



# Résumé

La finalité de ce mémoire est de comparer plusieurs méthodes de tarification en assurance-crédit en faisant appel aux techniques classiques de l'actuariat. Il s'inscrit à la suite des précédents travaux internes menés par les actuaires Tristan Quatre et Antoine Kahn.

L'entreprise accueillante, AXA Assurcrédit, est une société anonyme dont le chiffre d'affaires 2018 ressort à 18,4 millions d'euros. Elle propose à ses clients « B to B » une assurance contre le risque d'impayés de leurs factures.

Le mémoire s'organise en plusieurs ensembles. Une première partie introductive décrivant la problématique, la démarche de résolution ainsi que les données disponibles. L'exercice consiste à déterminer la prime pure la plus appropriée pour toute entreprise souhaitant se couvrir contre les défauts de paiement de ses clients.

Les parties suivantes détaillent les trois méthodes étudiées qui sont d'usage fréquent mais rarement appliquées à cette garantie. Tout d'abord la théorie de la crédibilité, et plus spécifiquement l'approche linéaire de Bühlmann-Straub pour un tarif plus « équitable ». Ensuite le modèle collectif, qui autorise la décomposition de l'espérance des sinistres en fréquence et coût, utilise des régressions linéaires généralisées. Enfin le modèle « Asymptotic Single Risk Factor » pour simuler la fonction des pertes des portefeuilles de limites de crédit.

Ce travail est une « preuve de concept » destinée à informer le Directoire sur les apports et les limites des méthodes actuarielles en terme de tarification. Le passage en production des tarifs soulève des questions ardues qui ne sont pas étudiées ici.

**Mots clés** : assurance-crédit, tarification, modèle de Bühlmann-Straub, modèle fréquence cout, distribution Poisson mélange, modèle linéaire généralisé, régression binomiale négative, risque de crédit, ASRF model, PD, EAD, LGD, portefeuille granulaire.

# Abstract

The purpose of this dissertation is to compare several pricing methods for the accounts receivable insurance using classic actuarial techniques. It follows the previous internal work by actuaries Tristan Quatre and Antoine Kahn.

The host company, AXA Assurcrédit, is a limited company whose 2018 turnover was 18.4 million euros. She offers its “B to B” customers insurance against the risk of unpaid invoices.

The dissertation is organized into several chapter. A first introductory part describing the problem, the resolution process as well as the available data. Exercise consists of determining the most appropriate pure premium for any company wishing to protect itself against bad debts.

The following parts detail the three methods studied which are frequently used but rarely applied to this guarantee. First of all, the theory of credibility, and more specifically the linear approach of Bühlmann-Straub for a more “fair” price. Afterwards the collective model, which allows the decomposition of the expected loss into frequency and severity, using generalized linear regressions. Finally the “Asymptotic Single Risk Factor” model to simulate the loss function of credit limits portfolios.

This work is a “proof of concept” intended to inform the Management Board on the contributions and the limits of actuarial methods in terms of pricing. The IT deployment of the tariffs raise difficult questions that are not studied here.

**Key words** : trade credit insurance, pricing, Bühlmann-Straub model, frequency severity model, Poisson mixture distribution, generalized linear model, negative binomial regression, credit risk, ASRF model, PD, EAD, LGD, granular portfolio.

## Remerciements

Je souhaite remercier Monsieur Lionel Stempert, Président du Directoire, ainsi que Monsieur Philippe Crespon, Directeur Technique, qui ont rendu possible ce travail.

Je remercie également Monsieur David Faure, professeur au CNAM, qui a accepté de superviser ce mémoire.

Je remercie l'ensemble des équipes d'AXA Assurcrédit pour le temps qu'ils m'ont consacré ainsi que pour toutes les explications qu'ils ont pu me fournir sur leur activités.

Je remercie le CNAM, ses professeurs, ses personnels administratifs, en particulier Ghislaine Erny, et les étudiants de la filière actuariat, sans qui ce travail ne pourrait pas exister.

Enfin je dédie ce mémoire à Justine qui a fait preuve d'une infinie patience à mes côtés.

# Note de synthèse

L'assurance-crédit est un contrat par lequel l'assureur couvre l'assuré contre le risque d'impayé occasionné par ses clients. Habituellement la prime s'exprime en pour mille du chiffre d'affaires assurable. Elle peut être modulée en fonction d'options spécifiques comme par exemple la garantie ordre à livrer. Le souscripteur prend également en compte la solvabilité des principaux clients de l'assuré pour la fixation du taux de prime.

Au quotidien, l'assuré déclare auprès de l'assureur-crédit ses nouveaux acheteurs afin d'obtenir des agréments, autrement dit des limites de crédit dénommées, qui correspondent aux montants couverts en cas de défaut de paiement. Les couvertures accordées varient en fonction de la zone géographique, du secteur d'activité ou de la notation de l'acheteur. Les limites font l'objet d'une gestion dynamique après la souscription. Les entreprises sont monitorées et leurs agréments peuvent faire l'objet de réductions automatiques ou par les arbitres, individuellement ou sur un secteur d'activité, en fonction de la conjoncture économique.

Dans cette étude nous nous intéressons uniquement au calcul des primes pures des portefeuilles d'acheteurs dénommés sans option particulière. Il est important de noter que la somme des agréments n'est pas strictement égale au chiffre d'affaires assurable. En effet il est possible qu'une limite corresponde à plusieurs facturations en cours d'année ou que la limite demandée est par souplesse plus importante que la facturation réelle.

Trois méthodes sont passées en revue dans ce travail qui s'entend comme une « preuve de concept ». Tout d'abord la théorie de la crédibilité et en particulier l'approche linéaire de Bühlmann-Straub est envisagée pour rendre plus équitable la prime. Ensuite la régression linéaire généralisée est utilisée pour déterminer les espérances conditionnelles des fréquences et coûts des sinistres dans une optique de segmentation du tarif. Enfin le modèle « Asymptotic Single Risk Factor » permet d'estimer la sinistralité moyenne comme la conséquence du risque de crédit des portefeuilles d'agréments.

## Le modèle de Bühlmann-Straub

Dans chaque classe tarifaire, définie par une tranche de chiffre d'affaires et un taux de prime, subsiste une certaine hétérogénéité de telle sorte que la plupart des assurés paieront un petit peu trop tandis que certains seront nettement sous tarifés. Afin d'éviter cette situation qui inciterait les « bons » risques à partir, l'assureur crédit a mis en place une participation bénéficiaire qui consiste à rembourser à l'assuré une partie de sa prime si la police est peu sinistrée et récompense l'effort effectué par l'entreprise dans la sélection de ses prospects.

Une alternative à la « PB » est la personnalisation du tarif en intégrant l'historique des sinistres dans le calcul de la prime qui s'effectuera « au mérite » en tenant compte de l'exposition de l'assuré. L'expérience des grandes entreprises devrait se montrer plus stable dans le temps et mériter une plus grande crédibilité.

A chaque ratio de sinistralité  $X_{ij}$  le modèle de Bühlmann-Straub associe un poids déterministe  $w_{ij}$  qui représente l'exposition au risque. Il utilise les hypothèses suivantes

- ✓ les vecteurs  $(\Theta_i, X_{i1}, \dots, X_{it})$  sont indépendants
- ✓ les observations  $X_{ij}$  sont indépendantes conditionnellement à  $\Theta_i$
- ✓ les deux premiers moments sont finis tels que

$$E[X_{ij}|\Theta_i] = \mu(\Theta_i) \qquad \text{Var}[X_{ij}|\Theta_i] = \frac{1}{w_{ij}}\sigma^2(\Theta_i)$$

Les variables aléatoires  $\Theta_i$  iid représentent le collectif de risques dans lequel on sélectionne les niveaux de risque  $\theta_i$  qui conditionnent le tirage des variables aléatoires  $X_{ij}$ . elles constituent en quelque sorte la synthèse de tous les facteurs de risque associés au contrat  $i$  qui n'ont pas été pris en compte dans la segmentation initiale car inobservables par l'assureur.

La prime individuelle  $\mu(\Theta_i) = E[X_{i,t+1}|\Theta_i]$  est approché par un estimateur qui est affine en les observations et minimise l'écart quadratique moyen

$$E \left[ \left( \mu(\Theta_i) - c_0 - \sum_{j=1}^t c_j X_{ij} \right)^2 \right]$$

La prime de crédibilité se présente comme une moyenne pondérée de la prime collective et de la moyenne individuelle par facteur de crédibilité  $z_i$

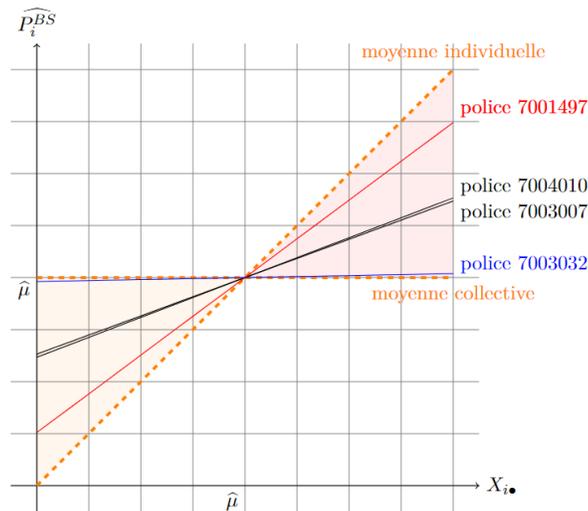
$$P_{cred}^i = \left(1 - \frac{w_{i\bullet}M^2}{\Sigma^2 + w_{i\bullet}M^2}\right)\mu + \left(\frac{w_{i\bullet}M^2}{\Sigma^2 + w_{i\bullet}M^2}\right)\frac{1}{w_{i\bullet}} \sum_{j=1}^t w_{ij}X_{ij} = (1 - z_i)\mu + z_iX_{i\bullet}$$

Ces résultats sont appliqués au rapport du montant de sinistres au montant des agréments.

numéro de police	ratios ( $\times 1000$ )			poids (en k€)				facteur de crédibilité	primes ( $\times 1000$ )		
	2012	2013	2014	2012	2013	2014	total		indi.	coll.	Bühl.
7001497	0.00	0.00	0.00	15748	17205	19430	52383	0.745	0.00	2.03	0.52
7001624	6.88	5.66	8.20	2470	3097	3048	8615	0.324	6.91	2.03	3.61
7002996	0.00	0.00	0.00	549	382	342	1273	0.066	0.00	2.03	1.90
7003007	0.00	5.19	3.10	3582	3467	3414	10463	0.368	2.73	2.03	2.29
7003032	0.00	0.00	0.00	186	77	77	340	0.019	0.00	2.03	2.00
7004010	10.44	0.00	0.00	3513	3636	3996	11145	0.383	3.29	2.03	2.51

Des assurés de tailles identiques sont tarifés selon des systèmes de prime proches (polices 7003007 et 7004010). Plus l'assuré possède un poids important plus sa droite est pentue et proche de la première bissectrice. Cela correspond à une prime équitable donnant plus de crédibilité à l'expérience personnelle mais à une moindre mutualisation (police 7001497).

A contrario les « petits » assurés sont considérés comme moins crédibles et se voient réclamer une prime plus proche de la prime collective (e.g. la police 7003032). Sur la figure ci-dessous la zone orange correspond à une contribution à la mutualisation, la zone rouge à une subvention du collectif.



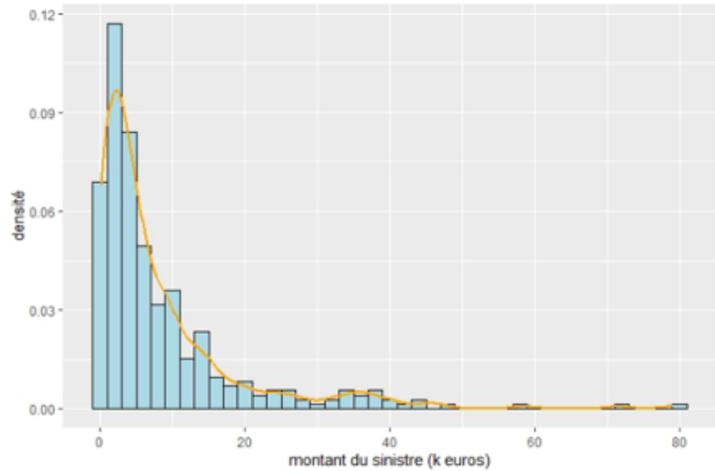
## Le modèle « fréquence - coût »

Dans le modèle « fréquence - coût » en univers segmenté, la prime pure est le produit des espérances de la fréquence et du coût des sinistres conditionnelles à l'information utile soit

$$E[S|X, Y] = E[N|X] \cdot E[C|Y]$$

Les effets des variables explicatives sur les niveaux de risque sont modélisés par des modèles de régression de type régression linéaire généralisée ou modèle additif généralisé. L'étude concerne 645 polices et 363 sinistres observés sur l'année 2015, pour une fréquence annuelle moyenne  $N$  de 0,563 et un coût annuel moyen  $C$  de 8400 euros.

k	nombre observé de polices avec k sinistres
0	475
1	87
2	35
3	20
4	12
5	5
6	7
7	1
8	3
total	645



Les estimateurs sans biais classiques  $\widehat{E}[N] = 0,563$  et  $\widehat{\text{Var}}[N] = 1,573$  laissent présager un faible ajustement de la loi de Poisson. Les modèles de mélange de lois forment une solution élégante pour prendre en compte cette forte surdispersion. Dans notre étude ils prolongent la loi de Poisson en introduisant une notion d'hétérogénéité  $\Theta_i$  au sein d'un groupe de contrats pour aboutir à une régression binomiale négative.

On dispose d'un échantillon de  $n$  observations  $i$  indépendantes de  $N$  et de  $p + 1$  variables tarifaires  $X_0, X_1, \dots, X_p$  non aléatoires avec  $(N_i, \Theta_i)$  vérifiant

- ✓ la distribution conditionnelle  $N_i|\Theta_i = \theta_i$  suit une loi de Poisson  $\mathcal{P}(\lambda_i\theta_i)$  avec  $\lambda_i > 0$
- ✓ la distribution marginale  $\Theta_i$  est une loi Gamma  $\gamma(\frac{1}{\alpha}, \frac{1}{\alpha})$  avec  $\alpha > 0$
- ✓ il existe  $p + 1$  réels  $\beta_0, \dots, \beta_p$  inconnus tels que

$$E[N_i|X_0, X_1, \dots, X_p] = \lambda_i = \exp(\beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i)$$

Ce schéma correspond à un cas particulier de mélange Poisson-Gamma pour lequel les deux premiers moments valent  $\mathbb{E}(N_i) = \lambda_i$  et  $\mathbb{V}(N_i) = \lambda_i + \alpha\lambda_i^2$  où  $\alpha$  quantifie le niveau de dispersion commun autour de la moyenne  $\lambda_i$ . On retiendra que pour tout  $k = 0, 1, \dots$

$$\mathbb{P}(N_i = k) = \frac{\Gamma(k + \frac{1}{\alpha})}{\Gamma(k + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{\alpha\lambda_i}{1 + \alpha\lambda_i}\right)^k \left(\frac{1}{1 + \alpha\lambda_i}\right)^{\frac{1}{\alpha}}$$

La fréquence est estimée via une régression binomiale négative tandis que le coût fait l'objet d'une régression linéaire généralisée de type Gamma avec un lien logarithme. Au final la prime pure de 2015 est le produit de trois facteurs explicatifs qualitatifs :  $x_1$  la taille du portefeuille,  $x_2$  la qualité du portefeuille et  $y_1$  le montant moyen des agréments. La prime de l'assuré prenant les modalités  $(i, j, k)$  s'écrit

$$\widehat{S^{FC}} = \exp(\alpha_0) \exp(\alpha_1^i) \exp(\alpha_2^j) \exp(\beta_0) \exp(\beta_1^k)$$

Dans les deux cas les coefficients sont estimés à l'aide du logiciel R :

```
glm.nb(formula = frequence ~ taille_portefeuille + qualite_portefeuille,
data = base4, model = TRUE, y = TRUE, init.theta = 0.9659438339, link = log)
```

Coefficients	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.6695	0.3451	-7.735	1.04e-14 ***
taille_portefeuille_C2	1.5398	0.4043	3.808	0.000140 ***
taille_portefeuille_C3	2.2077	0.3767	5.861	4.59e-09 ***
taille_portefeuille_C4	2.8641	0.3685	7.772	7.75e-15 ***
taille_portefeuille_C5	3.5140	0.3989	8.810	< 2e-16 ***
taille_portefeuille_C6	4.1567	0.4239	9.806	< 2e-16 ***
qualite_portefeuille_C2	-0.5666	0.2201	-2.574	0.010059 *
qualite_portefeuille_C3	-0.9204	0.2058	-4.473	7.72e-06 ***
qualite_portefeuille_C4	-1.6326	0.4383	-3.724	0.000196 ***

```
glm(formula = cout ~ moyenne_agrement, family = Gamma(link = "log"), data =
sinistres)
```

Coefficients	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	8.3966	0.1533	54.784	< 2e-16 ***
moyenne_agrementC_2	0.6284	0.1833	3.428	0.000679 ***
moyenne_agrementC_3	0.7744	0.1954	3.963	8.91e-05 ***
moyenne_agrementC_4	0.9772	0.2429	4.023	7.01e-05 ***

## Le modèle « risque de crédit »

Une approche alternative considère que l'assureur détient indirectement un portefeuille de prêts sur les clients des assurés. Ils correspondent aux engagements de l'assureur et sont matérialisés par les limites de crédit à un instant donné. En cas de défaillance d'un acheteur, l'assureur subit un sinistre considéré comme le non remboursement d'un crédit « virtuel » et on s'intéresse à la distribution des pertes du portefeuille et à son espérance

$$L_n = \sum_i = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \mathbf{1}_{D_i} \text{ avec } E[L_n] = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot PD_i$$

Le risque de crédit est quantifié par les banques en effectuant une distinction entre les pertes attendues  $EL$  et les pertes inattendues  $UL$ . Elles sont déterminées à partir du modèle « Asymptotic Single Risk Factor », originellement élaboré par Vasicek, qui représente le logarithme du rendement de l'actif  $i$  standardisé sous la forme

$$X_i = \sqrt{\rho} Y + \sqrt{1 - \rho} Z_i$$

Le facteur de risque systémique  $Y$  s'interprète comme une variable représentant l'état global de l'économie et le facteur de risque idiosyncratique  $Z_i$  représente le risque individuel de chaque débiteur. Les variables  $Y$  et  $Z_i$  sont mutuellement indépendantes et suivent une loi normale centrée réduite. La deuxième hypothèse du modèle ASRF stipule que le portefeuille est infiniment granulaire ce qui correspond à l'idée qu'il n'existe pas d'exposition dépassant une petite proportion de l'exposition totale. Elle s'exprime par les deux conditions suivantes

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n EAD_i = \infty \text{ et } \sum_{n=1}^{\infty} \left( \frac{EAD_n}{\sum_{i=1}^n EAD_i} \right)^2 < \infty$$

En appliquant une version de la loi forte des grands nombres Gordy montre qu'en l'absence de concentrations en terme d'exposition, la fonction de perte converge presque sûrement vers son espérance conditionnelle au facteur systémique soit  $\lim_{n \rightarrow \infty} L_n = E[L_n|Y]$  et la valeur en risque d'un portefeuille infiniment granulaire se décompose en la somme des contributions en risque des différentes créances autrement dit

$$\lim_{n \rightarrow \infty} VaR_{\alpha}(L_n) = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \Phi \left( \frac{\Phi^{-1}(PD_i) + \sqrt{\rho} \Phi^{-1}(\alpha)}{\sqrt{1 - \rho}} \right)$$

## Les hypothèses des modèles

Dans le cadre du calcul de la prime pure, le choix d'un modèle comporte un risque de spécification. En particulier les hypothèses retenues doivent être vérifiées. Cette confrontation à la réalité n'est pas toujours aisée, directe et complète selon la méthode utilisée.

En premier lieu, nous avons exploité le modèle de Bühlmann-Straub puisque ce dernier tient compte du poids propre à chaque assuré. Cependant en assurance-crédit, l'indépendance des vecteurs  $(\Theta_i, X_{i1}, \dots, X_{it})$  est discutable puisqu'un acheteur en difficulté peut être présent dans au moins deux portefeuilles et générer des sinistres simultanés en cas d'impayés. Toutefois les niveaux de risque  $\Theta_i$  qui reflètent les politiques de credit management individuelles des entreprises assurées garantissent une certaine indépendance des défauts.

Les hypothèses du modèle linéaire généralisé sont peu restrictives. L'estimation par le maximum de vraisemblance suppose que les observations sont indépendantes. Comme vu précédemment cela n'est pas entièrement vérifié dans le cadre de l'assurance-crédit. Toutefois pour cette approche c'est la qualité d'ajustement globale et la performance en prédiction qui sont examinées sur la base des différences entre observations et estimations.

Dans la dernière partie la distribution des pertes du portefeuille de prêts virtuels que l'assureur détient sur les clients des assurés est une variable aléatoire

$$L_n = \sum_{i=1}^n L_i = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \mathbb{1}_{D_i}$$

Dans le cadre de ce mémoire les quantités suivantes sont supposées déterministes

- ✓ l'exposition au défaut  $EAD_i = E[\widetilde{EAD}_i]$
- ✓ la perte en cas de défaut  $LGD_i = E[\widetilde{LGD}_i] \in [0, 1]$

Enfin dans la formule de l'espérance, la probabilité de défaillance est « Point In Time »

$$E[L_n] = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot PD_i$$

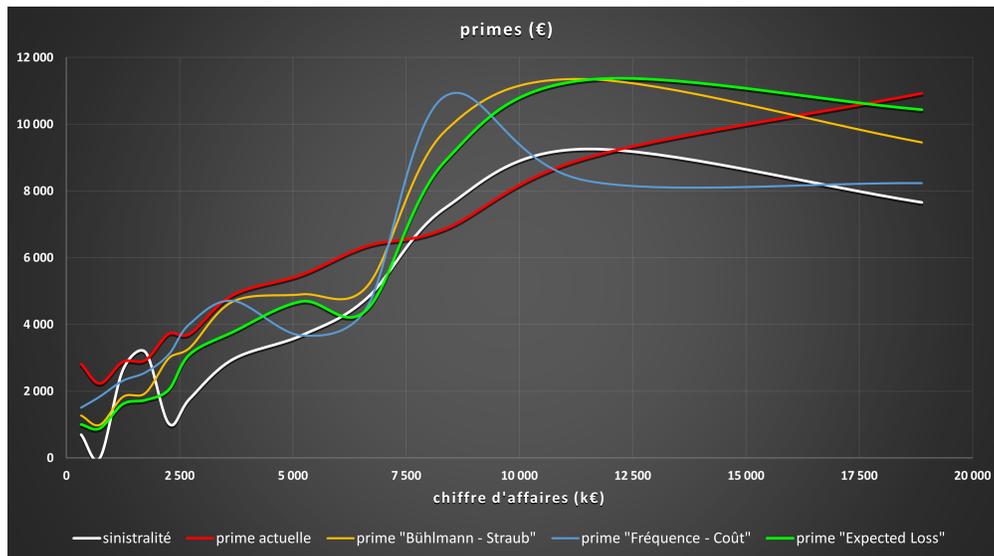
Les  $PD_i$  tentent de capter les variations du cycle économique et ainsi évoluent avec lui.

## Comparaison des tarifs

Le tableau suivant nous indique que les primes couvrent largement la sinistralité observée avec quelques écarts si l'on segmente les contrats selon le chiffre d'affaires. Par contre l'erreur quadratique moyenne doit être regardée avec des pincettes puisque le modèle « fréquence - coût » fait l'objet d'un sur-apprentissage. La prime actuelle, qui est une prime de marché fixée par les souscripteurs qui ont une vision quotidienne des taux pratiqués par la concurrence, est globalement acceptable mais bénéficierait d'une linéarisation.

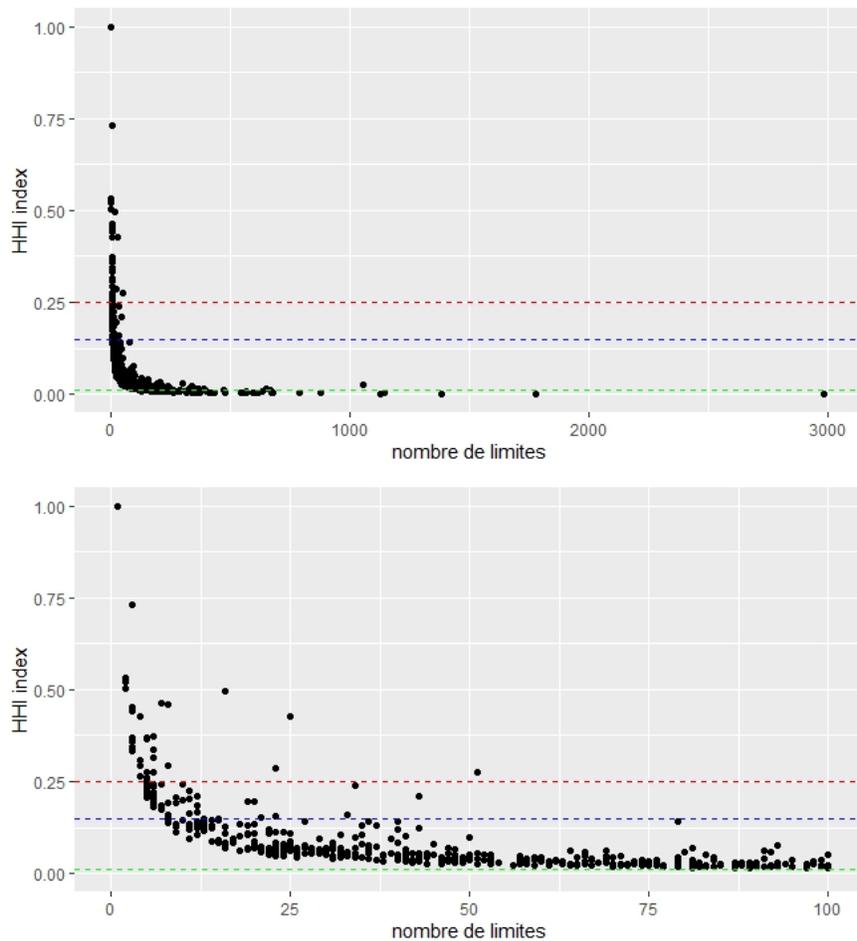
La prime de « Bühlmann-Straub » n'est pas adaptée à la souscription puisqu'elle demande la connaissance d'un historique assez profond et l'hypothèse d'indépendance des contrats n'est pas respectée. cette approche est intéressante pour construire des variables synthétique de type « bon ou mauvais » payeur qui pourront intervenir dans un modèle linéaire généralisé.

montant total sur le portefeuille (en €)	mean square error
sinistres	2 643 416
prime actuelle	3 694 261
prime « Bühlmann-Straub »	3 440 487
prime « Bâle 2 »	3 758 107
prime « fréquence - coût »	3 102 278



Les principaux écueils de la prime « fréquence - coût » concernent sa mise à jour et la faible sensibilité à un retournement de la conjoncture économique, à l'instar de la théorie de la crédibilité. Enfin le choix des variables explicatives est limité. En outre il semble peu probable que les caractéristiques d'un assuré influencent notablement le risque d'impayé des acheteurs.

L'origine des sinistres se trouvant principalement dans les portefeuilles d'acheteurs, la prime « Bâle 2 » est parfaitement adaptée. Cette approche est très conservatrice et extrêmement simple à mettre en oeuvre. Toutefois cette simplicité, à l'instar de sa robustesse, n'est qu'apparente et repose sur la qualité et la fréquence de l'estimation des paramètres bâlois. Par ailleurs l'étude nous enseigne que les portefeuilles d'acheteurs sont pour la très grande majorité granulaires (selon d'Herfindahl-Hirschman) ce qui permet de profiter des résultats relatifs à la perte exceptionnelle et le calcul des valeurs en risque.



Cette étude peut être prolongée dans plusieurs directions : la précision des paramètres bâlois doit être améliorée pour mieux refléter le risque d'impayé des factures commerciales. Un monitoring glissant sur douze mois des taux de défaut par note de solvabilité recommandé pour ajuster les  $PD$  et capter l'évolution de la sinistralité. Une meilleure caractérisation de l'exposition au défaut pourrait être traitée à travers une enquête auprès des assurés pour mieux comprendre l'utilisation des limites de crédit. Enfin le taux de recouvrement mériterait d'être différencié en segmentant les assurés en groupes homogènes de risque.

# Liste des tableaux

1.1	Le cycle de l'assurance-crédit . . . . .	3
1.2	Historique des défaillances des entreprises françaises . . . . .	6
4.1	Représentation des variables aléatoires . . . . .	22
4.2	Erreurs quadratiques moyennes . . . . .	26
4.3	Nombre de sinistres observés de 2012 à 2014 . . . . .	27
4.4	Montant de sinistres observés de 2012 à 2014 . . . . .	28
4.5	Primes de Bühlmann . . . . .	30
5.1	Ratios de sinistralité observés de 2012 à 2014 . . . . .	36
5.2	Primes de Bühlmann-Straub . . . . .	37
5.3	Couverture de la sinistralité 2015 . . . . .	38
5.4	Prime « Bühlmann-Straub » . . . . .	38
5.5	Prime « Bühlmann-Straub » . . . . .	39
7.1	Fréquence $N$ de sinistres observée . . . . .	49
7.2	Fréquence $N$ de sinistres simulée avec la loi de Poisson . . . . .	50
7.3	Fréquence $N$ de sinistres simulée avec la loi binomiale négative . . . . .	51
7.4	Test d'adéquation de la loi binomiale négative . . . . .	52
7.5	Nombre d'agréments . . . . .	58
7.6	Qualité des agréments . . . . .	59
7.7	Chiffre d'affaires assurable . . . . .	59
7.8	Zone géographique . . . . .	60
7.9	Secteur d'activité . . . . .	60
7.10	Sélection des variables pas à pas . . . . .	61
7.11	Résidus et points influents . . . . .	61
8.1	Densité et histogramme empiriques du coût des sinistres . . . . .	64
8.2	Sélection d'une distribution continue . . . . .	65

8.3	Analyses graphiques . . . . .	65
8.4	Quotité garantie . . . . .	66
8.5	Montant moyen des agréments ayant une notation inférieure à 7 . . . . .	67
8.6	Résidus et points influents . . . . .	68
8.7	Couverture de la sinistralité 2015 . . . . .	69
8.8	Prime « fréquence - coût » par CA . . . . .	70
8.9	Prime « fréquence - coût » vs prime actuelle . . . . .	70
9.1	Probabilité de défaut par note de solvabilité . . . . .	79
9.2	Indice d'Herfindahl-Hirschman . . . . .	82
9.3	Couverture de la sinistralité 2015 . . . . .	83
9.4	Prime « bâloise » par CA . . . . .	83
9.5	Prime « bâloise » vs prime actuelle . . . . .	84

# Table des matières

Résumé . . . . .	i
Abstract . . . . .	ii
Remerciements . . . . .	iii
Note de synthèse . . . . .	iv
Liste des tableaux . . . . .	xiii
Table des matières . . . . .	xv
Introduction . . . . .	xix
<b>I Présentation du sujet du mémoire</b>	<b>1</b>
<b>1 L'assurance-crédit</b>	<b>2</b>
1.1 Le crédit inter-entreprises . . . . .	2
1.2 L'assurance-crédit . . . . .	3
1.3 L'offre d'AXA Assurcrédit . . . . .	4
1.4 Le prix de la couverture . . . . .	5
1.5 Cycles économiques et défaillances . . . . .	6
<b>2 Organisation de l'étude</b>	<b>7</b>
2.1 Les données à disposition, retraitements . . . . .	8
2.1.1 Le fichier des assurés . . . . .	8
2.1.2 Le fichier des agréments . . . . .	9
2.1.3 Le fichier du coût des sinistres . . . . .	9
2.1.4 Le fichier de la fréquence des sinistres . . . . .	9
2.2 Les différentes approches tarifaires . . . . .	10
2.2.1 Le modèle collectif . . . . .	10
2.2.2 Le modèle de crédibilité . . . . .	11
2.2.3 Le modèle de risque de crédit . . . . .	11
2.3 L'échantillonnage . . . . .	12

<b>3</b>	<b>Éléments de tarification</b>	<b>13</b>
3.1	Le coût de l'assurance . . . . .	14
3.1.1	Définition de la prime pure . . . . .	14
3.1.2	Chargement de sécurité . . . . .	15
3.2	Segmentation des primes . . . . .	16
3.2.1	La répartition des risques . . . . .	16
3.2.2	Les limites de la segmentation . . . . .	17
3.3	Le modèle collectif . . . . .	18
<b>II</b>	<b>Les modèles de crédibilité</b>	<b>19</b>
<b>4</b>	<b>Tarification par l'expérience</b>	<b>20</b>
4.1	Le modèle d'hétérogénéité . . . . .	21
4.1.1	Aléa moral, prime et équité . . . . .	21
4.1.2	Formalisation du problème . . . . .	22
4.1.3	Prévision en crédibilité . . . . .	23
4.2	Crédibilité affine . . . . .	24
4.2.1	Notations et hypothèses . . . . .	24
4.2.2	Calcul de la prime . . . . .	25
4.2.3	Estimation des paramètres . . . . .	26
4.2.4	Erreurs d'approximation . . . . .	26
4.3	Applications en assurance-crédit . . . . .	27
4.3.1	Nombre de sinistres . . . . .	27
4.3.2	Montant de sinistres . . . . .	28
4.3.3	Système de primes . . . . .	29
<b>5</b>	<b>Le modèle de Bühlmann-Straub</b>	<b>31</b>
5.1	Notion de volume de risque . . . . .	32
5.2	Hypothèses et notations . . . . .	33
5.3	Calcul de la prime . . . . .	34
5.4	Estimation des paramètres . . . . .	35
5.5	Application à l'assurance-crédit . . . . .	35
5.5.1	Le ratio de sinistralité retenu . . . . .	35
5.5.2	les facteurs de crédibilité . . . . .	36
5.5.3	Le tarif par expérience . . . . .	37
5.6	Discussion sur la mise en oeuvre du modèle . . . . .	40

<b>III</b>	<b>Les modèles « fréquence - coût »</b>	<b>41</b>
<b>6</b>	<b>Le modèle linéaire général</b>	<b>42</b>
6.1	Famille exponentielle . . . . .	43
6.1.1	Définition . . . . .	43
6.1.2	Exemple . . . . .	43
6.2	Modèle linéaire généralisé . . . . .	44
6.2.1	Composantes du modèle . . . . .	44
6.2.2	Estimation des paramètres . . . . .	45
6.3	Inférence et outils de diagnostics . . . . .	45
6.3.1	Sélection de modèles . . . . .	46
6.3.2	Qualité d'ajustement . . . . .	46
6.3.3	Analyse des résidus . . . . .	47
6.3.4	Points influents . . . . .	47
<b>7</b>	<b>La fréquence de sinistres</b>	<b>48</b>
7.1	La distribution du nombre de sinistres . . . . .	49
7.1.1	Les données observées . . . . .	49
7.1.2	Ajustement d'une loi de Poisson . . . . .	49
7.1.3	Ajustement d'une loi binomiale négative . . . . .	51
7.1.4	Test d'adéquation du $\chi^2$ . . . . .	52
7.2	Les distributions Poisson-mélange . . . . .	53
7.2.1	Formulation générale . . . . .	53
7.2.2	Loi Gamma à un paramètre . . . . .	54
7.2.3	Loi inverse gaussienne à un paramètre . . . . .	55
7.3	Régression binomiale négative . . . . .	56
7.4	Régression Poisson Inverse Gaussienne . . . . .	57
7.5	Description des variables tarifaires . . . . .	58
7.5.1	La taille du portefeuille . . . . .	58
7.5.2	La qualité du portefeuille . . . . .	59
7.5.3	Le chiffre d'affaires assurable . . . . .	59
7.5.4	La couverture géographique des agréments . . . . .	60
7.5.5	Le secteur d'activité de l'assuré . . . . .	60
7.6	Application de la régression binomiale négative . . . . .	61
7.6.1	Sélection des variables . . . . .	61
7.6.2	Modélisation retenue . . . . .	62

<b>8</b>	<b>Le coût moyen des sinistres</b>	<b>63</b>
8.1	Choix d'une distribution . . . . .	64
8.2	Variables tarifaires . . . . .	66
8.2.1	La quotité garantie . . . . .	66
8.2.2	Le montant moyen des agréments . . . . .	67
8.2.3	Le secteur d'activité de l'assuré . . . . .	67
8.3	Modélisation du coût . . . . .	68
8.3.1	Estimation des coefficients . . . . .	68
8.4	Le tarif « fréquence - coût » . . . . .	69
8.5	Discussion sur la mise en oeuvre du modèle . . . . .	71
<b>IV</b>	<b>Un modèle issu de la finance</b>	<b>72</b>
<b>9</b>	<b>Le coût du risque de crédit</b>	<b>73</b>
9.1	La distribution des pertes . . . . .	74
9.2	Le modèle « Asymptotic Single Risk Factor » . . . . .	75
9.2.1	Définition et hypothèses . . . . .	75
9.2.2	Le portefeuille infiniment granulaire . . . . .	76
9.3	Simulation numérique des pertes . . . . .	77
9.3.1	Méthodes de Monte-Carlo . . . . .	77
9.3.2	Algorithme pour la distribution des pertes . . . . .	78
9.4	Paramètres pour l'assurance-crédit . . . . .	79
9.4.1	La probabilité de défaut . . . . .	79
9.4.2	La perte en cas de défaut . . . . .	79
9.4.3	L'exposition au défaut . . . . .	80
9.5	Granularité des portefeuilles . . . . .	80
9.5.1	Le portefeuille d'AXA Assurcrédit . . . . .	80
9.5.2	Les portefeuilles des assurés . . . . .	82
9.6	Le tarif « bâlois » . . . . .	83
	<b>Conclusion</b>	<b>xix</b>
	<b>Bibliographie</b>	<b>xxi</b>

# Introduction

Le contrat d'assurance-crédit est un contrat par lequel un assureur couvre ses assurés contre le risque d'impayé occasionné par leurs clients. Ainsi à chaque client correspond une limite d'encours garanti déterminé par l'assureur-crédit.

Aujourd'hui la prime s'exprime en pour mille du chiffre d'affaires annuel assurable avec un taux de prime décroissant avec la taille de l'assuré. Axa Assurcrédit souhaite refondre sa tarification pour être plus sélective, plus pertinente et plus réactive dans les risques qu'elle accepte de couvrir.

La réussite de ce projet passe par la mobilisation d'un plus grand nombre de variables tarifaires à travers des modèles statistiques plus sophistiqués. Une meilleure segmentation des assurés et une meilleure prise en compte de la sinistralité passée donnera plus de latitude à l'assureur dans la bonne sélection des risques qu'il accepte de couvrir.

Ce mémoire prolonge des travaux précédents sur la mise en oeuvre des modèles linéaires généralisés et l'utilisation de l'approche fondée sur les notations internes pour le traitement du risque de crédit dans le milieu bancaire. Il explore l'utilité et le bien-fondé des modèles de crédibilité linéaire.

# Première partie

## Présentation du sujet du mémoire

# Chapitre 1

## L'assurance-crédit

### Sommaire

---

1.1	Le crédit inter-entreprises . . . . .	2
1.2	L'assurance-crédit . . . . .	3
1.3	L'offre d'AXA Assurcrédit . . . . .	4
1.4	Le prix de la couverture . . . . .	5
1.5	Cycles économiques et défaillances . . . . .	6

---

### 1.1 Le crédit inter-entreprises

L'usage du crédit client remonte au Moyen Âge avec la création du commerce à grande échelle et en particulier dans l'Europe du Sud, région très marquée par le catholicisme et soumise à l'interdiction religieuse du prêt avec intérêt. Cette pratique incontournable de tout négoce permettait aux acheteurs de tenir compte des délais d'acheminement des marchandises et de vérifier la qualité des biens fournis.

De nos jours le crédit inter-entreprises naît des décalages de paiement auxquels consentent les entreprises dans le cadre de leurs relations commerciales. En 2017, pour la France, celui-ci était évalué à plus de 607 milliards d'euros. Il existe un déséquilibre<sup>1</sup> en faveur des créances clients, ce qui constitue une source de financement pour ces derniers. Le crédit fournisseur trouvant son fondement dans la nécessité pour chaque entreprise de développer son chiffre d'affaires.

---

1. <https://entreprises.banque-france.fr/en/node/157722> « Alerte au crédit inter-entreprises! »

La loi n° 2008-776 du 4 août 2008 dite de "Modernisation de l'Économie" a fixé, en matière de crédit inter-entreprises, des délais de paiement maximaux de 60 jours nets à compter de la date d'émission de la facture ou de 45 jours fin de mois et a revu le taux des pénalités de retard ainsi que les sanctions attachées au non-respect de ces dispositions.

Des retards et défauts de paiement trop fréquents peuvent engendrer des problèmes de trésorerie jusqu'à compromettre la viabilité de l'entreprise voire dans le pire des cas provoquer un effet domino sur l'ensemble de l'économie.

C'est pourquoi ce risque d'impayés justifie complètement l'intérêt des instruments de couverture du poste client comme l'affacturage, produit bancaire très usité par les entreprises, ainsi que l'assurance-crédit.

## 1.2 L'assurance-crédit

Dans la section précédente nous avons vu que le risque de contrepartie associé au crédit inter-entreprises procède de l'éventualité que les clients ne règlent pas leurs fournisseurs aux échéances fixées contractuellement après livraison des marchandises.

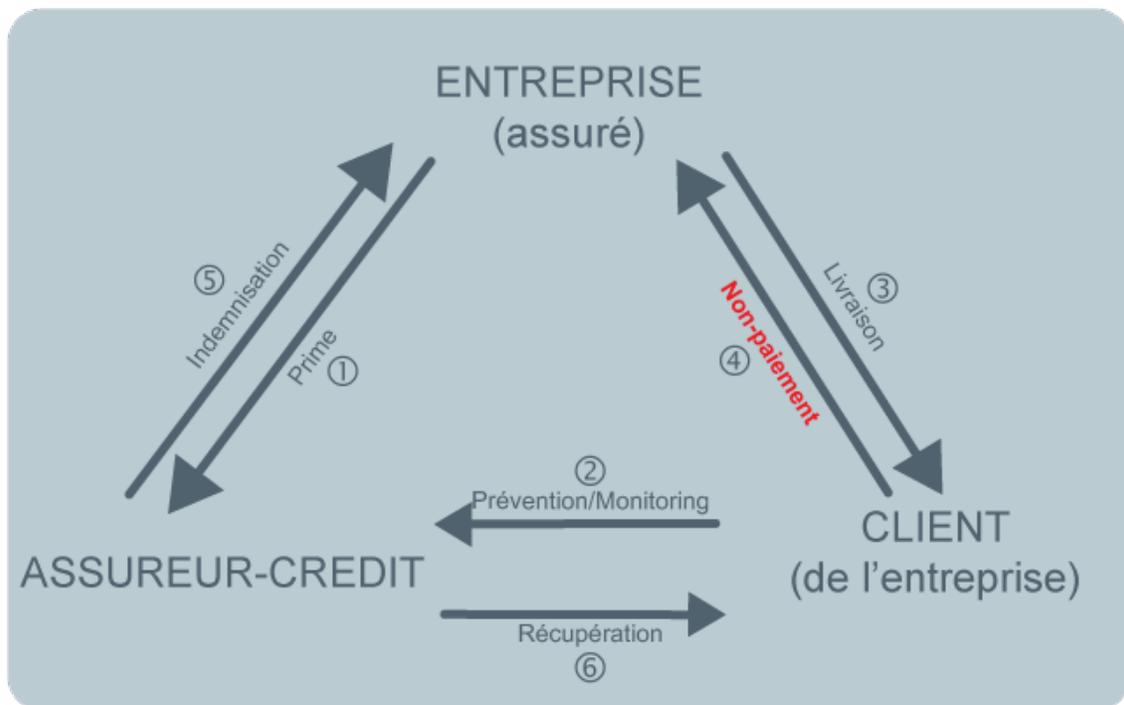


TABLE 1.1 – Le cycle de l'assurance-crédit

Face à cette contingence, un assureur-crédit proposera à ses assurés, contre paiement d'une prime (étape 1), de les couvrir contre le risque d'impayés de leurs acheteurs (étape 4) en les indemnisant (étape 5). L'assurance-crédit établit une relation tripartite (cf. figure 1.1) qui comprend des services accessoires de gestion du risque de crédit comme l'information sur la solvabilité financière des clients et des prospects (étape 2) ou le recouvrement des factures impayées (étape 6).

Au quotidien, l'assuré déclare auprès de l'assureur-crédit ses nouveaux acheteurs afin d'obtenir des agréments, autrement dit des limites de crédit dénommées, qui correspondent aux montants couverts en cas de défaut de paiement.

Il s'agit d'agréments monitorés en permanence. Ce suivi permet d'une part à l'assuré d'avoir une vision actualisée sur la santé financière de ses clients et d'autre part à l'assureur de gérer de façon dynamique sa sinistralité.

En complément des limites de crédits dénommées qui font l'objet d'une demande d'agrément de l'assuré auprès de l'assureur, il existe aussi des limites de crédits non dénommées, éventuellement surveillées, que nous n'étudierons pas. La garantie peut comprendre des options supplémentaires comme les garanties « ordres à livrer », « risque de fabrication », « risque politique » ou encore « risque de catastrophe naturelle ».

L'assurance-crédit est bien une activité d'assurance dans la mesure où il existe un aléa concernant la réalisation du sinistre et son niveau d'intensité. On remarquera que du côté des primes, une incertitude demeure aussi puisque le montant final dû par l'assuré, en général indexé sur son chiffre d'affaire, est inconnu en début de période.

### 1.3 L'offre d'AXA Assurcrédit

Le marché de l'assurance-crédit est relativement oligopolistique. Trois acteurs contrôlent environ 76% du marché mondial en 2018 :

- ✓ Euler Hermes, filiale du groupe Allianz, avec 34% ;
- ✓ Atradius, filiale du groupe Catalana Occidente, avec 25% ;
- ✓ Coface, filiale du groupe Natixis, avec 17%.

Dans un marché français de l'assurance-crédit mature et extrêmement concurrentiel, focalisé sur le prix de la couverture, Groupama est spécialisée sur le secteur de l'agro-alimentaire. AXA Assurcrédit, qui est détenue par AXA et Coface, s'adresse préférentiellement aux PME. Coface, également assureur-crédit, apporte ses moyens techniques et son expertise métier sur cette garantie qui relève de la branche d'activités 14.

AXA Assurcrédit distribue ses produits à travers le réseau des agents du groupe AXA France mais également par l'intermédiaire de courtiers généralistes ou spécialisés. La gamme comprend une offre forfaitaire appelée Kup et une offre sur mesure appelée Globale.

L'assureur prend à sa charge un pourcentage d'indemnisation appelé quotité garantie généralement inférieur à 100%, afin de se protéger de l'aléa moral. Par ailleurs pour chaque police existe une limite de décaissement (ou maximum d'indemnité annuelle) qui plafonne l'indemnisation annuelle à un multiple de la prime, le plus souvent 30 fois.

## 1.4 Le prix de la couverture

Le coût de la couverture est fixé par le souscripteur qui tient compte dans son analyse « du prix du marché » et de la qualité du portefeuille de l'assuré après l'étude d'une sélection de sa clientèle. L'analyse du souscripteur est multicritère et intègre le prix des garanties supplémentaires que souhaite le prospect.

Dans la suite on se concentre sur les portefeuilles d'agréments du produit Globale. Cette offre fait l'objet soit d'une prime forfaitaire déterminée en début d'année soit d'un taux de prime et d'un montant minimum à payer en forme d'acompte. Ce montant sera ensuite ajusté en fonction du chiffre d'affaires réel de l'assuré.

L'objectif du mémoire est de fixer un prix, appelé prime pure, qui correspond au montant attendu des sinistres du portefeuille d'agréments. Il s'agit d'un exercice complexe puisque la composition du portefeuille évolue en permanence en fonction des demandes de l'assuré et des réponses de l'assureur et du monitoring permanent de la solvabilité des acheteurs. L'étude est réalisée avec une vision figée des portefeuilles en début de période.

## 1.5 Cycles économiques et défaillances

Il est souvent reproché à l'assurance-crédit d'amplifier les mouvements de défiance entre entreprises en temps de crise. Cette procyclicité interroge sur la capacité de la garantie à remédier à des chocs exogènes massifs. Ainsi les décisions des assureurs-crédit de procéder pendant la crise sanitaire à des réductions de garanties et à des modifications des conditions contractuelles ont pu susciter l'incompréhension et entraîner des difficultés supplémentaires pour les entreprises.

Néanmoins la fédération française de l'assurance estime infondées les critiques quant au caractère procyclique de l'activité d'assurance-crédit : « l'assurance-crédit n'est que le reflet de la situation économique à un moment précis, l'assureur-crédit prenant ses décisions en fonction de celle-ci, et ne peut servir d'amortisseur des risques systémiques qui touchent l'ensemble de l'économie. »

Le « mur des faillites » initialement craint durant la crise de la covid-19 ne s'est heureusement pas réalisé, loin de là, puisque la mise en place de toute une palette de dispositifs a conduit à une chute de 38% des défaillances d'entreprise en 2020.



TABLE 1.2 – Historique des défaillances des entreprises françaises

Au niveau individuel, les causes d'une défaillance sont multiples mais peuvent être regroupées en deux grandes catégories : celles relevant des caractéristiques intrinsèques de l'entreprise (âge, taille, mode de gouvernance, stratégie prix, stratégie produit, salaires ...) et celles relevant de son environnement (conjuncture économique locale, nationale et internationale, évolution du cadre réglementaire, de la fiscalité, comportement des concurrents...).

# Chapitre 2

## Organisation de l'étude

### Sommaire

---

<b>2.1</b>	<b>Les données à disposition, retraitements</b>	<b>8</b>
2.1.1	Le fichier des assurés	8
2.1.2	Le fichier des agréments	9
2.1.3	Le fichier du coût des sinistres	9
2.1.4	Le fichier de la fréquence des sinistres	9
<b>2.2</b>	<b>Les différentes approches tarifaires</b>	<b>10</b>
2.2.1	Le modèle collectif	10
2.2.2	Le modèle de crédibilité	11
2.2.3	Le modèle de risque de crédit	11
<b>2.3</b>	<b>L'échantillonnage</b>	<b>12</b>

---

La disponibilité et la qualité des données sont des enjeux majeurs dans la réalisation des études actuarielles dont elles déterminent souvent les limites.

Nous verrons par la suite que la qualité et le volume des informations sauvegardées par l'assureur a un effet important sur la finesse du tarif. La conservation des données doit faire l'objet d'une réflexion très en amont pour envisager l'étude et la mise en place de modèles actuariels robustes et efficaces.

De par l'originalité de son organisation, AXA Assurcrédit doit être très attentive à cette question « Data » et intégrer dans sa politique de gestion des données un chapitre relatif aux modèles actuariels.

## 2.1 Les données à disposition, retraitements

Cette section présente les informations structurées dont nous disposons pour mettre en oeuvre les méthodes de tarification étudiées dans la suite. Les fichiers, au format csv, ont été constitués à partir de différentes extractions et d'un ensemble de programmes écrits en langage python. Il s'agit d'un travail préliminaire de traitements des données relativement lourd mais nécessaire pour fournir la matière principale de l'étude.

L'obtention de ces informations a également été un travail de longue haleine car l'informatique d'AXA Assurcrédit est opérée par son actionnaire COFACE. Cela explique pourquoi, dans le contexte difficile de l'épidémie de la Covid-19, le rafraîchissement des bases n'a pas été envisagée au cours de cette longue étude. D'une part pour ne pas monopoliser des ressources précieuses et d'autre part en raison de la particularité de l'environnement économique durant cette période marquée par un fort interventionnisme de l'Etat. Les années de pandémie demeurent atypiques et constituent une rupture par rapport au passé.

Nous disposons des périodes de 2012 à 2015. L'année d'observation de référence est 2015. Elle est jugée représentative d'une année typique en terme d'assurance-crédit hors « crise ». La période étant courte et le coût du sinistre « connu » à l'avance aucune mise en « as - if » n'est opérée.

### 2.1.1 Le fichier des assurés

Le fichier des assurés, pour une année particulière, présente des informations d'identification et de description, en particulier les variables suivantes

- ✓ l'identifiant de la police
- ✓ la date de début de la police
- ✓ la date de fin de la police
- ✓ le montant de la prime
- ✓ la limite de décaissement
- ✓ la quotité sur les agréments
- ✓ l'identifiant de l'assuré
- ✓ le secteur d'activité de l'assuré
- ✓ le chiffre d'affaires de l'assuré

### **2.1.2 Le fichier des agréments**

Ce fichier regroupe, pour une année particulière et pour chaque assuré présent dans le fichier précédent, l'ensemble des agréments accordés sur ses acheteurs. Les limites de crédit étant évolutives en fonction de la demande de l'assuré et de la solvabilité des acheteurs, seule les demandes initiales sont conservées. Les principales variables retenues sont

- ✓ l'identifiant de la police
- ✓ l'identifiant de l'acheteur
- ✓ la notation de l'acheteur
- ✓ le montant initial de l'agrément
- ✓ le secteur d'activité de l'acheteur
- ✓ la zone géographique de l'acheteur

### **2.1.3 Le fichier du coût des sinistres**

Ce fichier recense les événements ayant laissé à l'assureur une charge de sinistre strictement positive brute de tous éléments, en particulier avant recouvrement. Il permet de constituer la base d'étude du coût des sinistres. Les variables les plus marquantes sont

- ✓ l'identifiant de la police
- ✓ l'identifiant de l'acheteur
- ✓ le montant initial de l'agrément
- ✓ la charge brute de sinistre

### **2.1.4 Le fichier de la fréquence des sinistres**

Cette dernière base est construite à l'aide des fichiers précédents et constitue la principale matière de l'étude. Elle contient des résumés statistiques sur les agréments, les sinistres, le profil de l'assuré ainsi que sur la garantie. On retiendra en particulier les variables suivantes

- ✓ l'identifiant de la police
- ✓ l'ancienneté de la police
- ✓ le montant de la prime
- ✓ la limite de décaissement
- ✓ la quotité sur les agréments
- ✓ l'identifiant de l'assuré

- ✓ le secteur d'activité de l'assuré
- ✓ le chiffre d'affaires de l'assuré
- ✓ le nombre d'agréments
- ✓ le montant des agréments
- ✓ le nombre de sinistres
- ✓ le montant brut des sinistres
- ✓ la répartition des agréments par notation
- ✓ la répartition des agréments par zone géographique

## 2.2 Les différentes approches tarifaires

Cette partie offre un aperçu des techniques de modélisation qui seront développées en détail dans les chapitres suivants afin de déterminer la prime pure à réclamer à l'assuré.

### 2.2.1 Le modèle collectif

Dans le modèle collectif, la charge de sinistres  $S$  est fonction du nombre de sinistres  $N$  et de leurs montants identiquement distribués suivant la variable aléatoire  $C$ . L'approche « fréquence - coût » procure à l'assureur la possibilité de segmenter le tarif à partir de l'information  $X$  qu'il détient. La prime pure se calcule comme le produit  $E[S|X] = E[N|X] \cdot E[C|X]$ .

Les effets des différentes caractéristiques des individus sur les deux variables d'intérêt, à savoir la fréquence et le coût des sinistres, sont mesurés à l'aide de régressions linéaires généralisées qui lient l'espérance conditionnelle aux modalités des variables tarifaires avec

$$E[Y|X = x] = g^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)$$

où  $Y$  est la variable à expliquer au moyen des variables explicatives  $X_0, X_1, \dots, X_p$  regroupées dans le vecteur  $X$  et observées sur l'échantillon d'étude, noté  $x$ , constitué par l'observation des  $p + 1$  vecteurs  $(x_0, x_1, \dots, x_p)$  sur les  $n$  assurés du portefeuille étudié.

## 2.2.2 Le modèle de crédibilité

La théorie de la crédibilité, et plus spécifiquement l'approche linéaire de Bühlmann-Straub, est utilisée pour rééquilibrer le tarif en fonction de la sinistralité passée de chaque assuré. La prime obtenue est une moyenne pondérée de la prime collective  $\mu$  et de la prime individuelle observée  $\bar{X}$ . La confiance accordée aux expériences individuelles se traduit par l'importance du facteur de crédibilité  $z$  dans le calcul de la prime qui s'exprime sous la forme

$$P_{cred} = (1 - z)\mu + z\bar{X}$$

Il s'agit d'un estimateur sans biais qui répartit de façon plus équitable le coût de la sinistralité entre les différents assurés sans renoncer à la mutualisation des risques.

## 2.2.3 Le modèle de risque de crédit

La troisième approche est directement inspirée des réglementations « bâloises » en vigueur dans les banques pour le traitement du risque de crédit. La limite de crédit accordée par l'assureur est assimilée à un prêt virtuel qu'il aurait directement octroyé à l'acheteur. Cette démarche occulte le rôle de l'assuré dans la relation tripartite.

Dans la situation où les hypothèses du modèle de Gordy s'appliquent au portefeuille de  $K$  crédits de l'assureur, la prime pure se formule simplement et classiquement comme

$$EL = \sum_{k=1}^K EAD_k \cdot LGD_k \cdot PD_k$$

et l'effort de modélisation porte sur l'estimation des différents paramètres.

Dans le cas où le portefeuille de crédits n'est pas assez granulaire, il y a lieu de recourir à des simulations Monte-Carlo de la fonction de pertes. Ces calculs utilisent un modèle « Asymptotic Single Risk Factor » qui représente le logarithme du rendement de l'actif standardisé  $X_i$  sous la forme

$$X_i = \sqrt{\rho}Y + \sqrt{1 - \rho}Z_i$$

où le facteur de risque systémique  $Y$  s'interprète comme une variable représentant l'état global de l'économie et le facteur de risque idiosyncratique  $Z_i$  représente le risque individuel.

## 2.3 L'échantillonnage

Le fichier de la fréquence des sinistres qui sert de pivot pour l'étude comprend 857 assurés. Dans un premier temps 10 assurés atypiques sont mis de côté. Soit ils subissent un sinistre « grave » soit ils montrent une sinistralité « sérielle » atypique très élevée.

Les points atypiques sont par définition des observations peu fréquentes c'est-à-dire des points qui ne suivent pas la distribution caractéristique du reste des données. Ils peuvent refléter des propriétés particulières du phénomène étudié ou peuvent tout simplement résulter d'erreurs de mesure ou de toutes autres anomalies ne devant pas être modélisées. Ainsi nous retirons de l'échantillon six individus qui ont présenté au moins un sinistre par mois sur la période de 2012 à 2015. La sinistralité relative à ces individus est la conséquence d'un « effet taille sériel » ; puisque certaines grosses entreprises traitent un grand nombre de factures avec un nombre important de clients.

La prédiction des sinistres graves est un enjeu majeur en assurance. Bien que leur fréquence soit faible, l'impact de ces sinistres sur le résultat est souvent significatif dans la mesure où ils sont potentiellement extrêmement coûteux. Les montants des sinistres les plus élevés présentent en général une distribution spécifique qui indique qu'ils relèvent de causalités distinctes des autres sinistres. Cette distinction suppose de définir un seuil à partir duquel une valeur est considérée comme extrême. Chez AXA Assurcrédit, le seuil des « graves » a été fixé à dire d'expert à 100 000 euros. Au nombre de quatre dans l'échantillon, ils ne font pas l'objet d'étude particulière puisque l'assureur maîtrise partiellement ce risque en fixant lui-même les montants des agréments et possède la possibilité de les résilier à tout moment.

Dans un deuxième temps, on ne retient que les 645 assurés qui présentent quatre années d'historique afin de pouvoir comparer les méthodes de tarification.

Enfin, cette étude exploratoire s'entend comme une « preuve de concept » et n'a pas de visée opérationnelle immédiate. En effet le passage en production des tarifs envisagés soulève des questions ardues relatives aux traitements automatisés des données et demeure hors sujet de ce mémoire. C'est pourquoi le fichier initial n'est pas découpé en échantillons d'apprentissage, de validation et de test. En outre la robustesse des méthodes n'est pas mesurée. Toutes les statistiques descriptives relatives aux variables sont présentées dans les chapitres suivants.

# Chapitre 3

## Éléments de tarification

### Sommaire

---

<b>3.1</b>	<b>Le coût de l'assurance . . . . .</b>	<b>14</b>
3.1.1	Définition de la prime pure . . . . .	14
3.1.2	Chargement de sécurité . . . . .	15
<b>3.2</b>	<b>Segmentation des primes . . . . .</b>	<b>16</b>
3.2.1	La répartition des risques . . . . .	16
3.2.2	Les limites de la segmentation . . . . .	17
<b>3.3</b>	<b>Le modèle collectif . . . . .</b>	<b>18</b>

---

Le fonctionnement de l'assurance se différencie d'une production ordinaire par l'inversion du cycle de production. Au contraire de la situation classique où le producteur d'un bien connaît son coût de fabrication et peut en conséquence lui fixer un prix de vente adéquat, l'assureur demande une prime d'assurance sans connaître le montant réel des sinistres que l'assuré est susceptible de subir.

Cette spécificité a deux incidences fondamentales sur le métier d'assureur. La première concerne la nécessité de mettre en place des outils mathématiques relativement sophistiqués afin d'évaluer le montant de la prime à demander à l'assuré.

La seconde, qui conditionne l'efficacité des méthodes actuarielles, est relative à la dépendance élevée de l'activité de tarification aux données que l'assureur recueille sur le risque couvert par le contrat. Cette collecte d'informations est de plus en plus contrariée par des obligations réglementaires contraignantes.

## 3.1 Le coût de l'assurance

Pour un assureur, l'objectif premier auquel il doit répondre, est de rester solvable. En particulier il doit mettre toutes les chances de son côté pour qu'en cas de sinistre il puisse indemniser ses assurés.

Dans cette partie les risques considérés sont à court terme et à fort aléa et sans modélisation de produits financiers. C'est le cas des assurances de dommages qui obligent l'assureur à indemniser l'assuré des conséquences d'un sinistre sur son patrimoine. Elles regroupent les assurances de choses et les assurances de responsabilité.

### 3.1.1 Définition de la prime pure

La prime pure est le prix du risque. Il s'agit du montant dont doit disposer l'assureur pour dédommager sans excédent ni déficit les assurés suite aux sinistres survenus.

L'aversion au risque conduit un individu à transférer partiellement ses risques chez l'assureur qui lui réclame la prime pure. Si  $S$  est la charge totale des sinistres relative à une police déterminée au cours d'une période d'assurance, le rôle de l'assurance est de substituer une constante  $c$  certaine à la variable aléatoire  $S$  que subirait l'assuré.

Cette proximité entre  $S$  et  $c$  est mesurée par une distance qui pénalise aussi bien les cas où  $c$  est inférieure à  $S$  que ceux où  $c$  est supérieure au montant des sinistres. L'exemple le plus naturel est la distance appelée « écart quadratique moyen » définie par

$$d(S, c) = \mathbb{E} [(S - c)^2] = \text{Var}[S] + (\mathbb{E}[S] - c)^2$$

La constante la plus proche de  $S$  au sens de la distance  $d$  est l'espérance  $\mathbb{E}[S]$ . Il faut remarquer que la distance séparant les dépenses aléatoires  $S$  de l'assureur à la prime pure  $\mathbb{E}[S]$  qu'il réclamera à l'assuré est la variance  $\text{Var}[S]$ .

Il s'agit d'une mesure du risque que prend l'assureur en acceptant de payer  $S$  contre  $\mathbb{E}[S]$ . Bien que l'assureur se garantisse de ne pas perdre d'argent en moyenne, ce critère n'est pas totalement sécurisant car il ne mesure pas les extrémités des distributions.

Ce chapitre s'appuie sur les livres de M.Denuit et A.Charpentier [DENU 2004] et [DENU 2005] ainsi que sur l'article de A.Charpentier et al. [CHAR 2015].

### 3.1.2 Chargement de sécurité

L'assureur est exposé à  $n$  risques représentés par les variables aléatoires positives  $X_1, \dots, X_n$  indépendantes et identiquement distribuées de moyenne  $\mu$  et de variance  $\sigma^2$  finies strictement positives. Chaque assuré paye une prime  $P$  identique.

D'après le théorème « de la limite centrée » la somme des sinistres  $S_n = X_1 + \dots + X_n$  converge en loi vers une variable gaussienne. Ce résultat indique que pour tout  $z$  réel

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left( \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right) = \Phi(z)$$

Il permet de calculer la probabilité de ruine  $\mathbb{P}_R$  lorsque le nombre de risques devient important. Cette quantité correspond à la probabilité pour l'assureur de ne pas être en mesure d'indemniser ses assurés autrement dit

$$\mathbb{P}_R = \mathbb{P}(S_n > nP) = \mathbb{P} \left( \frac{S_n - n\mu}{\sigma\sqrt{n}} > \frac{nP - n\mu}{\sigma\sqrt{n}} \right)$$

avec lorsque  $n$  est grand

$$\lim_{n \rightarrow +\infty} \mathbb{P}_R = 1 - \Phi \left( (P - \mu) \frac{\sqrt{n}}{\sigma} \right)$$

Ainsi trois cas se présentent suivant le niveau de la prime  $P$  réclamé à l'assuré

- ✓ si  $P < \mu$  alors  $\lim_{n \rightarrow +\infty} \mathbb{P}_R = 1$ , l'assureur tend vers la faillite
- ✓ si  $P = \mu$  alors  $\lim_{n \rightarrow +\infty} \mathbb{P}_R = \frac{1}{2}$
- ✓ si  $P > \mu$  alors  $\lim_{n \rightarrow +\infty} \mathbb{P}_R = 0$ , l'assureur devrait survivre

Ainsi un chargement du type  $(1 + \eta)\mu$  avec  $\eta > 0$  permet de limiter la probabilité de ruine pour un grand nombre d'assurés. C'est l'effet de mutualisation.

Ce principe fondamental en assurance semble paradoxalement antagoniste à la nécessité de segmenter les tarifs sur la base de groupes de risques homogènes.

Segmenter consiste à différencier les primes en fonction d'un certain nombre de caractéristiques des risques afin de parvenir à une meilleure concordance entre le coût qu'un assuré met à la charge de la mutualité et la prime qu'il doit payer pour la couverture offerte.

## 3.2 Segmentation des primes

Dans un article de 2015, A.Charpentier indique que l'assurance repose fondamentalement sur l'idée que la mutualisation des risques entre les assurés est possible. Toutefois cette mutualisation, qui peut être vue comme une relecture actuarielle de la loi des grands nombres, n'a de sens qu'au sein d'une population de risques homogènes. Cette condition impose aux assureurs de segmenter. A l'extrême, avec l'explosion du nombre de données, et donc de variables tarifaires possibles, certains assureurs évoquent l'idée d'un tarif individuel, semblant remettre en cause l'idée même de mutualisation.

### 3.2.1 La répartition des risques

On désigne par  $\Omega$  l'ensemble des caractéristiques de l'assuré influençant le risque  $S$  et  $X$  le sous-ensemble des variables explicatives connues et utilisées par l'assureur. Les tableaux suivants illustrent les trois situations de transfert de risque de l'assuré vers l'assureur en fonction du niveau de segmentation mis en place.

Dans le cas d'une tarification sans segmentation, l'assureur qui réclame la prime  $E[S]$  est à l'équilibre financier en moyenne mais il porte à sa charge l'intégralité du risque  $\text{Var}[S]$ .

	assuré	assureur
dépense	$E[S]$	$S - E[S]$
espérance	$E[S]$	0
variance	0	$\text{Var}[S]$

Si l'assureur possédait une connaissance parfaite des risques pour segmenter sa clientèle, la prime serait la variable aléatoire  $E[S|\Omega]$ . L'assureur demeurerait toujours à l'équilibre mais ne supporterait plus qu'une partie de la variance  $\text{Var}[S] = \text{Var}[E[S|\Omega]] + E[\text{Var}[S|\Omega]]$ .

	assuré	assureur
dépense	$E[S \Omega]$	$S - E[S \Omega]$
espérance	$E[S]$	0
variance	$\text{Var}[E[S \Omega]]$	$E[\text{Var}[S \Omega]]$

En réalité l'assureur ne dispose que d'une information imparfaite. La prime pure  $E[S|X]$  lui garantit encore d'être à l'équilibre et il accepte désormais une part de risque intermédiaire puisque  $\text{Var}[S] = \text{Var}[E[S|X]] + E[\text{Var}[S|\Omega]] + E[\text{Var}[E[S|\Omega]|X]]$ .

	assuré	assureur
dépense	$E[S X]$	$S - E[S X]$
espérance	$E[S]$	0
variance	$\text{Var}[E[S X]]$	$E[\text{Var}[S X]]$

Le terme additionnel au détriment de l'assureur s'interprète comme le risque dû au manque d'information et traduit aussi le fait que la segmentation utilisée ne crée pas de classes réellement homogènes.

### 3.2.2 Les limites de la segmentation

Dans une situation où mutualisation et segmentation se confrontent, les « mauvais risques » sont attirés par les tarifs mutualisés car plus avantageux pour eux. Les assureurs qui mutualisent sont alors contraints d'augmenter leurs primes avec le temps pour compenser les pertes endurées par les « mauvais risques ». La fuite des « bons risques » ne permet plus d'assurer l'équilibre financier. On entre ainsi dans une « spirale de la segmentation » pour se prémunir du danger de la sélection adverse qui conduit en théorie à la faillite.

Au contraire les « bons risques » sont attirés par les tarifs segmentés. Néanmoins les assureurs qui segmentent tarifient sur de plus petits ensembles d'individus, ce qui conduit à une plus forte volatilité de la sinistralité de leurs portefeuilles. La limite de la segmentation réside dans la connaissance des risques du portefeuille pour que la loi des grands nombres et le théorème central limite puissent toujours s'appliquer. Les segmentations fines sont synonymes de modèles complexes qui augmentent entre autres les risques opérationnels et génèrent un effet haussier sur les tarifs.

Si les tarifs individualisés récompensent les bons comportements, les individus les plus exposés aux risques, ou simplement les moins chanceux, se retrouvent quant à eux contraints de subir des primes élevées voire des exclusions alors que leurs mauvais comportements ne sont pas nécessairement intentionnel. Ainsi une segmentation mal maîtrisée peut conduire à des discriminations involontaires.

### 3.3 Le modèle collectif

Le modèle collectif ne considère pas les polices individuellement mais le portefeuille dans son ensemble. Ainsi la charge de sinistres totale est exprimée en fonction du montant de chaque sinistre et non plus en fonction du montant des sinistres générés par chaque police.

On définit une classe  $C$  de contrats, dont le cardinal est compris entre 1 et le nombre de contrats du portefeuille. Les contrats sont indistincts dans cette classe. La charge de sinistres  $S$  de la classe  $C$  est fonction du nombre de sinistres  $N$  et de leurs montants  $Y$ . Plus précisément  $N$  est la variable aléatoire à valeurs entières représentant le nombre de sinistres affectant la classe  $C$  sur une période de temps.

La suite de variables aléatoires réelles  $(Y_i)_{i \geq 0}$  représente les montants individuels des sinistres dans l'ordre de survenance avec par convention  $Y_0 = 0$  lorsque  $N = 0$ . Dans le modèle collectif, la charge de sinistres est la somme aléatoire

$$S = \sum_{i=0}^N Y_i$$

Pour une branche d'assurance et au sein d'une classe d'individus homogène, on peut supposer que ces derniers auront des sinistres dont les lois sont proches et idéalement de manière indépendante. Le modèle collectif repose sur les hypothèses suivantes

- ✓ indépendance entre le nombre et le coût des sinistres c'est-à-dire que les variables aléatoires  $N$  et  $(Y_i)_{i \geq 0}$  sont supposées indépendantes ;
- ✓ indépendance et stationnarité des montants de sinistre, les variables  $(Y_i)_{i \geq 0}$  sont indépendantes et de même loi et surtout ne varient pas avec l'effet du temps.

Si les deux premiers moments de  $N$  et de  $Y$  existent alors ceux de  $S$  valent

$$E[S] = E[N] \cdot E[Y] \text{ et } \text{Var}[S] = E[N] \cdot \text{Var}[Y] + \text{Var}[N] \cdot E[Y]^2$$

Enfin en univers segmenté, la prime pure se calcule comme la fréquence multipliée par le coût moyen sachant les variables tarifaires. Le tarif se décompose en deux composantes

$$E[S|X] = E[N|X] \cdot E[Y|X]$$

## Deuxième partie

### Les modèles de crédibilité

# Chapitre 4

## Tarification par l'expérience

### Sommaire

---

<b>4.1</b>	<b>Le modèle d'hétérogénéité</b>	<b>21</b>
4.1.1	Aléa moral, prime et équité	21
4.1.2	Formalisation du problème	22
4.1.3	Prévision en crédibilité	23
<b>4.2</b>	<b>Crédibilité affine</b>	<b>24</b>
4.2.1	Notations et hypothèses	24
4.2.2	Calcul de la prime	25
4.2.3	Estimation des paramètres	26
4.2.4	Erreurs d'approximation	26
<b>4.3</b>	<b>Applications en assurance-crédit</b>	<b>27</b>
4.3.1	Nombre de sinistres	27
4.3.2	Montant de sinistres	28
4.3.3	Système de primes	29

---

L'objectif de l'assureur au moment d'établir un tarif, consiste à charger un montant de primes suffisant pour payer les sinistres futurs. Il pourra ensuite répartir les primes entre les assurés qui sont regroupés selon certaines de leurs caractéristiques. Néanmoins dans chaque groupe subsistera une certaine forme d'hétérogénéité de telle sorte que la plupart des assurés paieront un petit peu trop tandis que certains seront nettement sous tarifés.

Afin d'éviter cette situation qui inciterait les « bons » risques à partir, il faudra encore affiner la distribution des primes à l'intérieur des groupes. C'est là qu'entrent en jeu les systèmes de tarification basés sur l'expérience et plus particulièrement les méthodes de crédibilité.

## 4.1 Le modèle d'hétérogénéité

### 4.1.1 Aléa moral, prime et équité

L'expression « aléa moral » sert à désigner une situation où un agent isolé d'un risque se comporte différemment que s'il était complètement exposé à celui-ci.

L'aléa moral est en premier lieu apparu dans le domaine des assurances. C'était la possibilité qu'un assuré augmente sa prise de risque comparativement à la situation où il aurait supporté entièrement les conséquences négatives d'un sinistre.

D'une façon plus générale et par extension, on nomme aussi aléa moral toute modification du comportement d'un cocontractant contraire aux intérêts des autres parties au contrat, comparée à la situation qui prévalait avant la conclusion du contrat ; comme par exemple la dégradation du travail d'un employé après la fin de sa période d'essai.

Un portefeuille d'assurance peu segmenté, qui propose le même montant de prime à tous les assurés, pose la question de la justesse du tarif, essentiellement pour des raisons de concurrence et de comportement des assurés. L'assureur sera alors tenté de personnaliser a posteriori les primes en intégrant l'information qui lui est fournie par l'historique des sinistres.

Cette mise à jour du tarif s'effectuera « au mérite » en récompensant les contrats peu sinistrés et en sanctionnant pécuniairement les plus coûteux. Cette approche plus équitable a pour objectif de limiter l'aléa moral sans toutefois renoncer au principe de mutualisation.

L'assurance des entreprises est particulièrement concernée par la théorie de la crédibilité. En effet par contraste avec d'autres branches comme l'assurance des particuliers, ce segment est relativement pauvre en données a priori et se prête donc mal à des segmentations fines.

La demande de primes en lien avec la notion de crédibilité apparaît au début du siècle dernier chez certains constructeurs automobiles comme General Motors ou Tucker, assurés chez Allstate pour les accidents de travail, qui constatent que leurs primes d'expérience sont sensiblement inférieures à celle du marché. En outre General Motors fait valoir que sa propre exposition au risque est suffisamment importante pour obtenir un tarif personnalisé.

### 4.1.2 Formalisation du problème

Le modèle d'hétérogénéité permet de décrire la situation à laquelle nous sommes confrontés. Au sein d'un groupe de  $n$  contrats aux caractéristiques visibles semblables subsiste une hétérogénéité cachée en terme de risque. Des différences se révèlent au cours de l'observation de  $t$  périodes successives. Chaque contrat  $i$  produit pour chaque période  $j$  une charge de sinistres aléatoire  $X_{ij}$ . On s'intéresse à la détermination de la prime pure à  $t+1$  soit  $E[X_{it+1}]$ .

Une réponse à cette question peut se formaliser sous la forme d'un modèle d'urne à urne. La première urne représente le collectif de risques dans lequel on sélectionne un élément  $\theta_i$  réalisation de la variable aléatoire  $\Theta_i$ . Le niveau de risque sélectionné conditionne le contenu de la deuxième urne dans laquelle on tire les variables aléatoires  $X_{ij}$ .

Les variables aléatoires en jeu relatives au  $n$  contrats sont présentées dans la table 4.1.

variables non observables	variables observées				
	1	...	$j$	...	$t$
$\Theta_1$	$X_{11}$	...	$X_{1j}$	...	$X_{1t}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\Theta_i$	$X_{i1}$	...	$X_{ij}$	...	$X_{it}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\Theta_n$	$X_{n1}$	...	$X_{nj}$	...	$X_{nt}$

TABLE 4.1 – Représentation des variables aléatoires

La variable aléatoire  $\Theta_i$  constitue en quelque sorte la synthèse de tous les facteurs de risque associés au contrat  $i$  qui n'ont pas été pris en compte dans la segmentation initiale car inobservables par l'assureur. L'hypothèse d'une unique distribution indique que les contrats diffèrent par leurs niveaux de risque tout en restant similaires au point de les regrouper dans un même portefeuille.

Le principe d'homogénéité temporelle des contrats correspond à l'idée que si le niveau de risque du contrat  $\theta_i$  est connu alors les charges de sinistres sont indépendantes et identiquement distribuées. Enfin on postule que l'historique d'un contrat n'a pas d'influence sur l'avenir d'un autre.

Ce chapitre s'appuie sur les cours de P.Théron [THER 2018], de V.Goulet [GOUL 2023] ainsi que sur l'ouvrage séminal de H.Bühlmann et A.Gisler [BUHL 2005].

### 4.1.3 Prédiction en crédibilité

En théorie, pour un nouvel assuré, l'assureur aurait le choix entre deux primes :

- ✓ la prime individuelle, si le niveau de risque  $\theta_i$  du contrat  $i$  était connu alors la meilleure prévision de ses sinistres futurs serait

$$P_{ind} = \mu(\theta_i) = E[X_{it+1} | \Theta_i = \theta_i]$$

- ✓ la prime collective, comme première approximation de la prime de risque  $\mu(\theta_i)$  inconnue, nous pouvons utiliser la moyenne des primes de risque  $\mu(\Theta_i)$

$$P_{coll} = \mu = E[\mu(\Theta_i)] = E[E[X_{it+1} | \Theta_i]] = E[X_{it+1}]$$

En pratique  $\theta_i$  n'est pas connu. L'assureur va donc réclamer la prime pure collective  $\mu$  qui est un montant certain et dont l'estimation ne pose pas de difficultés. Toutefois cette approximation identique pour tous les contrats reste inéquitable.

Si l'assureur ne peut observer directement le niveau de risque de l'assuré  $i$ , l'historique de sinistralité à sa disposition noté  $\mathbf{X}_i = (X_{i1}, \dots, X_{it})$ , lui fournit une information sur  $\theta_i$ . Il est important de noter que du point de vue de l'assureur, la méconnaissance des risques individuels, implique que la prime individuelle recherchée est une variable aléatoire

$$P_{ind} = \mu(\Theta_i) = E[X_{it+1} | \Theta_i]$$

Enfin la meilleure estimation de la prime  $\mu(\Theta_i)$  au regard de l'erreur quadratique moyenne et compte tenu de l'information  $\mathbf{X}_i$  est la prime de Bayes définie par

$$P_{Bayes} = E[\mu(\Theta_i) | X_{i1}, \dots, X_{it}]$$

Il s'agit d'une prime certaine à la différence de la prime individuelle.

Toutefois le calcul de la prime bayésienne est difficile à mener et exige la connaissance des lois des variables aléatoires mises en jeu. L'approche proposée par Hans Bühlmann apporte une solution pragmatique pour déterminer une prime satisfaisante au regard de l'information disponible.

## 4.2 Crédibilité affine

La prime de Bayes, optimale au sens des moindres carrés, admet rarement une expression explicite. Pour cette raison en 1967 H.Bühlmann a proposé de se restreindre aux primes qui dépendent linéairement des observations passées.

### 4.2.1 Notations et hypothèses

On se place dans l'espace des fonctions aléatoires de carré intégrable  $L^2$ . Le modèle de Bühlmann, relatif au portefeuille de  $n$  contrats similaires  $i$  observés sur  $t$  périodes  $j$ , repose sur les hypothèses générales suivantes

- ✓ les variables  $\Theta_i$  sont indépendantes et identiquement distribuées
- ✓ les vecteurs  $(\Theta_i, X_{i1}, \dots, X_{it})$  sont indépendants
- ✓ conditionnellement à  $\Theta_i = \theta$  les  $X_{ij}$  sont indépendants et identiquement distribués
- ✓ les moments  $\mu(\theta) = E[X_{ij}|\Theta_i = \theta]$  et  $\sigma^2(\theta) = \text{Var}[X_{ij}|\Theta_i = \theta]$  existent et sont finis

Le portefeuille est caractérisé par ses paramètres de structure qui apparaissent naturellement dans le calcul de la prime

- ✓ la moyenne a priori  $\mu = E[X_{ij}] = E[\mu(\Theta_i)]$
- ✓ la variance a priori  $\sigma^2 = \text{Var}[X_{ij}] = \Sigma^2 + M^2$
- ✓ la variance intra  $\Sigma^2 = E[\text{Var}[X_{ij}|\Theta_i]] = E[\sigma^2(\Theta_i)]$  ou variabilité due au hasard
- ✓ la variance inter  $M^2 = \text{Var}[E[X_{ij}|\Theta_i]] = \text{Var}[\mu(\Theta_i)]$  ou hétérogénéité

On souhaite approcher la prime individuelle  $\mu(\Theta_i) = E[X_{it+1}|\Theta_i]$  par un estimateur qui est affine en les observations et utilise toute l'information disponible c'est-à-dire qu'il minimise l'écart quadratique moyen

$$E \left[ \left( \mu(\Theta_i) - a - \sum_{i=1}^n \sum_{j=1}^t b_{ij} X_{ij} \right)^2 \right]$$

## 4.2.2 Calcul de la prime

Par le fait de l'indépendance des contrats, on démontre que la prime de l'assuré  $i$  sera une fonction de ses seules observations  $X_{ij}$ . Sa détermination se ramène à la recherche des constantes  $c_0, c_1, \dots, c_t$  qui minimisent l'écart quadratique moyen

$$E \left[ \left( \mu(\Theta_i) - c_0 - \sum_{j=1}^t c_j X_{ij} \right)^2 \right]$$

L'atteinte du minima conduit à deux coefficients qui ne dépendent pas de l'indice  $i$

$$c_0 = \frac{\Sigma^2}{\Sigma^2 + tM^2} \mu = (1 - z_t) \mu$$

$$c_k = \frac{M^2}{\Sigma^2 + tM^2} = \frac{z_t}{t}$$

Le résultat se présente comme une moyenne pondérée de la prime collective et de la moyenne individuelle. Il s'agit donc d'une prime de crédibilité qui s'écrit

$$P_{cred}^i = \left(1 - \frac{tM^2}{\Sigma^2 + tM^2}\right) \mu + \left(\frac{tM^2}{\Sigma^2 + tM^2}\right) \frac{1}{t} \sum_{j=1}^t X_{ij} = (1 - z_t) \mu + z_t \bar{X}_i$$

L'estimateur  $P_{cred}^i$  converge presque sûrement vers  $\mu(\Theta_i)$  avec  $E[P_{cred}^i] = E[\mu(\Theta_i)]$ . Ainsi la prime de crédibilité répartit de façon plus équitable la sinistralité entre les différents assurés sans renoncer à la mutualisation des risques.

Le calcul de la prime fait apparaître une quantité appelée facteur de crédibilité qui ne dépend pas de l'assuré

$$z_t = \frac{tM^2}{\Sigma^2 + tM^2}$$

Sans expérience passée  $z_0 = 0$  et la prime est la moyenne collective  $\mu$ .

Le facteur de crédibilité croît avec le nombre d'années d'observations et l'hétérogénéité du portefeuille, soit  $M^2 > 0$ , et accorde alors plus de poids à la prime d'expérience  $\bar{X}_i$ . A contrario la crédibilité diminue avec l'importance des fluctuations de sinistralité observées au sein du portefeuille par le fait du hasard.

### 4.2.3 Estimation des paramètres

Les statistiques suivantes sont des estimateurs sans biais et convergents des paramètres structuraux

$$\hat{\mu} = \frac{1}{nt} \sum_{i=1}^n \sum_{j=1}^t X_{ij}$$

$$\widehat{\Sigma}^2 = \frac{1}{n(t-1)} \sum_{i=1}^n \sum_{j=1}^t (X_{ij} - \frac{1}{t} \sum_{j=1}^t X_{ij})^2$$

$$\widehat{M}^2 = \frac{1}{n-1} \sum_{i=1}^n (\frac{1}{t} \sum_{j=1}^t X_{ij} - \frac{1}{nt} \sum_{i=1}^n \sum_{j=1}^t X_{ij})^2 - \frac{1}{t} \widehat{\Sigma}^2$$

En substituant les valeurs ci-dessus dans les expressions du facteur et de la prime de crédibilité, on obtient l'estimateur empirique de crédibilité linéaire.

### 4.2.4 Erreurs d'approximation

Le tableau 4.2 donne par ordre croissant l'erreur quadratique moyenne commise en remplaçant la prime individuelle inconnue  $P_{ind} = \mu(\Theta)$  par les différentes primes envisageables

$P_{Bayes} = E[\mu(\Theta) X]$	$E[\text{Var}[\mu(\Theta) X]]$
$P_{cred} = (1-z)\mu + z\bar{X}$	$(1-z)M^2 = z\frac{\Sigma^2}{t}$
$P_{coll} = E[\mu(\Theta)] = \mu$	$\text{Var}[\mu(\Theta)] = M^2$
$\bar{X}$	$E[\frac{\sigma^2(\Theta)}{t}] = \frac{\Sigma^2}{t}$

TABLE 4.2 – Erreurs quadratiques moyennes

Par définition, la prime bayésienne est la meilleure prévision de la prime de risque que l'on puisse calculer. Le tableau nous indique que la prime collective n'est pas optimale à partir du moment où des données de sinistres deviennent disponibles.

Enfin le principal intérêt de la prime de crédibilité réside dans la simplicité de sa mise en oeuvre. Il n'est pas nécessaire de spécifier les lois de structure qui ajoutent un risque de modèle. En contrepartie la précision de la prime de crédibilité est moindre que celle de la prime de Bayes. On remarquera que les erreurs de la prime collective et de la moyenne individuelle sont liées à la prime de crédibilité au facteur de crédibilité près.

## 4.3 Applications en assurance-crédit

La méthode de Bühlmann est appliquée respectivement au nombre et au montant de sinistres en utilisant la fonction « cm » du paquet « actuar » du logiciel R.

L'historique de sinistres couvre les années de 2012 à 2014. Sur cette période les 645 assurés subissent en moyenne chaque année 374 sinistres pour un montant total de 3 628 170 euros. Cela correspondant à une sinistralité moyenne annuelle par assuré de 5625 euros. Chaque assuré connaîtra en moyenne 0,58 sinistre par an pour un montant moyen de 9701 euros.

### 4.3.1 Nombre de sinistres

L'extrait du portefeuille présenté en table 4.3 montre une forte hétérogénéité puisque certaines polices ne sont pas sinistrées (e.g. la police 7001497) et à l'inverse d'autres cumulent les sinistres (e.g. la police 7001624).

numéro de police	nombre de sinistres			primes			prime pure (en euros)
	2012	2013	2014	individuelle	collective	Bühlmann	
7001497	0	0	0	0.00	0.58	0.13	1213
7001624	3	3	2	2.67	0.58	2.22	21505
7002996	0	0	0	0.00	0.58	0.13	1213
7003007	0	1	1	0.67	0.58	0.65	6286
7003020	1	0	0	0.33	0.58	0.39	3749
7003032	0	0	0	0.00	0.58	0.13	1213
7003050	2	2	3	2.33	0.58	1.96	18968
7003064	0	0	1	0.33	0.58	0.39	3749
montant annuel moyen des sinistres de 2012 à 2014							3 628 170

TABLE 4.3 – Nombre de sinistres observés de 2012 à 2014

Pour le nombre de sinistres, les calculs menés sous R fournissent les estimations suivantes

$$\begin{aligned} \hat{\mu} &= 0,579845 & \widehat{\Sigma}^2 &= 0,701292 \\ \widehat{M}^2 &= 0,850473 & \widehat{z}_3 &= 0,784397 \end{aligned}$$

Le facteur de crédibilité  $\widehat{z}_3 \approx 0,784$ , qui apparaît dans l'estimation de la prime de Bühlmann  $(1 - \widehat{z}_3)\hat{\mu} + \widehat{z}_3\overline{X}_i$ , est très élevé. Cela indique une forte hétérogénéité des risques.

Dans ce premier modèle de crédibilité, les meilleurs risques, ceux qui n'ont connu aucun sinistre sur les trois dernières années se verront réclamer une prime de 0,13 plutôt que 0,58

résultant de la mutualisation par l'espérance. En multipliant cette prime par le montant moyen par sinistre, estimé à 9701 euros, la police 7001497 paiera une prime pure de 1213 euros soit presque cinq fois moins que la prime collective de 5625 euros.

A contrario, les plus grands risques seront pénalisés. Par exemple la police 7001624 paiera 2,22 au lieu de 0,58 soit presque quatre fois la prime collective. Néanmoins, par effet de mutualisation, sa prime de crédibilité de Bühlmann reste inférieure à sa prime individuelle qui vaut 2,67 sur les trois dernières années. Autrement dit l'assuré paiera 21 505 euros plutôt que 25 869 euros, correspondant à sa sinistralité moyenne passée, mais toujours plus que la prime moyenne de 5625 euros.

La prime de Bühlmann estimée, définie par  $\widehat{P}_i^B = (1 - \widehat{z}_3)\widehat{\mu} + \widehat{z}_3\overline{X}_i$  est une fonction affine de la moyenne individuelle  $\overline{X}_i$ . Cela implique que la sinistralité moyenne annuelle, égale à 3 628 170 euros, est conservée par cette nouvelle distribution des coûts.

### 4.3.2 Montant de sinistres

Si l'on s'applique à étudier directement le montant de sinistres par assuré, les conclusions sont similaires. Au regard de la prime collective de 5625 euros, les « bons » risques sont récompensés tandis que les « mauvais » risques sont sanctionnés.

numéro de police	montant de sinistres			primes			prime pure (en euros)	
	2012	2013	2014	individuelle	collective	Bühlmann		
7001497	0	0	0	0	5625	2285	2285	
7001624	17003	17525	25000	19843	5625	14067	14067	
7002996	0	0	0	0	5625	2285	2285	
7003007	0	18000	10597	9532	5625	7945	7945	
7003020	2119	0	0	706	5625	2704	2704	
7003032	0	0	0	0	5625	2285	2285	
7003050	27312	19259	40440	29004	5625	19507	19507	
7003064	0	0	5250	1750	5625	3324	3324	
montant annuel moyen des sinistres de 2012 à 2014								3 628 170

TABLE 4.4 – Montant de sinistres observés de 2012 à 2014

Toutefois le facteur de crédibilité est moindre pour le montant que pour le nombre de sinistres à savoir 59,3% contre 78,4%.

Pour le montant de sinistres, les calculs menés sous R fournissent les estimations suivantes

$$\begin{aligned}\widehat{\mu} &= 5625,07 & \widehat{\Sigma}^2 &= 194260119 \\ \widehat{M}^2 &= 94647660 & \widehat{z}_t &= 0,5937707\end{aligned}$$

Dans l'estimation de la prime de Bühlmann, la moyenne individuelle est pondérée à 59,4% contre 40,6% pour la moyenne collective. Ce second modèle accorde donc beaucoup moins de crédibilité à l'expérience de l'assuré.

Dans le modèle de crédibilité basé sur le montant de sinistres présenté en table 4.4, les meilleurs risques, ceux qui n'ont causé aucun sinistre sur les trois dernières années (e.g. la police 7001497) se verront réclamer une prime de 2285 euros au lieu de 5625 euros en moyenne. C'est presque le double du premier modèle, soit 1213 euros, et cela s'explique par la baisse de presque 20 points du facteur de crédibilité. En effet le rapport des complémentaires à 1 des facteurs de crédibilité est égale au rapport des primes lorsque la moyenne individuelle est nulle. Cet écart se lit sur l'axe des ordonnées du schéma 4.5.

A contrario, les plus grands risques (e.g. la police 7003050) paieront 19507 euros au lieu de 5625 euros en moyenne soit presque quatre fois la prime collective. Néanmoins, par effet de mutualisation, la prime de crédibilité de Bühlmann reste inférieure à la prime individuelle qui vaut 29004 euros.

### 4.3.3 Système de primes

Le schéma 4.5 nous indique que la prime de Bühlmann opère une nouvelle répartition, plus équitable, autour de la moyenne collective  $\widehat{\mu}$ ; plus précisément

- ✓ si  $\overline{X}_i < \widehat{\mu}$  alors  $\overline{X}_i < \widehat{P}_i^B < \widehat{\mu}$ , la prime de l'assuré contribue à la mutualisation mais reste inférieure à la prime collective;
- ✓ si  $\overline{X}_i = \widehat{\mu}$  alors  $\overline{X}_i = \widehat{P}_i^B = \widehat{\mu}$ , la prime de l'assuré est inchangée;
- ✓ si  $\widehat{\mu} < \overline{X}_i$  alors  $\widehat{\mu} < \widehat{P}_i^B < \overline{X}_i$ , la prime de l'assuré bénéficie de la mutualisation mais reste supérieure à la prime collective.

Se pose alors la question de faire un choix parmi les deux primes proposées. La prime de Bühlmann calculée à partir du nombre de sinistres est plus équitable car elle se rapproche plus de la diagonale que la prime de Bühlmann calculée à partir du montant de sinistres.

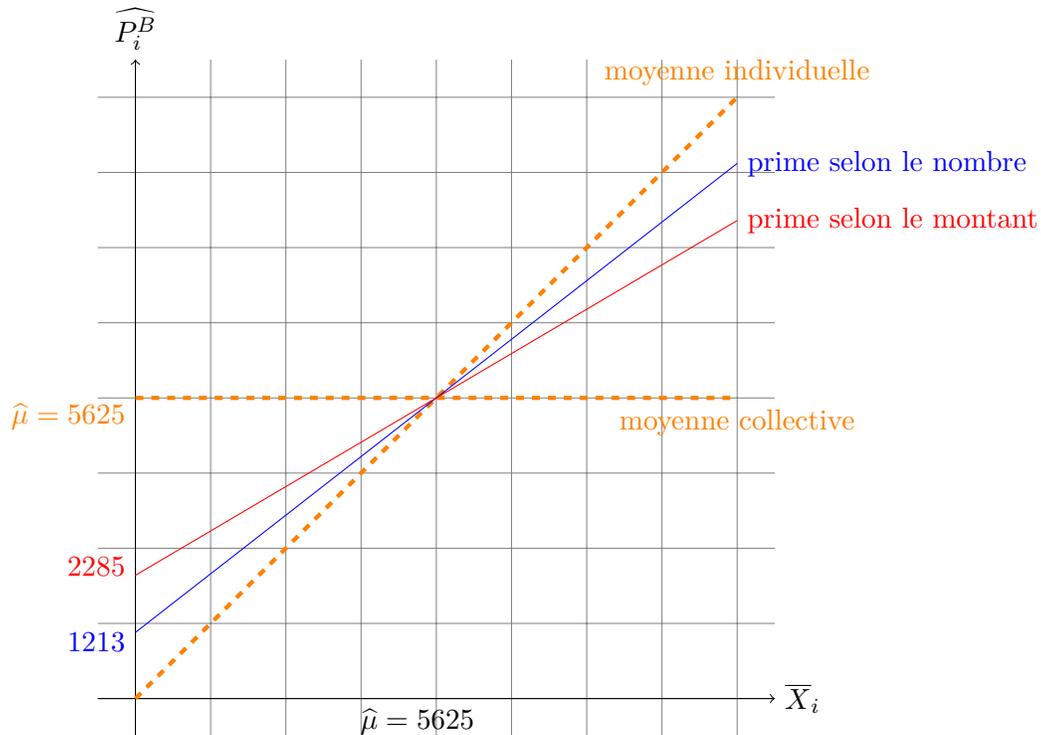


TABLE 4.5 – Primes de Bühlmann

En conclusion le modèle de Bühlmann offre la possibilité de personnaliser la prime d'un assuré autour de la moyenne collective en intégrant son historique personnel selon « la crédibilité des données » qui augmente avec leur hétérogénéité. Toutefois ce modèle présente plusieurs écueils pour la garantie assurance-crédit, notamment

- ✓ deux assurés présentant le même historique de sinistralité se voient attribuer la même prime même si leurs tailles diffèrent ;
- ✓ l'évolution de la taille des portefeuilles d'acheteurs d'une année sur l'autre n'est pas prise en compte.

Cette première illustration met en évidence la nécessité d'introduire la notion de « volume de risque » afin d'affiner la prédiction. C'est l'objet du chapitre suivant.

# Chapitre 5

## Le modèle de Bühlmann-Straub

### Sommaire

---

<b>5.1</b>	<b>Notion de volume de risque</b>	<b>32</b>
<b>5.2</b>	<b>Hypothèses et notations</b>	<b>33</b>
<b>5.3</b>	<b>Calcul de la prime</b>	<b>34</b>
<b>5.4</b>	<b>Estimation des paramètres</b>	<b>35</b>
<b>5.5</b>	<b>Application à l'assurance-crédit</b>	<b>35</b>
5.5.1	Le ratio de sinistralité retenu	35
5.5.2	les facteurs de crédibilité	36
5.5.3	Le tarif par expérience	37
<b>5.6</b>	<b>Discussion sur la mise en oeuvre du modèle</b>	<b>40</b>

---

Le modèle de Bühlmann–Straub est une extension du précédent qui prend en compte l'exposition au risque. Cette approche est intéressante dans les situations où les tailles des contrats varient beaucoup entre assurés et au cours du temps.

En assurance-crédit, l'exposition d'une entreprise de cent millions d'euros de chiffre d'affaires n'est pas comparable à celle d'une entreprise de taille plus modeste. Or à sinistralité égale, le modèle de Bühlmann attribuerait la même crédibilité aux deux entreprises ; ce qui n'est pas intuitif.

Toutes autres choses étant égales par ailleurs, il paraît naturel de penser que l'expérience de la plus grande entreprise devrait se montrer plus stable dans le temps et par conséquent mériter une plus grande crédibilité dans un système de tarification basée sur l'expérience.

## 5.1 Notion de volume de risque

Certaines branches d'assurance ont en commun le fait que la prime pure croît avec la taille de l'affaire assurée et se calcule comme le produit d'un taux par une mesure de volume appropriée, par exemple la masse salariale d'une entreprise en assurance collective ou le nombre de véhicules dans une flotte en assurance auto.

Dans le cadre de l'assurance-crédit, une mesure de volume pourrait être le chiffre d'affaires assurable qui fait l'objet à l'heure actuelle d'une déclaration par l'assuré mais n'est pas toujours présent dans les bases informatiques.

On imagine la situation où des regroupements  $i$  d'assurés observés au temps  $j$  porte sur  $w_{ij}$  contrats  $k$  de volume unitaire produisant des charges de sinistres  $\xi_{ijk}$ . Ces contrats élémentaires, supposés indépendants, sont suffisamment homogènes en terme de risque pour que la classe porte un risque aléatoire unique  $\Theta_i$  avec

$$\begin{aligned} \mathbb{E}[\xi_{ijk}|\Theta_i] &= \mu(\Theta_i) \\ \text{Cov}[\xi_{ijk}, \xi_{ijl}|\Theta_i] &= \delta_{kl}\sigma^2(\Theta_i) \end{aligned}$$

Par la suite, on s'intéresse au ratio de sinistralité de la classe  $i$  au temps  $j$  défini comme

$$X_{ij} = \frac{1}{w_{ij}} \sum_{k=1}^{w_{ij}} \xi_{ijk}$$

Ses deux premiers moments conditionnels au risque  $\Theta_i$  valent

$$\begin{aligned} \mathbb{E}[X_{ij}|\Theta_i] &= \mu(\Theta_i) \\ \text{Var}[X_{ij}|\Theta_i] &= \frac{1}{w_{ij}}\sigma^2(\Theta_i) \end{aligned}$$

Ainsi à chaque temps  $j$ , pour chaque classe de contrats  $i$ , on observe le ratio  $X_{ij}$  associé à  $w_{ij}$  unités de la mesure de volume retenue. Cette dernière quantité  $w_{ij}$  peut être vue comme l'exposition au risque du contrat  $i$ .

Cette formulation conduit naturellement aux hypothèses du modèle de Bühlmann–Straub. Dans le cadre de l'assurance-crédit, on prendra pour poids  $w_{ij}$  le nombre ou le montant des agréments et pour charge de sinistres  $\xi_{ijk}$  le nombre ou le montant de sinistres.

## 5.2 Hypothèses et notations

On considère l'espace des fonctions aléatoires de carré intégrable  $L^2$ . A chaque observation du ratio de sinistralité  $X_{ij}$ , le modèle de Bühlmann–Straub associe un poids déterministe  $w_{ij}$  qui représente l'exposition au risque. Le modèle porte sur un portefeuille de  $n$  contrats similaires  $i$  observés sur  $t$  périodes  $j$ . Il utilise les hypothèses suivantes

- ✓ les variables  $\Theta_i$  sont indépendantes et identiquement distribuées
- ✓ les vecteurs  $(\Theta_i, X_{i1}, \dots, X_{it})$  sont indépendants
- ✓ les observations  $X_{ij}$  sont indépendantes conditionnellement à  $\Theta_i$
- ✓ les deux premiers moments sont finis tels que

$$E[X_{ij}|\Theta_i] = \mu(\Theta_i) \qquad \text{Var}[X_{ij}|\Theta_i] = \frac{1}{w_{ij}}\sigma^2(\Theta_i)$$

La variance de l'observation  $X_{ij}$  conditionnellement à  $\Theta_i$  est inversement proportionnelle à son poids. Ainsi une grande exposition au risque entraîne une expérience plus « stable ».

Le portefeuille possède plusieurs paramètres de structure le caractérisant et qui devront être également estimés afin de calculer la prime de crédibilité

- ✓ la moyenne a priori  $\mu = E[X_{ij}] = E[\mu(\Theta_i)]$
- ✓ la variance a priori  $\sigma^2 = \text{Var}[X_{ij}] = \frac{1}{w_{ij}}\Sigma^2 + M^2$
- ✓ la variance intra  $\Sigma^2 = E[\sigma^2(\Theta_i)] = w_{ij} E[\text{Var}[X_{ij}|\Theta_i]]$  ou variabilité due au hasard
- ✓ la variance inter  $M^2 = \text{Var}[\mu(\Theta_i)] = \text{Var}[E[X_{ij}|\Theta_i]]$  ou hétérogénéité

Dans la suite nous aurons recours aux notations ci-dessous

$$\begin{aligned} X_{i\bullet} &= \frac{1}{w_{i\bullet}} \sum_{j=1}^t w_{ij} X_{ij} & w_{i\bullet} &= \sum_{j=1}^t w_{ij} \\ X_{\bullet\bullet} &= \frac{1}{w_{\bullet\bullet}} \sum_{i=1}^n w_{i\bullet} X_{i\bullet} = \frac{1}{w_{\bullet\bullet}} \sum_{i=1}^n \sum_{j=1}^t w_{ij} X_{ij} & w_{\bullet\bullet} &= \sum_{i=1}^n w_{i\bullet} = \sum_{i=1}^n \sum_{j=1}^t w_{ij} \end{aligned}$$

### 5.3 Calcul de la prime

On souhaite approcher la prime individuelle  $\mu(\Theta_i) = E[X_{i,t+1}|\Theta_i]$  par un estimateur qui est affine en les observations et utilise toute l'information disponible. Du fait de l'indépendance des contrats, la prime de l'assuré  $i$  sera une fonction de ses seules observations  $X_{ij}$ . Sa détermination se ramène à la recherche des constantes  $c_0, c_1, \dots, c_t$  qui minimisent l'écart quadratique moyen

$$E \left[ \left( \mu(\Theta_i) - c_0 - \sum_{j=1}^t c_j X_{ij} \right)^2 \right]$$

L'atteinte du minima conduit à deux coefficients qui sont logiquement propres à chaque assuré  $i$

$$c_0 = \frac{\Sigma^2}{\Sigma^2 + w_{i\bullet} M^2} \mu = (1 - z_i) \mu$$

$$c_j = \frac{w_{ij} M^2}{\Sigma^2 + w_{i\bullet} M^2} = z_i \frac{1}{w_{i\bullet}}$$

La solution qui minimise l'écart quadratique est une moyenne pondérée de la prime collective et de la moyenne individuelle. Il s'agit donc d'une prime de crédibilité qui s'écrit

$$P_{cred}^i = \left(1 - \frac{w_{i\bullet} M^2}{\Sigma^2 + w_{i\bullet} M^2}\right) \mu + \left(\frac{w_{i\bullet} M^2}{\Sigma^2 + w_{i\bullet} M^2}\right) \frac{1}{w_{i\bullet}} \sum_{j=1}^t w_{ij} X_{ij} = (1 - z_i) \mu + z_i X_{i\bullet}$$

L'estimateur  $P_{cred}^i$  converge presque sûrement vers  $\mu(\Theta_i)$  avec  $E[P_{cred}^i] = E[\mu(\Theta_i)]$ .

Le calcul de la prime met en évidence une quantité spécifique à chaque assuré  $i$  qui est son facteur de crédibilité

$$z_i = \frac{w_{i\bullet} M^2}{\Sigma^2 + w_{i\bullet} M^2}$$

Sans expérience passée i.e. les poids sont nuls, on a  $z_i = 0$  et la prime est la moyenne collective  $\mu$ . Le facteur de crédibilité croît avec le temps et l'hétérogénéité du portefeuille - soit  $M^2 > 0$  - et accorde alors plus d'importance à la prime d'expérience. A contrario la crédibilité diminue avec l'importance des fluctuations de sinistralité observées au sein du portefeuille.

Tout l'intérêt de ce modèle réside dans le facteur de crédibilité individuel  $z_i$  qui croît avec le poids des observations en convergeant vers l'unité, et accorde plus de crédit à l'expérience individuelle puisque  $P_{cred}^i = (1 - z_i) \mu + z_i X_{i\bullet}$ .

## 5.4 Estimation des paramètres

Les statistiques suivantes sont des estimateurs sans biais et convergents des paramètres structuraux à partir desquels se calcule l'estimateur empirique de crédibilité linéaire

$$\widehat{\Sigma^2} = \frac{1}{n(t-1)} \sum_{i=1}^n \sum_{j=1}^t w_{ij} (X_{ij} - X_{i\bullet})^2$$

$$\widehat{M^2} = \frac{w_{\bullet\bullet}}{w_{\bullet\bullet}^2 - \sum_{i=1}^n w_{i\bullet}^2} \left( \sum_{i=1}^n w_{i\bullet} (X_{i\bullet} - X_{\bullet\bullet})^2 - (n-1) \widehat{\Sigma^2} \right)$$

$$\widehat{\mu} = X_{\bullet\bullet} = \frac{1}{w_{\bullet\bullet}} \sum_{i=1}^n w_{i\bullet} X_{i\bullet}$$

Par ailleurs, on peut démontrer qu'en théorie de la crédibilité, l'estimateur linéaire de  $\mu$  à variance minimale est plutôt

$$\mu^* = \frac{1}{z_{\bullet}} \sum_{i=1}^n z_i X_{i\bullet}$$

Formellement  $\mu^*$  n'est pas un estimateur puisqu'il dépend des paramètres  $\Sigma^2$  et  $M^2$ . Le logiciel *R* utilise par défaut ce pseudo-estimateur dans la fonction *cm* du package *actuar*.

## 5.5 Application à l'assurance-crédit

### 5.5.1 Le ratio de sinistralité retenu

Un chiffre d'affaires assurable important fait l'objet en général d'une couverture par un grand nombre d'agréments. Ainsi l'unité de volume opportune est le montant des agréments et on s'intéresse au rapport du montant de sinistres au montant des agréments.

Plus précisément on considère comme volume de risque le poids  $w_{ij} = A_{ij}$  qui est la somme des agréments (en  $k\text{€}$ ) donnés par l'assureur et comme charge de sinistres  $\xi_{ijk} = S_{ijk}$  dont la somme correspond au total des sinistres de l'assuré  $i$  l'année  $j$ . Ainsi le ratio de sinistralité étudié s'écrit

$$X_{ij} = \frac{1}{w_{ij}} \sum_{k=1}^{w_{ij}} \xi_{ijk} = \frac{S_{ij}}{A_{ij}}$$

## 5.5.2 les facteurs de crédibilité

L'extrait du portefeuille présenté dans la table 5.1 montre une forte hétérogénéité puisque certaines polices ne sont pas sinistrées (e.g. la police 7001497) et à l'inverse d'autres présentent des ratios de sinistralité élevés (e.g. la police 7001624).

numéro de police	ratios ( $\times 1000$ )			poids (en $k\text{€}$ )				facteur de crédibilité	primes ( $\times 1000$ )		
	2012	2013	2014	2012	2013	2014	total		indi.	coll.	Bühl.
7001497	0.00	0.00	0.00	15748	17205	19430	52383	0.745	0.00	2.03	0.52
7001624	6.88	5.66	8.20	2470	3097	3048	8615	0.324	6.91	2.03	3.61
7002996	0.00	0.00	0.00	549	382	342	1273	0.066	0.00	2.03	1.90
7003007	0.00	5.19	3.10	3582	3467	3414	10463	0.368	2.73	2.03	2.29
7003020	1.11	0.00	0.00	1908	2150	2242	6300	0.260	0.34	2.03	1.59
7003032	0.00	0.00	0.00	186	77	77	340	0.019	0.00	2.03	2.00
7003050	6.57	4.32	9.07	4156	4455	4457	13068	0.421	6.66	2.03	3.98
7004010	10.44	0.00	0.00	3513	3636	3996	11145	0.383	3.29	2.03	2.51

TABLE 5.1 – Ratios de sinistralité observés de 2012 à 2014

Pour ce ratio de sinistralité les calculs menés sous  $R$  fournissent les estimations suivantes

$$\hat{\mu} = 2,03 \qquad \widehat{\Sigma^2} = 67656,68 \qquad \widehat{M^2} = 3,76$$

Dans le modèle Bühlmann-Straub chaque assuré possède son propre facteur de crédibilité qui dépend de ses pondérations et se calcule comme

$$z_i = \frac{w_{i\bullet} M^2}{\Sigma^2 + w_{i\bullet} M^2}$$

Ainsi des assurés de tailles comparables seront tarifés selon des systèmes de prime de Bühlmann proches (e.g. les polices 7003007 et 7004010).

En outre plus l'assuré possède un poids important plus sa droite est pentue et proche de la première bissectrice. Cela correspond à une prime plus équitable donnant plus de crédibilité à l'expérience personnelle mais également à une moindre mutualisation (e.g. la police 7001497).

A contrario les « petits » assurés seront considérés comme moins crédibles et se verront réclamer une prime plus proche de la prime collective. Par ailleurs pour un « petit » assuré, si sa moyenne individuelle est éloignée de la moyenne collective, il peut être soit fortement bénéficiaire soit fortement contributeur à la mutualisation (e.g. la police 7003032).

Sur la figure 5.2 la zone orange correspond à une contribution à la mutualisation contrairement à la zone rouge qui correspond à une subvention de la mutualisation.

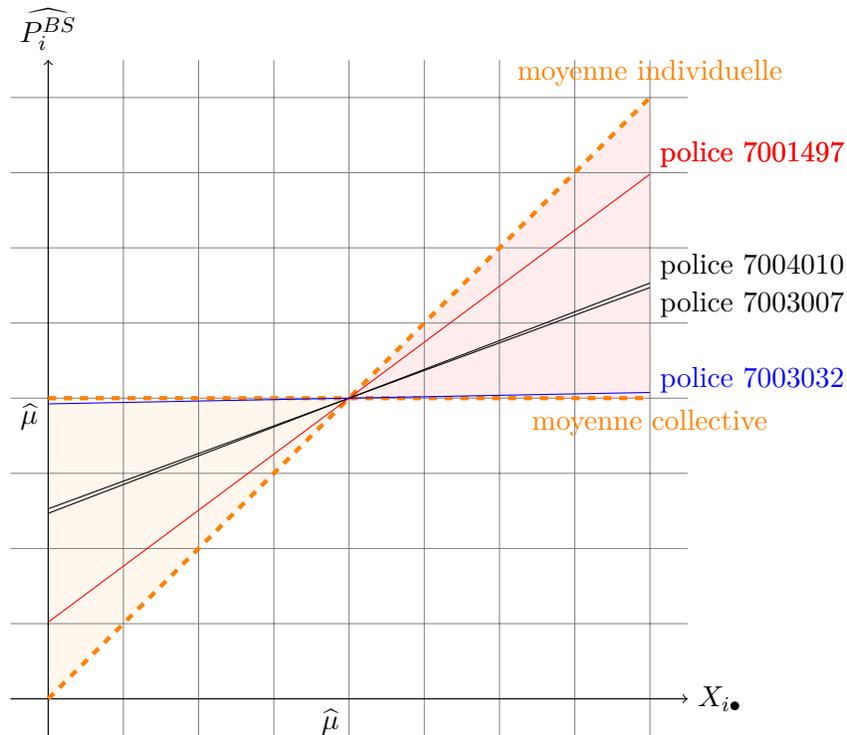


TABLE 5.2 – Primes de Bühlmann-Straub

### 5.5.3 Le tarif par expérience

Dans le modèle de Bühlmann-Straub la prime est une combinaison linéaire des moyennes individuelles et collective pondérées par le facteur de crédibilité, en résumé

$$X_i^{BS} = (1 - z_i)\mu + z_i X_{i\bullet} \text{ où } z_i = \frac{w_{i\bullet} M^2}{\Sigma^2 + w_{i\bullet} M^2} \text{ avec } X_{it+1} = \frac{S_{it+1}}{A_{it+1}}$$

Au final la prime pure pour l'année 2015 se calcule comme le produit

$$\widehat{S_{i2015}^{BS}} = [(1 - \hat{z}_i)\hat{\mu} + \hat{z}_i X_{i\bullet}] A_{i2015}$$

Selon le tableau 5.3 la prime prime de Bühlmann-Straub est suffisante pour couvrir les sinistres. Elle est plus équitable et globalement moins chère que la prime actuelle qui est une prime de « marché ».

montant total sur le portefeuille (en €)	mean square error
sinistres	2 643 416
prime actuelle	3 694 261
prime « Bühlmann-Straub »	3 440 487
prime « Bâle 2 »	3 758 107
prime « fréquence - coût »	3 102 278

TABLE 5.3 – Couverture de la sinistralité 2015

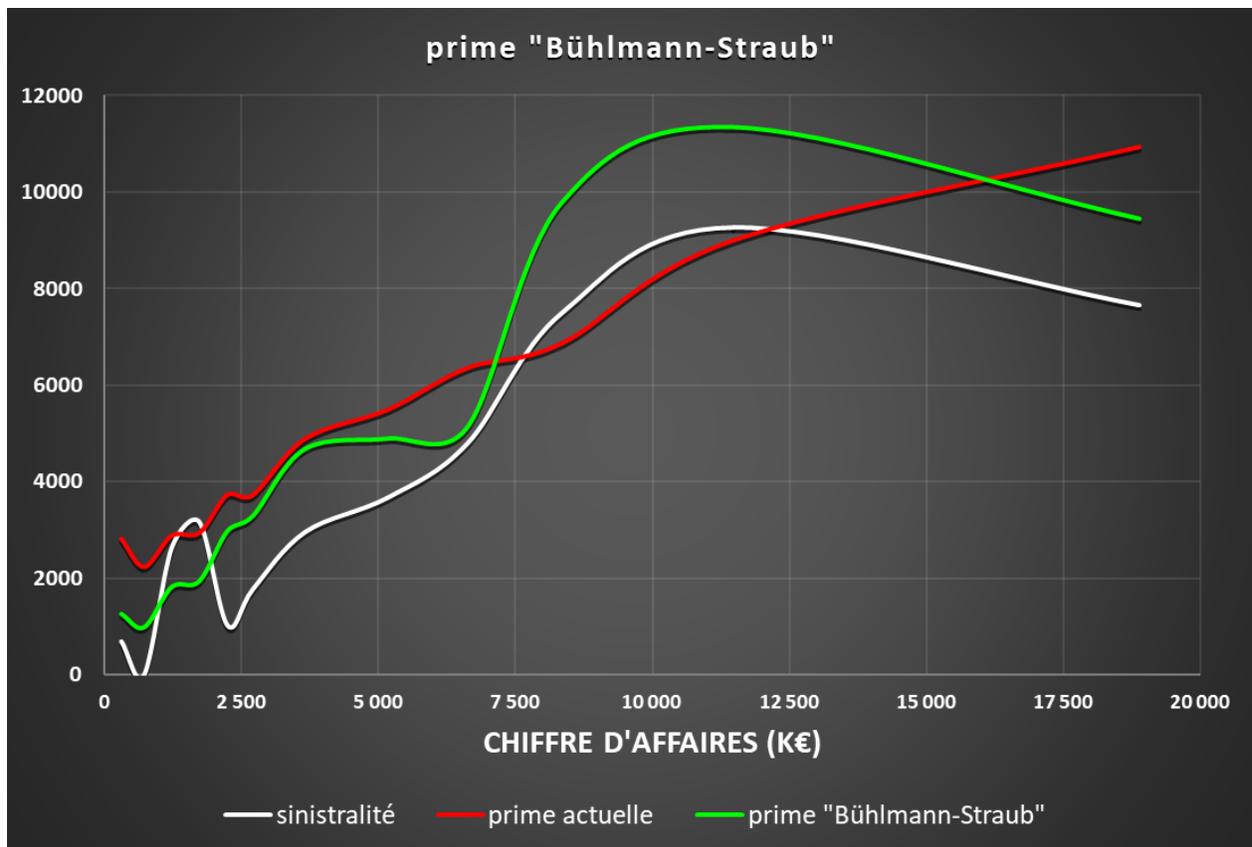


TABLE 5.4 – Prime « Bühlmann-Straub »

La représentation de la prime par classes de chiffres d'affaires montrerait que le modèle n'est pas forcément pertinent sur les petites entreprises. Toutefois cette représentation qui correspond à un usage du métier n'est pas entièrement satisfaisante car les petites entreprises possèdent un portefeuille de clients limités ce qui limite la portée des méthodes de crédibilité.

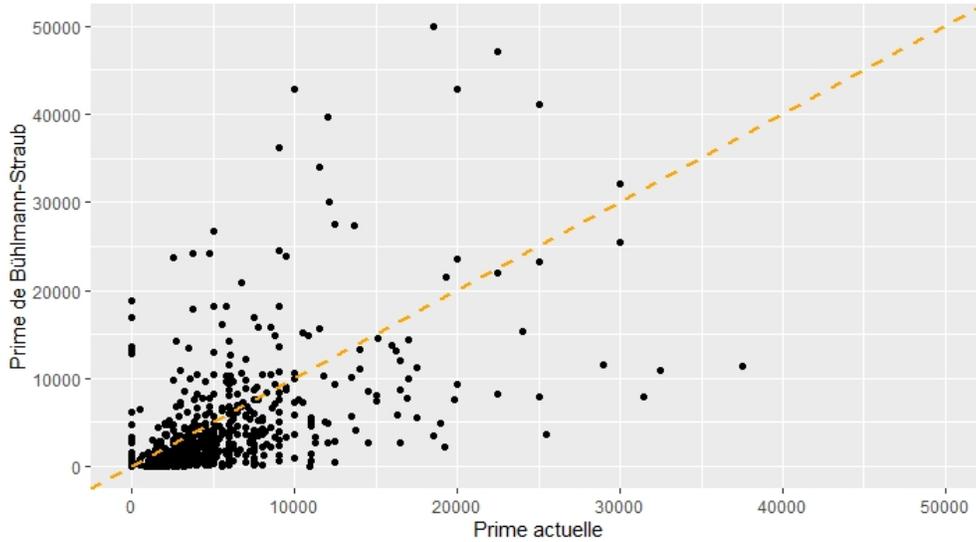


TABLE 5.5 – Prime « Bühlmann-Straub »

La tarification actuelle est basée sur des taux de marché décroissants en fonction du chiffre d'affaires déclaré alors que la prime de Bühlmann-Straub est personnalisée en fonction de la taille de l'assuré et de son historique de sinistres. Naturellement, les deux primes ne coïncident pas. Dans chaque tranche de chiffre d'affaires il existe des portefeuilles plus ou moins risqués qui ne devrait pas bénéficier du même taux de prime sauf à raisonner en moyenne et ne pas s'autoriser un sélection plus fines des assurés.

Le modèle de Bühlmann–Straub est une généralisation du précédent qui tient compte de l'exposition au risque des contrats. Ceci est particulièrement important dans les situations où la taille des contrats varie beaucoup. Cette approche permet de rendre la prime plus équitable en tenant compte de la sinistralité passée et de la « crédibilité » de chaque assuré.

Les résultats obtenus sont convenables néanmoins l'hypothèse d'indépendance des vecteurs  $(\Theta_i, X_{i1}, \dots, X_{it})$  est fragile. En effet un acheteur peut être présent dans plusieurs portefeuilles. Néanmoins l'acheteur « fragilisé » peut choisir de régler en priorité certaines factures et les fournisseurs traitent plus ou moins activement leurs impayés.

On remarquera enfin que cette approche est intéressante pour construire des variables synthétique de type « bon ou mauvais » payeur qui pourront intervenir dans un modèle linéaire généralisé par exemple.

## 5.6 Discussion sur la mise en oeuvre du modèle

De prime abord le modèle de « Bühlmann-Straub » paraît applicable avec intérêt à l'assurance-crédit. La prime de l'assuré ou du prospect  $i$  pour l'année  $j$  se calcule par une formule extrêmement simple  $[(1 - \hat{z}_i)\hat{\mu} + \hat{z}_i X_{i\bullet}] A_{ij}$ .

Pour un prospect sa crédibilité est nulle soit  $\hat{z}_i = 0$  et la somme de ses agréments pour l'année à venir  $A_{ij}$  est approchée soit par une déclaration de chiffre d'affaires assurable futur soit à partir de l'échantillon d'acheteurs communiqué par le prospect à l'assureur.

La principale critique de ce modèle porte sur le respect de l'hypothèse d'indépendance mutuelle des vecteurs  $(\Theta_i, X_{i1}, \dots, X_{it})$ . En effet dans le cadre de l'assurance-crédit les ratios de sinistralité  $X_{ij}$  ne sont pas totalement indépendants puisque au numérateur la sinistralité  $S_{ij}$  est la somme des impayés des acheteurs. Un acheteur en difficulté ou indélicat l'est probablement auprès de plusieurs de ses fournisseurs assurés auprès de la compagnie.

Par ailleurs dans une économie cyclique, les défauts des entreprises ont tendance à survenir plus fréquemment lorsque la situation de l'économie globale ou d'un secteur spécifique se dégrade. Ces défauts ne sont pas indépendants.

D'un autre côté, il est certain qu'un assuré qui est attentif à la gestion de ses factures connaîtra moins d'événements négatifs, ce qui est susceptible d'être capté par le modèle toutes choses égales par ailleurs. Cette approche est intrinsèquement moins pertinente que dans le suivi des accidents de travail dans une usine où l'assuré maîtrise son environnement.

Bien qu'on montre au chapitre précédent que cette prime est suffisante; elle doit encore être linéarisée pour se conformer aux usages du métier. Sa sensibilité au « choc de défaillances » doit être aussi étudiée.

En conclusion, l'utilisation de la prime de « Bühlmann-Straub » comme système de tarification nécessite des études complémentaires. Cependant elle propose un outil pour repérer les « bons ou mauvais risques » et quantifier un éventuel bonus ou malus dans le cadre d'une négociation commerciale sur un portefeuille nonobstant la dépendance des risques.

Elle permet également de créer des indicateurs synthétiques dans une approche de « feature engineering » : facteur de crédibilité, prime et volume de risque.

## Troisième partie

### Les modèles « fréquence - coût »

# Chapitre 6

## Le modèle linéaire général

### Sommaire

---

<b>6.1</b>	<b>Famille exponentielle</b>	<b>43</b>
6.1.1	Définition	43
6.1.2	Exemple	43
<b>6.2</b>	<b>Modèle linéaire généralisé</b>	<b>44</b>
6.2.1	Composantes du modèle	44
6.2.2	Estimation des paramètres	45
<b>6.3</b>	<b>Inférence et outils de diagnostics</b>	<b>45</b>
6.3.1	Sélection de modèles	46
6.3.2	Qualité d'ajustement	46
6.3.3	Analyse des résidus	47
6.3.4	Points influents	47

---

Les modèles linéaires généralisés constituent une extension du modèle linéaire permettant de regrouper dans un cadre unifié des variables non normales. Cette approche désormais classique a été développée initialement par J.Nelder et R.Wedderburn en 1972.

Ce chapitre propose de brefs rappels sur cette approche qui cherche à exprimer l'espérance d'une variable réponse en fonction d'une combinaison linéaire des variables explicatives. Elle sera utilisée dans les chapitres qui suivent pour déterminer et segmenter la prime pure.

Ce chapitre s'appuie sur les livres de J.Hardin [HARD 2018], de E.Ohlsson [OHLS 2010], de P.Jong [JONG 2008] ainsi que sur les cours de F.Planchet [PLAN 2004].

## 6.1 Famille exponentielle

### 6.1.1 Définition

La distribution de probabilités d'une variable aléatoire  $Y$  appartenant à la famille exponentielle de paramètre d'intérêt  $\theta$  et de paramètre de dispersion  $\phi$  se présente sous la forme

$$f_Y(y, \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right)$$

avec les hypothèses suivantes

- ✓  $a$  est une fonction dérivable non nulle
- ✓  $b$  est deux fois dérivable et de dérivée inversible
- ✓  $c$  est dérivable et ne dépend pas du paramètre naturel  $\theta$

On vérifie que  $E[Y] = b'(\theta)$  et  $\text{Var}[Y] = a(\phi)b''(\theta)$ . Les deux moments sont liés par la fonction variance  $\text{Var}[Y] = a(\phi)v(E[Y])$ .

### 6.1.2 Exemple

La loi Gamma fait partie de la famille exponentielle car sa distribution se reformule en

$$\frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^\nu y^{\nu-1} \exp\left(-\frac{\nu y}{\mu}\right) = \exp\left(\frac{-\frac{y}{\mu} - (-\ln(-(-\frac{1}{\mu})))}{\frac{1}{\nu}} + \left(\frac{1}{\frac{1}{\nu}} - 1\right) \ln y - \ln \Gamma\left(\frac{1}{\frac{1}{\nu}}\right) + \left(\frac{1}{\frac{1}{\nu}}\right) \ln\left(\frac{1}{\frac{1}{\nu}}\right)\right)$$

en effectuant les identifications suivantes

- ✓ pour le paramètre d'intérêt  $\theta = -\frac{1}{\mu}$  et  $b(\theta) = -\ln(-\theta)$
- ✓ pour le paramètre de dispersion  $\phi = \frac{1}{\nu}$  et  $a(\phi) = \phi$
- ✓  $c(y, \phi) = \left(\frac{1}{\phi} - 1\right) \ln y - \ln \Gamma\left(\frac{1}{\phi}\right) + \frac{1}{\phi} \ln \frac{1}{\phi}$

Les deux premiers moments de  $Y$  se recalculent comme  $E[Y] = \mu = b'(\theta) = -\frac{1}{\theta} = -\frac{1}{-\frac{1}{\mu}}$  et  $\text{Var}[Y] = a(\phi)b''(\theta) = \frac{1}{\nu} \frac{1}{\theta^2} = \frac{\mu^2}{\nu}$ ; on en déduit  $v(\mu) = \mu^2$ .

## 6.2 Modèle linéaire généralisé

### 6.2.1 Composantes du modèle

La modélisation a pour objectif de prédire la variable à expliquer  $Y$  à l'aide des variables explicatives  $X_0 = \mathbb{1}_n, X_1, \dots, X_p$  observées sur un échantillon déterministe de taille  $n$  et regroupées dans la matrice  $\mathbf{X}_{n \times p+1}$  des observations.

Les variables  $Y_i$  sont supposées indépendantes sachant  $\mathbf{X}$  et possèdent un paramètre de dispersion  $\phi$  commun. Un modèle linéaire généralisé est caractérisé par trois éléments

- ✓ la composante aléatoire qui identifie la distribution de probabilités de la variable à expliquer  $Y_i$  et appartient à la famille exponentielle soit

$$f_{Y_i}(y_i, \theta_i, \phi) = \exp\left(\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right)$$

- ✓ la composante déterministe, produit du vecteur des paramètres inconnus  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T$  et du vecteur des observations  $x_i = (1, x_{i1}, \dots, x_{ip})^T$  de l'individu  $i$ . Elle est notée  $\eta_i$  et est appelée également « prédicteur linéaire »

$$\eta_i = (\mathbf{X}\boldsymbol{\beta})_i = x_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

- ✓ **La fonction de lien**  $g$ , supposée strictement monotone et dérivable, qui lie les composantes aléatoire et déterministe. Plus précisément l'espérance conditionnelle  $\mu_i$  de  $Y_i$  relativement aux variables explicatives est une fonction du prédicteur linéaire

$$g(\mu_i) = \eta_i$$

Pour l'assureur cette approche propose une solution simple et efficace pour segmenter la prime pure au regard de l'information qu'il détient sur les assurés. La fonction de lien logarithme permet d'obtenir un tarif multiplicatif puisque la prime se décompose en produit d'exponentielles

$$E(Y_i|x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) = \exp(\beta_0) \exp(\beta_1 x_{i1}) \dots \exp(\beta_p x_{ip})$$

## 6.2.2 Estimation des paramètres

Les coefficients  $\beta_0, \dots, \beta_p$  sont estimés à partir de l'échantillon de données en utilisant la méthode du maximum de vraisemblance. L'annulation en  $\beta_j$  des dérivées partielles de la log-vraisemblance  $\mathcal{L}$  conduit aux équations suivantes

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \mathcal{L}_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = 0$$

qui n'admettent pas en général de solutions explicites et sont résolues par des méthodes numériques itératives. En notant  $\mathcal{L}'$  le gradient et  $\mathcal{L}''$  la matrice hessienne, un développement en série de Taylor permet de construire des suites convergentes vers  $\beta$ , en particulier

- ✓ l'algorithme de Newton-Raphson

$$\beta_{n+1} = \beta_n - \{\mathcal{L}''(\beta_n)\}^{-1} \mathcal{L}'(\beta_n)$$

- ✓ l'algorithme du score de Fisher

$$\beta_{n+1} = \beta_n - \{E[\mathcal{L}''(\beta_n)]\}^{-1} \mathcal{L}'(\beta_n)$$

Le calcul des  $\beta_0, \dots, \beta_p$  est indépendant de l'estimation du paramètre de nuisance  $\phi$ . Dans la suite nous supposons que  $\phi$  est connu, ce qui simplifie la détermination des lois asymptotiques des estimateurs.

## 6.3 Inférence et outils de diagnostics

Sous certaines conditions de régularité de la densité de probabilité, l'estimateur du maximum de vraisemblance possède les propriétés suivantes

- ✓  $\beta_n$  converge en probabilité vers  $\beta$  et par conséquent est asymptotiquement sans biais
- ✓  $\beta_n$  converge en loi vers une loi gaussienne  $\mathcal{N}(\beta, \mathcal{I}^{-1})$

$\mathcal{I}$  est la matrice d'information de Fisher évaluée en les vraies valeurs des paramètres qui s'écrit

$$\mathcal{I} = -E[\mathcal{L}''(\beta)] = E[\mathcal{L}'(\beta)\mathcal{L}'(\beta)^T]$$

### 6.3.1 Sélection de modèles

En présence de plusieurs modèles, se pose la question d'en retenir un. La première situation rencontrée est celle où un modèle  $\mathcal{M}_P$  à  $p$  paramètres est emboîté dans un modèle  $\mathcal{M}_G$  à  $g$  paramètres par exemple par retrait d'un facteur explicatif. L'apport du facteur explicatif peut être testé sur la base de la différence des déviées standardisées ou de manière équivalente à l'aide du rapport des vraisemblances

$$T = D_P - D_G = 2(\ln \mathcal{L}_G - \ln \mathcal{L}_P) = \ln \left( \frac{\mathcal{L}_G}{\mathcal{L}_P} \right)^2 \underset{H_0}{\sim} \chi_{g-p}^2$$

On rejettera au niveau  $\alpha$  le modèle  $\mathcal{M}_P$  au profit du modèle  $\mathcal{M}_G$  si la valeur observée de  $T$  est supérieure au quantile  $\chi_{n-p,1-\alpha}^2$ .

Pour un jeu de données particulier, le critère AIC est particulièrement approprié pour comparer des modèles construits avec des distributions différentes. Le modèle  $M$  retenu sera celui qui minimise l'Akaike Information Criterion  $AIC_M = -2 \ln \mathcal{L}_M + 2p$ .

### 6.3.2 Qualité d'ajustement

L'adéquation d'un modèle aux données se mesure par la proximité entre estimations et observations. L'hypothèse nulle  $H_0 = \{\text{le modèle est adéquat}\}$  est rejetée avec un niveau de confiance fixé  $\alpha$  petit lorsque les statistiques de test prennent des valeurs trop importantes, les deux plus classiques étant

- ✓ le test de Pearson

$$\chi_{\text{Pearson}}^2 = \sum_{i=1}^n \frac{(y_i - \widehat{\mu}_i)^2}{\text{Var}(\widehat{\mu}_i)} \underset{H_0}{\sim} \chi_{n-p-1}^2$$

qui rejette le modèle au seuil  $\alpha$  si  $\chi_{\text{Pearson}}^2$  est supérieur au quantile  $\chi_{n-p-1,1-\alpha}^2$

- ✓ le test du rapport des vraisemblances ou de la déviance

$$D_M = 2(\ln \mathcal{L}_S - \ln \mathcal{L}_M) \underset{H_0}{\sim} \chi_{n-p-1}^2$$

Le modèle estimé  $M$  est comparé au modèle saturé  $S$  qui est caractérisé par la solution  $\widehat{\mu}_i = y_i$ , sur la base de la déviance standardisée qui est le logarithme du carré du rapport des vraisemblances. Naturellement  $D_M$  est positif et petit pour un modèle de bonne qualité. On rejettera au niveau  $\alpha$  le modèle étudié si la valeur observée de  $D_M$  est supérieure au quantile  $\chi_{n-p-1,1-\alpha}^2$ .

### 6.3.3 Analyse des résidus

L'analyse des résidus permet de détecter des individus atypiques ou de détecter des effets non linéaires. Ils peuvent être standardisés. Les plus fréquemment utilisés sont

- ✓ les résidus de Pearson

$$rp_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{Var}[\hat{\mu}_i]}}$$

- ✓ les résidus de déviance

$$rd_i = \text{signe}(y_i - \hat{\mu}_i) \sqrt{d_i}$$

Les résidus de déviance sont définis à partir d'un terme  $d_i$  représentant la contribution de l'observation  $y_i$  à la déviance  $D$ .

Ces deux résidus peuvent être normalisés par un terme  $\sqrt{h_i}$  où  $h_i$  est le levier associé à l'observation  $i$ . On vérifiera comme pour le modèle linéaire, qu'il n'existe pas de structure inattendue dans les résidus, en moyenne ou en variance.

### 6.3.4 Points influents

Les points influents sont des points qui interviennent sur le modèle de telle sorte que si on les enlève, alors l'estimation des coefficients sera fortement changée.

La mesure la plus classique d'influence est la distance de Cook. Il s'agit d'une distance entre le coefficient estimé avec toutes les observations et celui estimé avec toutes les observations sauf une

$$D_i = \frac{1}{2}(\beta - \beta_i)^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\beta - \beta_i)$$

et où  $\mathbf{W}$  est la matrice diagonale de pondération

$$\mathbf{W}_{ii} = \frac{1}{\text{Var}(Y_i)} \frac{\partial \mu_i^2}{\partial \eta_i}$$

Cet indicateur prend simultanément en compte l'effet levier et l'importance du résidu de chaque observation. Le graphe de ces valeurs est donc plus synthétique et interprétable en tenant compte du graphe des résidus et de celui des termes diagonaux de la matrice de projection.

# Chapitre 7

## La fréquence de sinistres

### Sommaire

---

<b>7.1</b>	<b>La distribution du nombre de sinistres . . . . .</b>	<b>49</b>
7.1.1	Les données observées . . . . .	49
7.1.2	Ajustement d'une loi de Poisson . . . . .	49
7.1.3	Ajustement d'une loi binomiale négative . . . . .	51
7.1.4	Test d'adéquation du $\chi^2$ . . . . .	52
<b>7.2</b>	<b>Les distributions Poisson-mélange . . . . .</b>	<b>53</b>
7.2.1	Formulation générale . . . . .	53
7.2.2	Loi Gamma à un paramètre . . . . .	54
7.2.3	Loi inverse gaussienne à un paramètre . . . . .	55
<b>7.3</b>	<b>Régression binomiale négative . . . . .</b>	<b>56</b>
<b>7.4</b>	<b>Régression Poisson Inverse Gaussienne . . . . .</b>	<b>57</b>
<b>7.5</b>	<b>Description des variables tarifaires . . . . .</b>	<b>58</b>
7.5.1	La taille du portefeuille . . . . .	58
7.5.2	La qualité du portefeuille . . . . .	59
7.5.3	Le chiffre d'affaires assurable . . . . .	59
7.5.4	La couverture géographique des agréments . . . . .	60
7.5.5	Le secteur d'activité de l'assuré . . . . .	60
<b>7.6</b>	<b>Application de la régression binomiale négative . . . . .</b>	<b>61</b>
7.6.1	Sélection des variables . . . . .	61
7.6.2	Modélisation retenue . . . . .	62

---

## 7.1 La distribution du nombre de sinistres

### 7.1.1 Les données observées

Dans ce chapitre, on considère le nombre  $N$  de sinistres indemnisés, par an et par assuré, relatifs aux agréments donnés par l'assureur. L'étude concerne 645 polices et 363 sinistres observés sur l'année 2015, pour une fréquence annuelle moyenne de 0,563 (cf. Table 7.1).

nombre observé de polices	
k	avec k sinistres
0	475
1	87
2	35
3	20
4	12
5	5
6	7
7	1
8	3
total	645

TABLE 7.1 – Fréquence  $N$  de sinistres observée

### 7.1.2 Ajustement d'une loi de Poisson

Pour modéliser des données de dénombrement, l'approche classique consiste à recourir à une loi de Poisson de paramètre  $\lambda > 0$ , définie par

$$\mathbb{P}(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

Cette distribution, appelée loi des événements rares, peut être considérée comme une approximation d'une loi binomiale lorsque le nombre d'essais  $n$  est assez grand et la probabilité d'occurrence  $p$  très petite avec  $np = \lambda$ .

Dans le cas de la loi de Poisson, il est remarquable que  $E[N] = \text{Var}[N] = \lambda$ .

En supposant que  $(n_1, \dots, n_n)$  est la réalisation d'un  $n$ -échantillon  $(N_1, \dots, N_n)$  de la variable aléatoire  $N$ , la méthode des moments et celle du maximum de vraisemblance conduisent au même estimateur sans biais du paramètre  $\lambda$  défini par

$$\hat{\lambda} = \frac{n_1 + \dots + n_n}{n} = \frac{475 \times 0 + 87 \times 1 + \dots + 3 \times 8}{645} = 0,563$$

Par ailleurs les estimateurs sans biais classiques de  $E[N]$  et  $\text{Var}[N]$  donnent

$$\widehat{E[N]} = \frac{1}{n} \sum_{i=1}^n n_i = 0,563$$

$$\widehat{\text{Var}[N]} = \frac{1}{n-1} \sum_{i=1}^n (n_i - \frac{1}{n} \sum_{i=1}^n n_i)^2 = 1,573$$

Ces valeurs largement différentes laissent présager un faible ajustement de la loi de Poisson aux données et la présence d'une forte surdispersion. La surdispersion est définie comme une variance observée supérieure à la variance théorique issue du modèle.

En simulant la sinistralité avec la loi de Poisson de paramètre  $\hat{\lambda} = 0,563$ , il apparaît clairement que cette démarche n'est pas entièrement satisfaisante.

k	nombre observé de polices	nombre simulé de polices
	avec k sinistres	avec k sinistres
0	475	367
1	87	207
2	35	58
3	20	11
4	12	2
5	5	0
6	7	0
7	1	0
8	3	0
total	645	645

TABLE 7.2 – Fréquence  $N$  de sinistres simulée avec la loi de Poisson

### 7.1.3 Ajustement d'une loi binomiale négative

En présence de surdispersion, il faut envisager l'utilisation de la loi binomiale négative de paramètres  $r > 0$  et  $0 < p < 1$  définie pour  $k = 0, 1, \dots$  par

$$\mathbb{P}(N = k) = \frac{\Gamma(k+r)}{\Gamma(k+1)\Gamma(r)}(1-p)^k p^r$$

Pour  $r$  entier la loi binomiale négative s'interprète comme la distribution du nombre  $k$  d'échecs avant d'obtenir  $r$  succès dans une série de  $k+r$  expériences de Bernoulli. Une autre paramétrisation utilise l'espérance  $\mu$  et le paramètre de surdispersion  $s$  inverse de  $r$  avec

$$\mathbb{E}[N] = r \frac{1-p}{p} = \mu \quad \text{et} \quad \text{Var}[N] = r \frac{1-p}{p^2} = \mu + s\mu^2$$

La méthode des moments conduit aux estimations suivantes

$$\hat{r} = \frac{\widehat{\mathbb{E}[N]}^2}{\widehat{\text{Var}[N]} - \widehat{\mathbb{E}[N]}} = 0,314 \quad \text{et} \quad \hat{p} = \frac{\widehat{\mathbb{E}[N]}}{\widehat{\text{Var}[N]}} = 0,358$$

L'ajustement est bien meilleur avec la loi binomiale négative dont les paramètres sont estimés par la méthode du maximum de vraisemblance. On obtient  $\hat{\mu} = 0,563$ ,  $\hat{r} = 0,277$  et  $\hat{p} = 0,330$ , valeurs différentes de celles obtenues par la méthode des moments.

k	nombre observé de polices	nombre simulé de polices
	avec k sinistres	avec k sinistres
0	475	475
1	87	88
2	35	38
3	20	19
4	12	11
5	5	6
6	7	4
7	1	2
8	3	2
total	645	645

TABLE 7.3 – Fréquence  $N$  de sinistres simulée avec la loi binomiale négative

### 7.1.4 Test d'adéquation du $\chi^2$

On considère un échantillon  $(X_1, \dots, X_n)$  d'une variable aléatoire discrète sur l'ensemble  $\{1, \dots, k\}$ . L'objectif est de tester l'hypothèse  $H_0$  « la loi de l'échantillon est égale à  $\mathbb{P}$  » contre l'hypothèse alternative  $H_1$  « la loi de l'échantillon est différente de  $\mathbb{P}$  » où  $\mathbb{P}$  désigne une loi discrète à  $p$  paramètres représentée par le vecteur  $(p_1, \dots, p_k)$ . Pour chaque  $i = 1, \dots, k$ , on considère la variable aléatoire  $N_i = \sum_{j=1}^n \mathbb{1}_{\{X_j=i\}}$  qui représente le nombre de variables de l'échantillon qui prennent la valeur  $i$ .

Pour déterminer si les écarts entre la distribution des effectifs observés et la distribution des effectifs théoriques sont significatifs ou imputables aux fluctuations d'échantillonnage, on recourt à la loi asymptotique de la quantité  $\chi^2$  sous  $H_0$

$$\chi^2 = \sum_{i=1}^k \frac{(np_i - N_i)^2}{np_i} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_{k-1-p}^2$$

On rejettera  $H_0$  au niveau  $\alpha$  si la valeur observée  $\chi^2$  est supérieure au quantile  $\chi_{k-1-p, 1-\alpha}^2$  sachant que cette situation correspond à des écarts  $np_i - N_i$  importants.

Les données ne nous permettent pas de rejeter l'hypothèse nulle «  $\mathbb{P}$  est la loi binomiale négative à deux paramètres  $\hat{r} = 0,277$  et  $\hat{p} = 0,330$  au risque  $\alpha = 0,01$  » puisque  $\chi^2 = 1,683$  est inférieur à  $\chi_4^2(0,99) = 13,276$ .

k	nombre observé de polices	nombre simulé de polices	$\chi_k^2$
	avec k sinistres	avec k sinistres	
0	475	475	0,000
1	87	88	0,011
2	35	38	0,237
3	20	19	0,053
4	12	11	0,091
5	5	6	0,167
>5	11	8	1,125
total	645	645	1,683

TABLE 7.4 – Test d'adéquation de la loi binomiale négative

## 7.2 Les distributions Poisson-mélange

Les modèles de mélange de lois forment une solution élégante pour prendre en compte la surdispersion. Ce prolongement de la loi de Poisson permet d'introduire une notion d'hétérogénéité au sein d'un groupe de contrats.

### 7.2.1 Formulation générale

Sur une période donnée, le portefeuille présente un nombre moyen de sinistres  $\lambda$ . Le modèle de Poisson peut être raffiné en rendant ce paramètre aléatoire plutôt que constant. A cet effet on introduit le couple  $(N, \Theta)$  de variables aléatoires et le réel  $\lambda > 0$  vérifiant les deux hypothèses suivantes

- ✓ la distribution conditionnelle  $N|\Theta = \theta$  est une loi de Poisson  $\mathcal{P}(\lambda\theta)$
- ✓ la distribution marginale  $\Theta$  a pour fonction de densité  $h$

En pratique l'assureur s'intéresse à la loi de  $N$  définie pour  $k = 0, 1, \dots$  par

$$\mathbb{P}(N = k) = \int_0^{+\infty} \mathbb{P}(N = k|\Theta = \theta)h(\theta)d\theta = \int_0^{+\infty} e^{-\lambda\theta} \frac{(\lambda\theta)^k}{k!} h(\theta)d\theta$$

Dans le cas particulier où  $\Theta$  suit une loi Gamma  $\gamma(a, b)$ , la distribution marginale s'écrit

$$\begin{aligned} \mathbb{P}(N = k) &= \int_0^{+\infty} e^{-\lambda\theta} \frac{(\lambda\theta)^k}{k!} \frac{b^a \theta^{a-1} e^{-b\theta}}{\Gamma(a)} d\theta \\ &= \frac{1}{\Gamma(k+1)\Gamma(a)} \lambda^k b^a \int_0^{+\infty} \theta^{k+a-1} e^{-(\lambda+b)\theta} d\theta \\ &= \frac{1}{\Gamma(k+1)\Gamma(a)} \left(\frac{\lambda}{\lambda+b}\right)^k \left(\frac{b}{\lambda+b}\right)^a \int_0^{+\infty} t^{k+a-1} e^{-t} dt \\ &= \frac{\Gamma(k+a)}{\Gamma(k+1)\Gamma(a)} \left(\frac{\lambda}{\lambda+b}\right)^k \left(\frac{b}{\lambda+b}\right)^a \end{aligned}$$

On reconnaît la loi binomiale négative  $\mathcal{BN}(a, \frac{b}{\lambda+b})$  avec  $E[N] = \lambda \frac{a}{b}$  et  $\text{Var}[N] = \lambda \frac{a}{b} (1 + \frac{\lambda}{b})$ .

Il s'agit du modèle général de mélange Poisson-Gamma dont nous retiendrons un cas particulier dans la suite.

## 7.2.2 Loi Gamma à un paramètre

Dans la suite de l'étude, on se place dans la situation très particulière où l'hétérogénéité au sein d'un groupe de contrats est contrôlée par un paramètre unique  $\alpha > 0$ .  $\Theta$  représente hétérogénéité non observable. On admet la séparabilité des deux variables aléatoires  $\lambda$  et  $\Theta$  autrement dit

$$E_\lambda[\Theta] = E[\Theta] \text{ et } \text{Var}_\lambda[\Theta] = \sigma_\lambda^2$$

Un calcul direct permet d'obtenir les deux premiers moments des lois Poisson-mélange. Si on utilise des notations extensives, l'espérance et la variance se décomposent en

$$E_\lambda[N] = E_\Theta[E_{\lambda,\Theta}[N]] = E_\lambda[\lambda\Theta] = \lambda E[\Theta]$$

et

$$\text{Var}_\lambda[N] = E_\Theta[\text{Var}_{\lambda,\Theta}[N]] + \text{Var}_\Theta[E_{\lambda,\Theta}[N]] = E_\lambda[\lambda\Theta] + \text{Var}_\lambda[\lambda\Theta] = \lambda E[\Theta] + \lambda^2 \sigma_\lambda^2$$

Dans la suite  $\Theta$  suit la loi Gamma  $\gamma(\frac{1}{\alpha}, \frac{1}{\alpha})$  telle que  $E[\Theta] = 1$  et  $\text{Var}[\Theta] = \alpha$ .

D'après la section précédente, la distribution de  $N$  est la loi binomiale négative  $\mathcal{BN}(\frac{1}{\alpha}, \frac{1}{1+\alpha\lambda})$  dont les deux premiers moments valent

$$E[N] = \lambda \text{ et } \text{Var}[N] = \lambda + \lambda^2 \alpha$$

Il existe une analogie avec le modèle d'urne à urne présenté dans le cadre de la théorie de la crédibilité. L'incertitude sur le risque réel de l'assuré n'affecte pas l'espérance mais ajoute de la surdispersion par rapport à la loi de Poisson.

Dans la littérature ce modèle de représentation de la fréquence  $N$  est appelée loi binomiale négative de type 2 et notée  $\mathcal{BN}_2$ . Cette construction offre un cadre pertinent pour segmenter efficacement le tarif à l'aide de la régression binomiale négative. On retiendra en conclusion que pour tout  $k = 0, 1, \dots$

$$\mathbb{P}(N = k) = \frac{\Gamma(k + \frac{1}{\alpha})}{\Gamma(k + 1)\Gamma(\frac{1}{\alpha})} \left( \frac{\alpha\lambda}{1 + \alpha\lambda} \right)^k \left( \frac{1}{1 + \alpha\lambda} \right)^{\frac{1}{\alpha}}$$

### 7.2.3 Loi inverse gaussienne à un paramètre

Cette section emprunte des éléments au cours en ligne « Tarification avancée : théorie et applications en R » de Jean-Philippe Boucher de l'Université du Québec à Montréal.

Une alternative à la loi Gamma est de supposer que  $\Theta$  suit une distribution inverse gaussienne  $\mathcal{IG}(\mu, \alpha)$  de paramètres  $\mu$  et  $\alpha$ , de premiers moments  $E[\Theta] = \mu$  et  $\text{Var}[\Theta] = \mu^3\alpha$  et de fonction de densité

$$h(\theta) = \frac{1}{\sqrt{2\pi\alpha\theta^3}} \exp\left(\frac{-(\theta - \mu)^2}{2\mu^2\alpha\theta}\right)$$

On s'intéresse à nouveau au cas particulier où l'hétérogénéité au sein du groupe de contrats est contrôlée par un paramètre unique  $\alpha > 0$  tels que  $E[\Theta] = 1$  et  $\text{Var}[\Theta] = \alpha$ , ce qui correspond à la loi  $\mathcal{IG}(1, \alpha)$  de densité

$$h(\theta) = \frac{1}{\sqrt{2\pi\alpha\theta^3}} \exp\left(\frac{-(\theta - 1)^2}{2\alpha\theta}\right)$$

La distribution marginale de  $N$  obtenue par ce mélange est appelée loi Poisson Inverse Gaussienne (PIG). Sa fonction de probabilités s'obtient en intégrant l'expression suivante

$$\begin{aligned} \mathbb{P}(N = k) &= \int_0^{+\infty} e^{-\lambda\theta} \frac{(\lambda\theta)^k}{k!} \frac{1}{\sqrt{2\pi\alpha\theta^3}} \exp\left(\frac{-(\theta - 1)^2}{2\alpha\theta}\right) d\theta \\ &= \frac{\lambda^k}{k!} \left(\frac{2}{\pi\alpha}\right)^{\frac{1}{2}} e^{\frac{1}{\alpha}} \left(\sqrt{1 + 2\alpha\lambda}\right)^{-(k-\frac{1}{2})} K_{k-\frac{1}{2}}\left(\frac{\sqrt{1 + 2\alpha\lambda}}{\alpha}\right) \end{aligned}$$

avec

$$K_u(t) = \int_0^{+\infty} x^{u-1} e^{-\frac{1}{2}t(x+\frac{1}{x})} dx$$

$K_u(t)$  est la fonction de Bessel modifiée de type 2.

Les différentes études comparatives menées sur les deux lois précédentes ont montré que la loi Poisson Inverse Gaussienne est plus adaptée que la loi Poisson Gamma pour modéliser les distributions présentant des queues épaisses à droite. La principale difficulté avec ses mélanges réside dans leur mise en oeuvre informatique.

Les chapitres précédents et suivants s'appuient sur les livres de A.Cameron [CAME 2014] et de Hilbe [HILB 2011] et [HILB 2014].

## 7.3 Régression binomiale négative

La régression binomiale négative rend possible la personnalisation du tarif en partitionnant le portefeuille en classes de risques homogènes sur la base de certaines caractéristiques observables des assurés. Toutefois les variables tarifaires sont loin d'expliquer complètement le risque propre des assurés.

La surdispersion résiduelle au sein de chaque classe sera prise en compte par l'inclusion d'un effet aléatoire représentant l'influence de facteurs de risque inobservables.

On dispose d'un échantillon de  $n$  observations indépendantes de  $N$  et de  $p + 1$  variables tarifaires  $X_0, X_1, \dots, X_p$ . Les régresseurs sont supposés non aléatoires et mesurables sans erreur.

Pour tout assuré  $i$  le couple de variables aléatoires  $(N_i, \Theta_i)$  vérifie

- ✓ la distribution conditionnelle  $N_i | \Theta_i = \theta_i$  suit une loi de Poisson  $\mathcal{P}(\lambda_i \theta_i)$  avec  $\lambda_i > 0$
- ✓ la distribution marginale  $\Theta_i$  est une loi Gamma  $\gamma(\frac{1}{\alpha}, \frac{1}{\alpha})$  avec  $\alpha > 0$
- ✓ il existe  $p + 1$  réels  $\beta_0, \dots, \beta_p$  inconnus tels que

$$\mathbb{E}[N_i | X_0, X_1, \dots, X_p] = \lambda_i = \exp(\beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i)$$

L'existence d'une distribution commune pour les  $\Theta_i$  signifie que les effets des variables cachées ne dépendent pas des variables observables.

Le schéma présenté au-dessus correspond à un cas particulier de mélange de type Poisson-Gamma et sous les deux premières hypothèses, il peut être montré que la loi marginale de  $N_i$  est la loi binomiale négative  $\mathcal{BN}_2(\frac{1}{\alpha}, \frac{1}{1+\alpha\lambda_i})$ .

Les deux premiers moments valent respectivement  $\mathbb{E}(N_i) = \lambda_i$  et  $\mathbb{V}(N_i) = \lambda_i + \alpha\lambda_i^2$  où le paramètre  $\alpha$  quantifie le niveau de dispersion commun autour de la moyenne  $\lambda_i$ . On retiendra que pour tout  $k = 0, 1, \dots$

$$\mathbb{P}(N = k) = \frac{\Gamma(k + \frac{1}{\alpha})}{\Gamma(k + 1)\Gamma(\frac{1}{\alpha})} \left( \frac{\alpha\lambda_i}{1 + \alpha\lambda_i} \right)^k \left( \frac{1}{1 + \alpha\lambda_i} \right)^{\frac{1}{\alpha}}$$

## 7.4 Régression Poisson Inverse Gaussienne

Les mélanges de loi sont un ensemble plus vaste que la seule classe des mélanges Poisson-Gamma. La régression Poisson Inverse Gaussienne constitue une approche complémentaire pour prendre en compte la surdispersion d'une loi à une queue épaisse à droite.

On dispose d'un échantillon  $n$  observations indépendantes de  $N$  et de  $p$  variables tarifaires  $X_1, \dots, X_p$ . Les régresseurs sont supposés non aléatoires et mesurables sans erreur.

Pour tout assuré  $i$  le couple de variables aléatoires  $(N_i, \Theta_i)$  vérifient

- ✓ la distribution conditionnelle  $N_i | \Theta_i = \theta_i$  suit une loi de Poisson  $\mathcal{P}(\lambda_i \theta_i)$  avec  $\lambda_i > 0$
- ✓ la distribution marginale  $\Theta_i$  est une loi inverse-gaussienne  $\mathcal{IG}(1, \alpha)$  avec  $\alpha > 0$
- ✓ il existe  $p + 1$  réels  $\beta_0, \dots, \beta_p$  inconnus tels que

$$\mathbb{E}[N_i | X_0, X_1, \dots, X_p] = \lambda_i = \exp(\beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i)$$

L'existence d'une distribution commune pour les  $\Theta_i$ , hypothèse forte, signifie que les effets des variables cachées ne dépendent pas des variables observables.

Le schéma présenté au-dessus correspond à un cas particulier de mélange de type Poisson-Inverse-Gaussienne (PIG).

Les deux premiers moments valent respectivement  $\mathbb{E}(N_i) = \lambda_i$  et  $\mathbb{V}(N_i) = \lambda_i + \alpha \lambda_i^2$ ; le paramètre  $\alpha$  quantifie le niveau de dispersion commun autour de la moyenne  $\lambda_i$ .

Les paramètres des deux modèles de régression précédents sont le plus souvent estimés, à l'instar des modèles linéaires généralisés, selon le principe du maximum de vraisemblance. Ces estimations utilisent sauf exception des méthodes numériques itératives classiques de type Newton-Raphson.

Les variables tarifaires  $X_0, X_1, \dots, X_p$  sont regroupées dans la matrice  $\mathbf{X}_{n \times p+1}$  des observations. Une variable catégorielle à  $k$  modalités sera représentée par  $k - 1$  variables indicatrices.

## 7.5 Description des variables tarifaires

L'étude des variables tarifaires est un travail long et fastidieux. Les variables continues sont transformées en variables catégorielles qui sont plus simples à appréhender par les utilisateurs. Par la suite le nombre de classes retenues pour chaque variable est volontairement limité. D'une part en raison d'un échantillon d'étude assez réduit qui ne permet pas de segmenter trop finement. D'autre part dans l'optique de conserver un tarif qui soit facile à mettre en oeuvre.

Pour chaque catégorie  $C_i$  de chaque variable susceptible de participer au modèle, sa fréquence moyenne de sinistres  $\lambda_i$  est estimée sur sa population.

Cinq variables sont étudiées, ce qui est faible et montre tout l'enjeu de recueillir et traiter des informations pertinentes sur l'assuré et la garantie tout en restant conforme au Règlement Général sur la Protection des Données par exemple.

### 7.5.1 La taille du portefeuille

Naturellement un portefeuille d'acheteurs de taille importante connaîtra toutes choses étant égales par ailleurs un grand nombre de sinistres. Cet effet « taille » est pris en compte à l'aide de la variable qui compte le nombre d'agréments demandés et obtenus par l'assuré.

La variable possède six classes qui présentent des fréquences moyenne de sinistres très différenciées.

classe	nombre de polices	nombre de sinistres	$\lambda_i$
$C_1$	181	9	0,050
$C_2$	119	29	0,244
$C_3$	146	67	0,459
$C_4$	133	115	0,865
$C_5$	42	71	1,690
$C_6$	24	72	3,000
total	645	363	0,563

TABLE 7.5 – Nombre d'agréments

## 7.5.2 La qualité du portefeuille

Lors d'une demande d'agrément, l'acheteur fait l'objet d'une étude de solvabilité et reçoit une notation interne. Les portefeuilles présentant un pourcentage élevé d'acheteurs bien notés sont moins sinistrés. La variable retenue est la proportion d'acheteurs dans le portefeuille dont le rating est supérieur ou égal à 6. Il n'a pas été possible d'obtenir beaucoup de classes en partant d'un découpage en déciles. Ce qui traduit un effet fortement non linéaire.

classe	nombre de polices	nombre de sinistres	$\lambda_i$
$C_1$	345	253	0,733
$C_2$	94	52	0,553
$C_3$	142	51	0,359
$C_4$	64	7	0,109
total	645	363	0,563

TABLE 7.6 – Qualité des agréments

## 7.5.3 Le chiffre d'affaires assurable

Le chiffre d'affaires assurable représente les factures de l'assuré à l'égard de ses clients privés. Il fait l'objet d'une déclaration annuelle afin d'ajuster le coût de la garantie par application du taux de prime fixé au contrat.

classe	nombre de polices	nombre de sinistres	$\lambda_i$
$C_1$	140	29	0,207
$C_2$	148	50	0,338
$C_3$	159	79	0,497
$C_4$	106	90	0,849
$C_5$	59	64	1,085
$C_6$	33	51	1,545
total	645	363	0,563

TABLE 7.7 – Chiffre d'affaires assurable

### 7.5.4 La couverture géographique des agréments

Pour diverses raisons certaines zones géographiques sont moins sûres que d'autres en terme d'échanges commerciaux et donc plus soumises au risque d'impayés. Une entreprise qui vend sur le territoire européen rencontrera moins de difficultés que celle qui commerce avec le reste du monde. La variable envisagée est l'indicatrice de l'existence d'au moins un acheteur situé hors Europe.

classe	nombre de polices	nombre de sinistres	$\lambda_i$
$C_1$	399	173	0,434
$C_2$	246	190	0,772
total	645	363	0,563

TABLE 7.8 – Zone géographique

### 7.5.5 Le secteur d'activité de l'assuré

Le secteur d'activité de l'assuré est une information relativement détaillée mais difficile d'emploi. Si certains secteurs connaissent par essence de plus grands risques d'impayés comme par exemple celui de la « Construction » du fait de l'existence d'acheteurs financièrement fragiles, cette donnée est plus difficile à exploiter pour les entreprises qui s'adressent à l'ensemble des secteurs d'activité.

La variable décrivant le secteur d'activité opèrent parfois des regroupements qui ne semblent pas naturels mais qui présentent des risques proches.

classe	nombre de polices	nombre de sinistres	$\lambda_i$
$C_1$	44	8	0,182
$C_2$	48	18	0,375
$C_3$	480	271	0,565
$C_4$	73	66	0,904
total	645	363	0.563

TABLE 7.9 – Secteur d'activité

## 7.6 Application de la régression binomiale négative

### 7.6.1 Sélection des variables

Le nombre  $N$  de sinistres indemnisés par an et par assuré relatifs aux agréments est modélisé à l'aide d'une régression binomiale négative. Les cinq variables explicatives sont introduites une à une en retenant à chaque étape la combinaison qui présente le plus faible critère d'information d'Akaike.

valeur de l'AIC	$x_1$	$x_1, x_3$	$x_1, x_3, x_5$
taille du portefeuille $x_1$	<b>1096</b>		
chiffre d'affaires $x_2$	1212	1105	1073
qualité du portefeuille $x_3$	1229	<b>1067</b>	
zone géographique $x_4$	1241	1095	1069
secteur d'activité $x_5$	1243	1091	<b>1066</b>

TABLE 7.10 – Sélection des variables pas à pas

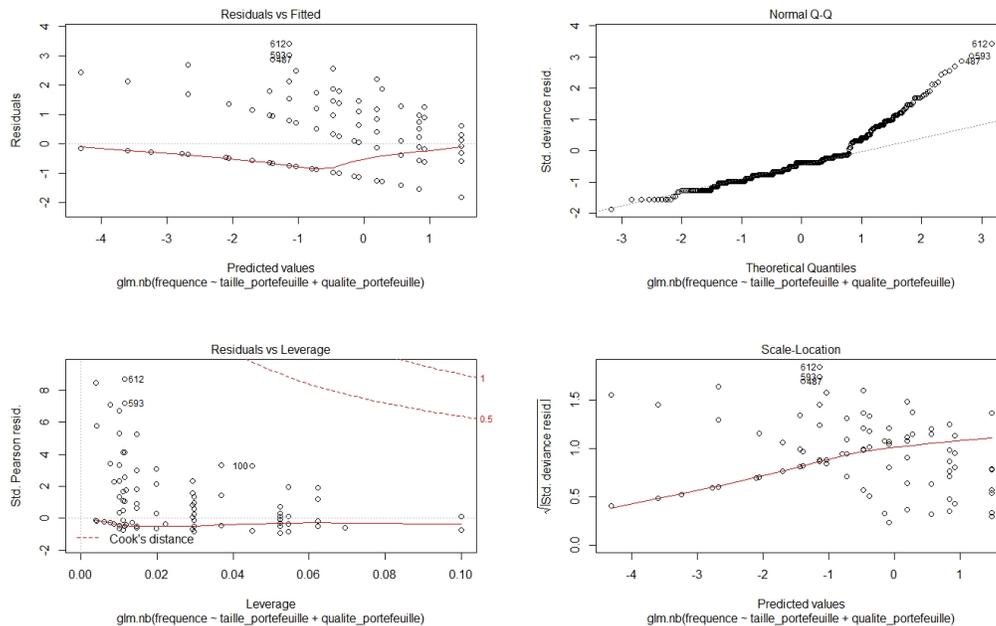


TABLE 7.11 – Résidus et points influents

A noter que l'analyse des résidus permet d'identifier trois assurés qui après examen sont des entreprises de taille importante mais non atypiques.

## 7.6.2 Modélisation retenue

Selon le tableau 7.10 les variables relatives au secteur d'activité et à la couverture géographique sont les moins discriminantes. Leurs apports ne sont pas significatifs relativement aux tests de Wald. En outre ces informations sont déjà prises en compte d'une façon ou d'un autre à travers la notation des acheteurs. Enfin il est assez clair que la variable relative au chiffre d'affaires assurable est redondante avec le nombre d'agréments. In fine deux variables sont retenues.

```
glm.nb(formula = frequence ~ taille_portefeuille + qualite_portefeuille,  
data = base4, model = TRUE, y = TRUE, init.theta = 0.9659438339, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.8225	-0.7467	-0.3658	-0.1640	3.3989

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-2.6695	0.3451	-7.735	1.04e-14	***
taille_portefeuille_C2	1.5398	0.4043	3.808	0.000140	***
taille_portefeuille_C3	2.2077	0.3767	5.861	4.59e-09	***
taille_portefeuille_C4	2.8641	0.3685	7.772	7.75e-15	***
taille_portefeuille_C5	3.5140	0.3989	8.810	< 2e-16	***
taille_portefeuille_C6	4.1567	0.4239	9.806	< 2e-16	***
qualite_portefeuille_C2	-0.5666	0.2201	-2.574	0.010059	*
qualite_portefeuille_C3	-0.9204	0.2058	-4.473	7.72e-06	***
qualite_portefeuille_C4	-1.6326	0.4383	-3.724	0.000196	***

(Dispersion parameter for Negative Binomial(0.9659) family taken to be 1)

Null deviance: 692.73 on 644 degrees of freedom

Residual deviance: 425.77 on 636 degrees of freedom

AIC: 1067.2

Number of Fisher Scoring iterations: 1

Theta: 0.966

Std. Err.: 0.198

2 x log-likelihood: -1047.200

# Chapitre 8

## Le coût moyen des sinistres

### Sommaire

---

<b>8.1</b>	<b>Choix d'une distribution . . . . .</b>	<b>64</b>
<b>8.2</b>	<b>Variables tarifaires . . . . .</b>	<b>66</b>
8.2.1	La quotité garantie . . . . .	66
8.2.2	Le montant moyen des agréments . . . . .	67
8.2.3	Le secteur d'activité de l'assuré . . . . .	67
<b>8.3</b>	<b>Modélisation du coût . . . . .</b>	<b>68</b>
8.3.1	Estimation des coefficients . . . . .	68
<b>8.4</b>	<b>Le tarif « fréquence - coût » . . . . .</b>	<b>69</b>
<b>8.5</b>	<b>Discussion sur la mise en oeuvre du modèle . . . . .</b>	<b>71</b>

---

L'analyse des coûts des sinistres est sensiblement plus compliquée que celle de la fréquence. D'une part en raison d'un échantillon plus restreint limité aux polices sinistrées et d'autre part, en théorie, en raison des durées de clôture des sinistres.

Toutefois en assurance-crédit le montant du sinistre est plafonné par le montant de l'agrément ainsi que par un multiple de la prime. Ainsi la notion de « graves » n'a pas réellement de sens puisque l'assureur maîtrise à tout moment le niveau de ses expositions.

Il s'agit là d'une particularité de cette garantie qui offre une certaine prévisibilité du montant des sinistres et la possibilité de le réduire à tout instant à l'initiative de l'assureur. Néanmoins tout assureur crédit doit être en mesure d'expliquer et motiver toute réduction ou résiliation de garantie,

## 8.1 Choix d'une distribution

La mise en oeuvre du modèle linéaire généralisé demande de choisir une distribution continue appartenant à la famille exponentielle. L'histogramme et la densité empiriques ci-dessous mettent en évidence l'existence d'une légère hétérogénéité sous-jacente. Cela justifie l'utilisation de variables tarifaires et suggère de tester l'ajustement aux données des lois continues classiques utilisées en actuariat.

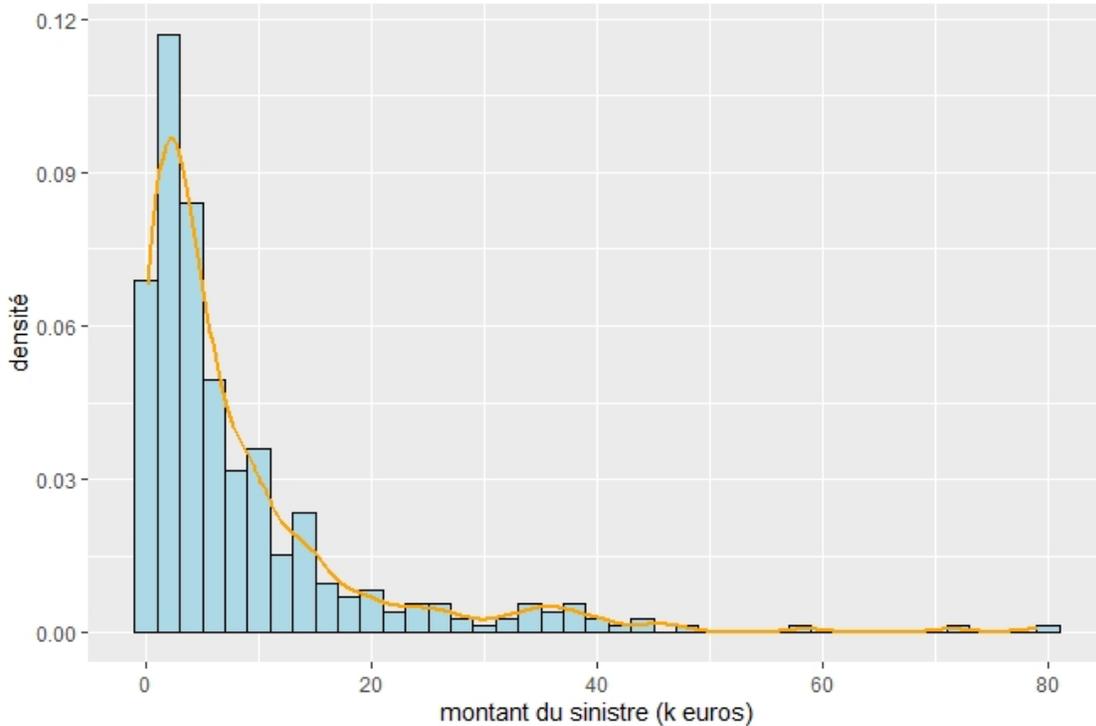


TABLE 8.1 – Densité et histogramme empiriques du coût des sinistres

La sélection d'une distribution appropriée repose sur des analyses graphiques comme la comparaison des fonctions de répartition ou des densités théoriques et empiriques. Les diagrammes « PP-plot » ou « QQ-plot » sont également utilisés.

Toutefois il s'agit de critères visuels subjectifs qui permettent seulement une classification plutôt grossière. Il est recommandé de recourir à des indicateurs numériques comme la distance de Kolmogorov-Smirnov ou le critère AIC qui prend en compte le nombre de paramètres à estimer.

Ce chapitre utilise les livres de A.Charpentier [CHAR 2014], de C.Partrat [PART 2004] et de G.Saporta [SAPO 2011].

Le tableau 8.2 résume ces deux derniers indicateurs pour un ensemble de distributions.

distribution	valeur de l'AIC	Kolmogorov-Smirnov
exponentielle	2273	0.1010
gamma	2268	0.0770
log-normale	2243	0.0433
Weibull	2261	0.0583
Pareto	2249	0.0318
inverse-gaussienne	2881	0.1271

TABLE 8.2 – Sélection d'une distribution continue

Les différents critères étudiés indiquent que la loi log-normale ou la loi de Pareto seraient les plus appropriées bien qu'elles n'appartiennent pas à la famille exponentielle.

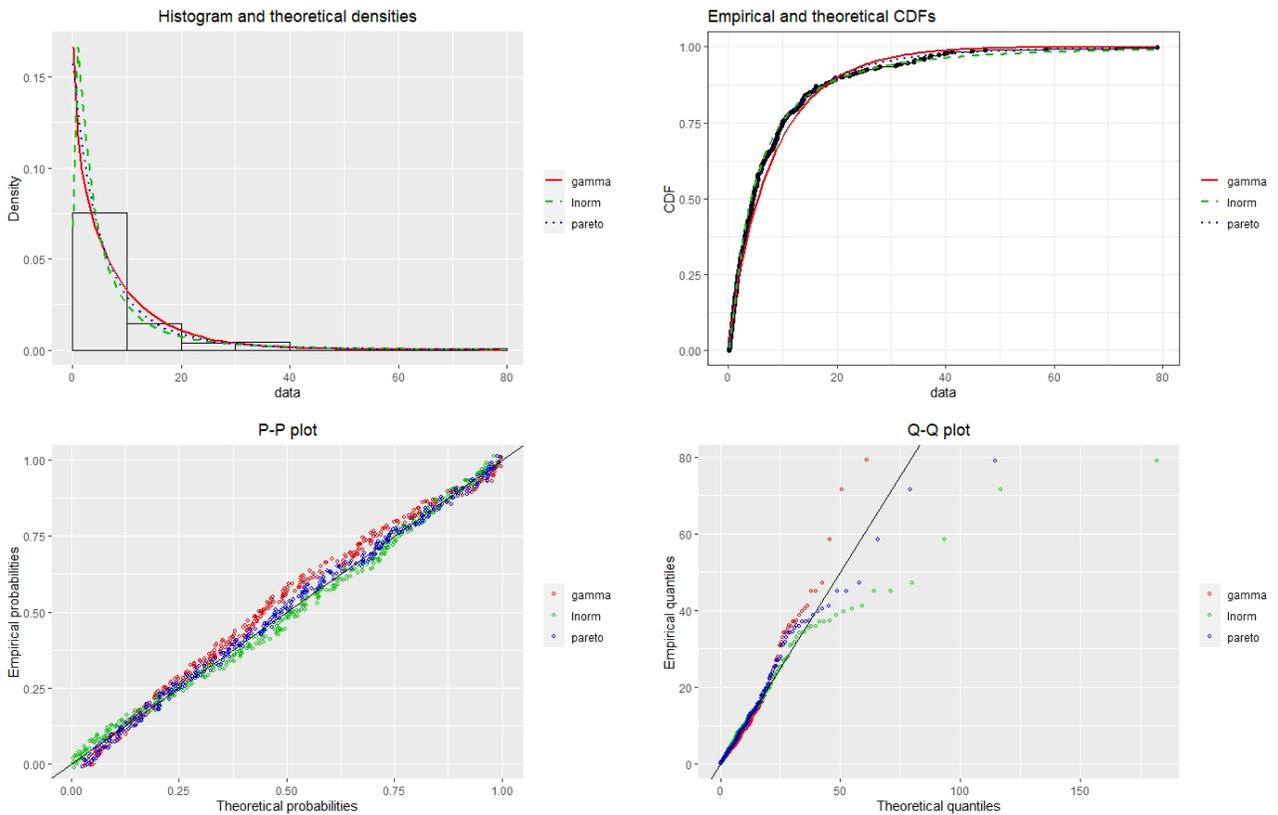


TABLE 8.3 – Analyses graphiques

Si on met en balance la précision des données qui n'est probablement pas optimale, il apparaît que l'utilisation de la loi Gamma constitue une alternative acceptable avec l'avantage d'une certaine facilité de mise en oeuvre.

## 8.2 Variables tarifaires

L'étude des variables tarifaires relatives à la sévérité s'avère dans les faits plus complexe que dans le cas de la modélisation de la fréquence. Au cours de l'analyse, il n'a pas été possible de montrer un lien évident entre le chiffre d'affaires assurable et le montant moyen des sinistres. La répartition géographique des agréments n'a pas non plus d'effet significatif sur le coût des sinistres.

En pratique les variables continues sont transformées en variables catégorielles plus simples à appréhender par les utilisateurs. Pour chaque catégorie  $C_i$  de chaque variable susceptible de participer au modèle son coût moyen des sinistres  $m_i$  est estimé.

### 8.2.1 La quotité garantie

La quotité garantie, appelée pourcentage d'indemnisation, correspond à la somme prise en charge par l'assureur sur une facture impayée. Lorsque la quotité garantie est importante le montant moyen des sinistres est élevé ce qui pourrait traduire l'existence d'un aléa moral chez l'assuré. Le test du rapport de vraisemblance montre que la variable est explicative.

classe	nombre	coût moyen
	de polices	$m_i$
$C_1$	48	4979
$C_2$	315	8920
total	363	8399

TABLE 8.4 – Quotité garantie

#### Analysis of Deviance Table

Model 1: cout ~ 1

Model 2: cout ~ quotite

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	362	503.19			
2	361	490.89	1	12.291	0.00633 **

## 8.2.2 Le montant moyen des agréments

Sans aller jusqu'à étudier la distribution des agréments pour chaque assuré; il apparaît que si la moyenne des agréments sur les acheteurs est élevée le coût moyen des sinistres est également important. Cela signifie que l'entreprise commerce en moyenne pour de plus gros montants susceptibles d'induire de plus gros impayés.

Cet « effet taille » se manifeste encore plus clairement si l'on se cantonne aux agréments les plus mal notés qui demeurent la source essentielle des sinistres. La variable explicative est ensuite découpée de proche en proche en quatre classes de montants moyens des agréments notés au dessous de 7 comme le présente le tableau ci-dessous.

Par ailleurs il se dégage que le coût moyen des sinistres est une fonction monotone décroissante du montant moyen des agréments de « mauvaise » qualité comme le montre le rapport des moyennes dans le tableau.

classe	nombre de sinistres	moyenne des montants moyen des agréments	coût moyen $m_i$	rapport des des moyennes
$C_1$	65	8820	4432	0.50
$C_2$	151	18263	8308	0.45
$C_3$	104	30863	9614	0.31
$C_4$	43	56553	11775	0.21
total	363	24718	8399	0.34

TABLE 8.5 – Montant moyen des agréments ayant une notation inférieure à 7

Il est important de noter que la pertinence de cette variable repose totalement sur le système de notation des entreprises utilisé pour accorder les agréments c'est-à-dire sur la qualité et l'homogénéité des algorithmes utilisés ainsi que sur le travail de arbitres en complément des cotations automatiques.

## 8.2.3 Le secteur d'activité de l'assuré

Le secteur d'activité de l'assuré est une information relativement détaillée mais difficile d'emploi. La facture moyenne pour un secteur d'activité n'a pas de lien direct avec la défaillance de l'acheteur sauf peut être dans des secteurs très spécialisés.

## 8.3 Modélisation du coût

### 8.3.1 Estimation des coefficients

La segmentation du coût moyen étant associée de façon multiplicative à celle de la fréquence, nous retenons dans cette étude prospective une approche parcimonieuse à l'instar de la tarification actuelle basée sur le chiffre d'affaires de l'assuré.

```
glm(formula = cout ~ moyenne_agrement, family = Gamma(link = "log"), data =
  sinistres)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.5762	-1.1616	-0.5189	0.1844	3.3008

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	8.3966	0.1533	54.784	< 2e-16 ***
moyenne_agrementC_2	0.6284	0.1833	3.428	0.000679 ***
moyenne_agrementC_3	0.7744	0.1954	3.963	8.91e-05 ***
moyenne_agrementC_4	0.9772	0.2429	4.023	7.01e-05 ***

(Dispersion parameter for Gamma family taken to be 1.526937)

Null deviance: 503.19 on 362 degrees of freedom  
Residual deviance: 473.97 on 359 degrees of freedom  
AIC: 7269.5

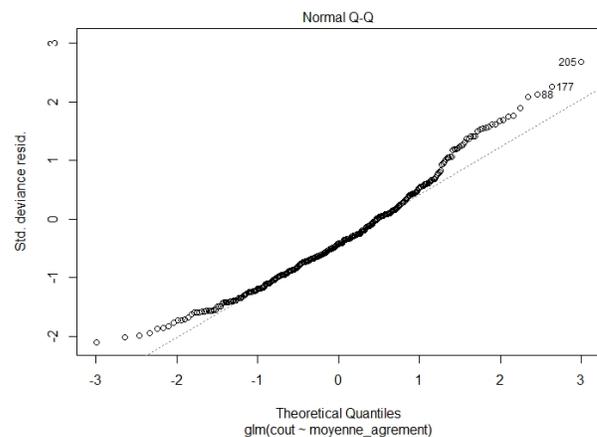
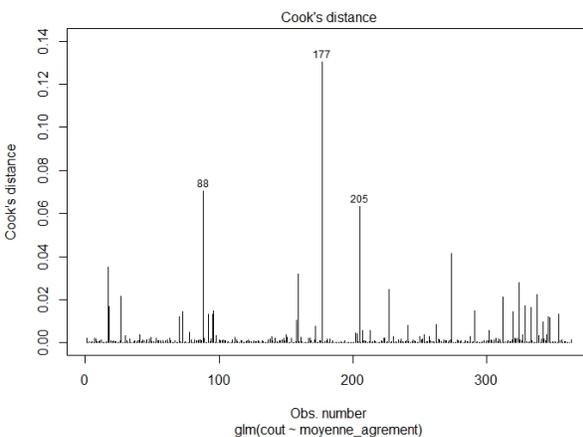


TABLE 8.6 – Résidus et points influents

## 8.4 Le tarif « fréquence - coût »

Dans le modèle « fréquence - coût » en univers segmenté, la prime pure est le produit des espérances de la fréquence et du coût des sinistres conditionnelles à l'information utile soit

$$E[S|X, Y] = E[N|X] \cdot E[C|Y]$$

En pratique il est assez courant de recourir à la fonction de lien logarithme qui transforme la combinaison linéaire des facteurs explicatifs en produit d'exponentielles. Ce choix rend le tarif multiplicatif au regard du risque de base  $\exp(\alpha_0) \exp(\beta_0)$ .

La prime pure de l'année 2015 se calcule comme un produit de trois facteurs explicatifs qualitatifs :

- ✓  $x_1$  la taille du portefeuille,
- ✓  $x_2$  la qualité du portefeuille,
- ✓  $y_1$  le montant moyen des agréments.

Lorsque l'assuré prend les modalités  $(i, j, k)$  sur les variables explicatives  $(x_1, x_2, y_1)$  sa prime vaut

$$\widehat{S_{2015}^{FC}} = \exp(\alpha_0) \exp(\alpha_1^i) \exp(\alpha_2^j) \exp(\beta_0) \exp(\beta_1^k)$$

Selon le tableau 8.7 la prime « fréquence - coût » est suffisante pour couvrir les sinistres. C'est la prime de plus faible montant total et celle qui présente la meilleure erreur quadratique moyenne. Cette prime identifie 96 segments d'assurés. En outre elle est relativement facile à mettre en oeuvre.

montant total sur le portefeuille (en €)	mean square error
sinistres	2 643 416
prime actuelle	3 694 261
prime « Bühlmann-Straub »	3 440 487
prime « Bâle 2 »	3 758 107
prime « fréquence - coût »	3 102 278
	132 468 048
	113 816 705
	348 141 027
	98 836 166

TABLE 8.7 – Couverture de la sinistralité 2015

La représentation de la prime par classes de chiffres d'affaires présente un profil peu régulier ce qui confirme que cette variable n'est pas le meilleur critère pour tarifer le risque bien que son utilisation demeure un standard commercial du secteur.

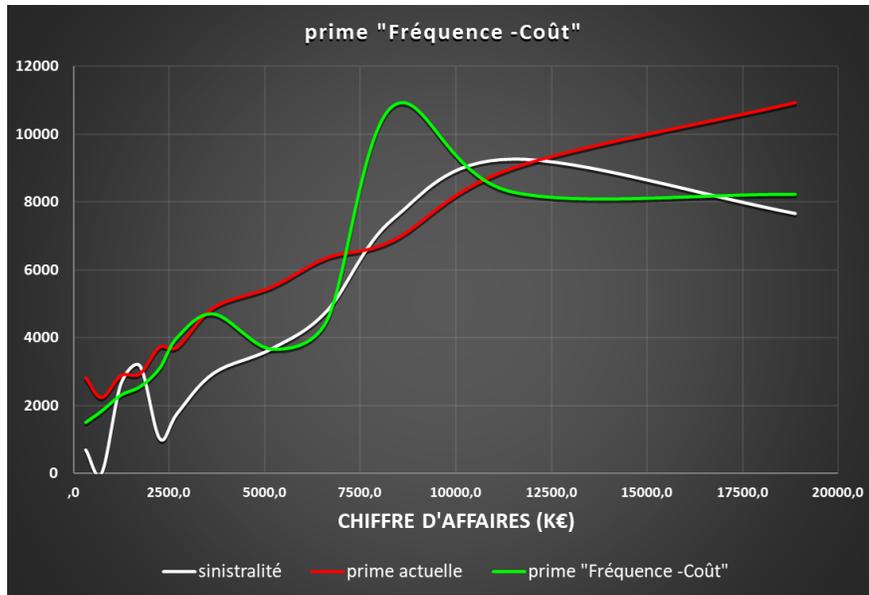


TABLE 8.8 – Prime « fréquence - coût » par CA

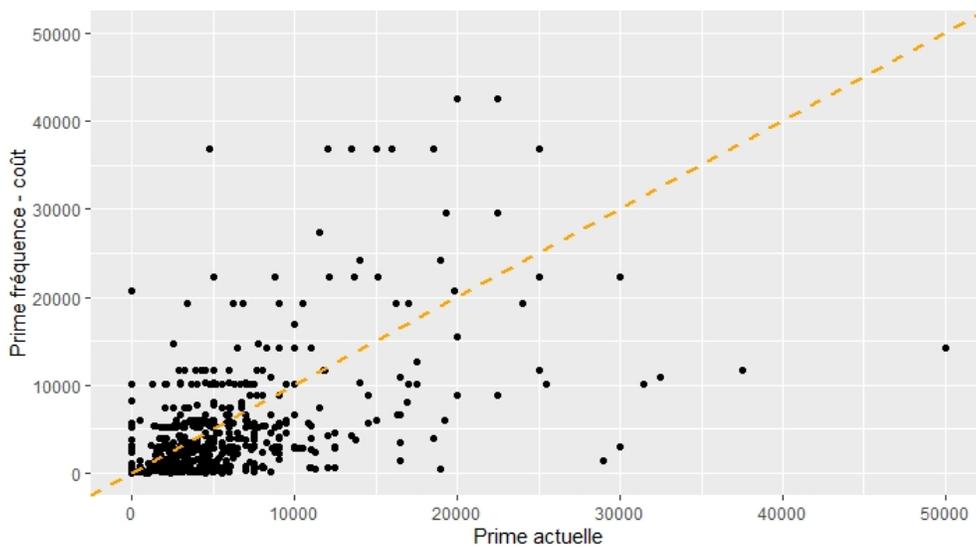


TABLE 8.9 – Prime « fréquence - coût » vs prime actuelle

La tarification actuelle est basée sur des taux de marché décroissants en fonction du chiffre d'affaires déclaré alors que la prime « fréquence - coût » est personnalisée en selon des caractéristiques propres à l'assuré. Naturellement, les deux primes ne coïncident pas.

## 8.5 Discussion sur la mise en oeuvre du modèle

Dans les applications actuarielles, une variable aléatoire symétrique normalement distribuée avec une variance fixe ne permet pas en général de décrire les situations correctement. C'est pourquoi l'utilisation des modèles linéaires généralisés est désormais préférée pour concevoir un tarif. Ils établissent une relation entre les caractéristiques de l'opération d'assurance et sa sinistralité. Ainsi la connaissance de ces caractéristiques permet de calculer et segmenter le tarif. Cette approche a priori se distingue du modèle de « Bühlmann-Straub » qui utilise des informations connues a posteriori sur la sinistralité.

Présentant des résultats aisément interprétables, les modèles linéaires généralisés font ainsi preuve d'une très bonne applicabilité opérationnelle. De plus, ces modèles paramétriques sont construits dans un cadre théorique leur permettant d'effectuer des tests statistiques capables d'évaluer leur qualité, jouissant alors d'une robustesse assez satisfaisante. L'approche « GLM » impose de faire une hypothèse sur la forme de la loi conditionnelle de la variable expliquée en fonction des explicatives. Cette hypothèse peut s'avérer fautive et on prend donc un risque de modèle difficile à mesurer.

En assurance crédit, les deux principales limites de cette approche sont de nature distincte. D'un point de vue théorique d'une part, l'hypothèse d'indépendance des observations n'est pas vérifiée. Ensuite dans une économie cyclique, les défauts des entreprises ont tendance à survenir plus fréquemment lorsque la situation de l'économie ou d'un secteur se dégrade. Cet écueil est aussi présent dans la mise en oeuvre de la théorie de la crédibilité.

D'autre part, d'un point de vue pratique, le choix des variables explicatives est limité et le portefeuille d'acheteurs d'un assuré n'est pas une de ses caractéristiques intrinsèques. Si le comportement d'un conducteur et le type de véhicule sont des facteurs déterminants d'un sinistre auto, il semble peu probable que les attributs d'un assuré influencent notablement le risque d'impayé de ses acheteurs. Par ailleurs, la conservation des données sur les entreprises est un véritable challenge au quotidien.

Plus vraisemblablement l'origine des sinistres se cache dans les portefeuilles d'acheteurs des assurés ; portefeuilles dont l'analyse et la synthèse sont des enjeux complexes aux solutions imparfaites. Ainsi la solution proposée au chapitre précédent est imparfaite et nous invite naturellement à dépasser la seule utilisation des trois variables retenues - taille du portefeuille, qualité du portefeuille et montant moyen des agréments - par un modèle financier plus direct.

## Quatrième partie

### Un modèle issu de la finance

# Chapitre 9

## Le coût du risque de crédit

### Sommaire

---

<b>9.1</b>	<b>La distribution des pertes</b>	<b>74</b>
<b>9.2</b>	<b>Le modèle « Asymptotic Single Risk Factor »</b>	<b>75</b>
9.2.1	Définition et hypothèses	75
9.2.2	Le portefeuille infiniment granulaire	76
<b>9.3</b>	<b>Simulation numérique des pertes</b>	<b>77</b>
9.3.1	Méthodes de Monte-Carlo	77
9.3.2	Algorithme pour la distribution des pertes	78
<b>9.4</b>	<b>Paramètres pour l'assurance-crédit</b>	<b>79</b>
9.4.1	La probabilité de défaut	79
9.4.2	La perte en cas de défaut	79
9.4.3	L'exposition au défaut	80
<b>9.5</b>	<b>Granularité des portefeuilles</b>	<b>80</b>
9.5.1	Le portefeuille d'AXA Assurcrédit	80
9.5.2	Les portefeuilles des assurés	82
<b>9.6</b>	<b>Le tarif « bâlois »</b>	<b>83</b>

---

Dans cette dernière partie nous considérons que l'assureur détient indirectement un portefeuille de prêts sur les clients des assurés. En cas de défaillance d'un acheteur, l'assureur subit un sinistre considéré comme le non remboursement d'un crédit « virtuel ».

Le calcul de la prime pure s'appuie sur le dispositif prudentiel bancaire destiné à mieux appréhender le risque de crédit dans le cadre du second accord de Bâle.

## 9.1 La distribution des pertes

L'évènement de défaut d'une entreprise du portefeuille sur la période  $T$ , le plus souvent une année, est noté  $D$  avec  $PD = \mathbb{P}(D = 1)$  sa probabilité de défaut inconditionnelle.

On retient l'hypothèse classique que l'actif de la firme  $(V_t)_{t \geq 0}$  suit un mouvement brownien géométrique. Si  $X$  désigne une loi normale centrée réduite alors

$$V_T = V_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}X\right)$$

Dans le modèle de Merton, le défaut intervient lorsque l'actif  $V_T$  est inférieur à la dette  $B$ , ce qui permet d'écrire la probabilité de défaut comme

$$PD = \mathbb{P}(\ln V_T < \ln B) = \Phi\left(\frac{\ln B - \ln V_0 - (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)$$

Dans la suite on considère un portefeuille de  $n$  firmes et pour chaque entreprise  $i$  les quantités suivantes sont supposées déterministes

- ✓ l'exposition au défaut  $EAD_i = E[\widetilde{EAD}_i]$
- ✓ la perte en cas de défaut  $LGD_i = E[\widetilde{LGD}_i] \in [0, 1]$

La distribution des pertes du portefeuille est la variable aléatoire

$$L_n = \sum_{i=1}^n L_i = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \mathbf{1}_{D_i} \text{ d'espérance } E[L_n] = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot PD_i$$

Le risque de crédit est quantifié et comptabilisé par les banques en effectuant une distinction entre deux types de pertes. D'une part les pertes attendues «  $EL$  » qui sont enregistrées au compte de résultat et qui correspondent à la moyenne des pertes anticipées soit  $E[L_n]$ .

D'autre part les pertes inattendues «  $UL$  » ou exceptionnelles qui doivent être absorbables par les fonds propres de la banque et qui sont définies comme les pertes au-delà des pertes attendues. Leur niveau et le moment auquel elles surviennent ne sont pas connus à l'avance.

## 9.2 Le modèle « Asymptotic Single Risk Factor »

### 9.2.1 Définition et hypothèses

Il a été originellement élaboré par Vasicek et repose sur la représentation du logarithme du rendement de l'actif  $i$  standardisé sous la forme

$$X_i = \sqrt{\rho_i} Y + \sqrt{1 - \rho_i} Z_i$$

Le facteur de risque systémique  $Y$  s'interprète comme une variable représentant l'état global de l'économie et le facteur de risque idiosyncratique  $Z_i$  représente le risque individuel de chaque débiteur. Les variables  $Y$  et  $Z_i$  sont mutuellement indépendantes et suivent une loi normale centrée réduite.

Par la suite, les  $\rho_i$  sont supposés constants et en s'appuyant sur les travaux de M. Dietsch nous retiendrons  $\rho = 0,01$ . Le coefficient  $\rho$  mesure le degré de dépendance entre les crédits du portefeuille. Il s'agit du coefficient de corrélation des actifs puisque

$$E[X_i X_j] = E[\rho Y^2 + (1 - \rho) Z_i Z_j + \sqrt{\rho} \sqrt{1 - \rho} Y (Z_i + Z_j)] = \rho$$

La première hypothèse du modèle ASRF implique que la probabilité de défaut conditionnelle à une réalisation  $y$  du facteur  $Y$  s'écrit

$$\begin{aligned} PD_i(y) &= \mathbb{P}(D_i = 1 | Y = y) \\ &= \mathbb{P}(X_i < \Phi^{-1}(PD_i) | Y = y) \\ &= \mathbb{P}(\sqrt{\rho} y + \sqrt{1 - \rho} Z_i < \Phi^{-1}(PD_i)) \\ &= \Phi\left(\frac{\Phi^{-1}(PD_i) - \sqrt{\rho} y}{\sqrt{1 - \rho}}\right) \\ &= \Phi(\zeta_i(y)) \end{aligned}$$

La distribution des pertes du portefeuille conditionnelle s'exprime alors comme

$$L_n(y) = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \mathbb{1}_{\{Z_i < \zeta_i(y)\}}$$

Cette description s'appuie sur les ouvrages de V. Brunel [BRUN 2014], de T. Roncalli [RONC 2005] et l'article de M. Dietsch [DIET 2003].

## 9.2.2 Le portefeuille infiniment granulaire

La deuxième hypothèse du modèle ASRF stipule que le portefeuille est infiniment granulaire. Cette notion de granularité asymptotique correspond à l'idée qu'il n'existe pas d'exposition dépassant une petite proportion de l'exposition totale. Elle s'exprime par les deux conditions suivantes

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n EAD_i = \infty \text{ et } \sum_{n=1}^{\infty} \left( \frac{EAD_n}{\sum_{i=1}^n EAD_i} \right)^2 < \infty$$

En appliquant une version de la loi forte des grands nombres, M.Gordy montre qu'en l'absence de concentrations en terme d'exposition, la fonction de perte converge presque sûrement vers son espérance conditionnelle au facteur systémique soit  $\lim_{n \rightarrow \infty} L_n = E[L_n|Y]$ .

Il montre également que dans un modèle mono-factoriel la valeur en risque d'un portefeuille infiniment granulaire se décompose en la somme linéaire des contributions en risque des différentes créances, autrement dit

$$\begin{aligned} \lim_{n \rightarrow \infty} VaR_{\alpha}(L_n) &= VaR_{\alpha}(E[L_n|Y]) \\ &= E[L_n|VaR_{1-\alpha}(Y)] \\ &= \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \Phi \left( \frac{\Phi^{-1}(PD_i) - \sqrt{\rho} VaR_{1-\alpha}(Y)}{\sqrt{1-\rho}} \right) \\ &= \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \Phi \left( \frac{\Phi^{-1}(PD_i) + \sqrt{\rho} \Phi^{-1}(\alpha)}{\sqrt{1-\rho}} \right) \end{aligned}$$

Si pour l'assurance-crédit l'espérance  $E[L_n] = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot PD_i = EL$  s'assimile à la prime pure, elle n'est pas suffisante pour se protéger des pertes exceptionnelles.

A cette fin et à l'image du banquier, l'assureur devrait disposer de fonds propres supplémentaires définis par

$$UL_{\alpha} = VaR_{\alpha}(L_n) - EL = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \left[ \Phi \left( \frac{\Phi^{-1}(PD_i) + \sqrt{\rho} \Phi^{-1}(\alpha)}{\sqrt{1-\rho}} \right) - PD_i \right]$$

Si l'on relâche l'hypothèse de granularité du portefeuille, il n'est plus possible de dériver une expression analytique du capital économique. Le recours aux méthodes de Monte-Carlo devient obligatoire.

## 9.3 Simulation numérique des pertes

Lorsque l'hypothèse d'infinie granularité du portefeuille n'est pas satisfaite, les mesures de risque sont approchées par des techniques numériques.

### 9.3.1 Méthodes de Monte-Carlo

Les simulations Monte-Carlo sont employées pour obtenir des approximations des mesures de risque relatives à la distribution des pertes du portefeuille de crédits. Elles s'appuient sur la loi forte des grands nombres que nous énonçons ci-dessous.

Soient  $(\xi^{(l)}, l \geq 1)$  une suite de variables aléatoires indépendantes et identiquement distribuées telles que  $E[\xi^{(1)}] < \infty$ . Pour  $n \geq 1$  la moyenne empirique est définie par

$$\widehat{S}_n = \frac{1}{n} \sum_{l=1}^n \xi^{(l)}$$

La loi forte des grands nombres stipule que

$$\mathbb{P} \left( \widehat{S}_n \xrightarrow[n \rightarrow \infty]{} E[\xi^{(1)}] \right) = 1$$

Soit  $\gamma$  une quantité à approcher. Supposons qu'il existe une suite  $(X^{(l)}, l \geq 1)$  de variables aléatoires indépendantes et identiquement distribuées « faciles à simuler » et une fonction  $f$  telle que  $\gamma = E[f(X^{(1)})]$ .

Alors pour tout tirage  $(X^{(l)}, l \geq 1)$  effectué par l'ordinateur, la moyenne empirique des  $(f(X^{(l)}), l \geq 1)$  converge en probabilité vers  $\gamma$  autrement dit

$$\mathbb{P} \left( \frac{1}{n} \sum_{l=1}^n f(X^{(l)}) \xrightarrow[n \rightarrow \infty]{} \gamma \right) = 1$$

La précision des calculs peut être déterminée à partir du théorème de la limite centrale. Soit  $(\xi^{(l)}, l \geq 1)$  une suite de variables aléatoires indépendantes et identiquement distribuées telles que  $E[\xi^{(1)}] < \infty$  et  $\text{Var}[\xi^{(1)}] < \infty$  estimé sans biais par  $\widehat{\sigma}_n^2$ . On montre que

$$\frac{\sqrt{n}}{\widehat{\sigma}_n} \left( \frac{1}{n} \sum_{l=1}^n f(X^{(l)}) - \gamma \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

### 9.3.2 Algorithme pour la distribution des pertes

Dans le cadre du modèle ASRF, l'algorithme de calcul de la distribution des pertes peut se décrire de la façon suivante

→ pour chaque exposition  $i = 1, \dots, n$  déterminer ses paramètres caractéristiques non aléatoires  $EAD_i, LGD_i, PD_i, \rho_i$

→ à chaque itération  $k = 1, \dots, K$

☆ générer l'exposition systémique  $y_k$  sachant que  $Y \sim \mathcal{N}(0, 1)$

☆ pour chaque exposition  $i = 1, \dots, n$

✎ générer l'exposition idiosyncratique  $z_{i,k}$  sachant que  $Z_{i,k} \sim \mathcal{N}(0, 1)$

✎ calculer le seuil de défaut  $\zeta_i(y_k) = \frac{\Phi^{-1}(PD_i) - \sqrt{\rho_i} y_k}{\sqrt{1 - \rho_i}}$

✎ calculer la perte  $L_{i,k} = EAD_i \cdot LGD_i \cdot \mathbf{1}_{\{z_{i,k} < \zeta_i(y_k)\}}$

☆ calculer la perte globale du portefeuille

$$L_{n,k} = \sum_{i=1}^n L_{i,k}$$

→ estimer la distribution des pertes empirique

$$\widehat{F}(l) = \frac{1}{K} \sum_{k=1}^K \mathbf{1}_{\{0 \leq L_{n,k} \leq l\}}$$

→ estimer la perte moyenne

$$\widehat{E}(L_n) = \frac{1}{K} \sum_{k=1}^K L_{n,k}$$

La vitesse de convergence de la méthode de Monte-Carlo est lente comparée à d'autres méthodes numériques. En fonction du nombre de limites  $n$  du portefeuille et du nombre d'itérations  $k$  liées à la précision recherchée, le calcul de la distribution des pertes peut être très long à réaliser sur un ordinateur de bureau. Ce paragraphe emprunte des notions au cours sur les méthodes de Monte Carlo J.Stoehr [STOE 2023].

## 9.4 Paramètres pour l'assurance-crédit

### 9.4.1 La probabilité de défaut

Afin d'attribuer à chaque acheteur une probabilité d'impayé, nous disposons de la note de solvabilité de la Coface appelée DRA pour « Default Rating Assesment ». Toutefois celle-ci doit être recalibrée car l'évènement probabilisé est la défaillance légale de l'entreprise plutôt que la survenance d'un impayé. Néanmoins il n'est pas surprenant que le DRA, qui est un score de défaillance légale, peut être utilisé comme un score d'impayé puisque les deux notions sont très proches. Beaucoup de défaillances sont la conséquence directe d'impayés.

Nous avons estimé les probabilités de survenance d'un sinistre par classe de DRA sur la période de 2012 à 2014.

	0	1	2	3	4	5	6	7	8	9	10
AXA Assurcrédit	8,00	8,00	4,00	2,00	1,00	0,75	0,50	0,25	0,25	0,20	0,10
Coface	25,00	25,00	10,00	4,00	2,00	1,30	0,70	0,40	0,15	0,05	0,00

TABLE 9.1 – Probabilité de défaut par note de solvabilité

L'assureur-crédit dispose à tout moment du droit de se désengager en révisant les montants garantis contrairement au banquier qui reste lié au débiteur sur une durée assez longue. C'est ce qui explique qu'en pratique les taux de défaut en assurance-crédit peuvent être plus faibles que les taux d'insolvabilité bancaire.

### 9.4.2 La perte en cas de défaut

La perte en cas de défaut, en anglais « Loss Given Default », est une variable représentant le pourcentage de l'encours de crédit non remboursé suite au défaut. Dans le cas de l'assurance-crédit, en cas de défaillance de l'acheteur, l'assureur reprend la créance et procède à son recouvrement. Si la compagnie arrive à récupérer une partie de ce qui est dû, elle se rembourse à hauteur de son engagement, puis reverse le solde à l'assuré.

L'utilisation d'une *LGD* qui ne varient pas suivant le cycle économique n'est pas réaliste. Par ailleurs dans les chapitres précédents le coût du sinistre qui est modélisé représente la charge brute de sinistre. Pour garantir la comparabilité des méthodes la *LGD* est fixée à 100% et garantit une approche prudente.

### 9.4.3 L'exposition au défaut

L'exposition au défaut est définie comme le montant de la dette au moment où le défaut intervient. Dans le cadre de l'assurance-crédit ce montant est difficile à déterminer à la signature du contrat. En effet l'assuré demande souvent à son assureur des agréments supérieurs à ce qu'il facture réellement afin d'avoir une certaine flexibilité et dans le même temps il émet plusieurs factures sur un même acheteur au cours de l'année. Enfin les montants d'agréments peuvent évoluer au cours du temps en fonction des besoins de l'assuré ou de la solvabilité de l'acheteur.

Selon des études internes, le taux d'utilisation de la limite, appelé également « use factor », est proche de 40%, valeur que nous retiendrons pour les calculs.

## 9.5 Granularité des portefeuilles

La mesure de concentration la plus utilisée est l'indice d'Herfindahl-Hirschman défini, pour un portefeuille de  $n$  accords, comme la somme des carrés des poids relatifs des agréments

$$HHI = \sum_{i=1}^n \left( \frac{EAD_i}{EAD} \right)^2 \text{ où } EAD = \sum_{i=1}^n EAD_i$$

Le niveau de diversification se mesure sur une échelle à quatre niveaux en prenant soin d'isoler les concentrations atypiques et élevées identifiables a priori

<i>HHI</i>	diversification
$\leq 0,01$	très élevée
$\leq 0,15$	élevée
$\leq 0,25$	modérée
$> 0,25$	faible

### 9.5.1 Le portefeuille d'AXA Assurcrédit

Le portefeuille d'AXA Assurcrédit est constitué de l'ensemble des limites de crédit octroyées sur les acheteurs de ses assurés. Les limites sur un même acheteur sont additionnées.

Les très fortes concentrations d'encours acheteurs doivent être étudiées spécifiquement à l'instar des assurés de tailles importantes qui méritent une tarification adaptée aux enjeux commerciaux et aux risques qui en découlent. Néanmoins il est assez difficile d'appréhender certaines concentrations qui découlent de liens financiers complexes pour constituer un groupe de facto comme cela peut être observé dans le secteur du BTP.

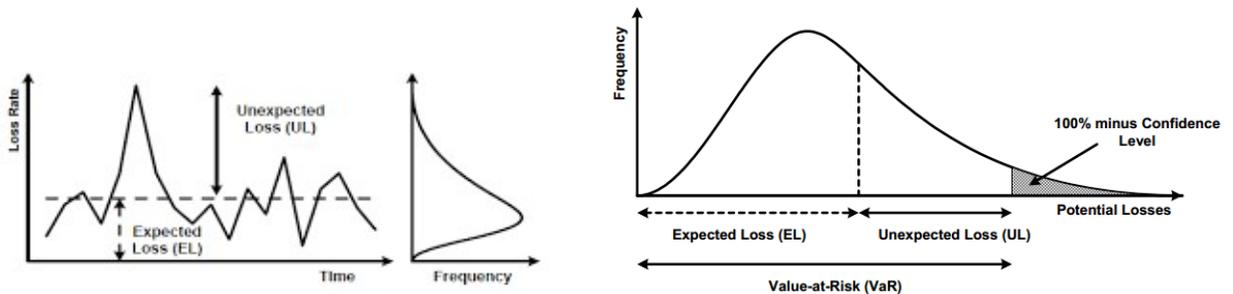
Ainsi l'indice de concentration  $HHI$  prend tout son sens, après retrait des groupes identifiés, sur une population de PME qui commercent entre elles.

Le calcul de l'indice  $HHI$  du portefeuille d'AXA Assurcrédit porte sur 71 473 limites de crédit sur 55 003 entreprises sachant que 3 279 limites n'ont pas d'acheteurs identifiés. Le montant total des encours accordés ressort à 2 542 570 000 euros. L'encours le plus important ne dépasse pas deux pour mille de l'encours total.

Avec un indice de 0,0001 le portefeuille est considéré comme infiniment granulaire et les résultats démontrés par M.Gordy relatifs aux pertes attendues et exceptionnelles sont valables. On trouvera plus de détails sur ce sujet dans l'ouvrage de E.Lütkebohmert [LUTK 2008].

Les établissements bancaires considèrent les pertes attendues, représentées dans les graphiques ci-dessous par la ligne en pointillé, comme une composante du coût de leur activité et l'intègrent au taux d'intérêt. Pour l'assureur il s'agit de la prime pure qui se calcule par la formule classique

$$EL = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot PD_i$$



Les pertes constatées au-delà de la moyenne demeurent exceptionnelles. La probabilité que les pertes subies excèdent, à un niveau de confiance donné, la somme des précédentes composantes est représentée par la surface noircie sur le graphique de droite.

Au niveau de confiance  $\alpha$  l'assureur devrait disposer de fonds propres supplémentaires définis par

$$UL_{\alpha} = VaR_{\alpha}(L) - EL = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot \left[ \Phi \left( \frac{\Phi^{-1}(PD_i) + \sqrt{\rho} \Phi^{-1}(\alpha)}{\sqrt{1 - \rho}} \right) - PD_i \right]$$

Tout comme la prime pure, les fonds propres supplémentaires sont additifs et pourraient être chargés à l'assuré. C'est la force de ce modèle sous l'hypothèse d'infinie granularité.

## 9.5.2 Les portefeuilles des assurés

Les portefeuilles d'acheteurs des assurés présentent des tailles très variables en nombre de limites et en terme d'encours avec une prédominance des petits portefeuilles. Il en résulte des indices de diversification assez hétérogènes avec naturellement une diversification qui croît avec le nombre d'acheteurs. Sauf de rares exceptions au-delà de 50 limites, les portefeuilles sont bien diversifiés (cf. graphiques 9.2).

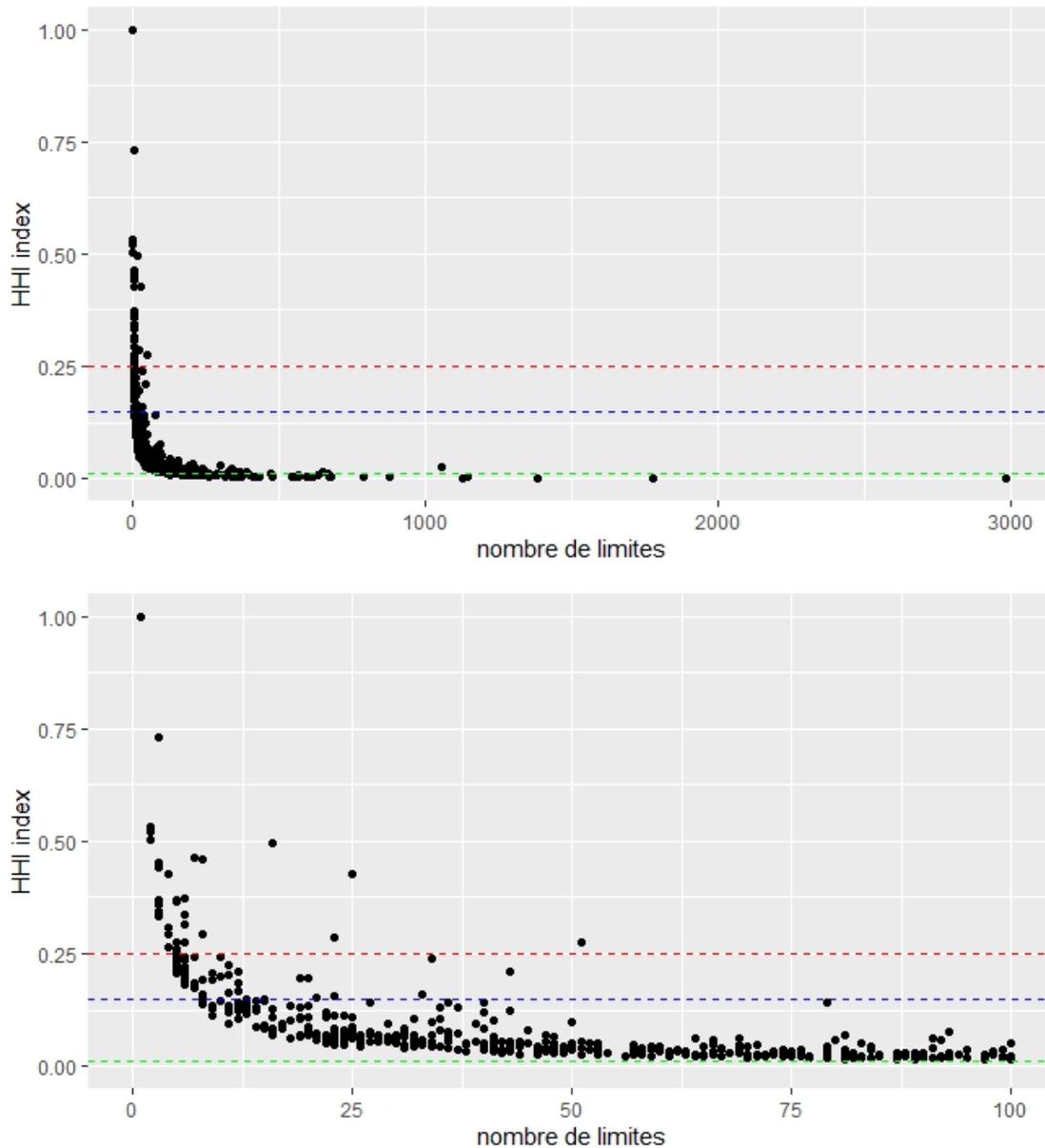


TABLE 9.2 – Indice d'Herfindahl-Hirschman

## 9.6 Le tarif « bâlois »

Dans le modèle ASRF granulaire la prime pure 2015 est individualisée. Elle se calcule très simplement comme le produit des différents paramètres de chaque acheteur :

$$EAD_i \cdot LGD_i \cdot PD_i$$

Selon le tableau 9.3 la prime « Bâle 2 » est suffisante pour couvrir les sinistres. Elle est légèrement plus chère que la prime actuelle et son erreur quadratique moyenne est la plus élevée de toutes.

montant total sur le portefeuille (en €)	mean square error	
sinistres	2 643 416	
prime actuelle	3 694 261	132 468 048
prime « Bühlmann-Straub »	3 440 487	113 816 705
prime « Bâle 2 »	3 758 107	348 141 027
prime « fréquence - coût »	3 102 278	98 836 166

TABLE 9.3 – Couverture de la sinistralité 2015

La représentation de la prime par classes de chiffres d'affaires paraît indiquer que le modèle n'est pas forcément pertinent sur les petites entreprises. Ce constat est à mettre en parallèle avec la faible granularité donc diversification des petits portefeuilles.

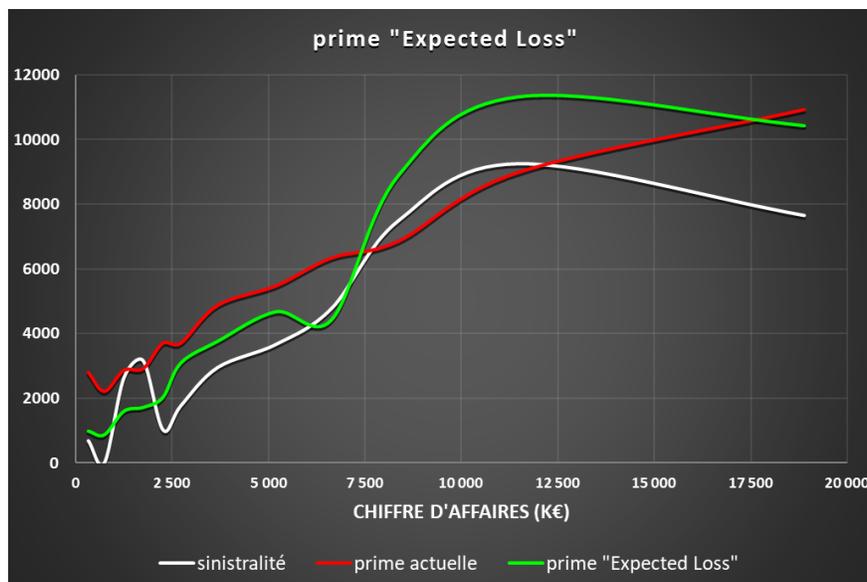


TABLE 9.4 – Prime « bâloise » par CA

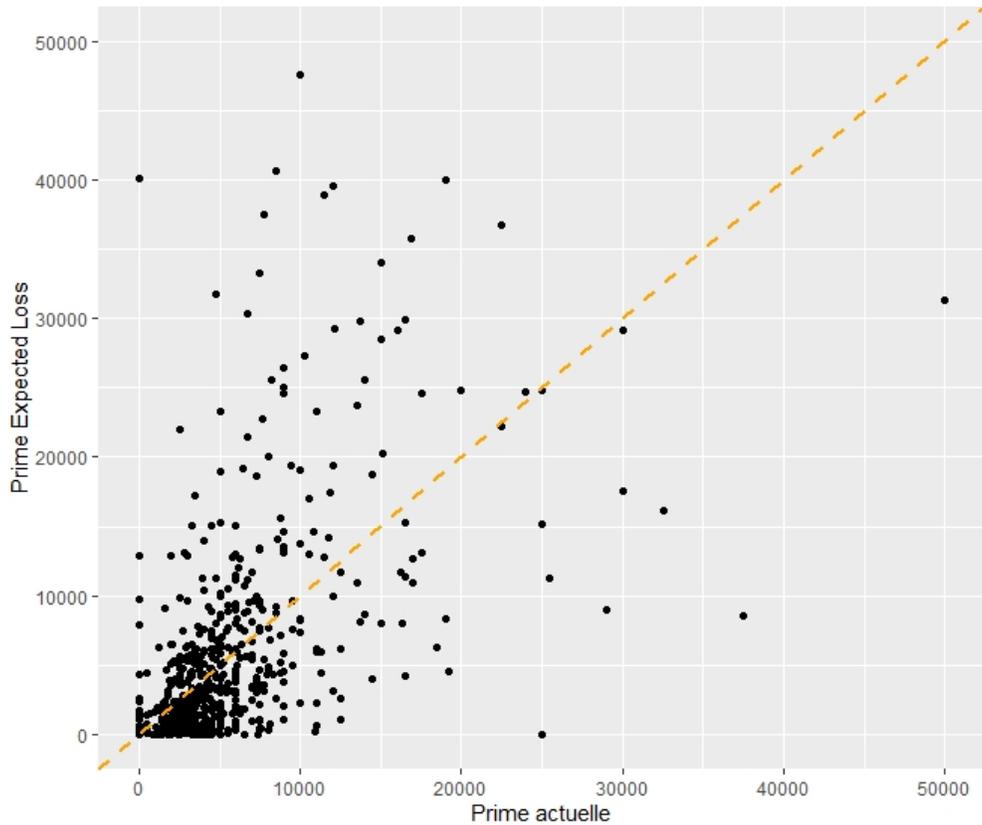


TABLE 9.5 – Prime « bâloise » vs prime actuelle

La tarification actuelle est basée sur des taux de marché décroissants en fonction du chiffre d'affaires déclaré alors que la prime de bâloise tient compte de l'exposition au défaut et des probabilités de défaut. Naturellement les deux primes ne coïncident pas.

Ce modèle propose une approche très conservatrice et extrêmement simple à mettre en oeuvre. Toutefois cette simplicité, à l'instar de sa robustesse, n'est qu'apparente et repose sur la qualité et la fréquence de l'estimation des paramètres qui reste une tâche complexe.

Par ailleurs cette approche ne peut pas capter une modification brusque et importante, à la hausse comme à la baisse, de la conjoncture économique qui est un des facteurs de risque sous-jacent. Une utilisation efficace du tarif bâlois doit s'accompagner d'un suivi mensuel glissant de la probabilité de défaut  $PD$  et dans une moindre mesure de la  $LGD$ . L'avantage de cette méthode est de proposer une évaluation permanente du risque de souscription qui peut dévier significativement de celui connu à la souscription.

# Conclusion

Dans les chapitres précédents nous avons montré que les tarifications de la garantie d'assurance-crédit qui utilisent la théorie de la crédibilité ou les modèles linéaires généralisés n'étaient que partiellement satisfaisantes. En effet les deux approches masquent en partie le fait que l'origine des sinistres se cache avant tout dans la composition des portefeuilles d'acheteurs. En effet la bonne gestion des factures impayées par les assurés ne garantit pas complètement l'indépendance des risques. La faillite d'un acheteur s'impose à tous.

Par ailleurs le profil de l'assuré ne détermine que faiblement à la sinistralité observée et les tentatives de résumer le risque global d'un portefeuille d'entreprises clientes par une variable synthétique explicative est un exercice périlleux. Autrement dit les hypothèses et les pré-requis de ces deux modèles ne sont pas complètement respectés et implique de considérer avec précautions les tarifs qui en découlent.

In fine la démarche retenue considère que l'assureur détient indirectement un portefeuille de prêts sur les clients des assurés. En cas de défaillance d'un acheteur, l'assureur subit un sinistre similaire au non remboursement d'un crédit virtuel. Le calcul de la prime pure s'appuie sur le dispositif prudentiel bancaire destiné à mieux appréhender le risque de crédit dans le cadre du second accord de Bâle qui fait l'objet de nombreux travaux en particulier sur les relations de dépendance et de contagion entre les défauts des contreparties.

Dans le cadre de cette étude, il faut retenir qu'après retrait des groupes qui font l'objet de tarifications personnalisées, le portefeuille de l'assureur est considéré comme infiniment granulaire et les résultats démontrés par M.Gordy relatifs aux pertes attendues et exceptionnelles sont valables. Il en découle que le tarif pour un assuré  $i$  dont le portefeuille est composée de  $n_i$  limites de crédit se calcule par la formule classique

$$EL_i = \sum_{j=1}^{n_i} EAD_j \cdot LGD_j \cdot PD_j$$

Néanmoins cette méthode simple en apparence cache quelques difficultés pratiques. La tarification n'est applicable et fiable qu'à la condition que les paramètres bâlois  $EAD_j$ ,  $LGD_j$  et  $PD_j$  soient correctement estimés et reflètent vraiment le risque du portefeuille.

Toutefois on gardera en tête que l'utilisation un modèle de dépendance à un seul facteur de risque commun avec des lois normales centrées réduites réduit la restitution de la réalité des dépendances entre les entreprises. Des études de backtesting doivent permettre de s'assurer de la cohérence et de la prudence d'un tel modèle.

La simulation a montré que la prime « bâloise » est prudente. Ce conservatisme repose en partie sur une connaissance fine du portefeuille d'acheteurs au moment de la souscription. A cet effet les études de clientèle doivent être les plus précises possible. Néanmoins le modèle retenu est dynamique et prend en compte la capacité de l'assureur à se désengager « en continu ». Ainsi l'assureur connaît à tout instant sa sinistralité moyenne potentielle.

Cette dernière méthode peut être prolongée dans plusieurs directions. En premier lieu la précision des paramètres bâlois doit être améliorée par des analyses plus fines pour mieux refléter le risque d'impayé des factures commerciales. Un monitoring glissant sur douze mois des taux de défaut par note de solvabilité est fortement recommandé pour ajuster les  $PD$  et capter l'évolution de la sinistralité. En complément une meilleure caractérisation de l'exposition au défaut pourrait être traitée à travers une enquête auprès des assurés pour mieux comprendre la demande et l'utilisation des limites de crédit. Enfin le taux de recouvrement mériterait d'être différencié en segmentant les assurés en groupes homogènes de risque.

D'un point de vue plus théorique, les risques de concentration sectorielle ou sur une contrepartie méritent une attention particulière. Par exemple une dégradation de la solvabilité des entreprises dans le secteur du bâtiment ne se propagera pas de la même façon dans toutes les autres branches de l'économie.

En synthèse ce mémoire montre que la garantie d'assurance-crédit est assimilable à un risque de crédit bancaire et que les portefeuilles d'acheteurs sont granulaires pour la majorité des assurés. Ainsi à l'exception de certaines concentrations sur des contreparties parfaitement identifiées et suivies dans le cadre de l'appétence aux risques, la prime d'un assuré se calcule de manière additive.

# Bibliographie

- [BRUN 2014] V.Brunel, B.Roger, Le Risque de Crédit, *Economica*
- [BUHL 2005] H.Bühlmann, A.Gisler, A course in Credibility Theory and its Applications, *Springer*
- [CAME 2014] A.Cameron, P.Trivedi, Regression Analysis of Count Data, *Cambridge*
- [CHAR 2014] A.Charpentier, Computational Actuarial Science with R, *CRC Press*
- [CHAR 2015] A.Charpentier et al., Segmentation et Mutualisation, les deux faces d'une même pièce, *Risques n° 103*
- [DENU 2004] M.Denuit, A.Charpentier, Mathématiques de l'assurance non-vie : Tome 1, Principes fondamentaux de théorie du risque, *Economica*
- [DENU 2005] M.Denuit, A.Charpentier, Mathématiques de l'assurance non-vie : Tome 2, Tarification et provisionnement, *Economica*
- [DIET 2003] M.Dietsch, De Bâle II vers Bâle III : les enjeux et les problèmes du nouvel Accord, *Revue d'économie financière*
- [GOUL 2023] V.Goulet, Théorie de la crédibilité avec R, *vigou3.gitlab.io*
- [HARD 2018] J.Hardin, J.Hilbe, Generalized Linear Models and Extensions, *Stata Press*
- [HILB 2011] J.Hilbe, Negative Binomial Regression, *Cambridge*
- [HILB 2014] J.Hilbe, Modeling Count Data *Cambridge*
- [LUTK 2008] E.Lütkebohmert, Concentration Risk in Credit Portfolios, *Springer*
- [OHLS 2010] E.Ohlsson, Non-Life Insurance Pricing with Generalized Linear Models *Springer*
- [PART 2004] C.Partrat, J.Besson, Assurance non-vie : Modélisation, Simulation, *Economica*
- [JONG 2008] P.Jong, G.Heller, Generalized Linear Models for Insurance Data, *Cambridge*
- [PLAN 2004] F.Planchet, Le modèle collectif, *www.ressources-actuarielles.net/*
- [RONC 2005] T.Roncalli, La Gestion des Risques Financiers, *Economica*
- [SAPO 2011] G.Saporta, Probabilités, Analyse des Données et Statistiques, *TECHNIP*
- [STOE 2023] J.Stoehr, Méthodes de Monte Carlo, *www.ceremade.dauphine.fr/ stoehr/*
- [THER 2018] P.Thérond, Théorie de la crédibilité, *www.therond.fr*