg est l'auteur du mémoire *Gestion actuarielle des produits dont les prestations sont fonction de la réserve (2016-2017)* rédigé dans le cadre du Master en Sciences Actuarielles de l'Université Catholique de Louvain sous la direction de Michel Denuit. Il a également co-écrit la seconde Detra Note *Quand Thiele et Cantelli aident l'actuaire à sortir de sa réserve... Ou comment gérer les contrats d'assurance-vie dont les prestations sont fonction des réserves (2017)*, disponible à cette adresse : <http://detralytics.com/detra-notes/>.

#### **Michel Denuit**

Michel Denuit est Directeur Scientifique chez Detralytics et Professeur de Sciences Actuarielles à l'Université Catholique de Louvain. Michel mène une carrière académique internationale depuis une

vingtaine d'années et a collaboré avec différents acteurs du marché de l'actuariat sur des projets techniques. Il a écrit et co-écrit divers ouvrages et publications métier. Une liste complète de ses publications est disponible à cette adresse : <https://uclouvain.be/en/directories/michel.denuit>.

## **Julien Trufin**

Julien Trufin est Directeur Scientifique chez Detralytics et Professeur de Sciences Actuarielles au département de Mathématiques de l'Université Libre de Bruxelles. Julien est un actuare qualifié de l'Institut des Actuaire en Belgique (IA|BE) et combine expérience en consultance et carrière académique, développée dans différentes universités telles que l'Université Laval (Canada), l'ULB et l'UCL. Il a écrit et co-écrit diverses publications métier. Une liste complète de ses publications est disponible à cette adresse : <http://homepages.ulb.ac.be/~jtrufin/>.



 **Detralytics**

*info@detralytics.eu*  
*Rue Belliard 2*  
*1040 Brussels*  
[www.detralytics.com](http://www.detralytics.com)