

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ANALYSE ET ÉVALUATION DE LA MÉTHODE DE TARIFICATION A
POSTERIORI AVEC LES DONNÉES DE PANEL

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

CARLOS FELIPPE ROSTAND KOETZ

JUIN 2018

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.10-2015). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Ce mémoire est la conclusion d'une étape d'études de ma vie qui a duré à peu près trois ans et demi. Il a été possible grâce au soutien de plusieurs personnes à qui je voudrais adresser mes sincères remerciements.

Tout d'abord, je veux remercier mon épouse madame Marceline De Mello, qui m'a donné tout le support familial, et mon enfant Oliver De Mello Koetz, qui m'a donné la motivation pour continuer et finaliser ce travail.

Je voudrais remercier monsieur Jean-Philippe Boucher, directeur de recherche de ce mémoire, pour son aide, temps et confiance qu'il m'a accordés. Je voudrais aussi remercier trois autres professeurs, qui m'ont fait des recommandations m'invitant à joindre à la communauté de l'Université du Québec à Montréal (UQAM), messieurs Luiz Ernesto Both, Éder Júlio Kinast et José Antônio Lumertz. Finalement, je veux aussi remercier la directrice du programme de français langue seconde de l'UQAM, madame Agnès Baron et madame Rachel Sparks qui m'ont soutenu dans mon travail pour la mise en forme de la langue française.

Enfin, j'adresse mes plus sincères remerciements à mes parents Carlos Tadeu Koetz et Angela Lizete Rostand Koetz, qui m'ont accompagné, aidé, soutenu et encouragé tout au long de la réalisation de ce mémoire.

TABLE DES MATIÈRES

LISTE DES TABLEAUX	vi
LISTE DES FIGURES	ix
RÉSUMÉ	x
INTRODUCTION	1
CHAPITRE I	
LES FONDEMENTS DE LA TARIFICATION ET LA BASE DE DONNÉES	4
1.1 La prime pure	4
1.2 La segmentation	6
1.3 La base de données	8
CHAPITRE II	
LA TARIFICATION AVEC LES DONNÉES TRANSVERSALES	19
2.1 Les modèles linéaires généralisés	19
2.1.1 Les modèles linéaires	20
2.1.2 La théorie des GLMs	21
2.2 La tarification par expérience et le système Bonus-Malus (SBM)	25
2.2.1 Les limitations des modèles de classification et l'historique de sinistre comme prédicteur	25
2.2.2 Historique des SBMs	26
2.2.3 L'objectif des SBMs	27
2.2.4 Les types de SBMs	28
2.2.5 Le SBM en détail	29
2.3 La structure mathématique d'un SBM standard	30
2.3.1 La structure d'un SBM	30
2.3.2 Le parcours des assurés dans le système	31
2.3.3 Les règles de transition	32

2.3.4	Les probabilités et la matrice de transition	34
2.3.5	Propriété stationnaire	35
2.3.6	Relativités	37
2.3.7	Interaction entre un système « <i>a priori</i> » (GLM) et un système « <i>a posteriori</i> » (SBM)	37
CHAPITRE III		
	LES MODÈLES DE BASE ET DE CONTRÔLE	39
3.1	Le modèle proposé	40
3.1.1	Problèmes de l'utilisation de la théorie SBM pour les données de panel	40
3.1.2	Inclusion d'une structure bonus-malus dans la tarification	42
3.2	Les paramètres du modèle	44
3.2.1	La distribution Poisson	44
3.2.2	L'estimation du niveau d'entrée l_1 de l'assuré selon le nombre d'années de permis	45
3.2.3	L'estimation du niveau des assurés avec expérience	47
3.2.4	La structure des relativités	50
3.3	La contrainte légale et le modèle de contrôle	51
3.3.1	Le modèle de contrôle	52
3.4	L'estimation de chaque modèle	54
3.4.1	Le parcours des assurés	56
3.5	Applications numériques	60
3.5.1	Le modèle de base	60
3.5.2	Le modèle de contrôle	65
CHAPITRE IV		
	ANALYSE DE LA CAPACITÉ PRÉDICTIVE ENTRE LES DEUX MO- DÈLES	75
4.1	Outils de comparaison classiques	76
4.1.1	Ajustement statistique	76

4.1.2	Le coefficient de variation	77
4.1.3	Le niveau relatif moyen stationnaire	85
4.2	Le ratio sinistre / prime	87
4.3	La compétition entre modèles	90
4.3.1	Mise en contexte	90
4.3.2	Résultats généraux	92
4.3.3	Analyse approfondie de la compétition	96
4.3.4	Résultats du marché	106
4.3.5	Décision	110
	CONCLUSION	113
	RÉFÉRENCES	119

LISTE DES TABLEAUX

Tableau	Page
1.1 Distribution des primes entre les assureurs et le choix de compagnie par l'assuré	8
1.2 Sommaire des variables	10
1.3 Distribution du nombre de polices d'assurances et de l'exposition des assurés	11
2.1 Sommaire des paramètres de distributions de la famille exponentielle	23
3.1 Historique de sinistres de l'assuré numéro 1 - modèle de base . . .	57
3.2 Historique de sinistres de l'assuré numéro 1 - modèle de contrôle .	59
3.3 Les β s calculés pour le modèle de base	61
3.4 relativités calculées pour le modèle de base	62
3.5 Les β s estimés pour le modèle de contrôle	66
3.6 Description du profil de l'assuré de base pour les deux modèles . .	66
3.7 Calcul de la prime de base de l'assuré avec le profil de base pour chaque modèle	68
3.8 Comparaison des β des modèles	69
3.9 relativités calculées pour le modèle de contrôle	71
4.1 Les coefficients de variation globaux des deux modèles	78
4.2 Les coefficients de variation globaux des modèles	80
4.3 Les coefficients de variation globaux des modèles, sans le modèle de classification <i>a priori</i>	82
4.4 Les coefficients de variation globaux des modèles avec et sans la segmentation	83

4.5	Les coefficients de variation du modèle de base par année	83
4.6	Les coefficients de variation du modèle de contrôle par année	84
4.7	Comparaison entre le NRMS des deux modèles	86
4.8	Comparaison entre les NRMS des nos modèles et les NRMS calculés par (Lemaire, 1995)	88
4.9	Sommaire de primes et sinistres avec le ratio sinistre / prime par modèle	89
4.10	Sommaire de la compétition entre les deux modèles/compagnies	93
4.11	évaluation du déplacement des assurés	95
4.12	évaluation relative du mouvement des assurés corrigé	95
4.13	Distribution des assurés selon la valeur de la prime et par modèle pour l'année 1	98
4.14	Distribution des sinistres selon la prime et le modèle pour la première année	98
4.15	Distribution de primes par bande de prime et par modèle pour la première année	100
4.16	Distribution des assurés par bande prime et par modèle pour l'année 1 pour ceux qui ont changé d'un modèle à l'autre.	101
4.17	Distribution des sinistres par bande de prime et par modèle pour l'année 1 pour les assurés qui ont décidé de changer de compagnie.	102
4.18	Distribution des primes par bande de prime et par modèle pour l'année 1 pour les assurés qui ont décidé de changer un modèle par l'autre.	103
4.19	Distribution des assurés par bande de prime et par modèle pour l'année 1 pour les assurés qui ont décidé de rester dans la même compagnie.	104
4.20	Distribution des sinistres par bande de prime et par modèle pour l'année 1 pour les assurés qui ont décidé de rester dans la même compagnie.	104

4.21	Somme des primes selon la tranche de prime et le modèle pour l'année 1 pour les assurés qui ont resté dans le même modèle. . . .	105
4.22	La prime et le sinistres (somme et moyenne) par bande de prime .	107

LISTE DES FIGURES

Figure	Page
1.1 Distribution de l'âge des assurés	12
1.2 Distribution de l'âge des véhicules	13
1.3 Distribution de la possession d'une maison	14
1.4 Distribution du nombre de conducteurs dans la police d'assurances	15
1.5 Distribution du statut matrimonial des assurés	16
1.6 Distribution d'infractions mineures au dossier	17
1.7 Distribution du NAP dans portefeuille	18
3.1 Parcours de l'assuré numéro 1 selon son historique de sinistres et la structure du modèle de base	58
3.2 Parcours de l'assuré numéro 1 dans le modèle de contrôle	59
3.3 Matrice de transition du modèle de base	63
3.4 Niveau d'entrée estimé des assurés pour le modèle de base	64
3.5 Graphique comparant les relativités des deux modèles	72
3.6 Matrice de transition du modèle de contrôle	73
3.7 Niveau d'entrée estimé pour le modèle de contrôle	74
3.8 Comparaison du niveau d'entrée estimé pour les deux modèles	74
4.1 Comparaison entre les coefficients de variation des deux modèles	84
4.2 L'équilibre actuariel du marché	109

RÉSUMÉ

La tarification d'assurance Incendie, Accidents et Risques Divers (IARD) est une importante tâche des actuaires depuis très longtemps. Le calcul a toujours été divisée entre « *a priori* » et « *a posteriori* ». Récemment, les actuaires essayent différentes façons d'utiliser ces approches ensemble. Boucher et Inoussa (2014) ont proposé une façon de les combiner grâce à un nouveau modèle qui estime tous les paramètres en même temps en utilisant les données d'un assureur de la province de l'Ontario pour l'illustrer. En conséquence, leur modèle a dû être modifié afin d'inclure une contrainte légale de cette région. Ce mémoire a pour objectif, parmi d'autres, d'analyser et d'évaluer l'efficacité de leur modèle. Pour ce faire, nous avons passé en revue les fondements et nous avons utilisé la même base de données qu'ils ont utilisée. La reproduction du modèle a été mise en place avec tous les détails de chaque composante du calcul. Ensuite, un modèle de contrôle a été fait pour comprendre en détail la perte de l'efficience à cause d'une telle imposition légale. Nous avons utilisé des outils d'évaluation proposés par Lemaire (1995) pour comparer le modèle principal avec d'autres modèles déjà connus. De plus, nous avons utilisé un nouvel outil pour vérifier la différence de comportement entre les modèles dans un environnement de compétition de marché. Les résultats obtenus vont dans la direction attendue depuis le début de cette recherche : la nouvelle méthode est plus efficace que les modèles traditionnels et l'inclusion de la contrainte légale dans le modèle principal a réduit sa performance.

Mot clés Tarification « *a priori* », tarification « *a posteriori* », assurance IARD, assurance auto, données transversales, données horizontales, données de panel, modèles linéaires généralisés, système de bonus-malus, Poisson, crédibilité, compétition, assurances, actuariat.

INTRODUCTION

La tarification en assurance Incendie, Accidents et Risques Divers (IARD) est une des plus importantes tâches dans le travail d'un actuare, et ce, depuis très longtemps. Deux méthodes ont toujours été en conflit afin de devenir la principale méthode de tarification en assurance : la tarification « *a priori* » et la tarification « *a posteriori* ». Dans la première, l'actuaire utilise les caractéristiques du risque de l'assuré afin d'établir une prime d'assurance, tandis que dans la deuxième, la tarification « *a posteriori* », l'actuaire utilise l'expérience des réclamations passées de l'assuré afin de fixer la prime d'assurance. Avec l'évolution des deux types de tarification, de nouveaux modèles de tarification sont apparus. D'ailleurs, des structures qui mélangent les deux méthodes deviennent de plus en plus courantes.

Le modèle proposé par (Boucher et Inoussa, 2014) marque une évolution des modèles courants dans le domaine de l'actuariat. Les auteurs proposent d'utiliser une combinaison des deux méthodes : l'incorporation d'un modèle « *a posteriori* », par exemple un système Bonus-Malus (noté SBM), dans un modèle de tarification par classification « *a priori* », par exemple en utilisant un GLM - « *Generalized Linear Models* », afin de réduire l'hétérogénéité résiduelle à l'intérieur de chaque classe. Dans leur proposition, les paramètres des deux parties sont toujours estimés ensemble, c'est-à-dire qu'ils ne sont pas faits un à la fois. L'élément clé de cette méthode est l'utilisation des données de panel au lieu des données transversales.

En pratique, ainsi que dans une bonne partie de la littérature actuelle, la modélisation de ce type de structure de tarification est différente de ce qui a été proposé par les auteurs, soit l'estimation simultanée des paramètres des modèles « *a priori*

» et « *a posteriori* ». La plupart des assureurs font cette modélisation en deux étapes. En premier lieu, ils estiment les paramètres de la classification « *a priori* » et ils choisissent le meilleur modèle selon la base de données disponible. En second lieu, en utilisant le modèle de tarification « *a posteriori* », ils complètent le modèle par classification « *a priori* » en estimant leurs paramètres. En tentant de corriger la structure qui combine les deux modèles dans une seule, ils incluent un facteur d'ajustement pour équilibrer l'équation de base de la tarification actuarielle, dans laquelle la somme des primes doit être égale à la somme des sinistres. Le résultat de cette approche n'est pas toujours le modèle le plus optimal, car l'ajustement est généralement faite avec l'estimation de la structure « *a priori* », qui considère la section « *a posteriori* » comme un ajustement et qui ne la considère pas comme une vraie composante du modèle.

Ainsi, ce mémoire a trois objectifs principaux :

1. Reproduire le modèle présenté par (Boucher et Inoussa, 2014) dans leur article : « *A Posteriori Ratemaking With Panel Data* » ;
2. Évaluer l'efficacité du modèle de (Boucher et Inoussa, 2014) selon des critères sélectionnés. Nous proposons de le comparer à un modèle de contrôle pour pouvoir quantifier les résultats des analyses. Ces analyses sont faites à l'aide des outils actuariels et statistiques d'évaluation des SBMs ;
3. Évaluer la contrainte légale incluse dans le modèle de (Boucher et Inoussa, 2014). Cette contrainte fait référence aux types de sinistres responsables et non responsables et elle est imposée par une loi de la province de l'Ontario. Elle protège les assurés ayant un sinistre non responsable d'avoir leurs primes augmentées par la compagnie d'assurance. Cette restriction a le potentiel d'influencer l'efficacité du modèle, et nous proposons un nouveau moyen d'évaluer cette contrainte.

Pour atteindre nos objectifs, nous passerons en revue les notions qui constituent la base du modèle dans les chapitres 1 et 2. Dans le chapitre 3, nous analyserons plus en détails l'article de (Boucher et Inoussa, 2014). Plus précisément, nous examinerons le modèle proposé ainsi que les résultats obtenus suite à l'application de ce modèle sur la base de données utilisée. Nous mènerons notre étude avec le modèle de contrôle. Nous estimons tous ses paramètres et sa structure optimale. Ainsi, le chapitre 3 peut être vu comme la réplique du modèle de (Boucher et Inoussa, 2014), avec tous les détails que nous avons utilisé pour y contruire.

Dans le chapitre 4, nous appliquerons les outils actuariels et statistiques afin d'évaluer l'efficacité des deux modèles étudiés. Pour mieux comprendre comment l'ajout de la contrainte légale affecte l'efficacité du modèle de (Boucher et Inoussa, 2014), nous introduisons un modèle de compétition afin d'observer le comportement des modèles quand ces deux cherchent à attirer les assurés à leur portefeuille. Ce chapitre 4 représente notre contribution académique par rapport au article étudié.

CHAPITRE I

LES FONDEMENTS DE LA TARIFICATION ET LA BASE DE DONNÉES

Dans ce chapitre, nous examinons tous les fondements de la tarification classique en assurance IARD. Les deux principaux éléments sont la prime pure et la segmentation. La prime pure représente l'espérance des sinistres, c'est-à-dire le coût moyen de réclamation, tandis que la segmentation représente la personnalisation de cette prime selon les caractéristiques reliées aux risques assurés. De plus, nous explorons la base de données utilisée en analysant toutes les variables présentes dans ce travail.

1.1 La prime pure

De nos jours, une des plus importantes tâches de l'actuaire est de bien tarifier le risque que les compagnies d'assurance assument. En passant des plus simples aux plus complexes, diverses techniques ont été développées afin de mieux estimer ces risques. Le sinistre est défini par :

$$S = \sum_{i=1}^N R_i. \quad (1.1)$$

Si nous supposons que N et R sont indépendantes et que les R_i sont des variables aléatoires i.i.d.¹, nous pouvons déterminer l'espérance du sinistre comme :

$$E[S] = E\left[\sum_{i=1}^N R_i\right], \quad (1.2)$$

$$E[S] = E[N]E[R]. \quad (1.3)$$

Cette équation est connue comme l'équation de la prime pure et elle est la base de la théorie du risque². Celle-ci s'exprime comme suit :

$$\text{Prime Pure} = E[S] = E[N]E[R], \quad (1.4)$$

dans laquelle :

- $E[S]$ est l'espérance des sinistres,
- $E[N]$ est l'espérance du nombre de sinistres (fréquence), et
- $E[R]$ est l'espérance du coût d'un sinistre (sévérité).

Dans la modélisation traditionnelle, nous pouvons choisir de modéliser la prime pure ($E[S]$) ou chacune de ses composantes, soit $E[N]$ et soit $E[R]$, pour ensuite les combiner. Le modèle présenté assure la complète solidarité entre tous les assurés du portefeuille car il n'y a pas de différence entre les primes payées par chacun des assurés. La solidarité pure est adéquate lorsque nous sommes en présence d'un groupe ayant des caractéristiques fortement homogènes ou dans les assurances sociales.

Pour une compagnie d'assurances, il n'est pas avantageux d'ajouter une totale solidarité dans les modèles de tarification. L'assureur doit plutôt charger une prime adéquate à chacun des assurés en fonction du risque qu'il représente. Étant donné

1. Indépendantes et identiquement distribuées.

2. Voir (Klugman *et al.*, 2008) ou (Newton L. Bowers *et al.*, 1997) pour plus de détails.

qu'il n'existe pas de modèle qui capte parfaitement les caractéristiques du risque avec une quantité raisonnable de variables, l'assureur travaille avec des estimations. Ainsi, nous ne pouvons pas éliminer complètement les différences entre la prime calculée et les sinistres survenus. Comme nous segmentons notre portefeuille, dans l'esprit de créer des classes d'assurés les plus homogènes possible, nous réduisons la crédibilité de toutes nos estimations car nous utilisons moins de données pour chaque tranche de segmentation. Il faut décider quel serait le niveau de crédibilité et de segmentation que nous voulons inclure dans nos calculs. Concernant ce niveau, il existe une balance entre la précision du modèle et sa crédibilité. Beaucoup de jugement de l'analyste est nécessaire pour mettre en place une structure de tarification d'assurance qui soit raisonnable.

1.2 La segmentation

L'équation suivante exprime de manière plus appropriée comment nous incluons la segmentation dans la prime pure :

$$\text{Prime Pure}|\mathbf{X} = E[S|\mathbf{X}] \quad (1.5)$$

$$E[S|\mathbf{X}] = E[N|\mathbf{X}]E[R|\mathbf{X}] \quad (1.6)$$

Le vecteur des caractéristiques $\mathbf{X} = (x_1, \dots, x_p)$ nous aide à segmenter le portefeuille par rapport à certaines caractéristiques du risque. Toutefois, nous n'utilisons pas toutes les variables pour segmenter, sinon nous instaurerions un surajustement du modèle, ce qui n'est pas préférable. Donc, malgré la segmentation du portefeuille, il y a encore de l'hétérogénéité à l'intérieur des classes d'assurés. Certaines techniques ont été développées pour corriger cet effet résiduel.

La mutualisation et la solidarité sont des fondements de la tarification en actuariat. Visant à réparer les conséquences du hasard, la mutualisation fait le transfert des primes entre les assurés. Chaque police d'assurances qui compose le portefeuille

possède des caractéristiques différentes comportant certains risques. Afin d'éviter le déséquilibre entre le montant que chaque assuré paye en prime et le risque que chacun représente pour la compagnie, un ajustement doit être fait. En conséquence, il faut ajuster la prime selon les caractéristiques de chaque assuré et le risque qui l'accompagne.

La mutualisation et la solidarité sont toujours présentes entre les assurés même dans un univers segmenté. Les actuaires sont capables de prévoir avec une certaine précision la quantité et la valeur des pertes, mais ils ne sont pas capables de dire quels individus subiront des sinistres. Malgré la segmentation, la mutualisation et la solidarité jouent un rôle fondamental.

Afin d'illustrer ceci, voici un exemple. Nous avons un portefeuille composé de quatre assurés $i = 1, 2, 3, 4$, et nous prévoyons $E[S] = 1000$ \$CA. Supposons que nous avons deux assureurs : l'assureur A ne tarifie pas par segmentation, il est dans un environnement complètement mutualiste et solidaire. La prime d'assurance que l'assureur A va charger à chacun de ses quatre assurés est de $P_i = \frac{E[S]}{4} = \frac{1000 \text{ \$CA}}{4} = 250$ \$CA. L'assureur B applique, dans son système de tarification, une technique de segmentation. Ainsi, il charge à chacun de ses quatre assurés les primes $P_1 = 100$ \$CA, $P_2 = 50$ \$CA, $P_3 = 500$ \$CA et $P_4 = 350$ \$CA (avec un total de 1000 \$CA).

Nous supposons qu'un potentiel assuré fait son choix uniquement en se basant sur la valeur de la prime d'assurance. Regardons les conséquences de la segmentation pour chacun des assureurs A et B. L'assureur A attire vers son portefeuille les assurés 3 et 4, tandis que 1 et 2 vont auprès de B. Ce dernier a fait le calcul suivant : $P_i^B = E[S_i]$. L'assureur A collecte 500 \$CA en prime, alors qu'il débourse 850 \$CA en sinistre. Il est évident qu'il fait du déficit. De l'autre côté, l'assureur B recueille 150 \$CA en primes et débourse les mêmes 150 \$CA en sinistres. Contrairement

à l'assureur A, sa situation n'est pas déficitaire.³ Le sommaire de ce mouvement est disponible dans le Tableau 1.1.

Tableau 1.1 Distribution des primes entre les assureurs et le choix de compagnie par l'assuré

Assuré (i)	Prime assureur A	Prime assureur B	Choix de l'assuré	Déficit
1	250 \$CA	100 \$CA	B	0 \$CA
2	250 \$CA	50 \$CA	B	0 \$CA
3	250 \$CA	500 \$CA	A	250 \$CA
4	250 \$CA	350 \$CA	A	100 \$CA

Comme nous le remarquons, la segmentation est un élément important pour faire face à la compétition et pour éviter l'antisélection. Lorsque nous segmentons, la prime d'assurances devient plus personnalisée parce qu'elle différencie les assurés. La non segmentation, ou la mauvaise segmentation, et la compétition causent l'antisélection, car les personnes à risque élevé vont s'assurer auprès des compagnies dans lesquelles l'écart entre la prime chargée et le risque apporté est le plus grand. La segmentation de la prime ajuste mieux le niveau de risque acquis par l'assureur. Si deux assureurs sont en compétition, la segmentation devient un facteur de distinction du point de vue du consommateur. La compagnie qui adapte ses primes par rapport aux risques évite l'antisélection.

1.3 La base de données

Nous présentons dans cette section la base de données utilisée pour la tarification actuarielle de ce travail. Nous nous concentrons sur l'exploration des variables de cette base pour avoir une meilleure vision de la composition de ces éléments. Chaque observation dans la base de données représente l'information annuelle

3. Toutefois, il pourrait souffrir des effets négatifs de la faillite de l'assureur A.

d'une police d'assurances liée à un assuré. Chaque police peut avoir jusqu'à six ans d'informations.

Dans les textes actuariels, on suppose que les polices d'assurance sur plusieurs années d'un même assuré sont indépendantes. Nous disons que la base de données est en coupe transversale. Lorsque nous supposons une dépendance entre les contrats d'un même assuré, nous parlons de données de panel. Cette approche est utilisée dans ce travail et cela a été un des éléments clés du modèle présenté par (Boucher et Inoussa, 2014) dans leur article. Heureusement, la base de données que nous possédons est la même que celle que (Boucher et Inoussa, 2014) ont utilisée. Cela nous permet d'avoir plus de certitude quant à la reproduction de leurs résultats.

La base de données est composée de 167 859 observations réparties sur six ans d'exposition. Nous y retrouvons 57 037 assurés ayant des informations individuelles pour chacune des années. Ces informations s'appellent les « caractéristiques du risque ». Elles se regroupent en 11 variables binaires (le vecteur pour l'assuré i au temps t $\mathbf{X}_{i,j,t} = (x_{i,1,t}, x_{i,2,t}, \dots, x_{i,11,t})$) faisant référence à six caractéristiques du risque, à savoir l'âge de l'assuré, l'âge du véhicule, la possession d'une maison, le nombre de conducteurs, le statut matrimonial et les infractions mineures de l'assuré. Le Tableau 1.2 présente une description détaillée de chacune des variables.

Cette base de données provient d'une compagnie d'assurances canadienne et les informations couvrent de la province de l'Ontario. En général, les bases de données développées par des compagnies d'assurances ont des centaines de variables. C'est pourquoi une présélection est toujours accomplie afin de garder les variables les plus significatives statistiquement. Bien que cela soit intéressant, nous ne nous occupons pas de la sélection de variables dans cette recherche. En conséquence, nous supposons que ces variables binaires sont statistiquement significatives et

Tableau 1.2 Sommaire des variables

Variable	Valeur	Description
$X_1 =$	1	= l'assuré est âgé entre 16 et 25 ans.
$X_2 =$	1	= l'assuré est âgé entre 26 et 60 ans.
$X_3 =$	1	= le véhicule est neuf.
$X_4 =$	1	= construction du véhicule : entre 1 et 3 ans.
$X_5 =$	1	= construction du véhicule : entre 4 et 5 ans.
$X_6 =$	1	= l'assuré a une maison.
$X_7 =$	1	= il existe un seul conducteur.
$X_8 =$	1	= il existe deux conducteurs.
$X_9 =$	1	= l'assuré est célibataire.
$X_{10} =$	1	= l'assuré est divorcé.
$X_{11} =$	1	= il n'a pas reçu d'infractions mineures.

que la présélection est déjà accomplie⁴.

La base de données présente une contrainte : il n'y a pas de nouveaux arrivants dans la matrice, c'est-à-dire qu'il n'existe pas d'assurés qui intègrent le portefeuille. Par contre, la sortie des assurés est possible. Pour cette raison, nous avons une réduction considérable d'assurés et d'exposition durant la période d'observation. Nous disons alors que la base de données est dans un environnement semi-fermé dans lequel l'entrée des assurés n'est pas possible mais leur départ l'est.

Étant donné que le nombre de polices d'assurances varie annuellement, nous avons

4. Il est important de noter que le but de ce travail n'est pas d'analyser quelles variables sont les plus significatives dans un modèle quelconque, comme un GLM, mais plutôt d'utiliser la même base de données que (Boucher et Inoussa, 2014) ont utilisé afin d'estimer les paramètres des modèles que nous étudions dans ce travail.

travaillé avec des informations relatives, ce qui est approprié lorsque nous travaillons avec des données dynamiques dans le temps. Nous présentons dans le Tableau 1.3 le nombre de polices actives ainsi que l'exposition annuelle.

Tableau 1.3 Distribution du nombre de polices d'assurances et de l'exposition des assurés

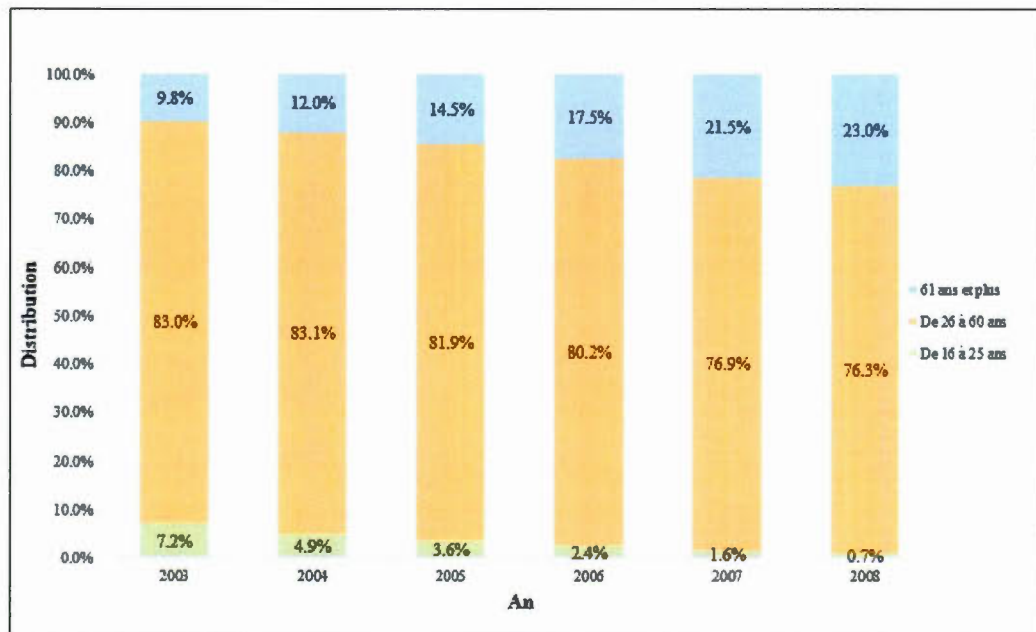
Temps	Année de début	Comptage	Exposition
	de la couverture		
1	2003	57 037	46 886
2	2004	42 627	38 655
3	2005	31 075	27 789
4	2006	22 055	19 396
5	2007	14 926	7 308
6	2008	139	67

Analysons brièvement les variables afin de mieux les comprendre avant de les utiliser dans un modèle. Pour chacun des assurés, le vecteur avec les caractéristiques varie d'une année à l'autre. Alors, nous sommes intéressés à la distribution de chaque caractéristique et à son impact sur les paramètres du modèle au fil des années. Notons que l'année 2008 est une année d'exposition résiduelle de l'année 2007. Par le fait même, elle n'est pas crédible et elle peut avoir beaucoup de variabilité.

Tous les graphiques suivants présentent la distribution des caractéristiques du risque. Nous commençons par la distribution de l'âge des assurés dans le graphique 1.1, ce qui comprend les variables binaires x_0 , x_1 et x_2 .

Nous avons observé deux tendances par rapport au changement de l'âge des assurés dans le graphique 1.1. La première tendance est relative à la réduction de

Figure 1.1 Distribution de l'âge des assurés

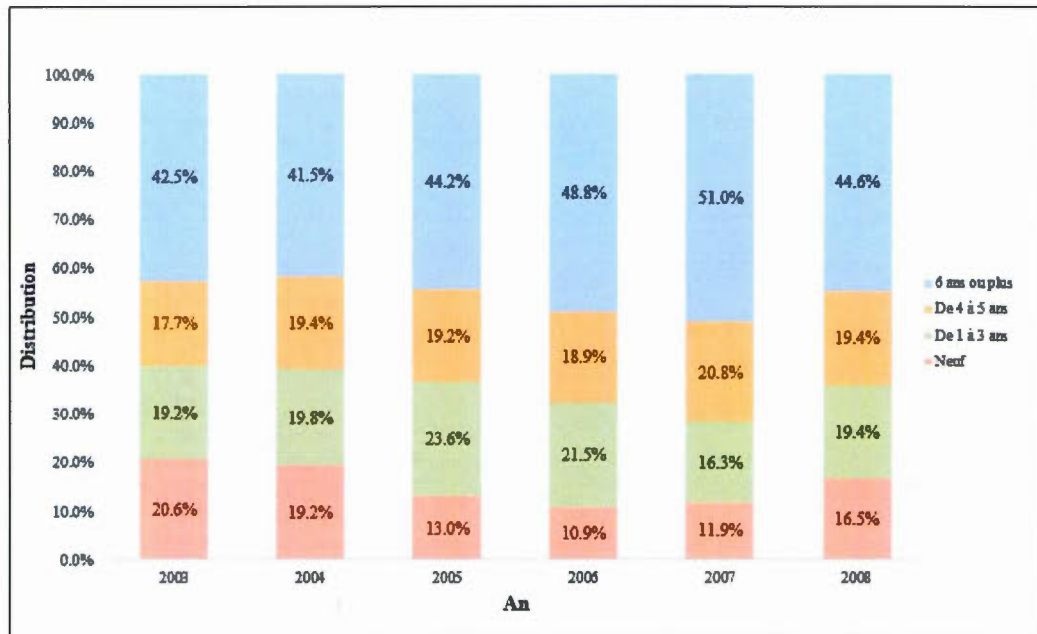


deux groupes d'assurés : les jeunes et le bloc intermédiaire. La deuxième est l'augmentation de la classe des assurés plus âgés.

En général, les jeunes assurés magasinent leurs assurances en fonction de la prime payée. Ils cherchent le prix le plus abordable, ce qui crée une propension à sortir du portefeuille d'assurances. Les assurés plus âgés n'ont pas nécessairement ce critère de sélection. Il existe une seule façon de changer de la classe moyenne vers la classe âgée, qui est à travers le vieillissement naturel des assurés et la reclassification de groupe d'âge.

La variable « âge du véhicule » est illustrée dans le graphique 1.2. Elle possède une importante caractéristique : la valeur peut changer librement de catégorie puisqu'elle n'est pas liée aux caractéristiques intrinsèques de l'assuré. Elle varie selon le désir de l'assuré, car nous observons souvent un changement de véhicule

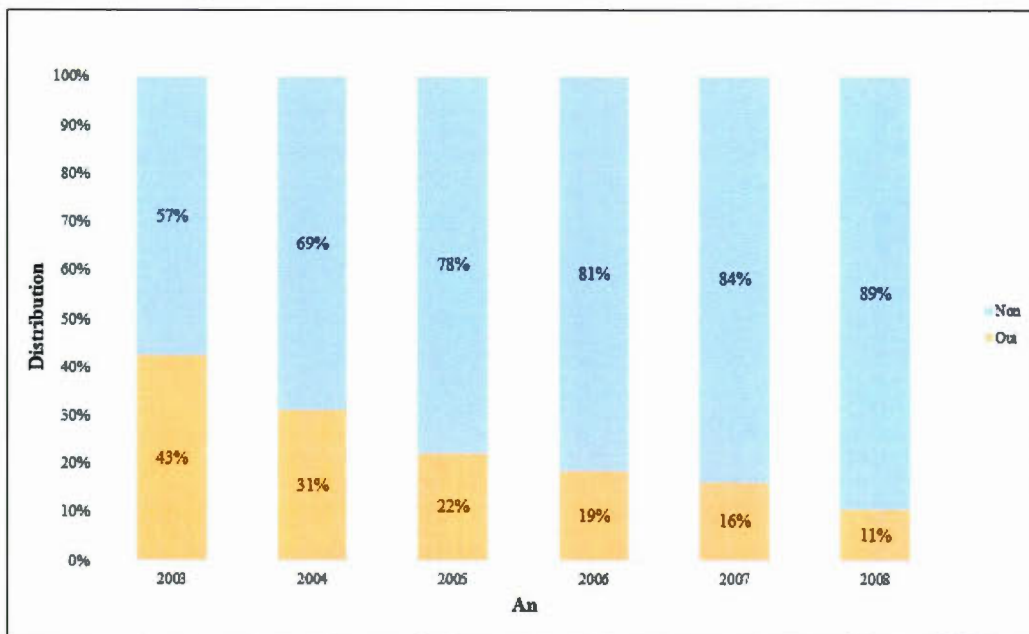
Figure 1.2 Distribution de l'âge des véhicules



chez les assurés. Ce phénomène est d'ailleurs bien illustré dans le graphique 1.2. Sous toute réserve, nous remarquons une tendance du vieillissement des véhicules. De toute façon, ces changements ne sont pas majeurs.

Le graphique 1.3 dévoile un changement majeur dans la caractéristique relative au fait d'être propriétaire. Nous remarquons une tendance à la baisse pour cette caractéristique. En ce moment, il n'est pas possible de tirer des conclusions logiques sur cette baisse importante. D'habitude, la tendance attendue serait le contraire puisque les probabilités de devenir propriétaire d'une maison augmente avec l'âge des individus. En plus, comme l'âge des véhicules, cette variable change librement de valeur car elle n'est pas directement liée aux caractéristiques de l'assuré. Toutefois, nous croyons que ce fait est relié aux assurés qui quittent la compagnie parce qu'un propriétaire devenant un locataire est moins probable.

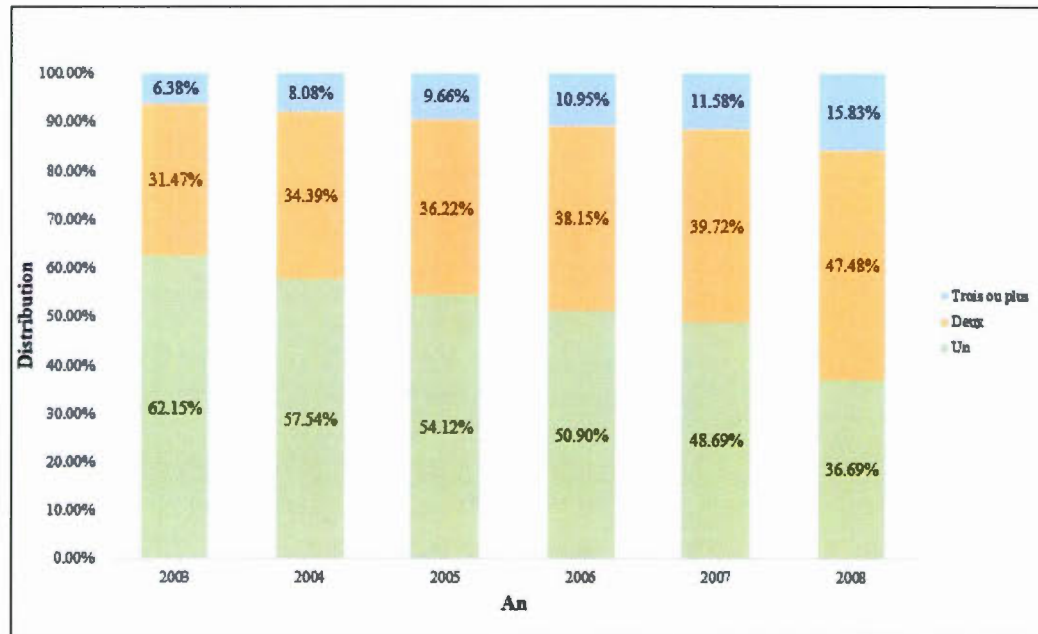
Figure 1.3 Distribution de la possession d'une maison



Le graphique 1.4 présente la distribution du nombre de conducteurs d'une police d'assurance. Il y a un changement pour les trois catégories possibles du nombre de conducteurs. Nous observons une réduction de presque la moitié de polices d'assurance ayant un conducteur, soit de 62% à 37%, et une augmentation de presque 50% du nombre de polices ayant deux conducteurs, soit de 31% à 47%. Le nombre de polices ayant trois conducteurs ou plus a presque triplé, en passant de 6% à 16%. Ces changements peuvent s'expliquer par l'ajout du conjoint ou de la conjointe et/ou de l'enfant au dossier de la police d'assurances.

La distribution du statut matrimonial des assurés est présentée dans le graphique 1.5. Il n'y a que trois variations possibles : de célibataire à marié, de marié à divorcé et de divorcé à marié. Ainsi, nous voyons une tendance à la hausse du nombre d'assurés mariés, soit de 69% à 78%. Le nombre de personnes divorcées est stable

Figure 1.4 Distribution du nombre de conducteurs dans la police d'assurances

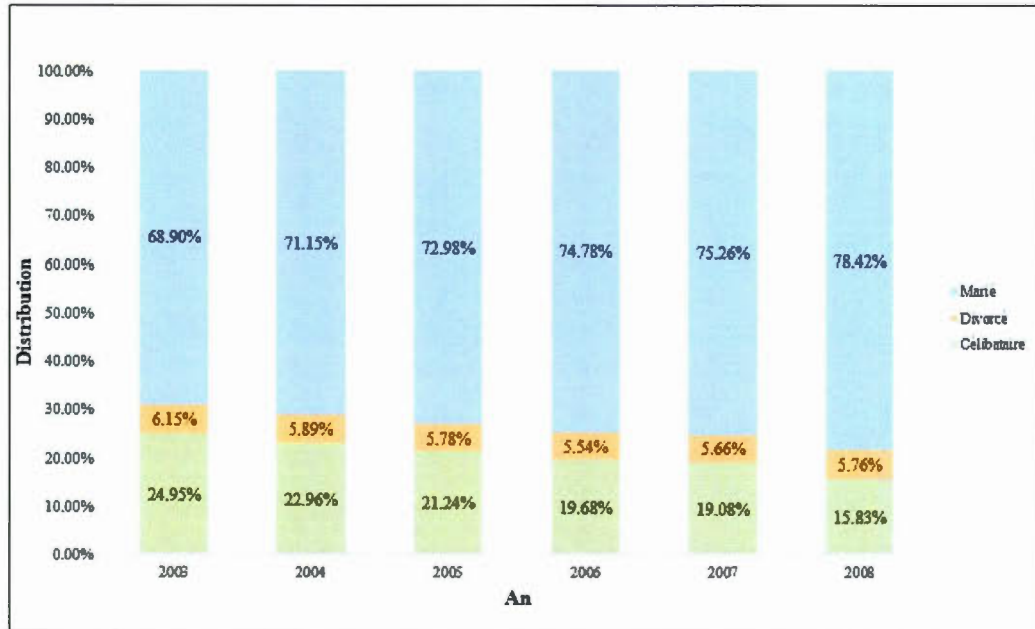


et le nombre d'assurés célibataires évolue en proportion inverse à l'augmentation du nombre de mariés, soit en passant de 25% à 16%.

Le graphique 1.6 présente la distribution des assurés ayant un dossier comportant des infractions mineures. Cette variable est liée directement à la personne assurée et les possibilités de changement sont binaires. Comme il n'existe pas de possibilité d'avoir de nouveaux arrivants, nous nous attendons à une croissance dans le nombre d'assurés ayant ce type de dossier. Par contre, la sortie des assurés du portefeuille est possible. Le chiffre de 11% en 2003 décroît à 1% en 2008. Cela illustre la tendance des assurés ayant ce type de dossier à quitter la compagnie.

Il existe diverses raisons pour lesquelles ces assurés décident de quitter la compagnie. Toutefois, l'explication la plus probable est la surcharge ou la pénalité sur la prime d'assurances payée, ce qui peut les inciter à quitter le portefeuille pour

Figure 1.5 Distribution du statut matrimonial des assurés



trouver des prix plus abordables chez les concurrents.

Une autre variable fondamentale dans ce travail est le nombre d'années depuis l'obtention du permis de conduire (NAP). Cette dernière est utilisée comme variable clé pour évaluer le niveau d'entrée de chaque assuré dans les modèles.

Dans le graphique 1.7, nous constatons qu'il existe une concentration d'environ 80% des assurés ayant entre 6 et 38 ans d'expérience de conduite. Le plus grand NAP observé est de 74 ans. Pour la modélisation, nous avons un seuil au NAP = 6. Plus précisément, au Canada, il est possible de consulter un historique de sinistre pour les six dernières années. 11,6% des assurés ont 6 ans ou moins d'expérience depuis l'obtention du permis de conduire et 88,4% ont 7 ans ou plus.

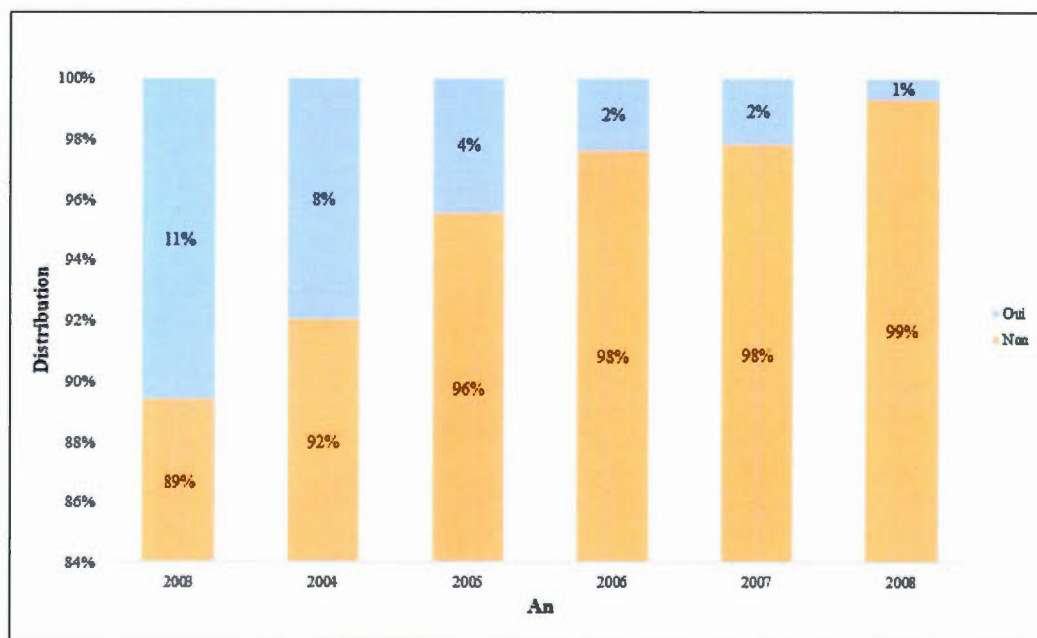
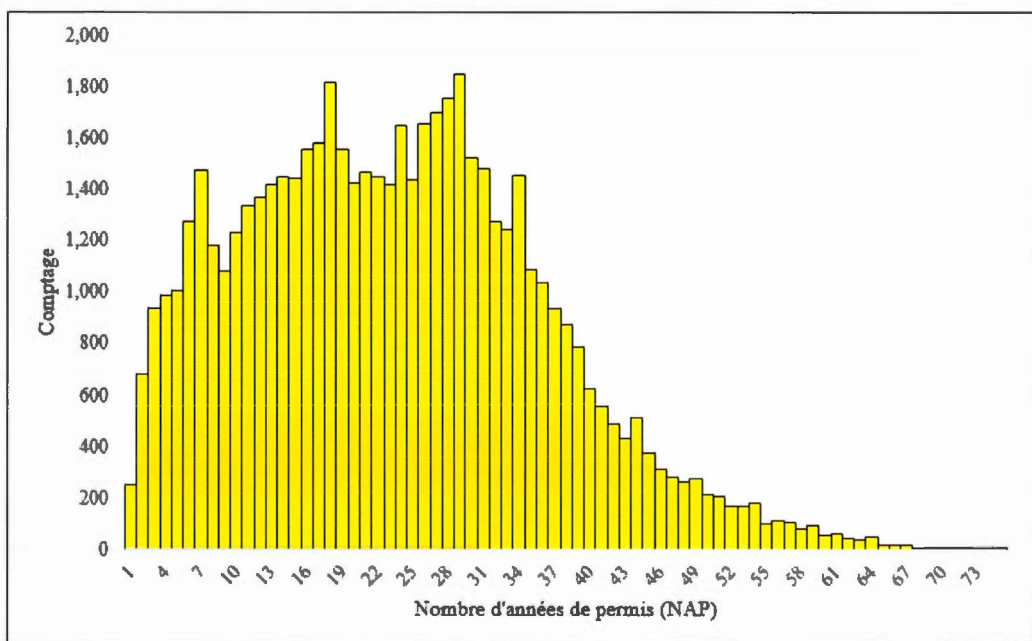
Figure 1.6 Distribution d'infractions mineures au dossier

Figure 1.7 Distribution du NAP dans portefeuille

CHAPITRE II

LA TARIFICATION AVEC LES DONNÉES TRANSVERSALES

Dans ce chapitre nous présentons la base technique et les modèles qui font partie du modèle principal, appelé modèle de base, et du modèle de contrôle utilisés dans ce travail. Cette base technique est composée des modèles linéaires généralisés, de la famille exponentielle linéaire et des systèmes de bonus-malus.

2.1 Les modèles linéaires généralisés

Les modèles linéaires généralisés, aussi appelés « *Generalized Linear Models* » (GLM), sont le résultat d'un travail de généralisation de la théorie des modèles linéaires. Les GLMs ont été présentés pour la première fois dans les travaux de (Nelder et Wedderburn, 1972). Suite à leur apparition et dû à l'augmentation rapide de la puissance computationnelle, beaucoup de progrès ont été observés dans divers champs d'études. L'actuariat en est un bon exemple.

En actuariat, les GLMs sont utilisés pour résoudre les questions liées à la tarification en assurances, à l'évaluation des portefeuilles, à l'analyse du comportement des assurés et à l'évaluation des réserves. Nous les utilisons souvent pour estimer la fréquence, la sévérité et la prime pure. Pour chacun de ces éléments, certaines spécificités du GLM sont utilisées.

En tarification, les GLMs permettent de répondre à certaines questions, par exemple,

est-ce que le nombre de sinistres par année d'un assuré est relié à l'âge ? Au sexe ? Au type de voiture ou à l'emplacement du stationnement ? Ainsi, les GLMs font la liaison entre les variables explicatives, ce que nous appelons les caractéristiques du risque, et le nombre de sinistres par année d'un assuré, ce qui correspond à ce que nous appelons variable réponse, à travers un modèle quelconque. Nous les utilisons afin de différencier les assurés dans un modèle de classification « *a priori* ». Toutefois, il faut se rappeler que même si les GLMs sont des modèles performants, le jugement de l'analyste est primordial afin d'évaluer les limites de l'interprétation.

2.1.1 Les modèles linéaires

Selon la théorie des modèles linéaires, le vecteur réponse des variables aléatoires $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n) \sim \mathcal{N}$ est lié aux variables explicatives de la façon suivante :

$$\mathbf{Y} = \mathbf{X}^T \beta + \epsilon, \quad (2.1)$$

où \mathbf{X} est la matrice des variables explicatives. Pour un assuré i , nous supposons que $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ sont indépendants et identiquement distribués. β sont les paramètres du modèle à estimer et le terme d'erreur ϵ est gaussien. De plus, $E[Y_i] = \mu_i = \mathbf{X}_i^T \beta$.

La dépendance linéaire entre la variable réponse, les variables explicatives, incluant leurs paramètres, et la structure additive du terme d'erreur sont des éléments fondamentaux de la théorie des modèles linéaires classiques. L'estimation des paramètres β et σ^2 peut se faire par maximum de vraisemblance ou de log-vraisemblance :

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2.2)$$

$$\hat{\sigma}^2 = \frac{1}{n - \dim(\beta)} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \hat{\beta})^2. \quad (2.3)$$

Comme les présente (Boland, 2007), les prédicteurs linéaires nous permettent de lier les variables explicatives aux variables réponses :

$$\hat{y}_i = \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip} \equiv \hat{\eta}_i \quad (2.4)$$

et le résidu généré par ce prédicteur est

$$\hat{\epsilon}_i = y_i - \mathbf{x}_i^T \hat{\beta} = y_i - \hat{\mu}_i. \quad (2.5)$$

Selon (Boland, 2007), le désavantage de la théorie linéaire est que $\hat{\eta}_i \in (-\infty, +\infty)$, ce qui permet d'avoir des valeurs négatives et positives sans aucune limite supérieure ou inférieure. Pour bien modéliser les variables d'intérêt en assurances, il faut avoir plus d'options comme variables réponses. C'est ainsi que les GLMs sont préférables aux modèles linéaires. En effet, le modèle linéaire classique est un cas particulier de GLMs¹.

2.1.2 La théorie des GLMs

La théorie des GLMs dispose d'outils très importants pour l'actuariat. L'essence de la généralisation d'un modèle linéaire est faite à travers une fonction g qui décrit la relation entre la moyenne μ de la variable aléatoire Y et la combinaison linéaire des variables explicatives $\beta_0 + x_1\beta_1 + \cdots + x_p\beta_p$. Dans cette théorie, nous faisons une liaison entre la moyenne μ_i et les observations Y_i comme suit :

$$g(E(Y_i)) = g(\mu_i) = \sum_{j=0}^p \beta_j x_{ij}. \quad (2.6)$$

La structure de la moyenne n'est plus uniquement linéaire car celle-ci est une combinaison linéaire additive des variables explicatives à l'intérieur d'une fonction. Grâce à cette distinction, les GLMs deviennent un outil très puissant. Selon

1. Pour plus de détails sur la théorie linéaire classique, voir (Fox, 2008).

(DeJong et Heller, 2008), la différence la plus marquante entre le modèle linéaire classique et les GLMs se situe dans la distribution de la variable réponse. Dans le premier cas, la variable réponse suit une distribution Normale et la relation entre la moyenne et les variables explicatives est directe. Alors que, dans le deuxième cas, la variable réponse est choisie dans la famille des distributions exponentielles linéaires, et la relation entre la moyenne et les variables explicatives est une fonction. Cela signifie que la variable réponse peut être hétéroscédastique : la variance fluctue avec la moyenne.

Nous pouvons résumer certains attributs des GLMs en trois traits caractéristiques :

- la relation entre la moyenne μ_i de la variable aléatoire Y_i et les variables explicatives x_{i1}, \dots, x_{ip} est faite en fonction de $g(\mu_i) = \beta_0 + x_1\beta_1 + \dots + x_p\beta_p$,
- la variance dépend de la moyenne : $Var(Y_i) = \frac{\phi}{\omega_i}V(\mu_i)$, où ϕ est le paramètre de dispersion, ω_i est le poids et $V(\mu_i)$ est la fonction de la variance qui indique la relation entre la moyenne et la variance, et
- la distribution de la variable réponse Y peut suivre n'importe quelle distribution de la famille exponentielle.

L'équation caractéristique de cette famille se traduit de cette façon :

$$f(y) = c(y, \phi) \exp \left\{ \frac{y\theta - a(\theta)}{\frac{\phi}{\omega_i}} \right\}, \quad (2.7)$$

où θ est le paramètre canonique et ω_i représente un poids que nous fixons à 1. Toujours en rapport à la famille exponentielle, $E(Y) = a'(\theta)$ et $Var(Y) = \phi a''(\theta)^2$, où a' et a'' sont la première et la deuxième dérivées. Nous pouvons sélectionner divers types de fonctions g pour mettre en rapport la moyenne et les variables explicatives, mais chaque distribution a un lien plus naturel, appelé

2. Toutes les démonstrations et dérivations des résultats sont présentées dans (DeJong et Heller, 2008).

forme canonique. Le tableau 2.1 présente les principales distributions et leurs caractéristiques.

Tableau 2.1 Sommaire des paramètres de distributions de la famille exponentielle

Distribution	Lien canonique (θ)	$V(\mu)$	ϕ	$a(\theta)$	$E(\theta)$
$\mathcal{B}(n, \pi)$ - Binomiale	$\ln(\frac{\pi}{1-\pi})$	$n\pi(1-\pi)$	1	$n\ln(1+e^\theta)$	$n\pi$
$\mathcal{P}(\mu)$ - Poisson	$\ln(\mu)$	μ	1	e^θ	μ
$\mathcal{N}(\mu, \sigma^2)$ - Normale	μ	1	σ^2	$\frac{1}{2}\theta^2$	μ
$\mathcal{G}(\mu, \nu)$ - Gamma	$-\frac{1}{\mu}$	μ^2	$\frac{1}{\nu}$	$-\ln(-\theta)$	μ

Les paramètres β sont sélectionnés de façon à maximiser la vraisemblance de la reproduction de l'échantillon d'observations y_1, \dots, y_n . La fonction de probabilité dépend de θ et ϕ . Si tous les y_i sont indépendants, nous avons l'équation suivante comme fonction conjointe de probabilité, qui représente la vraisemblance de l'échantillon (y_1, \dots, y_n) :

$$f(y; \theta_i, \phi) = \prod_{i=1}^n f(y_i; \theta_i, \phi). \quad (2.8)$$

Puisque travailler avec une fonction multiplicative est complexe, nous utilisons le logarithme de l'équation (2.8). Les valeurs de θ et ϕ sont celles qui maximisent la fonction de vraisemblance et de logvraisemblance :

$$l(\theta_i, \phi) \equiv \sum_{i=1}^n \ln f(y_i; \theta_i, \phi). \quad (2.9)$$

L'estimation par maximum de logvraisemblance possède certaines propriétés. Les plus importantes sont les suivantes : l'estimateur $\hat{\theta}$ est non-biaisé, signifiant que $E[\hat{\theta}] = \theta$. De plus, l'estimateur est invariant et convergent, et il possède une variance minimale parmi les estimateurs sans biais.

2.1.2.1 La variable « *Offset* »

En assurances, il arrive souvent qu'un contrat cesse avant la fin prévue. La vente de la voiture ou l'annulation unilatérale du contrat en sont des exemples. Alors, lorsque nous appliquons le modèle GLM, il faut considérer ces éventualités. Il existe ce que nous appelons une variable « *Offset* » pour considérer la fin du contrat avant la date prévue. Elle prend en considération l'exposition au risque de chaque observation dans la base de données.

Si nous considérons μ comme la moyenne d'une variable de comptage y , le nombre de sinistres par année, nous avons :

$$g\left(\frac{\mu}{n}\right) = x'\beta, \quad (2.10)$$

où n représente l'exposition.

Si nous considérons la fonction g comme le logarithme, nous avons alors :

$$\begin{aligned} \ln\left(\frac{\mu}{n}\right) &= x'\beta \\ \ln \mu &= \ln n + x'\beta \\ \mu &= n \exp(x'\beta), \end{aligned} \quad (2.11)$$

où $\ln n$ est le terme d'« *Offset* », cette variable peut être vue comme un autre régresseur avec $\beta = 1$.

2.1.2.2 La distribution Poisson

Il faut à présent introduire la distribution Poisson dans les GLMs. Cette distribution est utilisée pour prendre en considération une variable de comptage avec $y \sim P(\mu)$ et $g(\mu) = x'\beta$. Sa forme canonique est logarithmique :

$$g(\mu) = \ln \mu = x'\beta, \quad (2.12)$$

où $\hat{\mu} = \exp(x'\hat{\beta}) > 0$. La distribution de Poisson est habituellement utilisée comme point de départ par un assureur pour modéliser le nombre de sinistres car il est assez simple de la manipuler et elle possède des propriétés statistiques assez directes d'interpréter et d'utiliser.

2.2 La tarification par expérience et le système Bonus-Malus (SBM)

Dans cette section, nous abordons différents aspects de la tarification par expérience et du système de bonus-malus. Nous voyons d'abord les limites du modèle de classification « *a priori* ». Ensuite, nous survolons l'histoire du SBM pour ensuite définir ses objectifs, ses types et son détail.

2.2.1 Les limitations des modèles de classification et l'historique de sinistre comme prédicteur

Il est problématique de ne considérer que les systèmes de classifications « *a priori* », car les classes ne sont pas encore complètement homogènes. Certains facteurs sont relativement faciles à observer comme l'âge du conducteur, le type de voiture, le sexe du conducteur etc. Mais, selon (Boland, 2007), d'autres facteurs sont difficiles à quantifier, comme la connaissance des conditions routières, l'observation des limites de vitesse, le bon jugement, la conduite sous l'influence de l'alcool ou de la fatigue et l'agressivité ou le stress au volant.

Selon (Denuit et Charpentier, 2005), lorsqu'il y a la présence de facteurs et/ou des variables non-observables dans le modèle de classification « *a priori* », il reste toujours une hétérogénéité résiduelle à l'intérieur de chacune des classes. Cette hétérogénéité résiduelle est un des défauts des modèles de classification.

Selon (Denuit et Charpentier, 2005) et (Boland, 2007), le nombre de sinistres réclamés par les assurés auprès de l'assureur est la variable la plus importante dans la prédiction du nombre de sinistres. Ces auteurs ont mentionné que diverses

études ont déjà démontré que le nombre de sinistres dans le passé est une variable plus importante que d'autres telles que l'âge, le sexe ou l'usage du véhicule. (Boland, 2007) a affirmé que, s'il fallait utiliser une seule variable pour tarifier, nous devrions mettre en place un système de tarification par expérience basé sur le nombre de sinistres observés.

De façon générale, les systèmes de tarification par expérience donnent des pénalités, une augmentation de la prime, à la suite d'une réclamation pour un sinistre et des rabais pour ceux qui n'en font pas. Ainsi, la valeur de la prime d'assurance est révisée annuellement par rapport au nombre de sinistres déclarés. Ces systèmes peuvent être utilisés seuls ou en complément avec d'autres systèmes de tarification par classification, ce qui réduit l'hétérogénéité résiduelle à l'intérieur des classes.

2.2.2 Historique des SBMs

La tarification par expérience est apparue afin de combler les lacunes de la tarification « *a priori* ». L'utilisation de la tarification par expérience n'est pas une nouveauté en actuariat. Son apparition est jumelée avec le développement de la Théorie de la Crédibilité, pratiquement au même moment que la formation de la « *Casualty Actuarial Society* » (CAS) en 1914.

Les systèmes bonus-malus sont apparus un peu plus tard, suite aux demandes faites par les assureurs aux actuaires de l'époque. Selon (Denuit et Charpentier, 2005), les premières études sont apparues en France au début des années 1960. Fréchet (1959), Delaporte (1959) et Thépaut (1959) ont été les premiers à présenter des travaux à ce sujet.

Selon (Denuit *et al.*, 2007) et (Boland, 2007), une des principales raisons d'adopter les SBMs, contrairement à la Théorie de la Crédibilité, est son accessibilité : le grand public comprend le fonctionnement des SBMs. De cette façon, les dé-

partements commerciaux des assureurs ont été très séduits par l'idée. Il est plus complexe d'expliquer le changement d'une prime d'assurance d'une année à l'autre en utilisant les équations de la Théorie de la Crédibilité. Même si la partie mathématique d'un SBM est aussi complexe que la mathématique de la Théorie de la Crédibilité, l'application des rabais ou surcharges sur la prime devient plus intuitive, car les règles sont définies au préalable. Ainsi, les assurés sont en mesure de mieux comprendre que les primes d'assurance varient selon l'expérience des sinistres d'une année à l'autre.

2.2.3 L'objectif des SBMs

L'objectif principal d'un SBM, comme celui des autres méthodes et des modèles de tarification d'assurance, est de déterminer la prime d'assurance. Les SBMs utilisent le nombre de sinistres passés de chaque assuré pour corriger la prime d'assurance déterminée par les modèles de tarification « *a priori* ». En ce sens, selon (Denuit et Charpentier, 2005), la fonction d'un SBM est de répartir le coût entre les bons et les mauvais conducteurs le plus équitablement possible.

Selon (Denuit et Charpentier, 2005), les SBMs ont trois buts principaux :

- pénaliser ou récompenser les assurés selon ses niveaux d'habilité et de responsabilité au volant ;
- ajuster le montant de la prime d'assurance au fil des années pour mieux représenter le risque que chaque assuré apporte au système ; et
- répondre aux demandes commerciales : avoir un processus d'ajustement annuel plus simple à comprendre. Lorsque les assurés déclarent des sinistres, ils verront une augmentation de leurs primes, tandis que ceux qui n'ont pas fait de réclamation obtiendront une réduction.

Selon (Denuit *et al.*, 2007), tous ces objectifs devraient être conciliés avec un

équilibre financier du système à long terme. L'introduction ou l'adoption d'un SBM dans un portefeuille déjà existant n'a aucun impact dans le montant de prime total collecté annuellement par la compagnie d'assurance. Si une différence est observée, nous pouvons nous attendre à avoir des erreurs dans le processus d'inclusion du SBM.

2.2.4 Les types de SBMs

D'après (Denuit et Charpentier, 2005), les deux types de SBMs les plus courants sont le système à classes (SBM traditionnel) et le système « à la française », aussi appelé par (Boland, 2007) « *No Claim Discount System* (NCD) ». Les principales différences entre le NCD et le SBM sont :

1. Dans les NCDs, les assurés joignent le système en payant une prime élevée. Au fil des années, ils reçoivent des rabais annuels s'ils ne font pas de réclamation. Dans un SBM, le niveau d'entrée dans le système est un des niveaux au milieu de l'échelle. Ainsi, l'assuré paye plus s'il fait des réclamations.
2. Dans le SBM, on retrouve plus de classes que dans le NCD.
3. Le SBM fonctionne sous le format des relativités appliquées à une prime de base, tandis que le NCD est basé sur la prime maximale, ou complète, sur laquelle des rabais prédéfinis sont appliqués.

Si nous détaillons cette dernière différence, nous retrouvons les équations simplifiées suivantes pour les SBM et pour les NCD.

$$P_l^{SBM} = r_l \times P_{entree}^{SBM} \quad (2.13)$$

$$P_j^{NCD} = (1 - D_j) \times P_{complete}^{SBM}, \quad (2.14)$$

où r_l est la relativité du niveau $l = \{1, 2, \dots, s\}$ du SBM et D_j est le rabais total de la classe $j = \{1, 2, \dots, k\}$ du NCD.

Selon (Boland, 2007), le NCD est un bon modèle pour l'assurance auto une fois qu'il introduit de façon systématique des rabais sur la prime. Ce système récompense les bons conducteurs avec des rabais considérables et décourage les assurés de réclamer de petits montants, ce qui réduit fortement les coûts administratifs des sinistres. Il faut noter, toutefois, que tous ces arguments sont aussi valables pour les SBMs.

2.2.5 Le SBM en détail

Vu que nous utilisons le SBM et non le NCD dans ce travail, nous approfondissons uniquement la partie mathématique des SBMs. En plus, nous utilisons plusieurs parties d'un des deux modèles que nous avons sélectionnés comme exemple dans les explications sur le fonctionnement d'un SBM. Ces modèles sont présentés en détail dans le chapitre suivant.

Commençons avec l'équation suivante :

$$P_i = r_l \times \lambda_i, \quad (2.15)$$

dans laquelle $P_l^{SBM} = P_i$ et $P_{entree}^{SBM} = \lambda_i$. Le l est le niveau d'entrée dans le système, il peut être fixe ou variable dans certaines conditions. r_l est la relativité assignée aux assurés selon leur expérience de sinistres de l'année précédente. Le niveau maximal s peut être fixe, par exemple $s = 10$, ou il peut être estimé avec l'ensemble de facteurs et paramètres du modèle. Alors, si nous le considérons comme fixé d'avance, nous pouvons choisir $l = (1, 2, \dots, 10)$. Il est fondamental qu'il soit fini³. Après, l'assuré voit sa prime d'assurance corrigée par la relativité à chaque année selon sa propre expérience de sinistres.

La relativité est une composante clé dans le mécanisme de fonctionnement d'un

3. Voir (Lemaire, 1995) pour une exception de SBM avec un nombre d'états non finis.

SBM. Selon (Denuit *et al.*, 2007), la difficulté réside dans l'estimation des relativités quand un système de tarification « *a priori* » est utilisé. De plus, les auteurs mentionnent que la classification « *a priori* » et les corrections « *a posteriori* » obtiennent des primes plus homogènes. Nous aurons toujours un jeu entre ce que les systèmes de classification peuvent absorber et l'hétérogénéité résiduelle que le SBM peut corriger. Ainsi, en ajoutant plus de variables significatives dans le système de classification « *a priori* », plus l'effet est faible du SBM et vice-versa.

Par ailleurs, (Pesonen, 1963) a utilisé la Théorie de la Crédibilité pour définir la distribution aléatoire des paramètres du risque aux SBMs. Selon (Norberg, 1976)⁴, après avoir défini le nombre de classes, le niveau d'entrée et les règles de transition, les relativités optimales peuvent être calculées en minimisant l'écart quadratique moyen entre l'espérance du sinistre et la prime.

2.3 La structure mathématique d'un SBM standard

Dans un univers non segmenté où le SBM est le mécanisme principal de la tarification, comme dans l'équation (2.15), le r_t sera évalué et calculé. Nous proposons donc $P_{entree}^{SBM} = P_{base}^{SBM}$ comme une valeur fixe pour tous les assurés du portefeuille. Ainsi, la prime de chaque assuré est déterminée exclusivement à partir de son expérience de sinistre et des règles de transition dans le système.

2.3.1 La structure d'un SBM

Comme les SBMs contiennent des grandes différences entre le nombre de paramètres nécessaires pour trouver le meilleur modèle selon certains types de tests, nous allons proposer une structure de modèle très semblable à la structure du modèle de contrôle présenté dans le chapitre 3. La structure est la suivante :

4. Cette approche est devenue la méthode classique de calcul des relativités par les actuaires.

- 10 niveaux avec $l = (1, 2, \dots, 10)$;
- $r_1 \leq r_2 \leq \dots \leq r_{10}$;
- la structure de déplacement est du type $-1/+X$, c'est-à-dire, l'assuré recule d'un niveau s'il ne déclare pas de sinistre ou il avance de X classes à chaque réclamation ;
- nous estimons la valeur de X à l'aide des données ;
- toutes les probabilités sont basées sur la distribution de Poisson ; et
- tous les assurés commencent à un certain niveau que nous estimons aussi à l'aide des données.

2.3.2 Le parcours des assurés dans le système

Dans la structure présentée ci-dessus, tous les assurés débutent au niveau $l = l^*$ de l'échelle. Le parcours est modélisé par les variables aléatoires (L_1, L_2, \dots) dont les valeurs varient entre $l = (1, \dots, 10)$ étant donné le niveau occupé L_k par l'assuré dans l'intervalle de temps $(k, k + 1)$. Comme l'intervalle de temps est une année, l'assuré peut uniquement changer de niveau au temps $k + 1$. Le changement de niveau est mis en place dans l'intervalle de temps $(k + 1, k + 2)$, par rapport au nombre de sinistres N_k déclarés à la compagnie d'assurance dans l'intervalle de temps $(k, k + 1)$. De plus, nous définissons (pour l'exemple) $L_0 = l_6$.

Alors, la variable L_k est définie de la manière suivante :

$$\begin{aligned}
 L_{k+1} &= \begin{cases} \max[L_k - 1, 0] & \text{si } N_k = 0 \\ \min[L_k + N_k \times \text{pen}, s] & \text{si } N_k \geq 1 \end{cases} \\
 &= \max[\min[L_k - (1 - I_k) + N_k \times \text{pen}, s], 0], \quad (2.16)
 \end{aligned}$$

où, dans notre exemple, $\text{pen} = +3$, $s = 10$ et

$$I_k = \begin{cases} 1 & \text{si } N_k \geq 1 \\ 0 & \text{sinon.} \end{cases}$$

Le parcours de chaque assuré dépend uniquement de l'état actuel du nombre de sinistres déclarés et de la règle de transition du SBM. Ce parcours est indépendant des niveaux occupés dans le passé. Cette propriété de l'indépendance conditionnelle fait partie des chaînes de Markov⁵. Les chaînes de Markov sont l'outil de base pour construire l'échelle d'un SBM. Selon (Denuit *et al.*, 2007), si le nombre de sinistres d'une année à l'autre est indépendant, le parcours de l'assuré dans l'échelle du SBM est composé par une chaîne de Markov.

2.3.3 Les règles de transition

Dans la situation illustrée, les règles de transition sont déjà définies : $-1/+X$. Nous pouvons formaliser la règle de la façon suivante :

$$t_{ij}(N_k) = \begin{cases} 1 & \text{si l'assuré passe du niveau } i \text{ au niveau } j \\ 0 & \text{sinon.} \end{cases} \quad (2.17)$$

Nous commençons toujours de l'état $i = j$. En format matriciel nous avons

$$\mathbf{T}(N_k) = \begin{pmatrix} t_{11}(N_k) & t_{12}(N_k) & \dots & t_{1s}(N_k) \\ t_{21}(N_k) & t_{22}(N_k) & \dots & t_{2s}(N_k) \\ \vdots & \vdots & \ddots & \vdots \\ t_{s1}(N_k) & t_{s2}(N_k) & \dots & t_{ss}(N_k) \end{pmatrix}.$$

Ainsi, $\mathbf{T}(N_k)$ est une matrice pour chaque $N_k \geq 0$ composée par des zéros et des uns. Chaque ligne a l'un (1) à l'endroit où il est permis de se déplacer pour chaque

5. Pour plus d'information à propos des propriétés des chaînes de Markov, voir (Ross, 2014).

$$\mathbf{T}(N_k = 2) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{T}(N_k \geq 3) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

2.3.4 Les probabilités et la matrice de transition

Dans notre démonstration, nous supposons $N_1, N_2, \dots \sim \text{Poisson}(\lambda)$. Nous incluons dans le parcours les niveaux $(L_1(\lambda), L_2(\lambda), \dots)$, où le λ est indiqué pour mettre en évidence la dépendance des sinistres. Il faut noter que L_k n'est pas dépendant de λ , mais il en est une fonction.

Soit $p_{l_1 l_2}(\lambda)$ la probabilité de se déplacer du niveau l_1 au niveau l_2 pour un assuré avec une fréquence annuelle de sinistre λ .

$$p_{l_1 l_2}(\lambda) = Pr[L_{k+1}(\lambda) = l_2 | L_k(\lambda) = l_1], \quad (2.18)$$

avec $l_1, l_2 \in (1, 2, \dots, s)$, $p_{l_1 l_2}(\lambda) \geq 0$ pour tous l_1 et l_2 et $\sum_{l_2}^s p_{l_1 l_2}(\lambda) = 1$. La matrice avec toutes les probabilités est définie comme :

$$\mathbf{P}(\lambda) = \begin{pmatrix} p_{11}(\lambda) & p_{12}(\lambda) & \dots & p_{1s}(\lambda) \\ p_{21}(\lambda) & p_{22}(\lambda) & \dots & p_{2s}(\lambda) \\ \vdots & \vdots & \ddots & \vdots \\ p_{s1}(\lambda) & p_{s2}(\lambda) & \dots & p_{ss}(\lambda) \end{pmatrix}$$

Pour notre illustration de -1/+3, contenant $p_{l_1 l_2}(\lambda) \sim Poisson(\lambda)$, nous avons :

$$\mathbf{P}(\lambda) = \begin{pmatrix} \exp(-\lambda) & 0 & 0 & \lambda \exp(-\lambda) & 0 & 0 & \frac{\lambda^2}{2} \exp(-\lambda) & 0 & 0 & 1 - \sum_1 \\ \exp(-\lambda) & 0 & 0 & 0 & \lambda \exp(-\lambda) & 0 & 0 & \frac{\lambda^2}{2} \exp(-\lambda) & 0 & 1 - \sum_2 \\ 0 & \exp(-\lambda) & 0 & 0 & 0 & \lambda \exp(-\lambda) & 0 & 0 & \frac{\lambda^2}{2} \exp(-\lambda) & 1 - \sum_3 \\ 0 & 0 & \exp(-\lambda) & 0 & 0 & 0 & \lambda \exp(-\lambda) & 0 & 0 & 1 - \sum_4 \\ 0 & 0 & 0 & \exp(-\lambda) & 0 & 0 & 0 & \lambda \exp(-\lambda) & 0 & 1 - \sum_5 \\ 0 & 0 & 0 & 0 & \exp(-\lambda) & 0 & 0 & 0 & \lambda \exp(-\lambda) & 1 - \sum_6 \\ 0 & 0 & 0 & 0 & 0 & \exp(-\lambda) & 0 & 0 & 0 & 1 - \exp(-\lambda) \\ 0 & 0 & 0 & 0 & 0 & 0 & \exp(-\lambda) & 0 & 0 & 1 - \exp(-\lambda) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(-\lambda) & 0 & 1 - \exp(-\lambda) \end{pmatrix},$$

où \sum_i égale à la somme des éléments des colonnes 1 à 9 dans la ligne i , $i = 1, 2, \dots, 6$:

$$\begin{aligned} \sum_1 &= \sum_2 = \sum_3 = 1 - \exp(-\lambda)(1 + \lambda + \frac{\lambda^2}{2}) \\ \sum_4 &= \sum_5 = \sum_6 = 1 - \exp(-\lambda)(1 + \lambda). \end{aligned}$$

Pour les lignes $i = 7, \dots, 10$, l'équation est directe et elle est présentée dans sa propre matrice.

2.3.5 Propriété stationnaire

Le comportement à long terme de la plupart des SBMs est basé sur la propriété d'ergodicité des chaînes de Markov. Une chaîne de Markov qui inclut une matrice de transition \mathbf{P} est ergodique s'il existe quelque $n_0 \geq 1$ pour lequel toutes les composantes de \mathbf{P}^{n_0} sont strictement positives. En d'autres mots, nous disons qu'une

matrice est ergodique si tous les états de la chaîne de Markov sont accessibles à partir de n'importe quel état initial dans un nombre fini d'étapes⁶. Si nous utilisons notre modèle illustratif avec $\lambda = 0.10$, nous obtenons :

$$\mathbf{P}(\lambda = 0.1) = \begin{pmatrix} 0.904837 & 0 & 0 & 0.090484 & 0 & 0 & 0.004524 & 0 & 0 & 0.000155 \\ 0.904837 & 0 & 0 & 0 & 0.090484 & 0 & 0 & 0.004524 & 0 & 0.000155 \\ 0 & 0.904837 & 0 & 0 & 0 & 0.090484 & 0 & 0 & 0.004524 & 0.000155 \\ 0 & 0 & 0.904837 & 0 & 0 & 0 & 0.090484 & 0 & 0 & 0.004679 \\ 0 & 0 & 0 & 0.904837 & 0 & 0 & 0 & 0.090484 & 0 & 0.004679 \\ 0 & 0 & 0 & 0 & 0.904837 & 0 & 0 & 0 & 0.090484 & 0.004679 \\ 0 & 0 & 0 & 0 & 0 & 0.904837 & 0 & 0 & 0 & 0.095163 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.904837 & 0 & 0 & 0.095163 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.904837 & 0 & 0.095163 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.904837 & 0.095163 \end{pmatrix}$$

et

$$\mathbf{P}^{70}(\lambda = 0.1) = \begin{pmatrix} 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \\ 0.673308 & 0.070812 & 0.078260 & 0.086491 & 0.028256 & 0.024146 & 0.018860 & 0.008828 & 0.006577 & 0.004462 \end{pmatrix}.$$

Nous pouvons voir que toutes les lignes sont égales. En plus,

$$\mathbf{P}^k(\lambda = 0.1) \simeq \mathbf{P}^{70}(\lambda = 0.1), \quad (2.19)$$

pour tous les entiers $k \geq 70$, avec une précision de six décimales. Cela veut dire que, peu importe la distribution initiale du portefeuille, nous aurons à long terme la distribution suivante :

$$\mathbf{P}^k(\lambda = 0.1) = (0.673308 \ 0.070812 \ 0.078260 \ 0.086491 \ 0.028256 \ 0.024146 \ 0.018860 \ 0.008828 \ 0.006577 \ 0.004462).$$

Nous pouvons alors la formaliser ainsi :

$$\lim_{n \rightarrow \infty} \mathbf{P}^{(n)}(\lambda) = \mathbf{\Pi}(\lambda) = \begin{pmatrix} \pi^T(\lambda) \\ \pi^T(\lambda) \\ \vdots \\ \pi^T(\lambda) \end{pmatrix}.$$

À long terme, nous utilisons annuellement la distribution stationnaire à deux fins : d'abord, nous l'employons pour planifier le paiement des sinistres et la collecte des primes. Ensuite, pour évaluer la prime de base, les relativités et l'équilibre financier du système.

6. Pour plus de détail voir (Ross, 2014).

2.3.6 Relativités

La relativité r_l s'applique aux assurés qui sont au niveau l . L'idée est de considérer les r_l comme un homologue à un facteur de risque Θ d'un assuré choisi au hasard dans le portefeuille. L'ajustement est généralement évalué par l'espérance au carré de la différence entre Θ et r_l . Ainsi, il faut minimiser $E[(\Theta - r_l)^2]$. La solution générale est donnée par l'équation :

$$r_l = \frac{\sum_k w_k \int_0^{+\infty} \theta \pi_l(\lambda_k \theta) dF_{\Theta}(\theta)}{\sum_k w_k \int_0^{+\infty} \pi_l(\lambda_k \theta) dF_{\Theta}(\theta)}. \quad (2.20)$$

Si le modèle n'a pas de classification « *a priori* », l'équation devient :

$$r_l = \frac{\int_0^{+\infty} \theta \pi_l(\bar{\lambda} \theta) dF_{\Theta}(\theta)}{\int_0^{+\infty} \pi_l(\bar{\lambda} \theta) dF_{\Theta}(\theta)} \quad (2.21)$$

et s'il existe une classification « *a priori* » avec la correction « *a posteriori* », nous obtenons⁷

$$r_l = \frac{\sum_k \lambda_k w_k \int_0^{+\infty} \pi_l(\lambda_k \theta) dF_{\Theta}(\theta)}{\sum_k w_k \int_0^{+\infty} \pi_l(\lambda_k \theta) dF_{\Theta}(\theta)}. \quad (2.22)$$

Pour plus de détails sur le développement de ces équations, voir (Denuit *et al.*, 2007) pages 184 à 215.

2.3.7 Interaction entre un système « *a priori* » (GLM) et un système « *a posteriori* » (SBM)

La problématique ci-dessus provient de l'interaction entre les deux types de systèmes de tarification. Nous remarquons que les relativités liées aux différents niveaux sont les mêmes, peu importe la classe du risque à laquelle l'assuré appartient. Ainsi, ces échelles pénalisent deux fois les assurés à haut risque : ils ont une

7. Pour plus de détails à propos de toutes les dérivations des équations, voir (Denuit *et al.*, 2007), (Denuit et Charpentier, 2005) et (Norberg, 1976).

prime plus élevée, basée sur leurs caractéristiques. De plus, il y a une forte probabilité qu'ils se déplacent vers les niveaux plus élevés, ce qui traduit une mauvaise expérience de sinistre. Pour les assurés à faible risque, la situation est l'inverse et ils sont doublement avantageés. Quand nous estimons tous les paramètres du système de classification *a priori* et du SBM ensemble, nous réduisons ce double effet généré quand nous estimons les deux parties séparément.

CHAPITRE III

LES MODÈLES DE BASE ET DE CONTRÔLE

Les données de panel, aussi appelées données longitudinales, sont constituées de l'observation d'un même assuré durant plusieurs périodes. Dans le cas de l'assurance, nous observons l'assuré sur plusieurs contrats annuels. Une dépendance entre le nombre de réclamations des contrats annuels d'un même assuré peut donc être supposée. Nous sortons ainsi de l'environnement plus classique des données transversales où les nombres de réclamations provenant de chacun des contrats d'assurance d'un seul assuré sont considérés indépendants. Malgré l'ajout d'une dépendance entre le nombre de réclamations des contrats d'un même assuré, nous supposons toujours une indépendance entre les différents assurés. Plus formellement, nous avons ainsi des vecteurs de variables aléatoires représentant toutes les réclamations d'un assuré i au cours de T_i années : $N_i = \{N_{i,1}, N_{i,2}, \dots, N_{i,T_i}\}$. Notre intérêt est donc de modéliser la distribution jointe $\Pr[N_{i,1}, \dots, N_{i,T_i}]$. Il faut aussi comprendre que les assurés ne sont pas tous observés durant le même nombre de contrats : certains assurés restent dans la même compagnie plusieurs années, alors que d'autres quittent après seulement une année. Mais, pour simplifier la notation, nous utiliserons T au lieu de T_i .

De nombreuses méthodes ont été proposées pour modéliser les données de panel de comptage. Nous pouvons nous référer à (Molenberghs et Verbeke, 2005) pour

un cadre général, ou encore à (Boucher et Guillén, 2009) pour des applications en actuariat. Toutefois, dans ce chapitre, à l'aide de la théorie vue au chapitre 2, nous proposons d'analyser en détails le modèle développé par (Boucher et Inoussa, 2014). Nous expliquerons ainsi l'idée de la méthode de tarification proposée.

3.1 Le modèle proposé

Les techniques employées par les actuaires en tarification IARD ont subi une évolution considérable au fil des ans. Un des facteurs les plus importants ayant grandement contribué à cette évolution est le développement des outils informatiques. En utilisant ces outils, les techniques actuarielles sont devenues beaucoup plus puissantes et précises.

3.1.1 Problèmes de l'utilisation de la théorie SBM pour les données de panel

En utilisant des données longitudinales, (Boucher et Inoussa, 2014) ont démontré que certains problèmes surviennent si on tente de calibrer un SBM de manière standard, c'est-à-dire en calculant les paramètres de la distribution conjointe pour ensuite calculer les valeurs des relativités d'un SBM. On aurait pu croire qu'en généralisant simplement les modèles de tarification de données transversales vus au chapitre 2, l'utilisation d'un SBM aurait pu être simple. Par exemple, en supposant que la dépendance entre les contrats d'un même assuré provient d'un effet aléatoire affectant tous les contrats, nous aurions ainsi la distribution jointe suivante :

$$\begin{aligned} \Pr[N_{i,1} = n_{i,1}, \dots, N_{i,T} = n_{i,T} | \mathbf{X}_{i,T}] &= \int \Pr[N_{i,1} = n_{i,1}, \dots, N_{i,T} = n_{i,T} | \theta_i, \mathbf{X}_{i,T}] g(\theta_i) d\theta_i \\ &= \int \left(\prod_{t=1}^T \underbrace{\Pr[N_{i,t} = n_{i,t} | \theta_i, \mathbf{X}_{i,t}]}_{\text{Dist. conditionnelle}} \right) \underbrace{g(\theta_i)}_{\text{Dist. effet aléatoire}} d\theta_i \quad (\mathcal{A}_i) \end{aligned}$$

où $\mathbf{X}_{i,T}$ est un vecteur de toutes les caractéristiques du risque de l'assuré i , du temps $t = 1$ à $t = T$. En supposant ainsi une distribution conditionnelle

Poisson($\lambda_{i,t}\theta$) avec un effet aléatoire Gamma(α, α), nous retrouvons une généralisation de la distribution Binomiale Négative de type 2. Quand nous ajoutons l'effet du temps pour considérer les données longitudinales, nous nous retrouvons avec une Binomiale Négative Multivariée (MVNB) :

$$\Pr[N_{i,1} = n_{i,1}, \dots, N_{i,T} = n_{i,T} | \mathbf{X}_{i,T}] = \left(\prod_{t=1}^T \frac{(\lambda_{i,t})^{n_{i,t}}}{n_{i,t}!} \right) \frac{\Gamma(n_{i,\bullet} + \nu)}{\Gamma(\nu)} \left(\frac{\nu}{\sum_{i=1}^T \lambda_{i,t} + \nu} \right)^\nu \left(\sum_{i=1}^T \lambda_{i,t} + \nu \right)^{-n_{i,\bullet}}, \quad (3.2)$$

avec $n_{i,\bullet} = \sum_{t=1}^T n_{i,t}$. En estimant le paramètre α de la loi gamma, les techniques classiques des SBMs pourraient, semble-t-il, être utilisées. Toutefois, on peut voir plus clairement l'erreur que cette approche générerait en analysant plus en détails l'équation de la distribution jointe de probabilité du nombre annuel de réclamations de tous les contrats de l'assuré i :

$$\Pr(N_{i,1} = n_{i,1}, \dots, N_{i,T} = n_{i,T} | \mathbf{X}_{i,T}) = \Pr(N_{i,1} | X_{i,1}) \times \Pr(N_{i,2} | n_{i,1}, X_{i,2}) \times \dots \times \Pr(N_{i,T} | n_{i,1}, \dots, n_{i,T-1}, X_{i,T}). \quad (3.3)$$

Nous pouvons voir clairement dans le développement de l'équation, que de $t = 2$ à $t = T$, une distribution prédictive est déjà supposée dans le modèle, c'est-à-dire $\Pr(N_{i,t} | n_{i,1}, \dots, n_{i,t-1} | X_{i,t})$. En conséquence, l'espérance prédictive, définie par $E[N_{i,t} | n_{i,1}, \dots, n_{i,t-1}, \mathbf{X}_{i,t}]$, correspondant finalement à une prime d'assurance utilisant les réclamations passées, est directement considérée et utilisée dans la distribution conjointe. Cela signifie qu'une tarification *a posteriori* utilisant l'expérience de sinistres est déjà supposée dans la tarification avec les données longitudinales et que les paramètres du modèles (donc les β de la classification *a priori*) sont estimés selon cette contrainte.

Nous voyons, de plus, que les SBMs et les relativités bonus-malus ne sont pas utilisés, et il n'est pas clair de savoir comment les inclure dans notre modèle sans le modifier substantiellement. Comme notre objectif est d'utiliser une prime

d'assurance prédictive incluant le SBM, il nous faut obligatoirement trouver une manière afin que les relativités SBM soient considérées et mises à l'intérieur de la structure de la distribution conjointe $(N_{i,1}, N_{i,2}, \dots, N_{i,T})$ du modèle. En d'autres mots, il nous faut trouver une manière d'inclure dans la prime de l'assuré i au temps $t + 1$, soit $P_{i,t+1} = \lambda_{i,t+1}r_t$, directement dans la distribution avant d'être utilisée pour estimer les paramètres.

3.1.2 Inclusion d'une structure bonus-malus dans la tarification

L'objectif principal de (Boucher et Inoussa, 2014) est de proposer une technique de calcul qui effectue la tarification *a priori* et *a posteriori* simultanément. Jusqu'à présent, la technique présentée dans la littérature actuarielle propose de la faire en deux étapes. La première étape est de modéliser la partie de tarification par classification (ou *a priori*) et dans la deuxième étape de travailler la partie de tarification par expérience (ou *a posteriori*) en considérant l'historique de sinistres. Généralement dans cette dernière étape c'est où est placé le facteur d'ajustement si nécessaire (*off-balance factor*).

Pour inclure les pénalités d'un SBM dans un modèle de données de panel, nous commençons par conditionner sur un niveau d'entrée l_1 . Pour des fins de simplifications, nous ignorons le $\mathbf{X}_{i,t}$ (le vecteur des caractéristiques de l'assuré i au temps t) pour présenter la distribution conjointe avec l'inclusion d'une telle structure bonus-malus. L'équation peut s'exprimer par :

$$\Pr(N_{i,1}, N_{i,2}, \dots, N_{i,T}) = \sum_{l_1=1}^s \Pr(N_{i,1}, N_{i,2}, \dots, N_{i,T} | L(1) = l_1) \Pr(L(1) = l_1), \quad (3.4)$$

où $L(t)$ est le niveau occupé par l'assuré au temps t , l_t est le niveau bonus-malus au temps t et s est le nombre de niveaux du SBM, avec $t = 1, \dots, T$. l_1 est le niveau d'entrée de l'assuré. Connaissant la structure de pénalité du SBM, le niveau

d'entrée l_1 dans le SBM et le nombre de réclamations de l'assuré à son premier contrat ($t = 1$), il devient possible de déterminer le niveau l_2 . Plus formellement, pour un assuré n'ayant que 2 contrats d'assurance, nous avons :

$$\begin{aligned}
\Pr(N_{i,1}, N_{i,2} | L(1) = l_1) &= \Pr(N_{i,1} | L(1) = l_1) \Pr(N_{i,2} | N_{i,1} = n_{i,1}, L(1) = l_1) \\
&= \Pr(N_{i,1} | L(1) = l_1) \\
&\quad \times \sum_{y=1}^s (\Pr(N_{i,2} | N_{i,1} = n_{i,1}, L(1) = l_1, L(2) = y) \\
&\quad \times \Pr(L(2) = y | N_{i,1} = n_{i,1}, L(1) = l_1)) \\
&= \Pr(N_{i,1} | L(1) = l_1) \\
&\quad \times \sum_{y=1}^s \Pr(N_{i,2} | L(2) = y) \Pr(L(2) = y | N_{i,1} = n_{i,1}, L(1) = l_1) \\
&= \Pr(N_{i,1} | L(1) = l_1) \Pr(N_{i,2} | L(2) = l_2).
\end{aligned}$$

Dans le premier passage de ligne, nous conditionnons la probabilité sur $N_{i,1} = n_{i,1}$. Dans le deuxième, nous décomposons $\Pr(N_{i,2} | N_{i,1} = n_{i,1}, L(1) = l_1)$ en considérant toutes les possibilités de niveaux en $L(2)$, c'est-à-dire tous les niveaux $y = l_2$ en sachant l_1 et $n_{i,1}$. Finalement, une seule possibilité demeure, soit que la combinaison de l_1 et $n_{i,1}$ donne un seul niveau possible $N_{i,2}$.

Ainsi, en généralisant pour plusieurs périodes à partir d'un certain niveau de bonus-malus que nous notons l_1 , il est possible de trouver les autres niveaux de bonus-malus si nous connaissons $n_{i,1}, n_{i,2}, \dots, n_{i,T}$. En d'autres mots, sachant l_t et $n_{i,t}$, nous pouvons déterminer l_{t+1} . En conséquence, on remarque que $\Pr(L(2) = y | n_{i,1}, L(1) = l_1) = 0$ pour toutes les valeurs y à l'exception de $y = l_2$. Ainsi, à l'inverse, $\Pr(L(2) = l_2 | n_{i,1}, L(1) = l_1) = 1$.

Si nous poursuivons le développement mathématique ci-dessus, nous obtenons le

résultat suivant :

$$\Pr(N_{i,1}, N_{i,2}, \dots, N_{i,T} | L(1) = l_1) = \prod_{t=1}^T \Pr(N_{i,t} | L(t) = l_t), \quad (3.5)$$

avec $L(t)$ le niveau occupé par l'assuré au temps t , avec $t = 1, \dots, T$.

Comme la structure d'un SBM est maintenant incluse dans la distribution jointe du nombre de réclamations pour l'assuré i , il est aussi possible d'inclure les relativités SBM. En supposant une distribution Poisson, nous pouvons intégrer les relativités r_l directement dans le paramètre de moyenne :

$$\Pr(N_{i,t} = n | L(t) = l, \mathbf{X}_{i,t}) = \frac{(\lambda_{i,t} r_l)^n e^{-\lambda_{i,t} r_l}}{n!}. \quad (3.6)$$

Alors, en regroupant les équations, nous obtenons :

$$\Pr(N_{i,1}, N_{i,2}, \dots, N_{i,T}) = \sum_{l_1=1}^s \prod_{t=1}^T \frac{(\lambda_{i,t} r_l)^n e^{-\lambda_{i,t} r_l}}{n!} \Pr(L(1) = l_1). \quad (3.7)$$

3.2 Les paramètres du modèle

Bien que nous avons esquissé de manière générale le modèle de tarification de données de panel avec les SBMs, il convient de voir plus en détails certains éléments de ce modèle. Dans cette section, nous détaillons les trois éléments ci-dessous que (Boucher et Inoussa, 2014) ont étudié en détails dans leur article.

1. La distribution Poisson ;
2. L'estimation du niveau d'entrée l_1 de l'assuré selon le nombre d'années de permis (NAP) ; et
3. La structure des relativités.

3.2.1 La distribution Poisson

Dans leur article, les auteurs ont se concentré en ajuster un modèle avec surdispersion. Dans ce cas, ils ont sélectionné la distribution Binomiale Négative (de type

2) et la distribution Poisson pour modéliser le nombre de réclamations. Même si la Binomiale Négative est appropriée, nous avons sélectionné la distribution Poisson comme base du travail. Cette distribution a été sélectionnée car il est simple de travailler avec elle. Elle est toujours la première distribution à être utilisée pour les données de comptage et elle possède quelques propriétés importantes qui sont utilisées dans ce travail. Notons, toutefois, qu'une généralisation serait assez directe pour utiliser une autre distribution de comptage, en modifiant l'équation (3.7) selon le besoin.

Nous nous retrouvons ainsi avec la distribution Poisson utilisant un lien logarithmique pour lier le score des régresseurs et la moyenne. Plus précisément, en incluant les relativités des SBM telles qu'exprimées dans l'équation (3.7), nous pouvons écrire :

$$\begin{aligned}\lambda_{it} &= d_{it} \exp(\mathbf{x}'_t \hat{\beta}) r_i^{(it)} \\ &= d_{it} \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1it} + \hat{\beta}_2 x_{2it} + \dots + \hat{\beta}_{11} x_{11it} + \dots) r_i^{(it)}.\end{aligned}\tag{3.8}$$

Nous avons ainsi inclus l'exposition au risque d_{it} et la relativité $r_i^{(it)}$ assignée à chacun des contrats selon l'expérience individuelle de chaque année.

3.2.2 L'estimation du niveau d'entrée l_1 de l'assuré selon le nombre d'années de permis

Pour construire le modèle, comme nous pouvons le voir avec l'équation (3.7), nous avons conditionné sur un niveau d'entrée l_1 . Pour compléter le modèle et définir correctement $\Pr(N_{i,1}, N_{i,2}, \dots, N_{i,T})$, il nous faut donc la distribution $\Pr(L(1) = l_1)$, ou, en d'autres mots, trouver le niveau d'entrée de chaque assuré. Il pourrait être simple de supposer que le SBM que nous étudions implique un niveau d'entrée unique pour tous les assurés étudiés dans la base de données. Ce ne serait toutefois pas réaliste. Le premier contrat d'un assuré dans la base de données n'est pas forcément le premier contrat d'assurance de ce conducteur.

Ainsi, plus généralement, un assuré ayant une grande expérience de conduite ne peut pas automatiquement avoir le même niveau d'entrée qu'un nouvel assuré ou qu'un assuré avec peu d'expérience. Il faut considérer cet élément dans le modèle.

Comme la majorité des assurés ont déjà un historique de conduite, capturé par la variable NAP indiquant le nombre d'années depuis l'obtention du permis de conduire, nous proposons de le prendre en considération pour bien classer ces assurés. De plus, notons qu'au Canada il est possible de consulter un historique des six dernières années pour tous les assurés. Par conséquent, un assuré changeant de compagnie d'assurance n'est pas considéré comme un nouveau conducteur et l'assureur a la possibilité de consulter les six dernières années du dossier de conduite. La base de données utilisée pour notre analyse possède cette information. Ainsi, nous sommes en mesure de retourner jusqu'à six ans avant la date d'entrée dans la base de données de la compagnie et d'utiliser cette information. Deux situations sont possibles :

1. Le premier cas, présente la situation d'un assuré ayant 6 ans ou moins d'expérience de conduite, c'est-à-dire une valeur de $NAP \leq 6$. Il nous suffit de supposer qu'il existe un niveau d'entrée précis pour tous les nouveaux conducteurs sans expérience de conduite. En d'autres termes, comme c'était le cas pour les SBM classiques (voir les chapitres précédents), notre structure de tarification par SBMs a un niveau précis d'entrée pour tous les nouveaux conducteurs. Il sera ainsi possible de déduire le niveau bonus-malus au temps $t = 1$ de tous les assurés ayant 6 ans ou moins d'expérience de conduite. En effet, à partir de ce niveau d'entrée et avec l'historique des sinistres des 6 dernières années, nous déterminons exactement le niveau bonus-malus de ce type d'assurés pour tous les contrats d'assurance observés dans la base de données.

2. L'autre cas, plus complexe, traite de tous les autres assurés, soit les assurés ayant un permis de conduire depuis plus de six ans, c'est-à-dire un $NAP > 6$. Un problème est que nous ne pouvons pas savoir à quel niveau bonus-malus était un assuré durant les années antérieures à l'historique. Pendant cette période pour laquelle on ne possède aucune donnée accessible, il est possible qu'un tel assuré ait été un mauvais conducteur ayant fait plusieurs réclamations, mais il est aussi possible qu'il n'ait jamais eu de réclamations. (Boucher et Inoussa, 2014) ont proposé une solution pour déterminer le niveau de classement probable de chaque assuré jusqu'au moment où l'historique des sinistres est disponible, ce que nous identifions comme $t = -6$, solution qu'il convient d'étudier en détails.

3.2.3 L'estimation du niveau des assurés avec expérience

La solution proposée par (Boucher et Inoussa, 2014) consiste à utiliser la distribution du niveau bonus-malus compte tenu du nombre d'années d'expérience de l'assuré depuis l'obtention du permis de conduire (NAP), soit 6 ans de moins que la valeur NAP inscrite au premier contrat de l'assuré dans la base de données. La distribution du niveau bonus-malus probable à être occupé par un assuré se fait en utilisant la matrice de transition élevée à une puissance égale au NAP de l'assuré six ans avant de rejoindre la compagnie d'assurance.

Plus précisément, la matrice de transition est la représentation de toutes les probabilités de mouvement d'un niveau à l'autre au sein d'un SBM. Lorsque nous élevons cette matrice à une puissance t , nous générons la probabilité de passer d'un niveau à l'autre après t années, en considérant tous les chemins possibles. Nous utilisons cette distribution pour évaluer la période sans aucune donnée du conducteur expérimenté jusqu'à six ans avant qu'il ne rejoigne l'entreprise, car, à partir de ce moment, nous utilisons l'historique des sinistres de l'assuré pour

identifier son niveau d'entrée.

La matrice de transition est calculée en introduisant une variante à la matrice de transition générale expliquée dans la sous-section 2.3.4. Cette transformation inclut le $\lambda_{i,t}$ qui représente la probabilité d'avoir un sinistre de l'assuré i au temps t et les r_l , $l = 1, \dots, s$ qui sont les relativités à être estimées par le SBM. En effet, les probabilités de transition sont calculées en utilisant :

$$\mathbf{P}_t(\lambda_{i,t}) = \begin{pmatrix} p_{11}(\lambda_{i,t}r_1) & p_{12}(\lambda_{i,t}r_1) & \dots & p_{1s}(\lambda_{i,t}r_1) \\ p_{21}(\lambda_{i,t}r_2) & p_{22}(\lambda_{i,t}r_2) & \dots & p_{2s}(\lambda_{i,t}r_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_{s1}(\lambda_{i,t}r_s) & p_{s2}(\lambda_{i,t}r_s) & \dots & p_{ss}(\lambda_{i,t}r_s) \end{pmatrix}. \quad (3.9)$$

Nous remarquons que les probabilités situées sur une même rangée dépendent toutes de la même valeur de relativité. Par contre, elles sont différentes d'une rangée à l'autre. Dans ce modèle, la probabilité qu'un assuré, qui est au niveau l_i , change à n'importe quel autre niveau dépend de la relativité du niveau qu'il occupe (r_i dans cet exemple). Nous comprenons que les probabilités d'avoir des sinistres sont reliées à la condition actuelle de l'assuré et que cette condition change selon plusieurs facteurs, comme l'acquisition de plus d'expérience de conduite à travers le temps. Cette proposition est différente de la proposition classique en actuariat dans laquelle il est supposé que la probabilité d'une personne d'avoir un sinistre est reliée à un effet aléatoire Θ invariant à travers le temps. Cela veut dire que cette personne est née et va mourir avec cet effet aléatoire. Les distributions classiques, comme la Binomiale Négative (du type 2) et la MVNB, utilisent cette dernière hypothèse dans leur constructions.

Avec la matrice ajustée (3.9), nous sommes en mesure d'obtenir la distribution de

$L(1)$ exprimée dans le vecteur $\mathbf{p}_{NAP_i}(\lambda_i)$:

$$\mathbf{p}_{NAP_i}(\lambda_i) = \mathbf{p}^1(\lambda_{i,t}) \left(\prod_{t=1}^{NAP_i} \mathbf{P}_t(\lambda_{i,t}) \right). \quad (3.10)$$

Nous introduisons $\mathbf{p}^1(\lambda_{i,t})$ qui est le vecteur de probabilité au temps $t = 1$ et qui est utilisé pour exprimer la condition d'entrée dans le système. Il est rempli de zéros à l'exception du niveau d'entrée où la valeur est égale à un. Plus généralement, nous pouvons le présenter pour un K quelconque comme :

$$\mathbf{p}^{(K)}(\lambda_{i,t}) = \begin{pmatrix} Pr[L(K, \lambda_{i,t}) = 1] \\ Pr[L(K, \lambda_{i,t}) = 2] \\ \vdots \\ Pr[L(K, \lambda_{i,t}) = s] \end{pmatrix}^T. \quad (3.11)$$

L'idéal serait d'obtenir toutes les informations $\mathbf{X}_{i,t}$ pour chacun des assurés i , pour $t = 1, \dots, NAP - 6$. Toutefois, ce n'est pas possible. Ainsi, nous ne savons pas où l'assuré habitait, quel type de véhicule il conduisait ou combien de kilomètres il a parcouru. Dans leur article, (Boucher et Inoussa, 2014) ont proposé une simplification dans la construction de la matrice de transition (3.9) en calculant un λ moyen pour tous les assurés. C'est à dire, $\lambda_{i,t} = \bar{\lambda}$ pour tous $i = 1, \dots, n$ et $t = 1, \dots, T_i$.

$$\mathbf{p}_{NAP}(\bar{\lambda}) = \mathbf{p}^K(\bar{\lambda}) \left(\prod_{t=1}^{NAP} \mathbf{P}_t(\bar{\lambda}) \right). \quad (3.12)$$

Toutefois, nous devons sélectionner un niveau bonus-malus pour l_1 et non une distribution de L_1 . Les auteurs ont proposé une méthode qui utilise la moyenne de la première relativité :

$$\bar{r}_{NAP} = \mathbf{p}_{NAP}(\bar{\lambda}) \mathbf{r}. \quad (3.13)$$

Le \bar{r}_{NAP} représente l'espérance de la relativité de l'assuré ayant un certain NAP et \mathbf{r} est le vecteur qui contient les relativités des s niveaux. Ensuite, pour déterminer

le niveau probable d'entrée de l'assuré qui possède plus de 6 ans d'expérience de conduite, nous appliquons la règle suivante :

$$r_j < \bar{r}_{\text{NAP}} \leq r_{j+1}. \quad (3.14)$$

Finalement, pour chaque valeur du NAP, nous sélectionnons le niveau $j+1$ comme le niveau de départ pour reconstruire l'expérience de l'assuré. Ce choix est plus conservateur que le choix du niveau j . Ce modèle était proposé dans l'article et nous l'avons retenu dans notre travail. Cependant, d'autres méthodes de calcul et de prises de décision auraient pu être choisies.

3.2.4 La structure des relativités

Au chapitre 2, nous avons vu que les SBM traditionnels donnaient la possibilité de générer une valeur de r_l pour chaque niveau du SBM. Dans le modèle que nous proposons, cela signifierait de devoir estimer s paramètres supplémentaires. En plus d'être coûteux en paramètres, cette approche fait apparaître des situations incohérentes où la relativité bonus-malus estimée pourrait être plus élevée pour des niveau SBM plus bas, par exemple des $\hat{r}_i > \hat{r}_j$ pour un $i < j$. Cette éventualité est à éviter : aucun assureur n'appliquerait un tel résultat, qui impliquerait une augmentation de primes à certains assurés n'ayant pas de réclamation et une diminution de primes à certains assurés ayant des réclamations.

Ainsi, une structure linéaire pour les relativités a été sélectionnée afin de lisser un quelconque type d'anomalie dans le cas de l'estimation statistique. De cette façon, les valeurs des relativités par rapport aux niveaux dépendront de paramètres δ et τ , et du niveau l . On aura donc des relativités toujours croissantes et paramétriques. La structure ici utilisée est la même que celle qui est proposée par (Gilde

et Sundt, 1989).

$$r_l = \begin{cases} 1 & \text{si } l = 1 \\ \delta + \tau(l - 1) & \text{sinon.} \end{cases} \quad (3.15)$$

Notons que toutes les relativités peuvent être vues comme des rapports entre le niveau occupé par l'assuré et le niveau 1, qui possède le plus gros rabais ou la relativité égale à un. Ces rapports sont des ajustements qui sont appliqués à la prime d'assurance pour la calibrer en fonction de l'historique de réclamations de l'assuré.

3.3 La contrainte légale et le modèle de contrôle

Pour la tarification d'assurances dans la province de l'Ontario au Canada, il faut distinguer deux types de sinistres : les sinistres responsables (R) et les sinistres non responsables (NR). Selon la Commission des services financiers de l'Ontario (FSCO, 2017), c'est à travers l'application de la loi sur les assurances et les règles de détermination de la responsabilité que les compagnies d'assurances déterminent la responsabilité d'un accident. La responsabilité est attribuée à chaque conducteur selon les circonstances de l'accident. L'assuré est considéré comme responsable si son degré de responsabilité est égal ou supérieur à 25%.

Prenons l'exemple d'un conducteur qui se fait emboutir à l'arrière de sa voiture. Selon la manière habituelle d'évaluer la responsabilité des conducteurs en Ontario, le conducteur ne serait pas considéré responsable. Toutefois, plusieurs scénarios sont possibles : la voiture ne roulait pas assez rapidement comparativement aux autres véhicules avec lesquels elle partageait la route, le conducteur a freiné trop brusquement, etc. Sans enlever la responsabilité à l'autre conducteur impliqué, nous pouvons croire non seulement que le conducteur de la voiture emboutie a une part de responsabilité, mais aussi que son comportement routier (rouler len-

tement, freiner brusquement, etc.) pourrait se répéter et causer de nouvelles collisions. Conséquemment, un accident considéré non responsable nous permettrait de prédire d'autres accidents non responsables ou même des accidents responsables.

Au niveau de la tarification, la province de l'Ontario interdit qu'une augmentation de prime soit associée à un sinistre non responsable. Par contre, elle n'oblige pas les assureur de réduire la prime. (Boucher et Inoussa, 2014) ont utilisé cette contrainte légale pour estimer leur modèle en accordant aux assurés ayant des sinistres non responsables le maintien du niveau occupé. Rappelons que les SBMs classiques s'expriment dans une forme $-1/+XX$, où le -1 représente la baisse d'un niveau pour chaque année sans réclamation et le $+XX$ représente le nombre de niveaux de pénalité par réclamation. Ainsi, le modèle avec une contrainte légale s'exprime maintenant selon une forme $-1/0/+XX$, où l'ajout du 0 représente l'inclusion des sinistres non responsables dans la structure du SBM du modèle de base. De plus, pour intégrer cette information dans la section de classification du modèle, nous avons défini deux jeux de paramètres : un premier pour caractériser les sinistres responsables ($\hat{\beta}^R$) et un autre pour caractériser les sinistres non responsables ($\hat{\beta}^{NR}$).

3.3.1 Le modèle de contrôle

Il est important de noter que la contrainte décrite ci-dessus est une imposition légale dans la province de l'Ontario. Nous croyons que les assurés ayant des sinistres non responsables sont, dans une certaine mesure, aussi dangereux que les assurés ayant des sinistres responsables. Dans certains cas, nous avons des assurés avec plusieurs sinistres non responsables dans leur dossier. Par exemple, un assuré n'est pas responsable, mais il génère des conditions pour un environnement favorable aux sinistres. Ainsi, nous pensons que les sinistres non responsables sont aussi prédictifs que les sinistres responsables. Comme actuaire, il faut comprendre

ce que cette contrainte implique dans l'estimation du modèle. Il est possible de déduire que les paramètres du modèle sont affectés par la contrainte, autant pour les régresseurs de la segmentation que les relativités bonus-malus.

Afin de pouvoir quantifier ces différences, il convient de comparer le modèle original à un modèle que nous qualifions de modèle de contrôle, qui ne suppose pas de contrainte légale. Ce modèle de contrôle ne distingue pas les sinistres responsables des sinistres non responsables et il nous permettra de faire toutes les comparaisons nécessaires pour quantifier les différences entre les deux. Ce faisant, nous démontrerons l'importance et l'impact de l'imposition de cette contrainte aux assureurs par le gouvernement. Ainsi, le modèle de contrôle que nous avons proposé nous permet de laisser les données nous indiquer ce qu'il en est, sans aucune contrainte.

Notons qu'une différence importante entre le modèle principal et le modèle de contrôle est que le modèle de contrôle n'utilise qu'un seul jeu de 12 paramètres dans la partie de classification pour décrire les sinistres contre deux jeux de 12 paramètres chacun pour le modèle de base (un jeu de 12 paramètres pour décrire les sinistres responsables et un autre jeu de 12 paramètres pour décrire les sinistres non responsables). Nous pensons qu'en possédant un seul jeu de paramètres, le modèle de contrôle sera moins pénalisé statistiquement en termes de nombre de paramètres à estimer, mais il devra utiliser ce seul bloc de paramètres pour décrire les deux types de réclamations. Nous avons deux possibilités : la première est que les réclamations responsables et non responsables sont vraiment différentes et le modèle de base peut donc mieux expliquer la distinction entre les sinistres avec ses deux groupes de paramètres. L'autre possibilité est qu'il n'y a pas assez de distinction entre les deux types de sinistres dans la pratique, ce qui permet à un seul ensemble d'être plus que suffisant pour décrire les réclamations sans avoir de redondance. Nous le vérifierons plus tard dans ce chapitre quand nous analyserons les β de chacun des modèles.

3.4 L'estimation de chaque modèle

Dans l'estimation de chaque modèle, nous avons testé toutes les possibilités de pénalités +XX (de 1 à 10). Au préalable, il faut toujours avoir un niveau d'entrée fixe pour tous les assurés. Pour chaque combinaison de pénalité +XX et de niveau d'entrée, nous avons obtenu un résultat de β (β_R et β_{NR} pour le modèle de base et β pour le modèle de contrôle), τ et γ (pour chaque structure de relativité de chaque modèle) qui a maximisé la logvraisemblance. Ainsi, nous sélectionnons l'ensemble avec la logvraisemblance maximale parmi toutes les possibilités testées pour chacun des modèles. Également, un des résultats de cette sélection est la structuration du niveau d'entrée estimé par chaque modèle pour les assurés avec un NAP plus grand que 6. Les graphiques 3.4 et 3.7 présentent cette règle d'entrée pour le modèle de base et celui de contrôle. Ils sont montrés et expliqués un peu plus tard dans ce chapitre.

Par ailleurs, dans le processus d'estimation, nous nous limitons à une échelle SBM de 10 niveaux, au maximum +XX = 10, et le niveau initial d'entrée est fixé de 1 à 10. En conséquence, nous aurions pu tester plus de niveaux, mais la rareté de l'information sur les assurés ayant un plus grand nombre de réclamations limite l'obtention de résultats plus stables et plus précis. De plus, (Boucher et Inoussa, 2014) ont également fixé 10 niveaux et, comme l'un de nos objectifs était de reproduire ce travail, nous supposons que le choix est adéquat.

Pour être capable d'estimer tous les paramètres des modèles, nous avons bâti des équations qui nous permettent de maximiser la logvraisemblance et qui contiennent tous les éléments spécifiques de chaque modèle. Nous commençons par l'équation de la logvraisemblance du modèle de base, qui contient la restriction légale sur les

types de sinistres, qui s'exprime comme ceci :

$$l(\hat{\beta}^R, \hat{\beta}^{NR}, \tau, \delta) = \sum_{i=1}^m \ln \left(\frac{(\lambda_{it}^R)^{n^R} e^{-(\lambda_{it}^R)}}{n^R!} \right) + \ln \left(\frac{(\lambda_{it}^{NR})^{n^{NR}} e^{-(\lambda_{it}^{NR})}}{n^{NR}!} \right), \quad (3.16)$$

où m est le nombre d'observations. Les λ_{it}^R et λ_{it}^{NR} sont structurés exactement comme dans l'équation 3.8 mais ils ne sont pas égaux. Ils possèdent une composante pour l'exposition de chaque assuré pour chaque année étudiée d_{it} , une autre composante pour prendre en considération la position dans l'échelle qui est la relativité respective de l'assuré $r_i^{(it)}$, mais ils sont séparés en responsables et non responsables. Afin de bien définir chacun des λ s, nous les présentons ci-dessous en détail :

$$\lambda_{it}^R = d_{it} \exp(\mathbf{x}'_t \hat{\beta}^R) r_i^{(it)} \quad (3.17)$$

et

$$\lambda_{it}^{NR} = d_{it} \exp(\mathbf{x}'_t \hat{\beta}^{NR}) r_i^{(it)}. \quad (3.18)$$

Les variables τ et δ estimées dans l'équation (3.16) sont à l'intérieur de la structure des relativités, $r_i^{(it)}$, comme présenté dans l'équation (3.15).

L'équation du modèle de contrôle est similaire, mais sans la restriction légale. Il n'y a qu'un seul ensemble de paramètres sinistres, puisqu'il n'y a pas de distinction entre les réclamations responsables et non responsables. Nous la présentons ci-dessous :

$$l(\hat{\beta}, \tau, \delta) = \sum_{i=1}^m \ln \left(\frac{(\lambda_{it})^n e^{-(\lambda_{it})}}{n!} \right). \quad (3.19)$$

Le λ_{it} est structuré comme dans l'équation (3.8). Il possède la composante pour l'exposition de chaque assuré, d_{it} , et la composante de position dans l'échelle du

SBM, $\tau_i^{(it)}$. Le λ du modèle de contrôle est présenté ci-dessous :

$$\lambda_{it} = d_{it} \exp(\mathbf{x}'_t \hat{\beta}) \tau_i^{(it)}. \quad (3.20)$$

Après avoir défini les équations, nous utilisons la maximisation de la logvraisemblance pour estimer les paramètres des modèles. Nous les avons listés ci-dessous :

- les β s du modèle de classification ; et
- le τ et le δ de l'équation des relativités de chaque modèle.

Chacun des modèles présente ses spécificités pour chacun des éléments énumérés ci-dessus. Le modèle de base comporte le β_R pour s'associer aux sinistres responsables et le β_{NR} pour être associé aux sinistres non responsables dans le système de segmentation.

Le format -1/0/+XX du modèle de base a été pris en compte, et seulement la pénalité +XX doit être estimée en fonction de la logvraisemblance maximale. Nous avons essayé toutes les combinaisons +XX de 1 à 10 et nous avons sélectionné celle avec la logvraisemblance maximale. Le modèle de contrôle n'a qu'un seul ensemble de bêtas pour représenter les sinistres, qu'ils soient responsables ou non responsables, dans le système de classification *a priori*. Pour estimer la pénalité dans le format -1/+XX du modèle de contrôle, nous avons fait une procédure similaire à celui du modèle de base. Les deux modèles utilisent la même équation (l'équation 3.15) pour le calcul des relativités. Cependant, chacun d'eux possède des paramètres τ et δ exclusifs pour cette équation. Ainsi, différents τ et δ ont été estimés pour chaque modèle.

3.4.1 Le parcours des assurés

Pour illustrer plus concrètement le fonctionnement des deux modèles, nous observons le parcours d'un assuré selon son expérience de sinistres. Pour ce faire, nous avons utilisé l'assuré numéro 1 de la base de données. Comme chacun des

modèles a ses propres règles, nous pourrons voir ses déplacements dans l'échelle bonus-malus.

3.4.1.1 Le modèle de base

Le tableau 3.1 présente l'historique des sinistres de cet assuré dans la perspective du modèle de base.

Tableau 3.1 Historique de sinistres de l'assuré numéro 1 - modèle de base

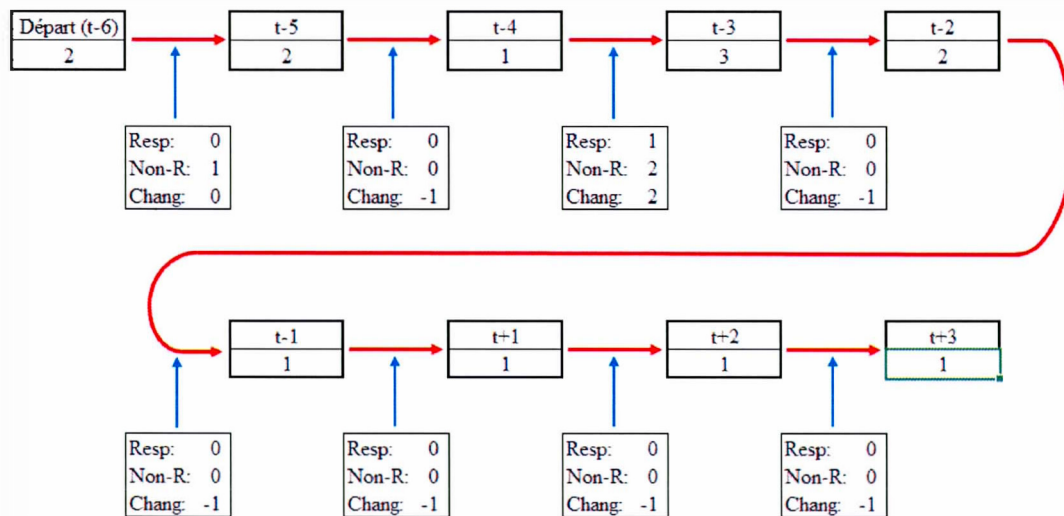
Sinistre	Passé						Présent		
	Temps \rightarrow	t=-6	t=-5	t=-4	t=-3	t=-2	t=-1	t=+1	t=+2
Responsable		0	0	1	0	0	0	0	0
Non-responsable		1	0	2	0	0	0	0	0

Comme il a son permis de conduire depuis 36 ans (NAP=36), nous en déduisons que cet assuré avait 30 ans d'expérience de conduite il y a 6 ans, au moment où nous sommes capable de vérifier son historique de sinistres. Nous lui assignons le niveau 2¹ pour la construction de son historique de bonus-malus à $t = -6$ ou lorsque l'assuré avait un NAP=30.

Le tableau 3.1² illustre l'historique des sinistres responsables et non responsables des six années avant qu'il soit un assuré de la compagnie. Nous y voyons aussi les deux années où il a été assuré.

-
1. Voir la section 3.5.1 pour l'explication au sujet de l'assignation du niveau 2 dans ce cas-ci.
 2. Nous tenons à souligner que l'année $t = 0$ n'existe pas, ce n'est qu'un instant dans le temps et non une année qui est dénombrable. Ainsi, l'assuré va de $t = -1$ à $t = +1$ directement. Par analogie, nous avons utilisé avant le Christ et après le Christ, où il n'y a pas d'année 0, mais 1 avant le Christ et 1 après le Christ.

Figure 3.1 Parcours de l'assuré numéro 1 selon son historique de sinistres et la structure du modèle de base



Dans la figure 3.1, les niveaux SBM de l'assuré sont présentés dans les carrés en gras. La trajectoire (en rouge) représente l'expérience de l'assuré au fil des ans. Le sommaire de cette expérience est présenté en dessous des flèches rouges. À l'année $t = -6$, l'assuré n'a pas eu de sinistre responsable, mais il a subi un sinistre non responsable. Cela signifie qu'il restera au même niveau bonus-malus pendant l'année $t = -5$ ³. L'année suivante, il n'a eu aucun sinistre responsable ou non responsable, et une réduction de niveau lui est garantie. Pour l'année $t = -4$, l'assuré a eu un sinistre responsable et deux sinistres non responsables. Dans ce cas-ci, le sinistre responsable le pénalise, et la règle du changement de niveau est déterminée par la multiplication du nombre de sinistres responsables par deux. Ainsi, il passe du niveau 1 au niveau 3. À l'année $t = -1$, nous voyons que, malgré

3. La structure SBM qui a maximisé la logvraisemblance c'est du format -1/0/+2. Voir la section 3.5.1 pour plus de détails sur le $+XX = +2$.

l'absence de sinistre, l'assuré ne change pas de niveau, car il a atteint le plus bas du système.

3.4.1.2 Le modèle de contrôle

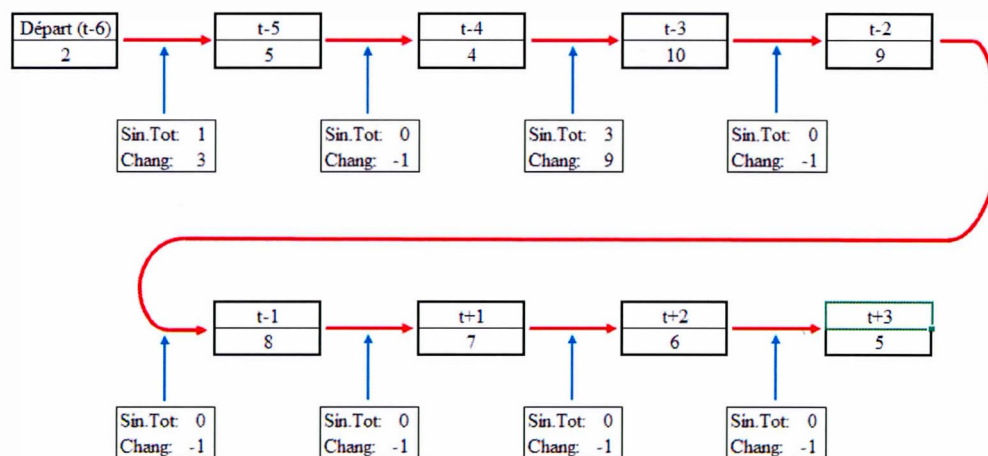
Le tableau 3.2 présente l'historique de sinistres de cet assuré dans la structure du modèle de contrôle.

Tableau 3.2 Historique de sinistres de l'assuré numéro 1 - modèle de contrôle

Temps \rightarrow	Passé						Présent	
	t=-6	t=-5	t=-4	t=-3	t=-2	t=-1	t=+1	t=+2
Nombre de sinistres	1	0	3	0	0	0	0	0

Comme dans le modèle de base, nous lui assignons le niveau 2⁴. Le tableau 3.2 illustre le même nombre d'années que le 3.1. Son parcours y est aussi.

Figure 3.2 Parcours de l'assuré numéro 1 dans le modèle de contrôle



Dans la figure 3.2, l'assuré a subi un sinistre, ce qui provoque un changement de

4. Voir la section 3.5.2 pour l'explication de cette assignation.

trois niveaux ou un changement = 3⁵. L'année suivante, il n'a eu aucun sinistre et une réduction de niveau lui est donc garanti (changement = -1). Entre les années $t = -4$ et $t = -3$, l'assuré a eu trois sinistres. Les trois sinistres le conduisent à un changement majeur de niveau. En appliquant la règle où le changement de niveau est déterminé par la multiplication du nombre de sinistres par trois, le changement de niveau est égal à 9 (sinistres \times 3 = 9). Ainsi, il passe du niveau 4 au niveau 10 (le niveau maximal). Quand l'assuré passe du temps $t = -2$ au temps $t = +2$, nous observons qu'il n'a eu aucun sinistre. Alors, il descend d'un niveau par année et arrive au niveau 5 à la fin de $t = +2$.

3.5 Applications numériques

En se référant à la base de données décrite au chapitre 1, à la théorie du chapitre 2 et aux modèles exposés au début de ce chapitre, nous présentons à présent les résultats de l'estimation du modèle de tarification par bonus-malus pour les données de panel de comptage pour chacun des deux types de modèles.

3.5.1 Le modèle de base

Le modèle de base ayant la meilleure logvraisemblance a une structure de pénalisation de format -1/0/+2. Cela signifie que, parmi tous les modèles possibles de forme -1/0/+XX, c'est ce modèle qui a généré le meilleur ajustement. Concrètement : 1) un assuré aura une pénalité de deux niveaux par réclamation responsable, 2) il descendra d'un niveau s'il n'a pas de réclamation dans l'année (responsable ou non responsable), et 3) il restera au même niveau s'il a au moins une réclamation non responsable (et aucune responsable). Dans le tableau 3.3,

5. Changement de niveau = sinistres \times 3 = 3. La structure SBM qui a maximisé la logvraisemblance est du format -1/+3. Voir la section 3.5.2 pour plus de détails sur le +XX = +3.

nous présentons l'estimation des β , les écarts-types des estimateurs, les intervalles de confiance (aussi utilisés mesurer la signifiante statistique des estimateurs à travers le Test de Wald) et la logvraisemblance de ce modèle. Le jeu de paramètres que nous avons estimé contient les β s de la distribution Poisson, un vecteur pour les sinistres responsables et un autre pour les sinistres non responsables.

Tableau 3.3 Les β s calculés pour le modèle de base

$\hat{\beta}_i$	Responsables				Non responsables			
	Est.	Écart-Type	Lim. Bas 95%	Lim. Haut 95%	Est.	Écart-Type	Lim. Bas 95%	Lim. Haut 95%
$\hat{\beta}_0$	-3.343	0.124	-3.582	-3.099	-3.352	0.109	-3.562	-3.138
$\hat{\beta}_1$	0.066	0.106	-0.138	0.274	0.207	0.104	0.007	0.411
$\hat{\beta}_2$	-0.001	0.064	-0.122	0.124	0.247	0.064	0.123	0.373
$\hat{\beta}_3$	0.193	0.059	0.079	0.309	0.237	0.052	0.138	0.338
$\hat{\beta}_4$	0.260	0.057	0.151	0.372	0.105	0.050	0.009	0.203
$\hat{\beta}_5$	0.161	0.059	0.048	0.276	0.057	0.052	-0.042	0.158
$\hat{\beta}_6$	0.184	0.047	0.094	0.276	0.029	0.042	-0.052	0.112
$\hat{\beta}_7$	-0.477	0.073	-0.617	-0.334	-0.331	0.068	-0.462	-0.198
$\hat{\beta}_8$	-0.380	0.074	-0.522	-0.234	-0.285	0.070	-0.418	-0.148
$\hat{\beta}_9$	0.125	0.057	0.015	0.237	0.098	0.051	0.001	0.198
$\hat{\beta}_{10}$	0.381	0.087	0.214	0.552	0.194	0.078	0.044	0.346
$\hat{\beta}_{11}$	-0.157	0.075	-0.302	-0.009	-0.081	0.067	-0.208	0.050
Logvraisemblance							-20,960.48759	

En utilisant le Teste de Wald (intervalle de confiance des estimateurs), nous observons que pour le côté des estimateurs des sinistres responsables l'âge de l'assuré n'a pas de relevance statistique, car dans l'intervalle de confiance des estimateurs le zéro est à l'intérieur. Dans le côté des estimateurs des sinistres non responsables la construction du véhicule entre 4 et 5 ans, la possession d'une maison par l'assuré, le fait d'être célibataires et la possession des infractions mineures n'ont pas aussi de relevance statistique.

Nous avons aussi obtenu $\hat{\tau} = 0.0836$ et un $\hat{\delta} = 1.29026$. Les estimations des relativités bonus-malus ont été calculées grâce à l'équation (3.15), et les valeurs obtenues sont indiquées au tableau 3.4.

Tableau 3.4 relativités calculées pour le modèle de base

Niveau l	relativité r_l
1	1.000
2	1.374
3	1.457
4	1.541
5	1.625
6	1.708
7	1.792
8	1.875
9	1.959
10	2.043

Par ailleurs, la figure 3.3 nous montre le résultat de l'estimation de la matrice de transition en utilisant la structure du modèle de base. Comme la probabilité de transition dépend des caractéristiques individuelles du risque et que nous n'avons pas construit une matrice pour chaque assuré, nous avons utilisé un λ moyen pour chaque type de sinistres dans la distribution Poisson pour construire une seule matrice. Nous avons calculé les estimateurs des λ s moyens pour les sinistres responsables et non responsables en utilisant la fréquence totale de la base de données. De cette façon, nous avons estimé le $\hat{\lambda}_R = 0.02873627$ pour les sinistres responsables, avec un intervalle de confiance à 95% de (0.027887395; 0.029585143), et le $\hat{\lambda}_{NR} = 0.03740854$ pour les sinistres non responsables, avec un intervalle de confiance à 95% de (0.036430; 0.038387). Nous avons utilisé la théorie de la moyenne d'échantillon comme estimateur comme présenté par (?) pages 448 à 453 pour construire les intervalles de confiance et nous voyons qu'ils sont statistiquement différentes.

Figure 3.3 Matrice de transition du modèle de base

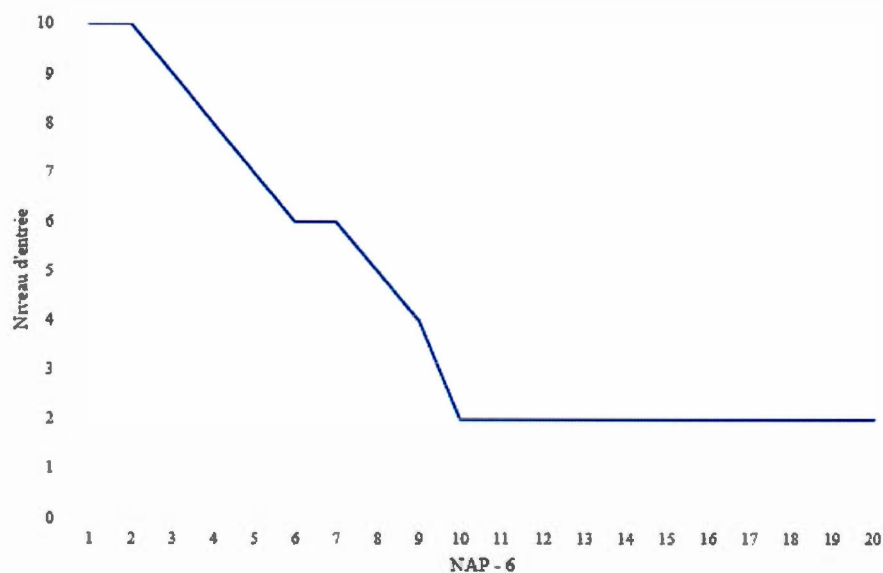
θ	1	2	3	4	5	6	7	8	9	10
1	0,971673	0	0,027922	0	0,000401	0	3,843E-06	0	2,761E-08	1,594E-10
2	0,913137	0,048155	0	0,03795	0	0,000749	0	9,858E-06	0	9,806E-08
3	0	0,908102	0,050883	0	0,040162	0	0,000841	0	0,0000117	1,24E-07
4	0	0	0,903094	0,05359	0	0,042364	0	0,000938	0	0,000014
5	0	0	0	0,898115	0,056274	0	0,0445551	0	0,00104	0,0000164
6	0	0	0	0	0,893162	0,058936	0	0,0467353	0	0,001166
7	0	0	0	0	0	0,888237	0,0615769	0	0,0489047	0,0012809
8	0	0	0	0	0	0	0,8833397	0,0641958	0	0,0524645
9	0	0	0	0	0	0	0	0,8784689	0,0667932	0,054738
10	0	0	0	0	0	0	0	0	0,873625	0,126375

Une fois que la matrice de transition a été construite, nous pouvons déterminer le niveau d'entrée des assurés qui ont plus de 6 ans de NAP. Nous utilisons la méthode développée au point 3.2.2, la matrice de transition et le NAP pour classer chaque assuré à son niveau probable d'entrée. Le graphique 3.4 résume le résultat du calcul du niveau d'entrée probable des assurés avec un NAP supérieur à six dans le modèle de base.

Pour les assurés avec un $NAP \leq 6$, nous pouvons accéder à l'historique complet des réclamations et reproduire avec exactitude le parcours de l'assuré jusqu'à la date de l'adhésion à la compagnie. Par exemple, lorsqu'un assuré potentiel appelle la société pour souscrire à une assurance, nous pouvons accéder rapidement à son historique de sinistres et déterminer le niveau exact auquel l'assuré entrerait. Cependant, dans le cas où l'assuré a un NAP supérieur à 6, nous n'avons aucun moyen de déterminer son niveau d'entrée dans la compagnie avec précision. De cette façon, nous utilisons les résultats du graphique 3.4 pour déterminer le niveau probable d'entrée de l'assuré. Par exemple, si une personne assurée potentielle avec $NAP = 10$ appelle la société pour souscrire à l'assurance, nous ne pouvons pas savoir exactement quel est son niveau d'entrée, puisqu'elle a un NAP plus grand que 6. De cette façon, nous obtenons $NAP - 6$ ($10 - 6 = 4$) et nous utilisons

la fonction décrite dans le graphique 3.4 du modèle de base pour lui assigner un niveau de départ. Nous pouvons donc déterminer le niveau probable d'entrée 6 ans avant que l'assuré rejoigne la compagnie et, à partir de ce point, utiliser l'historique des sinistres disponible pour compléter le parcours et déterminer le niveau d'entrée de l'assuré dans la compagnie. Notez que ce niveau a été établi sur la base d'une estimation et d'informations précises.

Figure 3.4 Niveau d'entrée estimé des assurés pour le modèle de base



Tout d'abord, nous constatons que le niveau d'entrée pour les personnes assurées avec un NAP supérieur à 6 n'est pas linéaire. Deux éléments soutiennent cette constatation. D'une part, la méthode choisie pour déterminer le niveau d'entrée utilise une règle par rapport à la relativité moyenne selon le NAP-6 et les relativités estimées pour le modèle.

De plus, nous vérifions que les deux premières années après NAP-6 sont considérées comme similaires en termes de risque, puisqu'elles sont classées au niveau 10. Nous avons observé une diminution du niveau d'entrée du NAP-6 de 2 à 6. Les niveaux

6 et 7 sont également classés et nous voyons une réduction jusqu'au NAP - 6 = 10. À partir de NAP - 6 supérieur à 10, le niveau d'entrée assigné aux assurés est le 2.

3.5.2 Le modèle de contrôle

Le modèle de contrôle ayant la meilleure logvraisemblance a une structure de pénalisation de format -1/+3. Tout comme pour le modèle de base, nous avons ajusté plusieurs dizaines de modèles, et c'est le modèle avec une pénalité de +3 qui générerait la meilleure logvraisemblance. Plus concrètement, cela signifie qu'un assuré aura une pénalité de trois niveaux par réclamation, peu importe s'il est responsable ou non de l'accident, et qu'il descendra d'un niveau s'il n'a pas de réclamation dans l'année. Dans le tableau 3.5 nous avons indiqué l'estimation des β s, les écarts-types et la logvraisemblance de ce modèle. Nous déjà remarquons que la seule variable qui n'a pas de signifiante statistique c'est la possession d'infractions mineures.

Maintenant que nous avons estimé les β des deux modèles, nous pouvons comparer comment le modèle de classification *a priori* pondère les caractéristiques de l'assuré et de son véhicule. Pour procéder à cette comparaison, il faut introduire le profil de base. Cet assuré est celui où tous les $X_i = 0$ à l'exception de $X_0 = 1$. Cela veut dire que le $\lambda_{i,t} = \exp(\beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \dots)$ devient $\lambda_{i,t} = \exp(\beta_0 X_0)$ pour un cas générique. Généralement, cet assuré de base représente le profil avec la plus grosse exposition parmi tous les profils de la base de données. Si nous regardons le tableau 1.2, nous pouvons structurer ce profil en déterminant le complément des informations qui y sont fournies. Ainsi, le tableau 3.6 présente les caractéristiques de l'assuré avec le profil de base. Il est à noter que ce profil de base est le même pour les deux modèles, car ils utilisent tous deux la même base de données et les mêmes X_i .

Tableau 3.5 Les β s estimés pour le modèle de contrôle

$\hat{\beta}_i$	Estimations	Écart-Type	Lim. Bas 95%	Lim. Haut 95%
$\hat{\beta}_0$	-2.765	0.086	-2.929	-2.597
$\hat{\beta}_1$	0.198	0.082	0.040	0.360
$\hat{\beta}_2$	0.117	0.049	0.023	0.212
$\hat{\beta}_3$	0.214	0.042	0.133	0.296
$\hat{\beta}_4$	0.170	0.040	0.093	0.248
$\hat{\beta}_5$	0.101	0.042	0.021	0.182
$\hat{\beta}_6$	0.105	0.033	0.041	0.171
$\hat{\beta}_7$	-0.356	0.053	-0.457	-0.253
$\hat{\beta}_8$	-0.306	0.053	-0.408	-0.201
$\hat{\beta}_9$	0.096	0.041	0.018	0.176
$\hat{\beta}_{10}$	0.270	0.061	0.153	0.388
$\hat{\beta}_{11}$	-0.086	0.057	-0.195	0.026
Logvraisemblance			-17,934.16189	

Tableau 3.6 Description du profil de l'assuré de base pour les deux modèles

Description du profil
Âgé 61 ans ou plus.
Véhicule est âgé de 6 ans ou plus.
L'assuré n'a pas de maison.
Il y a 3 conducteurs ou plus.
L'assuré est marié.
Il a reçu des infractions mineures.

Nous commençons par incorporer ces informations dans chacun des modèles et d'extraire la prime de base de chacun. Les équations liées à ce calcul sont similaires mais ne sont pas égales. Pour le modèle de base, la prime est calculée sous forme agrégée puisqu'elle comporte deux composantes : $\text{Prime} = \lambda^R + \lambda^{NR} = \exp(\beta_0^R) + \exp(\beta_0^{NR})$. Pour le modèle de contrôle, le calcul est direct : $\text{Prime} = \exp(\beta_0)$. Rappelons qu'à ce stade du travail, nous nous intéressons uniquement à la comparaison entre les modèles de classification *a priori* et nous ne prenons pas en compte les relativités des SBM de chacun. Le tableau 3.7 montre les détails du calcul.

Dans la comparaison du prix de base de chaque modèle, nous constatons :

- la prime de l'assuré de base du modèle de base est de 0.07034, et un intervalle de confiance à 95% de cette prime est (0.05622, 0.08844) ;
- la prime de l'assuré de base du modèle de contrôle est de 0.06299, et un intervalle de confiance à 95% de cette prime est (0.05344, 0.07450) ;
- la prime de base du modèle de contrôle est inférieure à celle du modèle de base, mais elles ne sont pas statistiquement différentes (à 95%) ; et
- la prime de base du modèle de base est composée de près de 50% (0,0353) du montant provenant de la structure des sinistres responsables et de 50% (0,0350) de la structure des sinistres non responsables.

Cette comparaison montre que, même si les pénalités sont plus sévères, le modèle de contrôle a une prime de base inférieure à celle du modèle de base, ce qui suscite une plus grande attractivité de la part des assurés à faible risque.

Pour compléter l'analyse de la partie de classifications de chacun des modèles, nous évaluons comment les régresseurs se comportent, en se rappelant qu'il y a plus de régresseurs que de variables. Il faut noter que le modèle de base est composé de deux β (β^R pour les sinistres responsables et β^{NR} pour les sinistres

Tableau 3.7 Calcul de la prime de base de l'assuré avec le profil de base pour chaque modèle

Modèle	Description	Formule	Valeur
Base	β_0^R	[1]	-3.34291
	Écart-type du β_0^R	[2]	0.12440
	Limite bas à 95% du β_0^R	[3]	-3.58162
	Limite haut à 95% du β_0^R	[4]	-3.09908
	Prime ^R	[5] =exp([1])	0.03533
	Limite bas à 95% de la Prime ^R	[6] =exp([3])	0.02783
	Limite haut à 95% de la Prime ^R	[7] =exp([4])	0.04509
	β_0^{NR}	[8]	-3.35229
	Écart-type du β_0^{NR}	[9]	0.10909
	Limite bas à 95% du β_0^{NR}	[10]	-3.56161
	Limite haut à 95% du β_0^{NR}	[11]	-3.13848
	Prime ^{NR}	[12] =exp([8])	0.03500
	Limite bas à 95% de la Prime ^{NR}	[13] =exp([10])	0.02839
	Limite haut à 95% de la Prime ^{NR}	[14] =exp([11])	0.04335
	Prime ^{Total}	[15] = [5] + [12]	0.07034
	Limite bas à 95% de la Prime ^{Total}	[16] = [6] + [13]	0.05622
	Limite haut à 95% de la Prime ^{Total}	[17] = [7] + [14]	0.08844
Contrôle	β_0	[18]	-2.76482
	Écart-type du β_0	[19]	0.08565
	Limite bas à 95% du β_0	[20]	-2.92918
	Limite haut à 95% du β_0	[21]	-2.59694
	Prime	[22] =exp([18])	0.06299
	Limite bas à 95% de la prime	[23] =exp([20])	0.05344
	Limite haut à 95% de la prime	[24] =exp([21])	0.07450

non responsables). Il sera donc nécessaire d'agrèger les résultats des β^R et β^{NR} pour les comparer aux β du modèle de contrôle. Lorsque chacun des modèles est utilisé en tarification, nous pouvons calculer le pourcentage de changement de prime selon chaque caractéristique (ou variable X_1, X_2, \dots, X_{10}) par rapport à l'assuré de base. Ainsi, le tableau 3.8 montre les β des deux modèles et l'effet de chaque régresseur.

Tableau 3.8 Comparaison des β des modèles

Variable X_j	Valeur	Description	modèle de base			modèle de contrôle			
			β^R	% chgmt (R)	β^{NR}	% chgmt (NR)	% chgmt (Total)	β_j	% chgmt in premium
X1	1	Si l'assuré est âgé entre 16 et 25 ans.	0.0662	7%	0.2071	23%	30%	0.1983	22%
X2	1	Si l'assuré est âgé entre 26 et 60 ans.	-0.0004	0%	0.2467	28%	28%	0.1168	12%
X3	1	Si le véhicule est neuf.	0.1929	21%	0.2370	27%	-8%	0.2139	24%
X4	1	Si le véhicule est âgé entre 1 et 3 ans.	0.2600	30%	0.1051	11%	41%	0.1698	19%
X5	1	Si le véhicule est âgé entre 4 et 5 ans.	0.1612	17%	0.0569	6%	23%	0.1007	11%
X6	1	Si l'assuré a une maison.	0.1842	20%	0.0294	3%	23%	0.1056	11%
X7	1	S'il y a un seul conducteur.	-0.1771	-38%	-0.3312	-28%	-66%	-0.3562	-30%
X8	1	S'il y a deux conducteurs.	-0.3797	-32%	-0.2847	-25%	-56%	-0.3059	-26%
X9	1	Si l'assuré est célibataire.	0.1246	13%	0.0985	10%	24%	0.0961	10%
X10	1	Si l'assuré est divorcé.	0.3812	46%	0.1937	21%	68%	0.2696	31%
X11	1	S'il n'a pas d'infractions mineures.	-0.1572	-15%	-0.0807	-8%	-22%	-0.0857	-8%

Dans la structure même du modèle de base, nous voyons déjà comment les deux types de sinistres ajustent différemment la prime de base. L'âge de l'assuré est le meilleur exemple. Dans le cas des réclamations responsables, l'âge de l'assuré ne semble pas avoir d'importance parce que les surcharges appliquées sont minimes ou nulles. Cependant, en ce qui concerne les réclamations non responsables, l'âge a un poids important dans l'ajustement de la prime de base, en appliquant des surcharges de 23% ou 28% selon le groupe d'âge. Cette différence de traitement dans le modèle de base apparaît également quand on prend en compte l'âge du véhicule, la possession d'une propriété, l'état civil de l'assuré et les infractions mineures.

En comparant les facteurs de pourcentage agrégés du modèle de base avec les facteurs respectifs du modèle de contrôle, nous remarquons que le modèle de base utilise des remises et des surcharges beaucoup plus larges par rapport à celles du

modèle de contrôle, ce qui génère des primes extrêmes. Par exemple, dans le modèle de contrôle, l'escompte maximum est de 30% et la surcharge maximale s'élève à 31%, tandis que, dans le modèle de base, nous avons une réduction maximale de 66% et une surcharge maximale de 68%.

Nous nous sommes questionnés à propos du besoin d'avoir un ou deux jeux de paramètres dans la sous-section 3.3.1. Cependant, comme notre but n'est pas d'évaluer en détails la partie de classification *a priori* du modèle, nous ne signalons que des observations à ce sujet. À nouveau, nous regardons les colonnes des facteurs de chaque β du tableau 3.8. Si les chiffres du modèle de base sont proches et sont similaires au chiffre du modèle de contrôle, nous pouvons dire qu'un seul jeu de paramètre est suffisant pour décrire la caractéristique. De cette façon, nous pouvons voir que quatre X_i pourraient être décrits par un seul jeu, soit : si le véhicule est neuf, si le véhicule est âgé entre 4 et 5 ans, s'il existe deux conducteurs et si l'assuré est célibataire. Pour tous les autres cas, deux jeux de paramètres semblent être nécessaires.

À noter que cette comparaison des surcharges ou rabais par variable est fait pour connaître l'impact réel de l'application concrète des modèles en tarification. L'objectif n'est pas de mesurer la différence statistique entre les estimateurs. Cette approche est plus axé pour résoudre les problèmes de décision qui les actuaires ont dans les compagnies d'assurance pour choisir des variables et des modèles à utiliser pour construire leurs tarifs.

Enfin, nous arrivons aux relativités du modèle de contrôle. Le tableau 3.9 présente les résultats de l'estimation, en utilisant les estimés de $\hat{\delta} = 1.13735056$ et $\hat{\tau} = 0.10002584$.

Les relativités des niveaux 2 à 10 du modèle de contrôle sont 0.10 plus élevées à chacun de ces niveaux. Du niveau 1 au niveau 2, cette distance est légèrement

Tableau 3.9 relativités calculées pour le modèle de contrôle

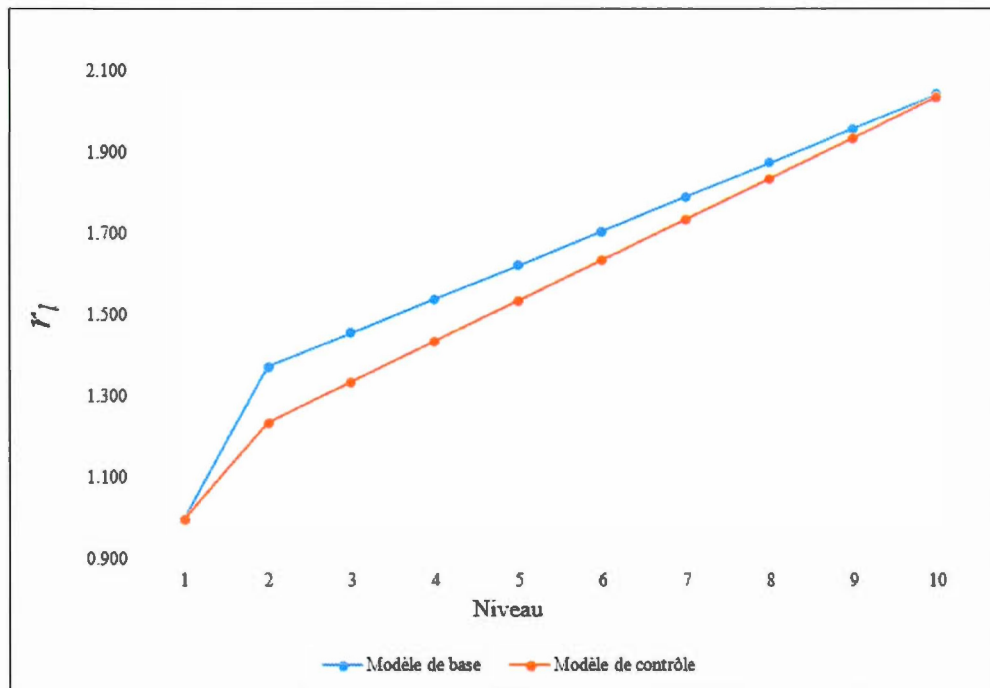
Niveau l	relativité r_l
1	1.000
2	1.237
3	1.337
4	1.437
5	1.537
6	1.637
7	1.738
8	1.838
9	1.938
10	2.038

supérieure, elle atteint près de 24%. Si nous comparons les relativités des deux modèles, le modèle de base a des relativités égales ou supérieures au modèle de contrôle à tous les niveaux. Les niveaux extrêmes sont égaux ou extrêmement proches. Cela se produit, car le modèle de base est moins sévère et doit donc trouver des moyens de compenser ce manque de sévérité de $-1/0/+2$ par rapport à $-1/+3$ du modèle de contrôle.

La figure 3.6 montre le résultat de l'estimation de la matrice de transition en utilisant la structure du modèle de contrôle. Ainsi, comme pour le modèle de base, nous avons calculé un λ moyen comme estimateur du paramètre de la distribution Poisson. La valeur de cet estimateur est $\hat{\lambda} = 0.0661448105$.

Nous avons utilisé la méthode développée au point 3.2.2, la matrice de transition du modèle de contrôle et le NAP pour classer le niveau d'entrée probable de chaque assuré. Le graphique 3.7 montre le résultat.

Figure 3.5 Graphique comparant les relativités des deux modèles



Nous pouvons voir des similarités dans le comportement de la fonction du niveau d'entrée des modèles. Par contre, lors de la comparaison des graphiques de niveau d'entrée du modèle de base et du modèle de contrôle dans le graphique 3.8, les NAP - 6 de 4 à 14 sont différents. Alors que dans le modèle de base, la courbe atteint le niveau 2 avec un NAP - 6 = 10, dans le modèle de contrôle, la courbe atteint le niveau 2 uniquement dans NAP - 6 = 14. De plus, le modèle de contrôle étant plus sévère en général, il est plus dur de déterminer les niveaux d'entrée entre les NAP-6 de 4 à 14. Par exemple, une personne avec NAP-6 = 8 sera classée avec un niveau d'entrée 5 dans le modèle de base et 6 dans le modèle de contrôle. Ainsi, le modèle de contrôle indiquera toujours un niveau égal ou supérieur au modèle de base comme niveau d'entrée de l'assuré avec NAP supérieur à 6. Il en résulte que le modèle de contrôle aura moins d'assurés aux niveaux inférieurs, dans le départ, que le modèle de base et, conséquemment, cela contribue à l'offre de meilleures

Figure 3.6 Matrice de transition du modèle de contrôle

θ	1	2	3	4	5	6	7	8	9	10
1	0.9359953	0	0	0.0619112	0	0	0.0020476	0	0	0.0000459
2	0.9214138	0	0	0	0.0754141	0	0	0.0030862	0	0.0000859
3	0	0.9153377	0	0	0	0.0809728	0	0	0.0035815	0.000108
4	0	0	0.9093016	0	0	0	0.086455	0	0	0.0042434
5	0	0	0	0.9033054	0	0	0	0.0918613	0	0.0048333
6	0	0	0	0	0.8973487	0	0	0	0.0971925	0.0054588
7	0	0	0	0	0	0.8914312	0	0	0	0.1085688
8	0	0	0	0	0	0	0.8855528	0	0	0.1144472
9	0	0	0	0	0	0	0	0.8797132	0	0.1202868
10	0	0	0	0	0	0	0	0	0.873912	0.126088

primes selon le type de risque que l'assuré apporte à la compagnie.

Figure 3.7 Niveau d'entrée estimé pour le modèle de contrôle

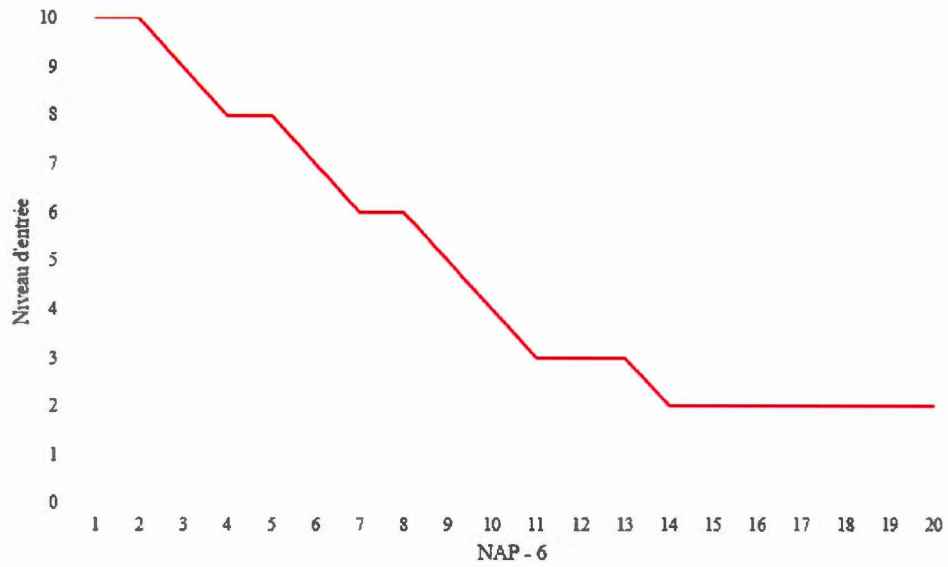
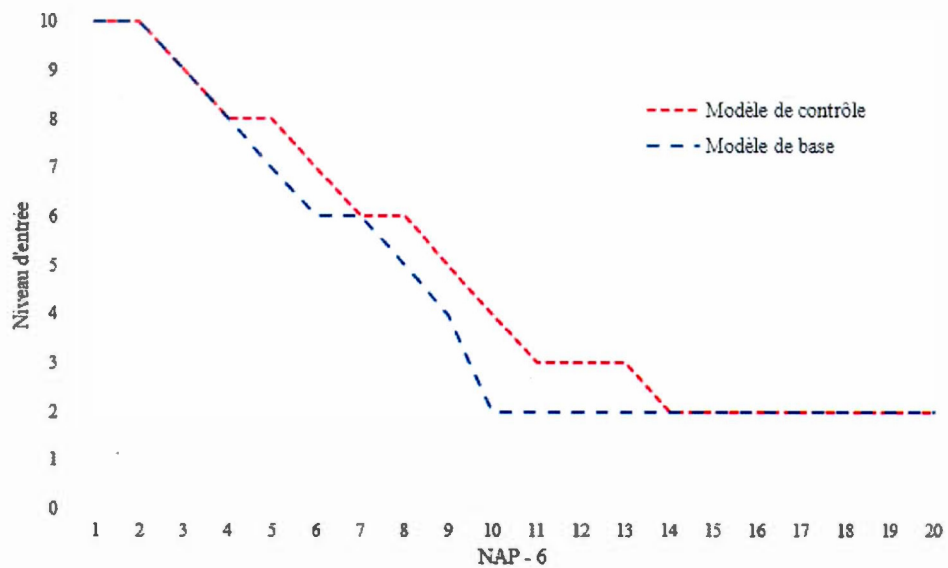


Figure 3.8 Comparaison du niveau d'entrée estimé pour les deux modèles



CHAPITRE IV

ANALYSE DE LA CAPACITÉ PRÉDICTIVE ENTRE LES DEUX MODÈLES

Dans ce chapitre du mémoire, nous comparons les modèles de base et de contrôle. Dans le monde des assurances, comme nous le mentionnons au chapitre précédent, il existe une divergence d'idées autour du sujet des accidents du type non responsables. Même si les assurés ne sont pas légalement coupables de la cause de ces accidents, certains assureurs croient qu'ils ont une partie de responsabilité dans ces événements, ne serait-ce qu'un degré de participation à la cause de l'accident.

L'analyse proposée a comme but comprendre l'influence des accidents non responsables dans la tarification en assurance automobile. Les évaluations et les comparaisons des deux modèles sont effectuées sous deux axes. Le premier axe d'analyse utilise certains outils statistiques, telles que la logvraisemblance et le critère d'information d'Akaike, et certaines unités de mesure proposées par l'ouvrage de (Lemaire, 1995) sur les bonus-malus. Le deuxième angle d'analyse est basé sur l'équilibre actuariel entre primes et sinistres pour chacun des modèles, à savoir ce qui se passerait si le modèle de contrôle et le modèle de base étaient en compétition dans un marché fermé. Nous comparons des profils d'assurés et nous identifions le modèle qui surpasse l'autre dans la course à la clientèle. Nous supposons un environnement fermé, c'est-à-dire qu'il n'y a pas de clients provenant de l'extérieur. Seuls les assurés présents dans la base de données au moment de

l'analyse sont considérés.

4.1 Outils de comparaison classiques

4.1.1 Ajustement statistique

La logvraisemblance est l'un des outils les plus utilisés pour évaluer l'ajustement des modèles. (Boucher et Inoussa, 2014) l'ont d'ailleurs utilisée comme instrument de comparaison entre les modèles qu'ils ont développés. Au chapitre précédent, nous y avons fait recours pour déterminer le type de paramétrisation offrant le meilleur ajustement statistique pour chaque modèle. En effet, pour obtenir la valeur des paramètres des modèles, nous avons essayé toutes les combinaisons possibles des niveaux de pénalité par sinistre déclaré, et tous les niveaux d'entrée possibles. Le modèle choisi demeure celui qui possède la combinaison qui génère la meilleure logvraisemblance.

Toutefois, cette statistique d'ajustement n'est pas comparable d'un modèle à l'autre. Le modèle de base modélise les deux types de sinistres séparément. Cela veut dire que ce modèle calcule la logvraisemblance pour la somme des deux modèles de probabilités, un pour chaque type de sinistre. Si nous utilisons une notation simplifiée, nous avons :

$$l(\beta^B) = \sum \ln(\Pr(N^R) + \Pr(N^{NR})). \quad (4.1)$$

Le modèle de contrôle, par contre, modélise la probabilité des sinistres agrégées de l'assuré sans aucun type de distinction. En notation simplifiée, nous donetons :

$$l(\beta^C) = \sum \ln(\Pr(S)), \quad (4.2)$$

avec $S = N^R + N^{NR}$. Pour que les deux soient comparables, le modèle de contrôle devrait modéliser la probabilité de toutes les combinaisons possibles qui donnent

S , ce qu'il ne fait pas. Il faut donc d'autres outils pour évaluer et comparer ces deux modèles. Ces outils sont présentés dans ce qui suit.

4.1.2 Le coefficient de variation

Lorsque la tarification en assurance ne considère pas les caractéristiques de chaque assuré, nous sommes dans un environnement parfaitement solidaire. Dans cette situation, il n'est pas possible de faire une distinction dans les primes d'assurance, ce qui signifie que la prime est la même pour tous les assurés. Alors, il n'y a aucune variabilité des primes d'assurances. À l'inverse, quand de nombreuses caractéristiques du risque sont considérées dans la tarification et que la prime est proportionnelle au risque assuré, le niveau de solidarité de la prime est réduit. La segmentation de la prime, à l'aide des caractéristiques des assurés, génère ainsi une variabilité de la prime d'assurance chargée aux assurés.

Selon (Lemaire, 1995), la meilleure manière de capturer et d'évaluer la solidarité, et donc la variabilité de la prime, est d'utiliser le coefficient de variation (CV) des primes chargées. L'équation (4.3) est ainsi utilisée pour calculer le coefficient de variation par année pour chacun des modèles, alors que l'équation (4.8) est utilisée pour calculer le coefficient de variation globale. Le coefficient de variation s'exprime ainsi :

$$CV_{prime}^t = \frac{\sigma_{prime}^t}{\bar{P}^t}, \quad (4.3)$$

avec

$$\bar{P}^t = \frac{\sum_{i=1}^{n_t} P_i^t}{\sum_{i=1}^{n_t} E_i^t}. \quad (4.4)$$

Les primes individuelles par année sont calculées comme suit :

$$P_i^t = \lambda_{it} \cdot r_{it}^t \quad (4.5)$$

$$= \exp(\mathbf{X}_{it}'\beta) \cdot r_{it}^t \quad (4.6)$$

et

$$\sigma_{prime}^t = \sqrt{\frac{\sum_{i=1}^{n_t} (P_i^t - \bar{P}^t)^2}{\sum_{i=1}^{n_t} E_i^t}}, \quad (4.7)$$

où \bar{P}^t représente la prime moyenne pour l'année t , σ_{prime}^t correspond à l'écart-type des primes pour l'année t , P_i^t représente la prime de l'assuré i pour l'année t et E_i^t est l'exposition de l'assuré i pour l'année t . Nous obtenons ainsi :

$$CV_{prime}^T = \frac{\sigma_{prime}^T}{\bar{P}^T} \quad (4.8)$$

avec

$$\bar{P}^T = \frac{\sum_{t=1}^6 \sum_{i=1}^{n_t} P_i^t}{\sum_{t=1}^6 \sum_{i=1}^{n_t} E_i^t} \quad (4.9)$$

et

$$\sigma_{prime}^T = \sqrt{\frac{\sum_{t=1}^6 \sum_{i=1}^{n_t} (P_i^t - \bar{P}^T)^2}{\sum_{t=1}^6 \sum_{i=1}^{n_t} E_i^t}} \quad (4.10)$$

où \bar{P}^T est la prime moyenne totale et σ_{prime}^T est l'écart-type total.

Le CV est une statistique connue, simple et facile à interpréter parce que, entre autres, elle n'a pas d'unité de mesure. Ces avantages facilitent son utilisation et son implémentation. Si nous nous fions à (Lemaire, 1995), nous concluons que le modèle le plus pénalisant est celui qui possède le CV le plus élevé.

Tableau 4.1 Les coefficients de variation globaux des deux modèles

modèle	CV Global
Base	0,4509332
Contrôle	0,4687446

Le tableau 4.1 présente les coefficients de variation globaux des deux modèles à l'étude. Nous observons que le modèle de contrôle est plus sévère que le modèle de base, car la valeur de son coefficient de variation global est plus élevée.

(Lemaire, 1995) a calculé et classé le CV pour 30 différents SBMs de divers pays dans le Tableau 5.1 de son ouvrage, illustré dans le tableau 4.2. Nous avons placé les nôtres dans ce tableau. Toutefois, avant de comparer les résultats, il est important de mettre en évidence quelques éléments :

1. Les modèles utilisés dans (Lemaire, 1995) sont évidemment différents des nôtres, (Lemaire, 1995) ayant utilisé la calibration classique des SBMs (voir section 3.2.7).
2. Parmi les différences entre les SBMs de (Lemaire, 1995) et les nôtres est le fait que l'auteur a utilisé une prime de base unique pour tous les assurés du modèle, tandis que nous avons segmenté la prime de chaque assuré selon ses caractéristiques.
3. (Lemaire, 1995) ne s'est pas intéressé à trouver la meilleure structure de pénalité pour les SBMs de chaque pays, tandis que nous l'avons fait. Nous remarquons qu'il est possible qu'une structure de pénalité différente et optimale ait pu être proposée par des analyses statistiques en fonction d'une base de données représentant l'historique de sinistre automobile d'un pays. Nous pouvons consulter la description précise des SBMs de chacun des pays dans l'annexe B aux pages 133 à 152 de (Lemaire, 1995).

Comme nous pouvons voir dans le tableau 4.2, nos modèles sont situés dans le haut du classement. Cela signifie que nos modèles permettent d'offrir des primes fortement différentes entre les assurés.

Il faut noter que dans les modèles que nous avons développés, une partie de la différenciation des primes provient des relativités bonus-malus, mais une autre partie provient de la segmentation *a priori*. De cette façon, pour pouvoir comparer de manière cohérente les résultats, nous avons calculé de nouveau le CV des modèles en éliminant la segmentation *a priori*. Il faut cependant indiquer que nous n'avons pas estimé de nouveau tous les paramètres des modèles : nous avons utilisé toute

Tableau 4.2 Les coefficients de variation globaux des modèles

Rang	Pays (ou système)	Le coefficient de variation
1	modèle de contrôle	0,4687
2	Suisse (nouveau)	0,4595
3	modèle de base	0,4509
4	Norvège (vieux)	0,3900
5	Kenya	0,3835
6	Finlande (nouveau)	0,3834
7	Suède	0,3769
8	Pays Bas	0,3523
9	Japon (nouveau)	0,3283
10	Taiwan	0,3162
11	Malaisie	0,3075
12	Denemark	0,3017
13	Suisse (vieux)	0,2700
14	Finlande (vieux)	0,2570
15	Allemagne (nouveaux)	0,2536
16	Hong Kong	0,2518
17	Royaume-Uni (sans protection)	0,2419
18	Luxembourg (nouveau)	0,2147
19	Belgique (nouveau)	0,2128
20	France	0,2049
21	Norvège (nouveau)	0,2049
22	Portugal	0,1956
23	Thaïlande	0,1925
24	Espagne	0,1533
25	Corée du Sud	0,1271
26	Japon (vieux)	0,1261
27	Royaume-Uni (protégé)	0,1260
28	Luxembourg (vieux)	0,1075
29	Italie (nouveau)	0,0934
30	Belgique (vieux)	0,0586
31	Brésil	0,0304
32	Italie (vieux)	0,0046

la structure déjà développée, mais nous avons enlevé l'impact des caractéristiques du risque. Le tableau 4.3 nous présente les résultats de ce nouveau calcul.

Après la correction du CV, il est possible de voir dans le tableau 4.3 que les coefficients de variation des modèles sont réduits. Nous nous y attendions. Ce qui est le plus important est la démonstration de la classification des modèles. Nos modèles demeurent dans la partie supérieur du tableau. Cela signifie que les modèles à l'étude sont capables de bien différencier la prime d'assurance.

Dans le cas sans segmentation *a priori*, le nouveau SBM de la Suisse est le plus sévère et il différencie le plus les primes. Or, si nous ajoutons la segmentation *a priori* pour différencier davantage les primes de nos modèles, comme nous pouvons le voir dans le tableau 4.2, le modèle de contrôle est plus sévère que celui de la Suisse. Notons que le modèle de base, même avec l'ajout d'une segmentation *a priori*, n'est pas encore aussi sévère que celui de la Suisse. Selon cet outil, le modèle de base différencie davantage la prime d'assurance comparativement aux autres pays, mais il est inférieur que le modèle de contrôle.

Dans le tableau 4.4, nous avons apparié les coefficients de variation de nos modèles avec l'inclusion d'une segmentation *a priori* et sans celle-ci. Une différence d'environ 11% des CVs est due à cette segmentation.

Le résultat le plus important de cette analyse est que les modèles proposés sont encore classés dans la partie la plus haute des tableaux. Cela nous assure qu'ils distinguent davantage les primes d'assurance. Dans le tableau 4.4, il est évident que le modèle de contrôle performe mieux que le modèle de base. Toutes les statistiques sont favorables et démontrent que le modèle de base est pénalisé par l'inclusion de la contrainte légale. Commençons maintenant à évaluer le CV par modèle. Dans le tableau 4.5, nous présentons les CVs du modèle de base.

Tableau 4.3 Les coefficients de variation globaux des modèles, sans le modèle de classification *a priori*

Rang	Pays (ou système)	Le coefficient de variation
1	Suisse (nouveau)	0,4595
2	modèle de contrôle	0,4221
3	modèle de base	0,4070
4	Norvège (vieux)	0,3900
5	Kenya	0,3835
6	Finlande (nouveau)	0,3834
7	Suède	0,3769
8	Pays Bas	0,3523
9	Japon (nouveau)	0,3283
10	Taïwan	0,3162
11	Malaisie	0,3075
12	Danemark	0,3017
13	Suisse (vieux)	0,2700
14	Finlande (vieux)	0,2570
15	Allemagne (nouveaux)	0,2536
16	Hong Kong	0,2518
17	Royaume-Uni (sans protection)	0,2419
18	Luxembourg (nouveau)	0,2147
19	Belgique (nouveau)	0,2128
20	France	0,2049
21	Norvège (nouveau)	0,2049
22	Portugal	0,1956
23	Thaïlande	0,1925
24	Espagne	0,1533
25	Corée du Sud	0,1271
26	Japon (vieux)	0,1261
27	Royaume-Uni (protégé)	0,1260
28	Luxembourg (vieux)	0,1075
29	Italie (nouveau)	0,0934
30	Belgique (vieux)	0,0586
31	Brésil	0,0304
32	Italie (vieux)	0,0046

Tableau 4.4 Les coefficients de variation globaux des modèles avec et sans la segmentation

Le coefficient de variation	CV (sans <i>a priori</i>)	CV (sans <i>a priori</i>)	Gain
modèle de contrôle	0,4221	0,4687	11,1%
modèle de base	0,4070	0,4509	10,8%

Tableau 4.5 Les coefficients de variation du modèle de base par année

Statistique	année					
	1	2	3	4	5	6
Prime moyenne	0,0687	0,0659	0,0633	0,0615	0,0599	0,0605
L'écart-type	0,0325	0,0248	0,0236	0,0231	0,0504	0,0527
Coefficient de variation	0,4730	0,3768	0,3738	0,3759	0,8416	0,8705

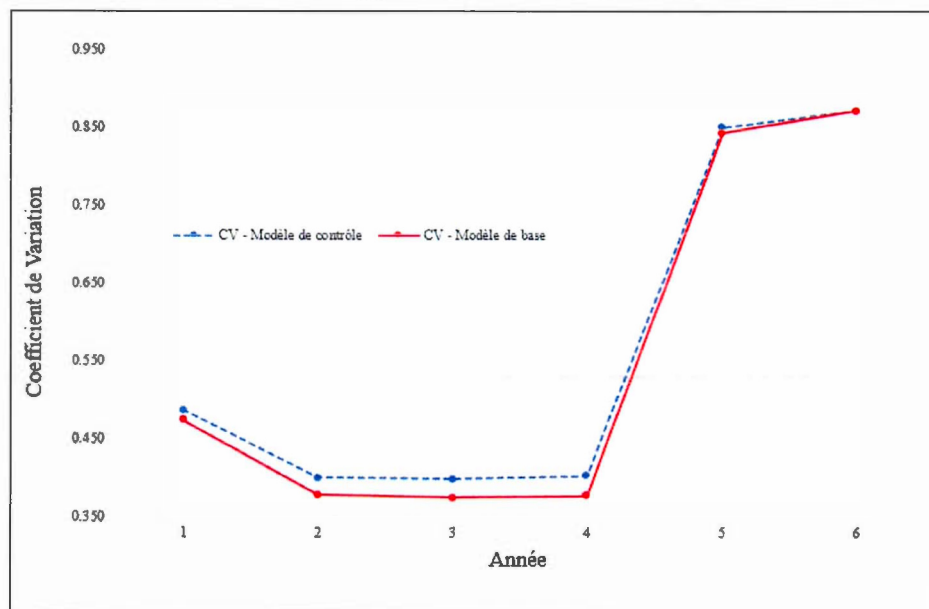
Au préalable, il faut noter que les deux dernières années d'exposition sont fortement réduites et cela génère des résultats hors des attentes. Dans le tableau 4.5, nous pouvons observer que les statistiques CV diminuent durant les quatre premières années, années durant lesquelles les bases de données ont une exposition crédible. Cette réduction illustre la tendance qu'ont les primes de se rapprocher et de réduire l'écart entre elles. Quand le temps passe, les assurés changent de niveaux dans le système et la majorité migrent aux positions ayant des primes moins chères. Cela garantit un rapprochement des valeurs des primes au fil des années.

Le tableau 4.6 nous présente les CVs du modèle de contrôle. Nous pouvons voir que les CVs du modèle de contrôle se comportent de façon similaire à ceux qui ont été observés dans le modèle de base. Dans les deux cas, nous constatons que les assurés vont aux niveaux inférieurs du système au fil du temps. Cette concentration dans les niveaux plus bas du système réduit l'écart de la valeur des primes. Pour mieux

Tableau 4.6 Les coefficients de variation du modèle de contrôle par année

Statistique	année					
	1	2	3	4	5	6
Prime moyenne	0,0688	0,0661	0,0628	0,0609	0,0591	0,0593
L'écart-type	0,0334	0,0264	0,0249	0,0245	0,0502	0,0516
Coefficient de variation	0,4856	0,3988	0,3970	0,4014	0,8489	0,8699

analyser toutes les informations à propos des coefficients de variation des modèles, nous avons élaboré le graphique 4.1.

Figure 4.1 Comparaison entre les coefficients de variation des deux modèles

Si nous utilisons le CV comme outil de base pour choisir le meilleur modèle entre les deux, nous pouvons nous baser sur ce que (Lemaire, 1995) nous indique : celui qui a le CV le plus élevé est le modèle qui distingue davantage les primes, ce qui le rend le meilleur modèle. Avec une meilleure différenciation, il est possible de

proposer des primes plus adaptées au risque de chaque assuré, de garantir plus de compétitivité des tarifs et d'attirer de meilleurs assurés en terme de risques pour le portefeuille. À long terme, cela augmenterait les chances de solvabilité de la compagnie d'assurances. Suite à cette constatation, le modèle de contrôle est meilleur que le modèle de base.

4.1.3 Le niveau relatif moyen stationnaire

Selon les mots de (Lemaire, 1995), le niveau relatif moyen stationnaire (NRMS) mesure la position du conducteur moyen, une fois que le SBM a atteint un état stable, sans définir ce qui est un état stable. Nous considérons alors que cet état stable signifie que le portefeuille est assez grand, inclut des assurés de toute sorte d'expérience de conduite (diverses valeurs de NAPs), et que les assurés sont distribués dans tous les niveaux du modèle. Nous retrouvons cette stabilité dans notre étude.

Une statistique NRMS plus élevée suggère une meilleure répartition des assurés dans les niveaux du SBM. En fait, plus ce nombre est faible, plus la concentration d'assurés dans les classes à rabais est élevée. Plus ce nombre est élevé, meilleur est le modèle, puisqu'il est capable de mieux répartir les assurés entre les niveaux et permet une plus grande variation des primes d'assurance. (Lemaire, 1995) a expliqué que cette statistique est comparable à des SBMs qui ont différents types de structures, différents nombres de niveaux et différentes structures de pénalisation. Le NRMS est calculé selon l'équation suivante :

$$NRMS = \frac{\text{Niveau Moyen Stationnaire} - \text{Niveau Minimal}}{\text{Niveau Maximal} - \text{Niveau Minimal}} \quad (4.11)$$

Le niveau moyen stationnaire (NMS) dépend uniquement de la matrice de transition du modèle et du nombre de niveaux.

$$NMS = \pi \mathbf{s} \quad (4.12)$$

où π est la ligne de la matrice stationnaire Π , basée sur les matrices de transitions \mathbf{T} et \mathbf{T}^c , dans les figures 3.3 et 3.6. Ces deux matrices sont utilisées pour estimer les vraies matrices de transition des deux modèles, car elles sont contruites d'une façon aussi approximative et avec les données de $t - 6$ jusqu'à $t + 6$, et elles sont utilisées pour parcourir ce qui se passe avec les assurés avant $t - 6$. Le vecteur-colonne s est une séquence numérique de 1 à 10 : $s = (1, 2, \dots, 10)$.

Le NRMS est toujours compris entre 0 et 1. Si la valeur est proche de zéro, il existe une très grosse concentration des assurés dans les classes de primes moins chère. À l'inverse, si la valeur est près de 1, il existe une concentration des assurés proche du niveau de prime le plus chère du système. Une valeur de NRMS plus au centre peut indiquer une meilleure distribution des assurés dans les niveaux du SBM. (Lemaire, 1995) a remarqué que pour tous les systèmes, il existe une tendance : il y a une réduction des primes au cours des années et une forte concentration des assurés se retrouve dans les classes les moins chères. Il explique que cette situation génère des indices généraux à la baisse, c'est-à-dire, la classe moyenne que les assurés occupaient est proche du niveau le plus bas du système.

Dans le tableau 4.7, nous présentons les NRMS du modèle de base et de contrôle. La méthode n'utilise que trois éléments pour faire son calcul : le $\bar{\lambda}$, la matrice de transition et la structure du SBM. Aucune relation avec les données n'est nécessaire pour obtenir le résultat.

Tableau 4.7 Comparaison entre le NRMS des deux modèles

modèle	NRMS
Base	1,22%
Contrôle	7,80%

Le tableau 4.7 précise que le modèle de contrôle distribue mieux les assurés dans les

niveaux du SBM et permet une plus grande diversification des primes d'assurance, ce qui est préférable. Plus précisément, cette différence majeure entre les deux modèles est due à la structure de pénalisation de chacun. Le modèle de base distingue les deux types de sinistres et il bouge l'assuré fautif seulement s'il est responsable de l'accident. Si l'assuré est considéré comme non responsable de l'accident, il ne change pas de niveau. De cette façon, le modèle de base garde une partie des assurés qui participent d'accidents dans les niveaux plus bas du système, tandis que le modèle de contrôle les envoie tous aux niveaux plus élevés du SBM. Cela se reflète dans les matrices de transitions T et T^c , qui sont une des variables de la statistique NRMS. Ainsi, cette statistique démontre que le modèle de contrôle est meilleur que le modèle de base par rapport à la distribution des assurés dans les niveaux du système. À titre de comparaison, le tableau 4.8 présente l'insertion des statistiques NRMS de nos modèles dans celles calculées par (Lemaire, 1995).

Nous vérifions à travers les résultats du tableau 4.8 que nos deux modèles ont des statistiques NRMS insérées avec celles des modèles testés par (Lemaire, 1995). Comme le calcul ne dépend que de trois éléments, la façon dont ils interagissent entre eux fait une différence significative. Le SBM le plus sévère est celui du Kenya où le conducteur est envoyé au niveau le plus élevé du système s'il fait une réclamation.

4.2 Le ratio sinistre / prime

Quand les actuaires font la tarification en assurance non-vie, leur premier objectif est d'avoir un modèle qui puisse récolter le montant nécessaire pour payer les sinistres chaque année, ce que nous appelons un équilibre actuariel. Cette analyse se fait selon un scénario libre de compétition. L'évaluation de ce rapport est appelée : ratio sinistre / prime.

Plus en détail, nous avons calculé la prime pour chaque assuré par modèle et par

Tableau 4.8 Comparaison entre les NRMS des nos modèles et les NRMS calculés par (Lemaire, 1995)

Rang	Pays (ou système)	NRMS
1	Kenya	28,79%
2	Espagne	25,67%
3	Malaisie	21,17%
4	Finlande (nouveau)	16,04%
5	Suède	14,20%
6	Pays Bas	11,78%
7	Royaume-Uni (protégé)	11,37%
8	Taiwan	9,55%
9	Finlande (vieux)	8,46%
10	Hong Kong	8,35%
11	Thaïlande	8,03%
12	Modèle de Contrôle	7,80%
13	Royaume-Uni (sans protection)	7,07%
14	Portugal	6,75%
15	Norvège (vieux)	6,61%
16	Suisse (nouveau)	6,47%
17	Allemagne (nouveaux)	5,85%
18	Japon (nouveau)	4,63%
19	Belgique (nouveau)	4,05%
20	Denemark	3,78%
21	Suisse (vieux)	2,90%
22	France	2,12%
23	Norvège (nouveau)	2,11%
24	Brésil	1,85%
25	Corée du Sud	1,37%
26	Luxembourg (nouveau)	1,36%
27	Italie (nouveau)	1,30%
28	Modèle de Base	1,22%
29	Luxembourg (vieux)	1,01%
30	Japon (vieux)	0,88%
31	Belgique (vieux)	0,74%
32	Italie (vieux)	0,01%

année. Nous avons évalué ces primes et nous les avons comparées avec les sinistres. Le Tableau 4.9 présente le sommaire de ces résultats et il inclut le ratio sinistre / prime pour chaque cas.

Tableau 4.9 Sommaire de primes et sinistres avec le ratio sinistre / prime par modèle

	année						Global
	1	2	3	4	5	6	
Sinistres totaux (\$)	1 752	1 263	830	556	224	2	4 627
Prime - Base (\$)	1 616,2	1 276,7	881,2	595,7	220,1	1,8	4 591,9
résultat - Base (\$)	-135,8	13,7	51,2	39,7	-3,9	-0,2	-35,1
Ratio sinistre / prime - Base (%)	108,4%	98,9%	94,2%	93,3%	101,8%	108,3%	100,8%
Prime - contrôle (\$)	1 616,8	1 279,3	875,0	589,3	216,6	1,8	4 578,8
résultat - contrôle (\$)	-135,2	16,3	45,0	33,3	-7,4	-0,2	-48,2
Ratio sinistre / prime - Contrôle (%)	108,4%	98,7%	94,9%	94,3%	103,4%	108,5%	101,1%

En observant les résultats du tableau 4.9, nous avons vérifié que les deux modèles se sont comportés comme prévu. Il existe des variations positives et négatives de ce rapport par année. Ces oscillations sont toujours attendues. Celles-ci sont observées dans le marché d'assurances. Ainsi, sans être en compétition, les modèles sont capables d'amasser suffisamment d'argent pour payer les réclamations. Le ratio sinistre / prime nous aide à comprendre qu'avant d'entrer en compétition, les modèles sont autonomes. Autrement dit, chacun peut survivre financièrement s'il n'y a pas de concurrence. Dans la section suivante, nous examinerons comment les modèles de base et de contrôle se comportent lorsqu'ils sont mis en compétition pour attirer les assurés. Lequel des deux survit ? Quel type d'assuré est attiré par chaque modèle ? Comment se comportent les primes et les sinistres à travers le temps entre les deux modèles ?

4.3 La compétition entre modèles

4.3.1 Mise en contexte

Nous pensons qu'il existe une autre manière d'évaluer et de comparer l'efficacité entre les deux modèles de tarification : à travers la simulation d'une compétition de marché où chaque modèle représente une compagnie d'assurances. Nous croyons que cet approche test aussi la capacité prédictive des modèles face à un environnement non parfait, c'est-à-dire face à un environnement non contrôlé et absent de compétiteurs. À notre connaissance, ce type d'évaluation des tarifications n'a jamais été utilisée ni étudiée dans la littérature actuarielle.

Comme nous l'avons remarqué jusqu'à présent, les deux modèles sont en mesure de déterminer les primes d'assurance suffisantes pour faire face au paiement des sinistres. Chaque modèle se comporte différemment en raison de ses spécificités et de ses contraintes. Dans un scénario de concurrence, ce qui est le plus important pour une compagnie d'assurance est de pouvoir attirer les personnes qui offrent moins de risques d'accident, de réduire ou d'éliminer l'antisélection et de contracter le montant le plus approprié de prime d'assurances par rapport au type de risque. D'une part, dans un modèle avec moins de restrictions, comme dans le cas du modèle de contrôle, il est possible d'avoir une prime d'assurance plus personnalisée et plus adaptée au type de risque. D'autre part, lorsqu'il existe une restriction légale, comme dans le cas du modèle de base, la valeur de la prime d'assurance de chaque personne assurée est moins précise par rapport au risque. Une certaine forme de compensation est nécessaire pour que le modèle puisse ramasser le montant nécessaire afin d'honorer ses engagements. La façon dont le modèle de base se comporte consiste à rééquilibrer les primes d'assurances en facturant un peu plus les assurés moins risqués pour compenser la restriction liée à l'impossibilité d'augmenter les primes des assurés qui ont des réclamations non responsables.

Jusqu'à présent, le modèle de contrôle est plus sévère par rapport à la structure des pénalisations que le modèle de base sur deux aspects :

1. Le modèle de contrôle a la possibilité de pénaliser les assurés qui ont des accidents non responsables. À l'inverse, le modèle de base, dû à une contrainte légale, ne peut pas utiliser les réclamations non responsables pour augmenter la prime d'assurance ;
2. Pour un accident responsable, le modèle de contrôle déplace l'assuré de trois niveaux (+3) sur l'échelle SBM, alors que le modèle de base ne suppose qu'un mouvement de deux niveaux (+2).

Deux autres éléments font varier la valeur des primes d'assurance : le profil de l'assuré à travers la structure de tarification *a priori* et la relativité de chacun des modèles. Les deux se complètent et travaillent ensemble pour déterminer la meilleure prime d'assurance pour un assuré. La valeur totale correspond approximativement à la même quantité d'argent récolté par le modèle de contrôle. Comme nous avons vu dans la section précédente, les deux modèles sont suffisamment capables d'être en équilibre actuariel dans un scénario sans compétition.

Pour mettre en place le système de compétition, nous avons émis une hypothèse de base : les assurés achètent toujours le produit d'assurance chez la compagnie d'assurance qui offre la prime la plus basse selon leurs caractéristiques du risque et leur niveau bonus-malus. Nous supposons aussi que les couvertures d'assurances, le marketing ou le niveau des services offerts par les deux compagnies, sont les mêmes. Du moins, ils ne sont pas considérés par les assurés lorsqu'ils choisissent leur compagnie. Chaque année, pour chaque assuré ayant un profil d'assurances précis et un historique de sinistre connu, nous étudierons le mouvement des assurés en vérifiant quel modèle offre la prime la plus basse. Nous ne considérerons que ces deux compagnies. Le marché est ainsi fermé, car il ne comprend que deux

assureurs sans la possibilité qu'un autre assureur intègre ce marché.

4.3.2 Résultats généraux

Nous commençons notre analyse par les résultats globaux de ce modèle de compétition. Le tableau 4.10 présente la synthèse de la comparaison des deux modèles.

En analysant les résultats du tableau 4.10, nous constatons que la société 1 n'est pas en mesure de payer la totalité des réclamations avec les primes d'assurance qu'elle recueille auprès de ses assurés. À l'inverse, la société 2 est en mesure d'honorer ses engagements avec un certain profit. Ce fait se produit dans toutes les années analysées. Le marché est déficitaire et cela est dû aux résultats de chaque compagnie. Il appartient à l'entreprise d'adapter ses primes afin de retrouver l'équilibre et d'obtenir un profit. Dans une situation réelle, après un an avec un ratio sinistre / prime de 316%, la société 1 aurait deux options : soit modifier radicalement ses tarifs et être supportée financièrement par ses partenaires, soit déclarer la faillite et quitter le marché. Dans le premier cas, le résultat d'une reprise financière est incertain. La société peut se rétablir ou avoir encore un déficit à la deuxième année. Dans le deuxième cas, comme le marché est fermé, les assurés qui se trouvaient dans la société 1 migreraient à la société 2 sans avoir le choix. La compétition disparue, cette société pourrait modifier ses tarifs sans avoir de contrainte. En quelque sorte, la compagnie 2 évaluerait son équilibre financier sur le long terme et le marché passerait du déficit au profit.

Si par hasard la concurrence dure toutes les six années, nous voyons clairement que la société 1 a été déficitaire avec un ratio sinistre / prime 322% et la société 2 a réalisé un gros bénéfice de 54% avec un ratio de sinistre / prime de 46%. Le marché affichait un déficit total de 9% (ratio sinistre / prime de 109%). Ce déficit était mesuré en fonction de la combinaison des activités des deux sociétés.

Tableau 4.10 Sommaire de la compétition entre les deux modèles/compagnies

année	modèle	Nombre d'assurés		Prime	Sinistre	résultat	Rapport	Condition
		Nombre	(%)	(\$)	(\$)	(\$)	Sinistre/Prime (%)	
[1]	[2]	[3]	[4]	[5]	[6]	[7] - [5] - [6]	[1] = [6] / [5]	[9]
année 1	Base	12 586	22%	714,0	2 254,0	-1 540,0	316%	Déficit
	contrôle	44 451	78%	2 287,9	1 183,0	1 104,9	52%	Profit
	marché	57 037		3 001,9	3 437,0	-435,1	114%	Déficit
année 2	Base	9 881	23%	604,4	1 739,0	-1 134,6	288%	Déficit
	contrôle	32 746	77%	1 772,9	799,0	973,9	45%	Profit
	marché	42 627		2 377,3	2 538,0	-160,7	107%	Déficit
année 3	Base	5 710	18%	335,9	1 124,0	-788,1	335%	Déficit
	contrôle	25 365	82%	1 298,1	558,0	740,1	43%	Profit
	marché	31 075		1 634,0	1 682,0	-48,0	103%	Déficit
année 4	Base	3 725	17%	209,2	794,0	-584,8	380%	Déficit
	contrôle	18 330	83%	903,7	368,0	535,7	41%	Profit
	marché	22 055		1 112,9	1 162,0	-49,1	104%	Déficit
année 5	Base	2 288	15%	70,9	312,0	-241,1	440%	Déficit
	contrôle	12 638	85%	339,7	133,0	206,7	39%	Profit
	marché	14 926		410,6	445,0	-34,4	108%	Déficit
année 6	Base	21	15%	0,7	3,0	-2,3	408%	Déficit
	contrôle	118	85%	3,1	0,0	3,1	0%	Profit
	marché	139		3,8	3,0	0,8	78%	Profit
Total	Base			1 935,2	6 226,0	-4 290,8	322%	Déficit
	contrôle			6 605,4	3 041,0	3 564,4	46%	Profit
	marché			8 540,5	9 267,0	-726,5	109%	Déficit

Nous concluons que, dans un contexte de concurrence, que si l'entreprise 1 utilise une contrainte légale dans sa tarification, elle se place dans une situation de désavantage concurrentiel, car, pour incorporer cette contrainte, elle doit changer la relation entre la valeur de la prime qu'elle charge et le risque qu'elle acquit pour chaque individu. Il en résulte des prix qui ne correspondent pas au vrai risque assumé par l'assureur. Toutefois, une compagnie ne va pas décider de inclure une telle contrainte dans son système de tarification sans une raison. Nous établissons un parallèle entre cette contrainte légale et les primes qui pardonnent. Ils travaillent de la même manière et, dans ce cas, l'entreprise inclut ce dernier dans son système de tarification pour pardonner et retenir les assurés qui ont des réclamations.

Pour comprendre un peu plus comment la concurrence a entraîné une défaillance du marché, la faillite de l'entreprise 1 et la survie de l'entreprise 2, nous proposons d'analyser l'évolution du déplacement des assurés dans le marché. Cela est fait en vérifiant les chiffres absolus, le comptage du nombre d'assurés pour chaque type de mouvement, et les chiffres relatifs, les indices en pourcentage. Le tableau 4.11 présente les résultats.

Tout d'abord, le modèle de contrôle possède environ 78% des assurés au début de la compétition. Cela signifie généralement que la prime d'assurance offerte par le modèle de contrôle à chaque assuré est plus avantageuse que celle du modèle de base pour la première année analysée. Ensuite, il y a plus de titulaires de polices d'assurances de la compagnie 1 qui passent à la compagnie 2 que l'inverse pendant toute la période. Dans notre modèle de compétition, la règle est claire, l'assuré choisit l'entreprise avec la prime d'assurance la plus basse. S'il y a un changement plus significatif des assurés qui passent du modèle de base au modèle de contrôle, c'est parce que le prix du modèle de contrôle est plus bas. Puis, le taux de renouvellement des contrats d'assurance est plus grand dans le modèle de

Tableau 4.11 évaluation du déplacement des assurés

	Formule	année 1	année 2	année 3	année 4	année 5	année 6
Nombre total d'assurés [1]		57 037	42 627	31 075	22 055	14 926	139
Nombre d'assurés dans le modèle de base [2]		12 586	9 881	5 710	3 725	2 288	21
Nombre d'assurés dans le modèle de contrôle [3]		44 451	32 746	25 365	18 330	12 638	118
Déplacement base pour contrôle [4]		1 833	2 572	1 023	631	1	
Déplacement contrôle pour base [5]		2 460	1 197	857	481	5	
Sortie du scénario à partir du modèle de base [6]		3 332	2 796	1 819	1 287	2 271	
Sortie du scénario à partir du modèle de contrôle [7]		11 078	8 756	7 201	5 842	12 516	
Reste dans le modèle de base [8]		7 421	4 513	2 868	1 807	16	
Reste dans le modèle de contrôle [9]		30 913	22 793	17 307	12 007	117	
Taux de changement base pour contrôle [10]	= [4] / [2]	14,6%	26,0%	17,9%	16,9%	0,0%	
Taux de changement contrôle pour base [11]	= [5] / [3]	5,5%	3,7%	3,4%	2,6%	0,0%	
Taux de sortie à partir du modèle de base [12]	= [6] / [2]	26,5%	28,3%	31,9%	34,6%	99,3%	
Taux de sortie à partir du modèle de contrôle [13]	= [7] / [3]	24,9%	26,7%	28,4%	31,9%	99,0%	
Taux de renouvellement à partir du modèle de base [14]	= [8] / [2]	59,0%	45,7%	50,2%	48,5%	0,7%	
Taux de renouvellement à partir du modèle de contrôle [15]	= [9] / [3]	69,5%	69,6%	68,2%	65,5%	0,9%	

contrôle que dans le modèle de base, car il offre des primes plus avantageuses à ses assurés à fin qu'ils restent dans son portefeuille.

Étant donné que nous ne travaillons pas avec des données provenant d'un réel marché de compétition, mais plutôt avec des données d'un réel assureur, nous devons ajuster l'effet produit par le départ de certains assurés. Les lignes 6 et 7 du tableau 4.11 représentent ces personnes qui sortent de la base de données utilisée. Comme cela ne fait pas partie des nos hypothèses, nous avons décidé de corriger cet effet et d'ajuster les indices en conséquence. De cette façon, nous avons produit le tableau 4.12, où nous présentons les résultats corrigés. Chaque année, nous utilisons uniquement les données des assurés qui changent ou qui restent de compagnie dans le scénario de compétition.

Tableau 4.12 évaluation relative du mouvement des assurés corrigé

	Formule	année 1	année 2	année 3	année 4	Moyenne
Taux de changement base pour contrôle [16]	= [10] / [10] + [14]	19,8%	36,3%	26,3%	25,8%	27,1%
Taux de changement contrôle pour base [17]	= [14] / [10] + [14]	7,3%	5,0%	4,7%	3,8%	5,2%
Taux de renouvellement à partir du modèle de base [18]	= [11] / [11] + [15]	80,2%	63,7%	73,7%	74,2%	73,0%
Taux de renouvellement à partir du modèle de contrôle [19]	= [15] / [11] + [15]	92,7%	95,0%	95,3%	96,2%	94,8%

Nous avons observé qu'en moyenne 27,1% des assurés passent du modèle de base au modèle de contrôle et que seulement 5,2% font l'inverse. Cela démontre à quel point le modèle de contrôle est plus attrayant et à quel point il est plus apte à retenir les titulaires de police. Selon nos estimations, 94,8% des titulaires de polices restent dans le modèle de contrôle chaque année alors que seulement 73%, demeurent dans le modèle de base. Cela confirme que le modèle de contrôle est plus efficace à fidéliser et à attirer les assurés que le modèle de base. En conséquence, cela assure plus de clients dans le portefeuille, ce qui représente une plus grande stabilité financière parce qu'il y a un plus grand nombre de personnes assurées qui paient une prime d'assurance. Cela influence grandement la réduction de la fluctuation du ratio sinistre / primes, ainsi que la précision des sinistres prévus.

4.3.3 Analyse approfondie de la compétition

Évaluons plus en détails pourquoi le modèle de base est déficitaire et moins attractif dans ce scénario concurrentiel par rapport au modèle de contrôle. Tout d'abord, nous analysons le comportement des primes, des sinistres et du ratio sinistre / prime selon certains critères. Nous utiliserons les résultats de nos observations pour étayer nos conclusions. Pour ce faire, nous évaluons la première année en détails, car elle a le plus grand nombre d'expositions et d'informations. Nous supposons que les années suivantes auront un comportement similaire. Dans chaque tableau analysé, nous avons divisé les assurés en dix sections. Les intervalles représentent un dixième de la prime d'assurance maximale qu'un assuré peut se permettre tout en tenant compte des deux modèles. Ces intervalles représentent également des niveaux de risque, c'est-à-dire qu'un assuré classé dans la première section est moins une menace qu'un assuré classé dans la dixième catégorie. Les bandes intermédiaires sont divisées et organisées de manière croissante par rapport au niveau de risque.

La présentation de cette analyse est faite par bandes incluant une bande totale. Elle comprend la répartition des assurés, des primes, des réclamations et du ratio sinistre / prime. Elle est divisée en :

- Analyse générale : nous allons vérifier les chiffres généraux.
- Changement de modèle : nous vérifierons les numéros des assurés qui ont décidé de changer de modèle.
- Renouvellement : nous vérifierons le total d'assurés qui ont décidé de rester dans la même compagnie.

4.3.3.1 Analyse générale

Dans ce qui suit, nous évaluons les chiffres généraux de la première année de la compétition par bande de prime et par modèle. Nous commençons par la distribution des assurés, ensuite nous approfondissons l'analyse des sinistres. Enfin, nous comprenons la répartition des primes et le ratio sinistre / prime. Le tableau 4.13 montre cette répartition.

Nous voyons que la répartition des assurés dans les trois premières bandes de prime est importante afin que le modèle puisse générer assez d'argent pour pouvoir payer les sinistres. Les assurés moins risqués se concentrent dans ces trois tranches et c'est la partie du portefeuille qui est la plus volumineuse. Le modèle de contrôle concentre 85% des assurés dans ces bandes, tandis que le modèle de base y concentre 75% de ses assurés. Cela nous montre que le modèle de base attire plus d'assurés à haut risque à son portefeuille que le modèle de contrôle.

Le tableau 4.14 présente la distribution des sinistres par bande de prime et par modèle. L'information fondamentale que nous pouvons extraire de ce tableau est que la compagnie 2 détient 78% des assurés, mais seulement 34% des réclamations

Tableau 4.13 Distribution des assurés selon la valeur de la prime et par modèle pour l'année 1

Bande de prime (\$)			N_{base}	%	Σ	$N_{\text{contrôle}}$	%	Σ	Total
0,0000	$< P \leq$	0,0268	2 102	17%	17%	7 085	16%	16%	9 187
0,0268	$< P \leq$	0,0537	3 181	25%	42%	15 763	35%	51%	18 944
0,0537	$< P \leq$	0,0805	5 281	42%	84%	17 891	40%	92%	23 172
0,0805	$< P \leq$	0,1074	1 495	12%	96%	3 074	7%	99%	4 569
0,1074	$< P \leq$	0,1342	449	4%	99%	525	1%	100%	974
0,1342	$< P \leq$	0,1611	69	1%	100%	92	0%	100%	161
0,1611	$< P \leq$	0,1879	8	0%	100%	19	0%	100%	27
0,1879	$< P \leq$	0,2148	1	0%	100%	2	0%	100%	3
0,2148	$< P \leq$	0,2416	0	0%	100%	0	0%	100%	0
0,2416	$< P \leq$	0,2685	0	0%	100%	0	0%	100%	0
Total			12 586			44 451			57 037

Tableau 4.14 Distribution des sinistres selon la prime et le modèle pour la première année

Bande de prime (\$)			modèle de base			modèle de contrôle			Total
			Resp.	Non- resp.	Total	Resp.	Non- resp.	Total	Global
0,0000	$< P \leq$	0,0268	19	77	96	52	4	56	152
0,0268	$< P \leq$	0,0537	28	416	444	59	4	63	507
0,0537	$< P \leq$	0,0805	104	807	911	235	10	245	1 156
0,0805	$< P \leq$	0,1074	241	311	552	439	37	476	1 028
0,1074	$< P \leq$	0,1342	95	96	191	168	84	252	443
0,1342	$< P \leq$	0,1611	24	26	50	48	22	70	120
0,1611	$< P \leq$	0,1879	2	6	8	12	6	18	26
0,1879	$< P \leq$	0,2148	0	2	2	2	0	2	4
0,2148	$< P \leq$	0,2416	0	0	0	1	0	1	1
0,2416	$< P \leq$	0,2685	0	0	0	0	0	0	0
Total			513	1 741	2 254	1 016	167	1 183	3 437

totales¹, tandis que la compagnie 1 ne compte que 22% des assurés et 66% du total des sinistres. Nous en concluons que la compagnie représentant le modèle de base a une forte propension à attirer les assurés avec plus de risque d'avoir des accidents. De plus, même dans les trois premières tranches de primes, le modèle de base a plus de sinistres que le modèle de contrôle. Il est clair qu'il existe une très forte concentration de sinistres non responsables sous le modèle de base, ce qui nous conduit à affirmer que ce type d'accident a beaucoup d'influence sur la décision du choix de modèle par les assurés. Nous nommons cela de l'antiselection. Elle se produit lorsqu'un facteur de risque dans l'assurance est mal calibré et elle permet l'existence d'une assurance dans un déséquilibre technique. Ce déséquilibre consiste en une prime inférieure à la valeur du risque calculé actuariellement. Dans le modèle de base, 77% des sinistres sont non responsables et seulement 23% sont des réclamations responsables. Dans le modèle de contrôle, les statistiques sont contraires, 85% des réclamations du type responsable et seulement 15% du type non responsable. Sur le total des sinistres non responsables, 91% sont liés au modèle de base, alors que 9% proviennent du modèle de contrôle. De toute évidence, l'absence de contrainte légale dans le modèle de contrôle a évité l'antiselection observée dans le modèle de base.

Le tableau 4.15 montre la somme des primes d'assurance, les moyennes et le ratio sinistre / prime par bande de primes. Nous pouvons observer que le modèle de base présente un déficit dans toutes les tranches, alors que le modèle de contrôle est rentable dans les trois premières bandes et il est déficitaire dans les autres. Malgré cela, le modèle de contrôle demeure rentable. Nous remarquons que la prime moyenne par tranche dans les deux modèles est semblable : le montant que chaque modèle perçoit par assuré à chaque bande est similaire. Comme elles sont

1. Comme nous modélisons la fréquence des sinistres comme le coût total, nous considérons la quantité totale des sinistres et le coût total des sinistres comme une seule chose.

Tableau 4.15 Distribution de primes par bande de prime et par modèle pour la première année

Bande de prime (\$)	Prime (\$)							Ratio Sinistre / Prime (%)		
	modèle de base		modèle de contrôle		Grand total		modèle de base	modèle de contrôle	Total Global	
	Somme	Moyenne	Somme	Moyenne	Somme	Moyenne				
	x 1000		x 1000		x 1000					
0,0000 < P ≤ 0,0268	28,8	13,7	93,8	13,2	122,6	13,3	333%	60%	124%	
0,0268 < P ≤ 0,0537	142,7	44,9	715,8	45,4	858,5	45,3	311%	9%	59%	
0,0537 < P ≤ 0,0805	341,7	64,7	1 121,5	62,7	1 463,2	63,1	267%	22%	79%	
0,0805 < P ≤ 0,1074	136,8	91,5	278,4	90,6	415,2	90,9	403%	171%	248%	
0,1074 < P ≤ 0,1342	52,6	117,2	61,5	117,2	114,1	117,2	363%	410%	388%	
0,1342 < P ≤ 0,1611	9,8	142,2	13,2	143,3	23,0	142,8	510%	531%	522%	
0,1611 < P ≤ 0,1879	1,4	171,2	3,2	170,1	4,6	170,4	584%	557%	565%	
0,1879 < P ≤ 0,2148	0,2	206,0	0,4	203,1	0,6	204,1	971%	492%	653%	
0,2148 < P ≤ 0,2416	0,0		0,0		0,0					
0,2416 < P ≤ 0,2685	0,0		0,0		0,0					
Total	714,0	56,7	2 287,9	51,5	3 001,9	52,6	316%	52%	114%	

similaires, le ratio très élevé de sinistre / prime du modèle de base est dû aux sinistres, qui sont nombreux. Nous confirmons ainsi l'importance des trois premières bandes pour la survie financière des modèles, puisqu'elles soutiennent le profit du modèle de contrôle et permettent au marché d'être légèrement déficitaire. Nous en concluons aussi que le modèle de base provoque l'antisélection et qu'il assume la plupart des sinistres non responsables. Ainsi, ce type de réclamations devrait être pris en considération de la même manière que les réclamations responsables afin d'éviter un revers financier du modèle dans un marché concurrentiel.

4.3.3.2 Le changement de compagnie

Nous évaluerons maintenant les chiffres des assurés qui ont décidé de changer de compagnie à la fin de la première année. Nous commençons par la distribution des assurés, puis nous approfondissons l'analyse des sinistres et, par la dernière, nous voyons le ratio sinistre / prime. Le tableau 4.16 présente la répartition des assurés.

Tableau 4.16 Distribution des assurés par bande prime et par modèle pour l'année 1 pour ceux qui ont changé d'un modèle à l'autre.

Bande de prime (\$)		N_{base}	%	Σ	$N_{\text{contrôle}}$	%	Σ	Total
0,0000	$< P \leq 0,0268$	165	9%	9%	234	10%	10%	399
0,0268	$< P \leq 0,0537$	468	26%	35%	411	17%	26%	879
0,0537	$< P \leq 0,0805$	1 028	56%	91%	1,159	47%	73%	2 187
0,0805	$< P \leq 0,1074$	137	7%	98%	565	23%	96%	702
0,1074	$< P \leq 0,1342$	26	1%	100%	84	3%	100%	110
0,1342	$< P \leq 0,1611$	9	0%	100%	6	0%	100%	15
0,1611	$< P \leq 0,1879$	0	0%	100%	1	0%	100%	1
0,1879	$< P \leq 0,2148$	0	0%	100%	0	0%	100%	0
0,2148	$< P \leq 0,2416$	0	0%	100%	0	0%	100%	0
0,2416	$< P \leq 0,2685$	0	0%	100%	0	0%	100%	0
Total		1 833			2 460			4 293

91% des personnes qui ont décidé de changer de la compagnie 1 pour la compagnie 2 sont dans les trois premières bandes. 73% font le chemin inverse. De cette façon, il est clair que le modèle de base transfère plus d'assurés jugés souhaitables, selon le niveau de risque, qu'il en reçoit. Cela soutient la tendance qu'ont les assurés moins risqués d'opter pour le modèle de contrôle.

Tableau 4.17 Distribution des sinistres par bande de prime et par modèle pour l'année 1 pour les assurés qui ont décidé de changer de compagnie.

Bande de prime (P)	modèle de base			modèle de contrôle			Grand Total
	Resp.	Non-Resp.	Total	Resp.	Non-Resp.	Total	
0,0000 < P ≤ 0,0268	1	0	1	0	0	0	1
0,0268 < P ≤ 0,0537	0	0	0	0	0	0	0
0,0537 < P ≤ 0,0805	5	0	5	6	0	6	11
0,0805 < P ≤ 0,1074	17	3	20	27	1	28	48
0,1074 < P ≤ 0,1342	9	1	10	9	6	15	25
0,1342 < P ≤ 0,1611	2	4	6	3	2	5	11
0,1611 < P ≤ 0,1879	0	0	0	0	0	0	0
0,1879 < P ≤ 0,2148	0	0	0	0	0	0	0
0,2148 < P ≤ 0,2416	0	0	0	0	0	0	0
0,2416 < P ≤ 0,2685	0	0	0	0	0	0	0
Total	34	8	42	45	9	54	96

En ce qui concerne le résultat du tableau 4.17, nous constatons que les réclamations influencent peu la décision de l'assuré par rapport au choix d'une compagnie ou d'une autre. Au total, 4 293 assurés ont changé de société. De ce dernier, nous ne pouvons observer que 96 sinistres, ce qui représente seulement 2,2% des cas.

Le ratio sinistre / prime des bandes les plus volumineuses est excellent, ainsi que le ratio total. Conséquemment, il existe peu de liens entre la décision de changer de compagnie et les réclamations. Nous confirmons que ce ne sont pas les réclamations qui font échanger les assurés de compagnie, mais plutôt le meilleur prix offert.

Tableau 4.18 Distribution des primes par bande de prime et par modèle pour l'année 1 pour les assurés qui ont décidé de changer un modèle par l'autre.

Bande de prime (\$)	Prime (\$)						Ratio Sinistre / Prime (%)		
	modèle de base		modèle de contrôle		Grand total		modèle de base	modèle de contrôle	Total Global
	Somme	Moyenne	Somme	Moyenne	Somme	Moyenne			
	x 1000		x 1000		x 1000				
0,0000 < P ≤ 0,0268	2,3	13,9	3,1	13,4	5,4	13,6	44%	0%	18%
0,0268 < P ≤ 0,0537	22,2	47,5	18,3	44,6	40,6	46,1	0%	0%	0%
0,0537 < P ≤ 0,0805	65,9	64,1	77,7	67,1	143,6	65,7	8%	8%	8%
0,0805 < P ≤ 0,1074	12,4	90,4	51,2	90,5	63,5	90,5	161%	55%	76%
0,1074 < P ≤ 0,1342	3,0	115,0	9,7	114,9	12,6	114,9	335%	155%	198%
0,1342 < P ≤ 0,1611	1,3	141,3	0,9	144,1	2,1	142,4	472%	578%	515%
0,1611 < P ≤ 0,1879	0,0		0,2	174,2	0,2	174,2		0%	0%
0,1879 < P ≤ 0,2148	0,0		0,0		0,0				
0,2148 < P ≤ 0,2416	0,0		0,0		0,0				
0,2416 < P ≤ 0,2685	0,0		0,0		0,0				
Total	107,0	58,4	161,1	65,5	268,1	62,4	39%	34%	36%

4.3.3.3 Des assurés qui sont restés dans la même compagnie

Dans cette partie, nous évaluons les chiffres des assurés qui ont décidé de rester dans la même compagnie après la première année de la compétition. Nous abordons la distribution des assurés, des sinistres, des primes et du ratio sinistre / prime.

Il est possible de voir dans le tableau 4.19 que 92% ont décidé de renouveler leur police d'assurances auprès du modèle de contrôle. Ces même personnes sont dans les trois premières bandes de prime. Ce ratio est de 79% pour le modèle de base. Dans ces bandes, nous avons la plus grande partie des assurés de chaque modèle. Cette différence de 13% est donc considérable. Le fait que le modèle de contrôle retient plus d'assurés est relié au fait d'avoir généralement une prime plus basse par profil dans ces sections. Cela est dû au fait que le modèle de contrôle offre de primes plus ajustées au risque et cela retient les assurés qui sont moins risqués.

À travers les informations du tableau 4.20, nous vérifions pourquoi le modèle de

Tableau 4.19 Distribution des assurés par bande de prime et par modèle pour l'année 1 pour les assurés qui ont décidé de rester dans la même compagnie.

Bande de prime (\$)			N_{base}	%	Σ	$N_{\text{contrôle}}$	%	Σ	Total
0,0000	$< P \leq$	0,0268	702	9%	9%	2 654	9%	9%	3 356
0,0268	$< P \leq$	0,0537	1 682	23%	32%	11 467	37%	46%	13 149
0,0537	$< P \leq$	0,0805	3 505	47%	79%	14 379	47%	92%	17 884
0,0805	$< P \leq$	0,1074	1 129	15%	95%	1 998	6%	99%	3 127
0,1074	$< P \leq$	0,1342	354	5%	99%	332	1%	100%	686
0,1342	$< P \leq$	0,1611	44	1%	100%	71	0%	100%	115
0,1611	$< P \leq$	0,1879	5	0%	100%	12	0%	100%	17
0,1879	$< P \leq$	0,2148	0	0%	100%	0	0%	100%	0
0,2148	$< P \leq$	0,2416	0	0%	100%	0	0%	100%	0
0,2416	$< P \leq$	0,2685	0	0%	100%	0	0%	100%	0
Total			7 421			30 913			38 334

Tableau 4.20 Distribution des sinistres par bande de prime et par modèle pour l'année 1 pour les assurés qui ont décidé de rester dans la même compagnie.

Bande de prime (P)	modèle de base			modèle de contrôle			Grand Total		
	Resp.	Non-Resp.	Total	Resp.	Non-Resp.	Total			
0,0000	$< P \leq$	0,0268	2	20	22	11	0	11	33
0,0268	$< P \leq$	0,0537	7	283	290	17	2	19	309
0,0537	$< P \leq$	0,0805	54	690	744	157	3	160	904
0,0805	$< P \leq$	0,1074	176	255	431	347	20	367	798
0,1074	$< P \leq$	0,1342	66	84	150	121	55	176	326
0,1342	$< P \leq$	0,1611	11	15	26	36	13	49	75
0,1611	$< P \leq$	0,1879	1	2	3	6	4	10	13
0,1879	$< P \leq$	0,2148	0	0	0	1	0	1	1
0,2148	$< P \leq$	0,2416	0	0	0	0	0	0	0
0,2416	$< P \leq$	0,2685	0	0	0	0	0	0	0
Total			317	1 349	1 666	696	97	793	2 459

contrôle a un taux de renouvellement plus élevé. Les raisons les plus fortes sont de deux ordres : il offre des primes plus avantageuses par rapport au modèle de base pour les trois premières bandes de prime et il pénalise les sinistres non responsables. Si nous revenons sur la deuxième raison, nous voyons que des 30 913 assurés qui sont restés dans le modèle de contrôle d'une année à l'autre, seuls 793 avaient des sinistres (2,6%). Tandis que, sur les 7 421 assurés du modèle de base, 1 666 ont présenté des réclamations, ce qui représente 22,5% des réclamations. Ceci nous montre à quel point le modèle de base attire les assurés susceptibles de faire des réclamations en raison de la structure de pénalisation de son SBM. De plus, le modèle de base comporte 1 349 réclamations non responsables alors que le modèle de contrôle n'en compte que 97. Cela renforce l'antisélection causé par l'inclusion de la contrainte légale. De cette façon, tous les assurés avec haute probabilité d'avoir des sinistres non responsables migrent au portefeuille du modèle de base. Finalement, l'inclusion des sinistres non responsables dans le modèle est aussi un excellent prédicteur des futures réclamations.

Tableau 4.21 Somme des primes selon la tranche de prime et le modèle pour l'année 1 pour les assurés qui ont resté dans le même modèle.

Bande de prime (\$)	Prime (\$)							Ratio Sinistre / Prime (%)		
	modèle de base		modèle de contrôle		Grand total		modèle de base	modèle de contrôle	Total	
	Somme	Moyenne	Somme	Moyenne	Somme	Moyenne			Global	
	x 1000		x 1000		x 1000					
0,0000 < P ≤ 0,0268	10,0	14,2	35,9	13,5	45,9	13,7	220%	31%	72%	
0,0268 < P ≤ 0,0537	78,0	46,4	536,8	46,8	614,7	47	372%	4%	50%	
0,0537 < P ≤ 0,0805	227,6	64,9	894,2	62,2	1 121,8	63	327%	18%	81%	
0,0805 < P ≤ 0,1074	103,5	91,7	180,8	90,5	284,3	91	416%	203%	281%	
0,1074 < P ≤ 0,1342	41,5	117,3	39,0	117,4	80,5	117	361%	452%	405%	
0,1342 < P ≤ 0,1611	6,2	142,0	10,1	142,9	16,4	143	416%	483%	457%	
0,1611 < P ≤ 0,1879	0,8	169,9	2,0	167,7	2,9	168	353%	497%	454%	
0,1879 < P ≤ 0,2148	0,0		0,0		0,0					
0,2148 < P ≤ 0,2416	0,0		0,0		0,0					
0,2416 < P ≤ 0,2685	0,0		0,0		0,0					
Total	467,7	63,0	1 698,8	55,0	2 166,5	57	356%	47%	114%	

En analysant les informations du tableau 4.21, nous remarquons que la prime moyenne du modèle de contrôle est plus basse que celui du modèle de base. Le modèle de contrôle n'a presque aucune réclamation dans les trois premières bandes et il a de bonnes quantités de primes, ce qui garantit un résultat optimal. Cet excellent résultat dans ces trois tranches détermine le profit général du modèle. C'est l'inverse pour le modèle de base. Il y a beaucoup des réclamations par rapport aux montants des primes, ce qui lui garantit des ratios sinistre / prime très élevés. Quand nous avons fait l'analyse du NRMS, dans le chapitre précédent, nous avons observé que le modèle de base avait une statistique basse par rapport au modèle de contrôle et par rapport aux autres SBM que nous avons comparés. Ceci indique que, dans la construction du modèle, la structure optimale du modèle de base concentrent les assurés dans les bandes inférieures de prime. Cela le désavantage face à la compétition dès qu'il charge davantage que le modèle de contrôle pour les bandes comprenant des assurés à faibles risques.

4.3.4 Résultats du marché

Dans cette sous-section, nous discutons les résultats obtenus dans le modèle de concurrence pour la première année. Puis, nous les utiliserons pour présenter une généralisation des résultats pour les années suivantes.

Nous commençons l'analyse des éléments du tableau 4.22 par la valeur moyenne des primes à chaque tranche. Ensuite, nous analyserons la valeur moyenne totale. D'abord, nous notons qu'il n'y a pas de valeur dans les deux dernières bandes puisqu'il n'y a pas d'exposition. Les valeurs moyennes sont croissantes dans les bandes de 1 à 8. La valeur moyenne des primes de la tranche 8 est environ 15 fois supérieure à celle de la première bande. Ainsi, le marché répartit le niveau de risque entre les bandes et celles-ci sont distinctes entre elles. La prime moyenne du marché se situe entre les tranches 2 et 3, ce qui est prévisible puisqu'il y a une

Tableau 4.22 La prime et le sinistres (somme et moyenne) par bande de prime

Bande de prime (\$)	Bande	N	Prime (\$)	Prime Moyenne (\$)* 1000	Sinistre (\$)	Sinistre Moyen (\$)* 1000	Ratio S / P (%)
0,0000 < P ≤ 0,0268	1	9 187	122,6	13,3	152	16,5	124%
0,0268 < P ≤ 0,0537	2	18 944	858,5	45,3	507	26,8	59%
0,0537 < P ≤ 0,0805	3	23 172	1463,2	63,1	1 156	49,9	79%
0,0805 < P ≤ 0,1074	4	4 569	415,2	90,9	1 028	225,0	248%
0,1074 < P ≤ 0,1342	5	974	114,1	117,2	443	454,8	388%
0,1342 < P ≤ 0,1611	6	161	23,0	142,8	120	745,3	522%
0,1611 < P ≤ 0,1879	7	27	4,6	170,4	26	963,0	565%
0,1879 < P ≤ 0,2148	8	3	0,6	204,1	4	1 333,3	653%
0,2148 < P ≤ 0,2416	9	0	0,0		1		
0,2416 < P ≤ 0,2685	10	0	0,0		0		
Total		57 037	3 002	52,6	3 437	60,3	114%

forte concentration d'assurés dans les trois premières tranches.

Nous procédons ensuite à l'examen du comportement des réclamations. Comme dans le cas des primes, le montant moyen de la réclamation augmente aussi. La réclamation moyenne de la huitième tranche est à peu près 81 fois plus élevée que la réclamation de la première tranche. Nous notons ici que la perte moyenne et la prime moyenne n'augmentent pas proportionnellement, ce qui nous montre comment le marché et les modèles qui le composent équilibrent la collecte des primes entre les bandes. C'est à dire que les assurés des groupes à plus faible risque absorbent une partie du coût des assurés à risque plus élevé, puisque le taux de croissance moyen des primes et des sinistres moyens ne sont pas égaux. De plus, le montant moyen de la réclamation se situe entre les bandes 3 et 4.

Dans l'analyse des primes et des sinistres moyennes, il est possible de comprendre pourquoi le marché a un résultat déficitaire. La réclamation moyenne est supérieure à la prime moyenne, ce qui entraîne un manque de ressources pour couvrir les coûts. De plus, en observant le ratio sinistre / prime du marché, nous remarquons que les bandes 2 et 3 présentent un résultat positif, mais qu'elles ne sont

pas en mesure de compenser le résultat négatif présent dans les autres bandes. Nous notons également que plus le niveau de risque est élevé, plus la perte est importante.

Par conséquent, dans l'analyse du marché dans ce modèle de concurrence, nous comprenons pourquoi le marché a présenté un résultat négatif. Le principal facteur est la restriction légale incluse dans le modèle de base qui cause l'antisélection des assurés. Cela crée une accumulation de mauvais risques et un manque de ressources qui se reflète finalement sur le marché en entier.

Il est possible de reformuler ce que nous avons observé dans ce modèle en une explication plus formelle portant sur l'équilibre actuariel dans un environnement de concurrence. L'équilibre actuariel est atteint lorsque la somme des primes est égale (ou supérieure) aux sinistres. Comme nous l'avons vu, le marché de ce modèle de concurrence est déficitaire, et nous avons pu identifier les causes de ce déficit. Alors, que pouvons-nous conclure de toute cette analyse empirique ? Comment pouvons-nous systématiser les résultats ?

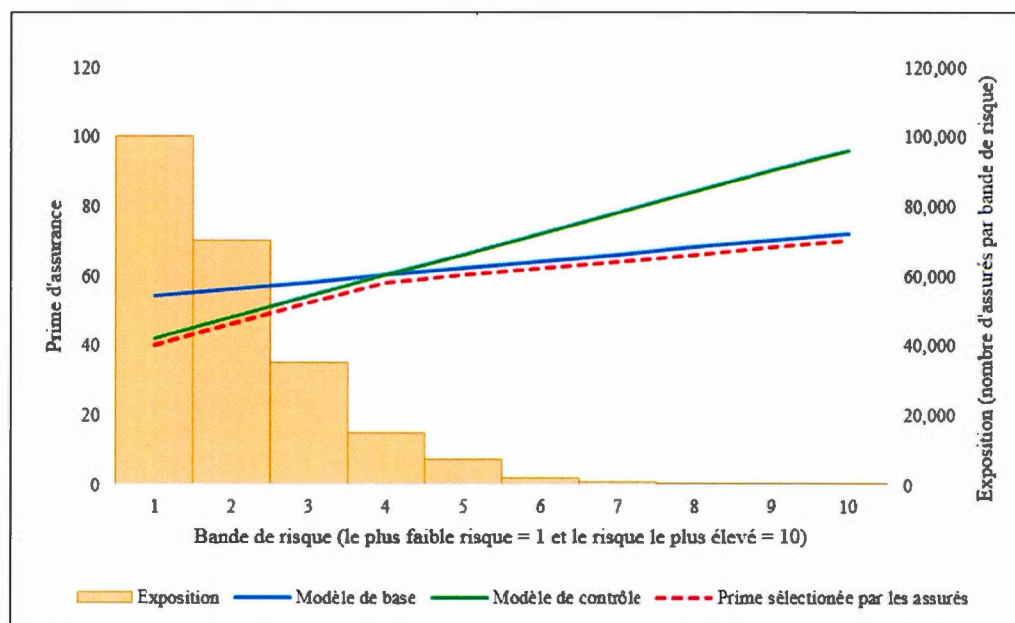
Pour formaliser les résultats observés, utilisons un scénario hypothétique représenté par le graphique 4.2. Ce graphique montre ce qui suit :

1. L'axe des abscisses représente une échelle de risque hypothétique de 1 à 10, où le niveau 1 représente la classe à risque le plus faible et le 10, la classe à risque le plus élevé. Les classes de risque intermédiaires sont organisées dans un ordre croissant. À chaque niveau, nous avons une quantité hypothétique d'expositions.
2. La ligne bleue représente la valeur de prime hypothétique moyenne pour un modèle avec la contrainte légale calculée isolément. En d'autres termes, sans tenir compte de toute forme de concurrence. Cette ligne représente le modèle de base. Le modèle hypothétique doit donc facturer un peu plus cher

dans les catégories à faible risque pour compenser le montant plus modeste facturé aux catégories à risque élevé.

3. La ligne verte représente la valeur de prime hypothétique moyenne pour un modèle sans aucune restriction légale, également calculée isolément. C'est-à-dire, sans tenir compte de toute forme de concurrence. Cette ligne représente le modèle de contrôle. Comme il est libre, ce modèle est en mesure de facturer la valeur actuariellement correcte de la prime par rapport au risque. C'est-à-dire que le montant facturé représente la correcte valeur de prime d'assurances sans aucune forme de compensation ou d'ajustement.
4. La ligne rouge représente le choix du consommateur par rapport à la valeur de la prime et le niveau de risque occupé. Elle est légèrement décalée vers le bas à des fins de visualisation.

Figure 4.2 L'équilibre actuariel du marché



Lorsque nous comparons les lignes bleues et vertes, nous faisons l'inclusion des deux modèles dans un marché concurrentiel. Ainsi, le modèle illustré par la ligne

bleue offre une prime hypothétique moyenne plus favorable pour les assurés dans les classes de risque 5 à 10, tandis que le modèle représenté par la ligne verte présente une meilleure prime hypothétique moyenne pour les classes de risque 1 à 3. Les primes hypothétiques moyennes des deux modèles sont semblables pour la classe 4. De cette façon, le premier modèle attirera davantage de personnes susceptibles de faire des réclamations, alors que le deuxième attirera plus d'individus moins susceptibles de faire des réclamations.

Le modèle illustré par la ligne verte concentre plus d'assurés dans les bandes 1 à 3, car il offre un meilleur prix pour le consommateur. Cela assure un revenu adéquat pour le risque que ces assurés représentent, il aura un résultat financier nul (équilibre) ou positif (profit). Le modèle représenté par la ligne bleue concentre plus d'assurés dans les tranches 5 à 10. En considérant que la prime prélevée par ce modèle sur ces bandes est insuffisante par rapport au risque supporté par l'assureur, il favorise l'antisélection. Un résultat déficitaire est attendu puisqu'il ne mobilise pas suffisamment de fonds pour rembourser les sinistres à payer.

4.3.5 Décision

À la lumière de l'outil d'évaluation présenté dans ce mémoire et des résultats qu'il a générés, nous voyons clairement que le modèle de contrôle est supérieur au modèle de base sous plusieurs aspects :

- La structure du modèle de contrôle est plus puissante pour pénaliser les assurés : elle impose un changement de trois niveaux ($+XX = +3$) pour tous types de sinistres. En revanche, la structure du modèle de base varie de deux ($+xx = +2$) pour des sinistres responsables et de zéro pour des sinistres non responsables. Le modèle de base pénalise insuffisamment en cas d'accident. La performance de la structure du modèle de contrôle évite l'antisélection.
- La prime de l'assuré de base du modèle de contrôle est inférieure à celle du

- modèle de base. De la même façon, le modèle de contrôle a généralement des primes inférieures pour les assurés à faible risque. Conséquemment, plus d'assurés à faible risque choisissent la compagnie qui utilise le modèle de contrôle.
- Le modèle de base ne reconnaît pas que les sinistres non responsables sont prédictifs. Cette non reconnaissance est une des causes de sa faible performance.
 - Le changement d'une compagnie à l'autre n'est pas relié aux sinistres, mais à la valeur de la prime offerte par le concurrent. Dans la majorité des cas, les assurés n'ont pas eu de sinistres.
 - La prime chargée par le modèle de base n'est pas la plus précise. À cause de la contrainte légale, ce modèle a besoin d'ajuster toutes ces primes. Cet ajustement fait en sorte qu'il ne performe pas bien dans un environnement concurrentiel.
 - Le grand volume d'exposition, avec la prime adéquate dans les trois premières bandes de primes, génère assez d'argent et de profit pour soutenir la bonne performance du modèle de contrôle malgré le fait que les bandes de 4 à 10 ne performant pas aussi bien. À l'inverse, le modèle de base n'a eu aucun bon résultat dans ces trois bandes de prime. Cela lui produit des résultats négatifs. Nous avons vu dans le chapitre antérieur que les indices de NRMS sont bas pour les deux modèles, ce qui nous amène à conclure que les premières bandes sont toujours les plus importantes pour l'équilibre financier de ces modèles.

Cet outil illustre sans aucun doute que la restriction légale imposée par la province de l'Ontario, présente dans le modèle de base, compromet le pouvoir prédictif de ce modèle. La restriction légale cache le fait que les réclamations du type non responsables sont fondamentales pour la bonne structuration et l'estimation du modèle. Elles sont autant prédictives que les réclamations responsables. De plus, le fait de ne pas considérer ces types de réclamations provoque toujours de l'antisélection, c'est-à-dire, ce modèle attire les assurés les plus risqués à son portefeuille.

Pour la question de l'évaluation de l'efficacité du modèle de base, il est clair que la restriction doit être toujours faite par le régulateur d'assurance de façon équitable pour tous les assureurs. C'est en effet la seule manière de garder l'équilibre technique du marché et de garantir qu'il n'y a pas un assureur en avantage ou en désavantage. Cela démontre clairement que l'inclusion de la contrainte légale réduit l'efficacité du modèle de base.

Une tendance actuelle, dans le marché d'assurances au Canada, est la police qui pardonne un certain nombre de réclamations de l'assuré et qui n'augmente pas sa prime par la suite. Ces réclamations sont des deux types. Cette manoeuvre à la mode n'est rien d'autre que la contrainte légale mise en place par l'assureur. Ainsi, comme la contrainte légale, cette offre commerciale réduit l'efficacité des modèles et peut engendrer des résultats financiers négatifs.

CONCLUSION

L'actuariat évolue grâce à des modèles, des méthodes et des techniques reliés au sujet de la tarification IARD. Au cours des dernières décennies, deux volets majeurs ont conduit la majorité des méthodes : la tarification « *a priori* » et « *a posteriori* ». La première utilise toujours les informations disponibles au moment du calcul des paramètres du modèle de chaque assuré, tandis que la deuxième se base sur les informations futures des sinistres de chacun des assurés. L'utilisation des informations disponibles au moment du calcul sert à estimer la prime la plus personnalisée possible, tandis que l'emploi des informations futures sert à corriger la prime d'assurances déjà en place. La combinaison de ces deux volets en une seule structure et l'estimation des paramètres du modèle en une seule étape sont récentes dans la littérature. Cela commence à être exploré, étudié et utilisé par les actuaires pour tarifier.

La façon habituelle, toutefois, est d'estimer les paramètres et d'optimiser chacun des volets en séparément. Cette approche est devenue de plus en plus questionnée par la communauté académique car estimer chacune des parties de façon optimale et les intégrer par la suite n'est pas équivalent à estimer tout en même temps. Le modèle proposé par (Boucher et Inoussa, 2014) est une solution pour réparer cela. Ils ont développé une façon d'estimer les paramètres et de sélectionner le meilleur modèle au même temps à l'aide des données de panel. L'inclusion de ces données dans l'élaboration de leur modèle a été la pièce clé de leur réussite. Comme leur approche est récent, peu a été fait pour évaluer l'efficacité et le comportement de leur modèle. De plus, ils ont inclus une contrainte légale qui ne s'applique que

dans la province de l'Ontario. Cette loi contraint les assureurs à ne pas augmenter la prime d'assurances suite à un accident non responsable.

Ce mémoire a eu trois objectifs principaux : le premier a été de reproduire le modèle présenté par (Boucher et Inoussa, 2014) dans leur article (« *A Posteriori Ratemaking With Panel Data* »). Le deuxième a été d'évaluer l'efficacité de leur modèle selon certains critères sélectionnés de (Lemaire, 1995) pour comparer différents SBMs. Le troisième a été d'évaluer la contrainte légale incluse dans le modèle. Cette restriction influence la performance du modèle, et nous avons proposé un nouveau moyen de l'évaluer. Afin de répondre aux objectifs proposés ci-dessus, nous avons pu explorer le modèle de (Boucher et Inoussa, 2014) grâce aux outils proposés dans ce travail.

D'abord, nous avons présenté dans le premier chapitre toutes les notions concernant la prime pure et la segmentation, ainsi qu'une analyse de la base de données utilisée. Cette base est celle que les auteurs ont utilisée dans leur recherche. Ces notions fondent la théorie des modèles de tarification de ce mémoire. Ensuite, dans le deuxième chapitre, nous avons présenté la technique retenue pour modéliser la classification *a priori* du modèle, soit les GLMs. Par la suite, nous avons approfondi la notion de tarification par expérience, tarification *a posteriori*. Nous avons introduit la technique des SBMs utilisée. Pour finaliser les fondements théoriques, nous avons exploré la base mathématique des SBMs, car une bonne partie du travail de (Boucher et Inoussa, 2014) est concentré sur ces techniques.

Dans le troisième chapitre, nous avons détaillé les composantes du modèle de (Boucher et Inoussa, 2014). En conclusion, les auteurs signalent que la tarification d'assurances en deux étapes ne doit plus être utilisée quand nous disposons d'une base de données de panel. Afin de démontrer cela, ils ont proposé une méthode qui consiste à avoir une distribution conjointe prédictive liée à une

structure d'espérance où les relativités du SBM du modèle sont déjà comprises. Les relativités sont dépendantes de la distribution conjointe des sinistres observés $(N_{i,1}, N_{i,2}, \dots, N_{i,T})$. Ils ont nommé ce modèle « Modèle de Panel SBM ». Dans ce mémoire, nous l'avons nommé « modèle de base ».

Lorsque nous avons exploré en détail chacune des parties de leur modèle, nous avons dû faire des estimations des paramètres, telles que les β s de la partie de classification « *a priori* », les relativités du SBM, la structure de pénalisation et le niveau d'entrée des assurés selon le nombre d'années de permis de conduire (NAP). Nous avons utilisé la logvraisemblance pour sélectionner le meilleur modèle de base. À l'intérieur de la structure du modèle de base, il y a l'inclusion de la contrainte légale. Cette contrainte impose qu'un assuré qui a subi des sinistres du type non responsable ne peut voir sa prime augmenter, toutefois elle n'indique pas qu'il faut avoir une réduction non plus. Ainsi, les auteurs ont fait une série d'ajustements pour l'intégrer dans leur modèle, et nous avons expliqué et reproduit son fonctionnement. Pour être capable de mesurer l'effet de cette contrainte, nous avons bâti un modèle de contrôle sans aucun type de restriction. Pour le construire, nous avons utilisé la même procédure que pour le modèle de base, à l'exception de la contrainte.

Nous avons observé différents comportements des deux modèles. La structure optimale du modèle de base est du type -1/0/+2 et celui du modèle de contrôle est -1/+3. Ce dernier modèle est plus sévère et pénalise plus que le premier, car il déplace l'assuré de trois niveaux par sinistre, alors que l'autre, de deux niveaux seulement si le sinistre est du type responsable. Nous avons aussi constaté que pour compenser cette différence dans la structure de pénalisation, les relativités du modèle de contrôle sont égales ou inférieures que celles du modèle de base : $r_l^c \leq r_l^b$, pour $l = (1, 2, \dots, 10)$, où « b » pour le modèle de base et « c » pour le modèle de contrôle. Cette compensation entre la structure de pénalisation et les

valeurs des relativités est toujours mise en relief dans la littérature et est cohérent avec ce que nous avons observé dans les deux modèles. Si la structure pénalisation était moins sévère, la valeur des relativités devrait être plus élevée et vice-versa.

De plus, nous avons constaté que le modèle de contrôle est plus conservateur en ce qui concerne l'assignation du niveau d'entrée des assurés qui ont six ans ou plus d'expérience de conduite lorsqu'ils deviennent clients de la compagnie. De plus, la prime de base du modèle de contrôle est moins chère que celui du modèle de base. Cela est possible parce que la structure de pénalisation du modèle de contrôle est plus sévère et ce modèle peut offrir une prime de base plus attrayante qui compense pour sa structure plus sévère. Quand un assuré déclare un sinistre, sa prime est ajustée à la hausse davantage dans le modèle de contrôle que dans l'autre modèle. Nous avons aussi constaté que les facteurs de la section de classification « *a priori* » du modèle de base ont de valeurs plus extrêmes. Cela veut dire que les rabais et les surcharges que le modèle de base apporte aux primes d'assurances sont plus importants que celles du modèle de contrôle.

Dans le quatrième chapitre, nous avons introduit deux outils actuariels-statistiques qui nous ont guidé dans l'évaluation de l'efficacité du modèle de base. Pour ce faire, nous avons considéré les résultats de l'ouvrage de (Lemaire, 1995) comme base de comparaison des résultats obtenus pour les modèles de base et de contrôle. Le premier outil, le coefficient de variation (CV), nous a montré que le modèle de contrôle différencie davantage la prime d'assurance que le modèle de base. En comparaison avec les modèles d'autres pays, les deux modèles se sont montrés plus performants que ceux des autres. Le deuxième outil, le niveau relatif moyen stationnaire (NRMS), nous a indiqué que le modèle de contrôle est capable de mieux distribuer les assurés dans son échelle que le modèle de base. Ce dernier modèle a mal performé même contre les SBMs des autres pays. Il était placé comme un des modèles qui distribue moins les assurés dans l'échelle, ce qui signifie qu'il

n'est pas assez sévère pour bien distribuer les assurés. Les résultats des outils actuariels-statistiques employés dans l'analyse ne sont pas surprenants, toutefois ces résultats nous ont confirmé que le modèle de contrôle est plus efficace que le modèle de base. Nous en concluons que l'ajout de la contrainte légale a affecté l'efficacité du système.

Nous avons aussi présenté le rapport sinistre / prime pour chacun des modèles dans le quatrième chapitre. Ce type de rapport sert à nous assurer que les primes calculées par les modèles sont suffisantes pour payer tous les sinistres. Les résultats sont comme attendus : les deux modèles sont en mesure d'honorer leurs obligations auprès des assurés et ils sont bien ajustés par rapport aux données utilisés. Cela veut dire qu'ils sont en équilibre actuarielle. Pour compléter toutes les analyses, après nous être assurés que les deux modèles sont en équilibre technique, quand isolés, nous avons proposé une nouvelle façon d'évaluer la contrainte légale à l'intérieur du modèle de base : une compétition de marché entre les modèles pour voir lequel est plus profitable. L'hypothèse sous-jacent est que les assurés vont s'assurer avec la compagnie qui offre la prime la plus basse et que cela a été le seul élément considéré dans leur choix.

Un des faits observés à nouveau, grâce à la compétition, est que le modèle de contrôle pénalise plus que le modèle de base et il présente une prime de base moins chère. Cela attire plus d'assurés à faible risque à son portefeuille. Par ailleurs, comme le modèle de base ne considère pas les sinistres non responsables comme un facteur de risque, la majorité de sinistres de ce type sont reliées aux assurés de la compagnie qu'utilise le modèle de base. Cela permet au modèle de contrôle d'éviter l'antisélection. Nous avons aussi constaté que le choix de l'assuré de changer de compagnie n'est pas liée aux sinistres, mais à la valeur de la prime. La majorité des assurés qui ont changé de compagnie n'ont eu pas de sinistres.

De plus, la prime calculée par le modèle de base est moins précise, car cette prime utilise un ajustement pour compenser la non inclusion des sinistres non responsables. Cette compétition démontre que le modèle de contrôle est plus efficace et plus puissant que le modèle de base, car il a fait du profit pendant toute la période de compétition. Le modèle de base a eu un déficit dans la même période. Le marché, composé de la combinaison des deux compagnies, a eu un déficit.

Après toutes les analyses empiriques, nous avons élaboré une proposition formelle sur le comportement du marché et des assurés. Dans notre généralisation, les assurés plus risqués choisissent la compagnie qui utilise le modèle de base et les assurés les moins risqués sélectionnent la compagnie qui utilise le modèle de contrôle. En conséquence, le marché n'atteint pas son équilibre, car le total perçu en primes demeure toujours inférieur au minimum nécessaire pour payer tous les sinistres générés par les assurés. Ainsi, le marché est en déficit, même si une des compagnies est profitable.

Finalement, après toutes les analyses et évaluations du modèle de base, nous pouvons conclure que l'inclusion de la contrainte légale, provenant de la province de l'Ontario, a affecté l'efficacité et la performance de ce modèle face à un modèle libre de cette contrainte. Tous les résultats l'indiquent. Pour les recherches futures, il est important d'inclure d'autres outils actuariels-statistiques pour l'évaluation de l'efficacité du modèle de base par rapport à d'autres aspects tels que la quantité des niveaux, les contraintes des différentes juridictions et le modèle de classification « *a priori* ». Des analyses plus approfondies en utilisant la compétition des modèles peut nous amener à de nouvelles méthodes d'évaluation et à de nouveaux modèles de tarification. Nous pouvons aussi examiner d'autres variations du modèle de (Boucher et Inoussa, 2014) ou, encore, l'améliorer.

RÉFÉRENCES

- Berkhout, P. et Plug, E. (2004). A bivariate poisson count data model using conditional probabilities. *Statistica Neerlandica*, 58(3), 349–364. Récupéré de <http://EconPapers.repec.org/RePEc:bla:stane:v:58:y:2004:i:3:p:349-364>
- Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin*, 4, 199–207.
- Boland, P. J. (2007). *Statistical and Probabilistic Methods in Actuarial Science*. Chapman and Hall/CRC.
- Boucher, J.-P. (2014). *Notes du cours MAT-8594 Statistique en actuariat*. UQÀM.
- Boucher, J.-P. et Guillén, M. (2009). A survey on models for panel count data with applications to insurance. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*, 103(2), 277–295.
- Boucher, J.-P. et Inoussa, R. (2014). A posteriori ratemaking with panel data. *ASTIN Bulletin*, 44(03), 587–612.
- DeJong, P. et Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press.
- Denuit, M. et Charpentier, A. (2005). *Mathématiques de l'assurance non-vie, Tome II : Tarification et Provisionnement*. Economica.
- Denuit, M., Maréchal, X., Pitrebois, S. et Walhin, J.-F. (2007). *Actuarial Modeling of Claim Counts : Risk Classification, Credibility and Bonus-Malus Systems*. A John Wiley and Sons, Inc., Publication.
- Explained, E. . S. (2015). Glossary :in-sample vs. out-of-sample forecasts. Récupéré de http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:In-sample_vs._out-of-sample_forecasts
- Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models* (deuxième éd.). Sage Publications, Inc.

- FSCO (2017). FSCO - financial services commission of ontario - after an auto accident : Understanding the claims process. Récupéré de https://www.fsco.gov.on.ca/en/auto/brochures/pages/brochure_claims.aspx
- Gilde, V. et Sundt, B. (1989). On bonus systems with credibility scales. *Scandinavian Actuarial Journal*, 1, 13–22.
- Klugman, S. A., Panjer, H. H. et Willmot, G. E. (2008). *Loss Models : From Data to Decisions* (troisième éd.). A John Wiley and Sons, Inc., Publication.
- Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers.
- Molenberghs, G. et Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer-Verlag New York.
- Nelder, J. A. et Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistic Society. Series A (General)*, 135(3), 370–384.
- Newton L. Bowers, J., Gerber, H. U., Hickman, J. C., James, D. A. et Nesbitt, C. J. (1997). *Actuarial Mathematics* (deuxième éd.). The Society of Actuaries.
- Norberg, R. (1976). A credibility theory for automobile bonus systems. *Scandinavian Actuarial Journal*, 1976(2), 92–107.
- Pesonen, M. (1963). A numerical method of finding a suitable bonus scale. *ASTIN Bulletin*, 1963(2), 102–108.
- Ross, S. M. (2014). *Introduction to Probability Models* (onzième éd.). Academic Press.
- Taylor, G. (1997). Setting a bonus-malus scale in the presence of other rating factors. *ASTIN Bulletin*, 27, 319–327.