UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MODELLING OF UNEARNED PREMIUM RISK WITH DEPENDENCE

DISSERTATION

PRESENTED

AS PARTIAL REQUIREMENT

TO THE MASTERS IN MATHEMATICS

BY

SÉBASTIEN JESSUP

APRIL 2020

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MODÉLISATION DU RISQUE LIÉ AUX PRIMES NON-ACQUISES AVEC DÉPENDANCE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

SÉBASTIEN JESSUP

AVRIL 2020

REMERCIEMENTS

Je souhaite d'abord et avant tout remercier mon directeur de recherche, Mathieu Pigeon, qui a su répondre à mes questions tout au long de la maîtrise, et qui a fourni une aide inestimable dans la rédaction d'un premier article scientifique. Merci également à Jean-Philippe Boucher qui nous a aidé à cerner le problème de recherche et qui a proposé des pistes de réflexion intéressantes tout au long du processus.

J'aimerais remercier l'Autorité des Marchés Financiers pour le soutien financier dans ce projet, ainsi que l'UQAM et le Laboratoire de mathématiques actuarielles et financières Quantact pour la bourse de maîtrise. Je remercie évidemment aussi la Compagnie d'assurance générale Co-operators et la Chaire Co-Operators en analyse des risques actuariels de l'UQAM pour les données mises à ma disposition.

Finalement, un énorme merci à mes parents, qui m'ont donné un sérieux coup de main au cours des études graduées, et sans qui ce parcours aurait été beaucoup plus difficile.

AVANT-PROPOS

Le projet de recherche présenté dans ce mémoire est le résultat d'un appel de projets lancé par l'Autorité des Marchés Financiers (AMF) au sujet de la prime non-acquise. Nous devions investiguer le risque lié à cette dernière, produire un article et le présenter dans au moins une conférence majeure.

La prime non-acquise n'ayant été que très peu explorée jusqu'à maintenant, ceci me permettait d'appliquer les connaissances acquises au cours de la maîtrise à un nouveau problème en étant financé par le Fonds pour l'éducation et la saine gouvernance de l'AMF.

En juillet et août 2019, j'ai pu présenter ma recherche dans le cadre du *Joint Statistical Meeting* à Denver, l'une des plus grandes conférences de statistiques en Amérique du Nord, ainsi qu'au *Actuarial Research Conference* à Indianapolis.

Le projet m'a donné l'opportunité d'écrire un premier article scientifique, qui vient tout juste d'être soumis. L'article a été coécrit par Mathieu Pigeon, mon directeur de recherche à la maîtrise, par Jean-Philippe Boucher, et par moi. Messieurs Pigeon et Boucher ont préparé les données, à partir desquelles j'ai fait une analyse quantitative préliminaire. Nous avons ensuite développé un modèle, avec lequel j'ai fait des simulations à partir des données fournies.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

RÉSUMÉ

La prime non-acquise, ou plus particulièrement le risque associé, n'a que récemment reçu de l'attention des régulateurs. Les pertes non-acquises se produisent après la date d'évaluation pour des contrats souscrits avant cette dernière. Étant donné qu'un modèle d'acquisition inadéquat et une modélisation approximative du passif des primes peut mener à une réserve inexacte pour le risque de la prime non-acquise, un modèle individuel non-homogène des pertes incluant la dépendance entre les couvertures est proposé pour établir une façon alternative d'évaluer ce risque. La survenance des pertes est analysée en termes de la saisonalité et de la fréquence d'implication de couvertures multiples. Des distributions homogènes et non-homogènes sont ajustées aux distributions marginales. Des copules sont ajustées par paire de couvertures en utilisant des méthodes basées sur le rang ainsi qu'une fonction de queue (*tail function*). Cette approche est appliquée à une base de données automobile récente d'Ontario.

Mot Clés : Risque de la prime non-acquise, Réserves actuarielles, Modélisation prédictive, Dépendance

ABSTRACT

Unearned premium, or more particularly the risk associated to it, has only recently received regulatory attention. Unearned losses occur after the evaluation date for policies written before the evaluation date. Given that an inadequate acquisition pattern of premium and approximate modelling of premium liability can lead to an inaccurate reserve around unearned premium risk, an individual nonhomogeneous loss model including cross-coverage dependence is proposed to provide an alternative method of evaluating this risk. Claim occurrence is analysed in terms of both claim seasonality and multiple coverage frequency. Homogeneous and heterogeneous distributions are fitted to marginals. Copulas are fitted to pairs of coverages using rank-based methods and a tail function. This approach is used on a recent Ontario auto database.

Key words: Unearned premium risk, Loss reserving, Predictive Modelling, Dependence

INTRODUCTION

À l'opposé de la plupart des entreprises, les assureurs font face à ce qu'on appelle un cycle de production inversé, c'est-à-dire qu'ils ne connaissent pas le véritable coût d'un contrat au moment de sa souscription. En effet, lorsqu'ils reçoivent la prime souscrite d'un assuré, il est possible que l'assuré ne fasse pas de réclamation, ou qu'il ait un sinistre générant des paiements sur une longue durée. Une des tâches principales d'un actuaire est de prévoir la valeur présente de tous les paiements futurs qu'un assureur pourrait verser pour indemniser ses assurés. Le montant correspondant à cette projection, qu'un assureur se doit de garder rapidement disponible, est la réserve actuarielle (Friedland, 2010).

Dans ce mémoire, je me concentre sur une composante particulière des réserves, soit les pertes liées à la prime non-acquise. Ce concept est introduit en détail au chapitre 3. La littérature sur ce sujet étant plutôt limitée, je propose une nouvelle approche pour l'évaluation des pertes du non-acquis. Je présente une modélisation individuelle des pertes prenant en compte la dépendance entre les couvertures afin de modéliser le risque lié à la prime non-acquise.

L'idée derrière l'utilisation d'un modèle individuel vient d'un simple exercice de réflexion par rapport aux limites des méthodes de projection des réserves utilisées dans l'industrie présentement. Les assureurs utilisent généralement des données agrégées afin de faire des projections sur les développements futurs des pertes à l'aide de triangles de développement, qui seront définis au chapitre 1. Ces données de pertes agrégées regroupent autant les pertes associées à la prime acquise que celles associées à la prime non-acquise.

Ainsi, avec les méthodes utilisées dans l'industrie, pour déterminer les pertes du non-acquis, on doit utiliser un proxy pour estimer quelle portion des pertes appartient au non-acquis, en l'occurrence le ratio de la prime non-acquise sur la prime souscrite (totale). Cette méthode est très approximative et ne fonctionne que si le ratio est exact. Par contre, avec une base de données individuelle, il est possible de seulement retenir les pertes du non-acquis, ce qui permet de se débarasser entièrement du besoin d'un proxy et d'évaluer directement les pertes du non-acquis, permettant intuitivement une meilleure projection de ces pertes.

La structure du mémoire est la suivante. Dans le chapitre 1, j'explique les différentes notions mathématiques nécessaires à la compréhension du modèle proposé. Ensuite, les chapitres 2 à 5 ainsi que la conclusion sont tirés d'un article soumis pour publication et illustrent respectivement : le contexte de l'étude, la problématique du risque lié à la prime non-acquise, le modèle proposé et son application sur une base de données réelle. La conclusion résume les éléments importants ainsi que des pistes futures de recherche.

CHAPTER I

NOTIONS PRÉALABLES

Dans ce chapitre, qui ne fait pas partie de l'article soumis pour publication, on explique certaines notions préalables qui sont mentionnées ou utilisées dans l'article. Le but est ici d'introduire suffisamment ces notions pour qu'un lecteur puisse bien suivre la discussion subséquente.

1.1 Introduction

Tel que mentionné précédemment, on ne connaît pas le développement des pertes au moment de leur survenance. Afin de bien comprendre ce à quoi un assureur peut s'attendre, on illustre un développement de sinistre typique à la figure 1.1. Lorsqu'un assuré a un accident, il peut y avoir un délai entre la survenance du sinistre et la déclaration de celui-ci. Durant cette période on parle d'une perte encourue mais non déclarée (*Incurred But Not Reported*, IBNR). S'ensuivent alors généralement un ou des paiements avant que le dossier soit fermé, phase durant laquelle on parle d'un sinistre déclaré mais non réglé (*Reported But Not Settled*, RBNS). Un sinistre qui a été déclaré mais pour lequel aucun paiement n'a été fait est un sinistre déclaré mais non payé (*Reported But Not Paid*, RBNP). Noter qu'il arrive qu'un dossier soit fermé sans aucun paiement. Comme on peut le voir sur la figure 1.1, la fin d'un contrat ne signifie pas que le dossier est fermé. Il arrive qu'une perte engendre des paiements sur plusieurs années.

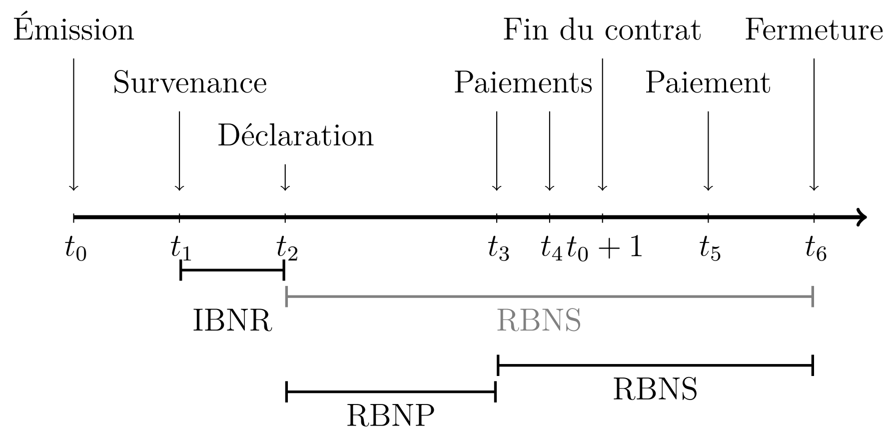


Figure 1.1: Développement typique d'un sinistre

Une autre visualisation possible du développement de sinistres lorsque ceux-ci surviennent est un diagramme de Lexis, comme on peut le voir à la figure 1.2, tirée de (Abdallah, 2016). L'axe vertical représente le temps écoulé depuis l'accident initial, soit les différents points $(\cdot)$ sur l'axe du temps calendaire. L'accident est déclaré au temps noté $(+)$ et fermé au temps noté $(\times)$. La barre verticale rouge représente la date d'évaluation, où l'assureur calcule la réserve (provision) nécessaire au paiement de toutes les réclamations restantes pour les contrats en vigueur, qu'elles soient rapportées ou non.

1.2 Réserve de sinistres

1.2.1 Triangles de développement

La façon classique d'organiser les données de développement de sinistre est dans un triangle de développement (*run-off triangle*). Celui-ci doit son nom à sa forme ; on y regroupe les pertes payées selon l'année d'accident et le laps de temps depuis

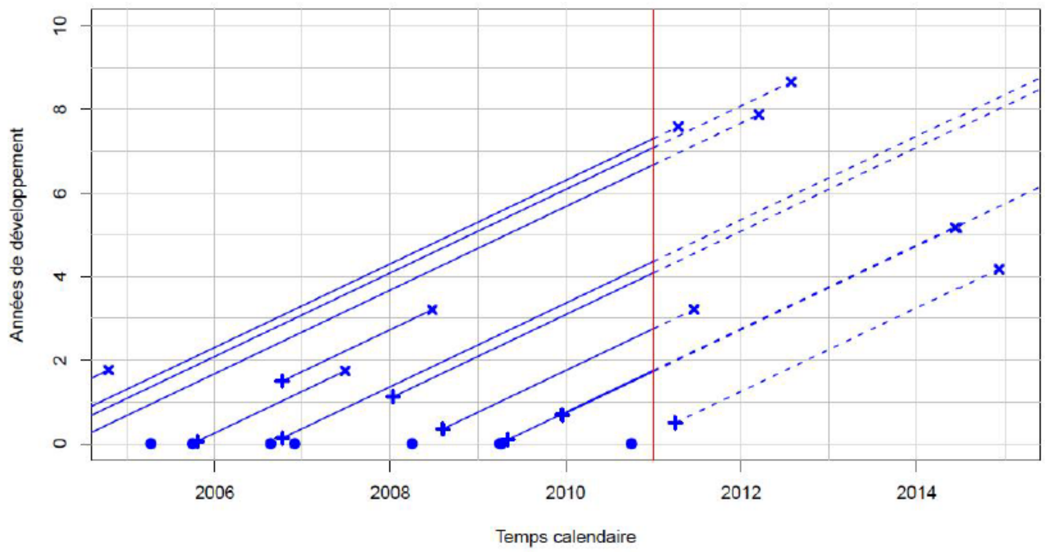


Figure 1.2: Diagramme de Lexis du développement de sinistres

l'accident, appelé période de développement, créant ainsi effectivement une forme de triangle.

Afin de comprendre les triangles, il suffit de considérer l'exemple du tableau 1.1.

Table 1.1: Exemple de triangle de développement

	Période de développement					
Année d'accident	**1**	**2**	**3**	**4**	**5**	**6**
1	P	P	P	P	P	O
2	P	P	P	P	O	O
3	P	P	P	O	O	O
4	P	P	O	O	O	O
5	P	O	O	O	O	O

(P) pour payé, (O) pour ouvert

Nous noterons habituellement par i les années d'accident, qui pour cet exemple vont de 1 à 5, et par j les périodes de développement, avec I l'année d'accident maximale, soit 5, et J la période de développement maximale. Supposons que nous sommes à la fin de l'année 5, et que les pertes peuvent prendre jusqu'à 6 ans avant d'être complètement payées, donc avec $J = 6$. Pour l'année 1, la majorité des pertes seront payées (P), il restera seulement une période de développement où il peut y avoir du développement supplémentaire (O, de *Outstanding*). À la fin de l'année 5, il y a au plus 12 mois d'écoulés, donc seulement la première période de développement sera payée. Les triangles servent principalement à projeter deux choses : les sinistres déclarés mais non réglés (RBNS, ou plus communément *case outstanding*), et les sinistres encourus mais non rapportés (IBNR), ou visuellement la portion dénotée O dans le triangle.

1.2.2 Méthodes déterministes

Méthode Chain Ladder

Plusieurs techniques existent pour utiliser les triangles afin de projeter le développement et amener les pertes à l'ultime, c'est-à-dire au stade où il n'y a plus de paiements supplémentaires. Une des techniques les plus connues est sans doute la méthode Chain Ladder, ou méthode de développement (Friedland, 2010). Celle-ci nécessite le calcul de facteurs de développement afin de cumuler les pertes à l'ultime.

Soient $Y_{i,j}$ les pertes pour la cellule (i,j) d'un triangle de développement, avec i et j définis comme précédemment, où $1 \leq i \leq I$ et $1 \leq j \leq J \leq I$. On considère les pertes cumulatives $C_{i,k}$ telles que

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

On suppose qu'il existe des facteurs de développement λ_j tels que

$$C_{i,j+1} = C_{i,j}\lambda_j.$$

On peut dès lors estimer ces facteurs de développement avec les pertes observées pour chaque année d'accident, de sorte que

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{J-j} C_{i,j+1}}{\sum_{i=1}^{J-j} C_{i,j}} \text{ , pour } j = 1, \ldots, J-1,$$

et ainsi projeter les pertes à l'ultime en multipliant la dernière diagonale payée dans le triangle par les facteurs trouvés, c'est-à-dire

$$\hat{C}_{i,J} = \left(\hat{\lambda}_{J-i+1} \times \cdots \times \hat{\lambda}_{J-1}\right) C_{i,J-1+1}.$$

L'estimation du montant de réserve $\hat{R}$ est alors

$$\hat{R} = \sum_{i=1}^{I} \hat{C}_{i,J} - C_{i,J-i+1}.$$

Autres méthodes

Plusieurs méthodes dérivées de la méthode Chain Ladder sont utilisées en industrie pour capturer le besoin de provisionnement en assurance. Bornhuetter & Ferguson (1972) proposent une approche qui combine la méthode Chain Ladder ainsi qu'une estimation externe basée sur le ratio $\frac{\text{pertes à l'ultime}}{\text{prime}}$ pour prendre en compte le fait que les années d'accident plus récentes n'ont pas autant d'information que les années plus anciennes, soit

$$\hat{C}_{i,J}^{BF} = C_{i,J-i} + \left(1 - \frac{1}{\prod_{j=J-i}^{J-1} \hat{f}_j}\right) \hat{\mu}_i.$$

Une autre méthode utilisée fréquemment est la méthode Cape Cod. Celle-ci ressemble à la méthode de Bornhuetter-Ferguson, à la différence près qu'elle utilise de l'information extérieure au triangle pour projeter les pertes à l'ultime. Une liste plus complète des différents modèles est disponible dans (Friedland, 2010).

1.2.3 Méthodes stochastiques

Modèle de Mack

Le modèle de Mack (1993) est une extension de la méthode Chain Ladder introduisant une notion stochastique aux projections des coûts à l'ultime.

Définition 1.2.1. Le modèle de Mack utilise trois hypothèses :
(H1) $\mathbb{E}[C_{i,j+1}|C_{i,1},\ldots,C_{i,j}] = C_{i,j}\lambda_j$, $1 \leq i \leq I$, $1 \leq j \leq I-1$;
(H2) $\{C_{i,1},\ldots,C_{i,I-1}\} \perp\!\!\!\perp \{C_{k,1},\ldots,C_{k,I-1}\}$, où $i \neq k$;
(H3) $\mathrm{Var}[C_{i,j+1}|C_{i,1},\ldots,C_{i,j}] = C_{i,j}\sigma_j^2$, $1 \leq i \leq I$, $1 \leq j \leq I-1$.

Si on définit la réserve pour l'année i comme $R_i = C_{i,J} - C_{i,J+1-i}$, la réserve totale est alors

$$\hat{R} = \sum_{i=1}^{I} \hat{R}_i = \sum_{i=1}^{I} \left(\hat{C}_{i,J} - C_{i,J-i+1}\right).$$

Il peut être démontré assez facilement que cet estimateur est sans biais, toutefois la preuve ne sera pas faite ici (voir (Wüthrich & Merz, 2008)). On cherche alors à trouver l'erreur quadratique moyenne (*mean squared error*, mse) de la réserve, soit

$$\begin{aligned} mse(\hat{R}_i) &= mse(\hat{C}_{i,J}) \\ &= \mathbb{E}\left[(C_{i,J} - \hat{C}_{i,J})^2\right] \\ &= \mathrm{Var}_1[C_{i,J}|D] + (\mathbb{E}[C_{i,J}|D] - \hat{C}_{i,J})^2, \end{aligned}$$

avec D toutes les données observées précédemment. L'intérêt de cette statistique définie dans (England & Verrall, 2002) est qu'elle permet de décomposer deux sources d'erreur, soit l'erreur stochastique ($\mathrm{Var}_1[C_{i,J}|D]$) provenant de l'incertitude du modèle, et l'erreur d'estimation ($(\mathbb{E}[C_{i,J}|D] - \hat{C}_{iJ})^2$) provenant de la variabilité des observations passées. En supposant que

$$\mathrm{Var}_1[C_{i,k+1}|C_{i,1},\cdots,C_{i,k}] = C_{i,k}\sigma_k^2, 1 \leq i \leq I, 1 \leq k \leq J-1,$$

on peut déduire un estimateur pour l'erreur quadratique de la réserve en supposant $I = J$, soit

$$\widehat{mse(\hat{R}_i)} = \hat{C}_{i,I}^2 \sum_{k=1+I-i}^{I-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left(\frac{1}{\hat{C}_{i,k}} + \frac{1}{\sum_{j=1}^{I-k} C_{j,k}} \right),$$

où

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

$$\hat{\sigma}_{J-1}^2 = \min\left(\frac{\hat{\sigma}_{J-2}^4}{\hat{\sigma}_{J-3}^2}, \min\left(\hat{\sigma}_{J-3}^2, \hat{\sigma}_{J-2}^2\right) \right).$$

Modèle de Poisson pour les réserves

Les modèles linéaires généralisés (*Generalised Linear Models*, GLM) sont un outil de modélisation d'une variable, disons Y, en fonction de variables explicatives $\mathbf{X}$ (ou variables indépendantes), qui forment un vecteur de dimension $(p \times 1)$ de caractéristiques influençant Y. La différence entre un GLM et une simple régression linéaire du type $Y = a + \mathbf{b}\mathbf{X} + \epsilon$, avec $\epsilon \sim N(0{,}1)$ un facteur d'erreur stochastique, est qu'on utilise une fonction de lien g telle que

$$\mathbb{E}[Y] = \mu = g^{-1}(\mathbf{X}^T \boldsymbol{\beta}),$$

où $\boldsymbol{\beta}$ est un vecteur de paramètres de dimension $(p \times 1)$. Noter qu'afin d'avoir une fonction de lien g, la distribution de Y doit appartenir à la famille exponentielle linéaire, qui est définie en annexe.

Il est possible de modéliser les triangles de pertes avec des GLMs. Dans un triangle incrémental, les réclamations $Y_{i,j}$ peuvent être modélisées selon l'année d'accident et la période de développement, de sorte que

$$\mathbb{E}[Y_{i,j}] = c_i e^{\gamma + \alpha_i + \delta_j},$$

où c_i est une mesure d'exposition, en l'occurrence la prime acquise pour chaque année d'accident (AA), γ est un facteur à l'origine, α_i est le facteur d'ajustement pour l'année d'accident, et δ_j le facteur pour la période de développement. Étant donné le facteur à l'origine γ, $\alpha_1 = 0$ et $\delta_1 = 0$. La prime acquise totale pour une année $c_i = P_i$ se compose de la prime non-acquise de l'année précédente, P_{i-1}^{NA}, ainsi que de la prime acquise de l'année en cours, P_i^A, donc

$$P_i = P_{i-1}^{NA} + P_i^A.$$

La distinction entre prime acquise et non-acquise sera couverte plus en détail au chapitre 3. Il est connu que les pertes incrémentales sont équivalentes en projettant à l'aide de la méthode Chain Ladder ou en supposant un GLM si $Y \sim Po(\lambda_{ij})$ (Charpentier, 2014). Ainsi, dans un modèle Poisson, en utilisant la prime acquise P_i comme base avec

$$Y_{ij} \sim Po(\lambda_{ij}),$$

on obtient

$$\mathbb{E}[Y_{i,j}] = \mu_{i,j} = P_i e^{\gamma+\alpha_i+\delta_j}.$$

Étant donné que la distribution de Poisson appartient à la famille exponentielle linéaire, qui est définie en annexe, la variance d'une variable suivant une loi de Poisson peut être exprimée en fonction d'un paramètre de dispersion $\phi_{i,j}$ et d'une fonction de variance V telle que

$$\mathrm{Var}[Y_{i,j}] = \phi_{i,j} V(\mu_{i,j}).$$

Encore une fois, l'approche stochastique permet de décomposer l'erreur quadratique moyenne en erreur stochastique et en erreur d'estimation. En définissant un vecteur $\boldsymbol{\beta} = [\gamma, \boldsymbol{\alpha}, \boldsymbol{\delta}]$ et un vecteur $\mathbf{Z}_{i,j} = [1, 0, \ldots, 0, 1, 0, \ldots, 0, 1, 0, \ldots, 0]$, où $\alpha_i = 1$ et $\delta_j = 1$, on obtient alors

$$\widehat{mse(\hat{R})} = \sum_{i+j>J+1} \phi_{i,j} V(\hat{\mu}_{i,j}) + \sum_{i+j>J+1, m+n>J+1} \hat{\mu}_{i,j} \hat{\mu}_{m,n} Z_{i,j} \mathbf{I}(\hat{\boldsymbol{\beta}})^{-1} Z_{m,n},$$

avec $\mathbf{I}$ la matrice d'information de Fisher (Lehmann & Casella, 2006), définie en annexe.

Autres modèles

Le modèle de Poisson est un cas spécifique d'une classe de modèles appelés Tweedie, qui viennent d'un modèle collectif, représenté par l'équation tel que la survenance des pertes N suit une distribution Poisson avec $N \sim Po(\lambda)$ et la sévérité X suit une distribution Gamma avec $X \sim Gamma(\alpha,\theta)$, de sorte que

$$Y = \sum_{i=1}^{N} X_i.$$

Si on considère la forme générale de la famille exponentielle linéaire définie à l'Annexe A, soit

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

les modèles Tweedie se caractérisent par le fait que $a''(\theta) = \mu^p$, avec $p \in (1,2)$. Le cas Poisson est le cas où $p = 1$; on retrouve ainsi l'équidispersion avec $\mathrm{Var}[Y] = \phi\mu^p = \mu$ puisque pour une distribution Poisson ϕ=1. On parle d'un modèle de Poisson surdispersé lorsque $\phi > 1$. Le modèle gamma est l'autre extrême où $p = 2$. Plus généralement, en supposant $p = 0$, on retrouve une distribution normale, et avec $p = 3$ une distribution inverse Gaussienne.

Il existe d'autres modèles tels que le modèle de Wright (1990) qui ressemble au modèle Tweedie, le modèle Lognormale (Kremer, 1982), ou encore le modèle Binomiale Négative (Verrall, 2000). L'idée revient toujours à trouver la distribution qui représente le mieux la distribution réelle des pertes.

1.3 Dépendance

En assurances, il est fréquent d'observer une dépendance entre des pertes, par exemple en assurance automobile où une collision peut entraîner des dommages à deux véhicules ou plus et qu'il est raisonnable de s'attendre à un lien entre les dommages de chaque véhicule. Lorsqu'il y a présence de dépendance, il est alors souhaitable de déterminer la distribution jointe des n variables dépendantes,

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \Pr(X_1 \leq x_1,\cdots,X_n \leq x_n).$$

Cette distribution peut être représentée par une copule C telle que

$$C(u_1,\cdots,u_n) = \Pr(U_1 \leq u_1,\cdots,U_n \leq u_n),$$

où $(u_1,\ldots,u_n) \in [0,1]^n$. On peut alors se servir du théorème de Sklar (1959), qui dit que toute distribution multivariée peut être exprimée à l'aide d'une copule C et des distributions marginales qui la composent, soit

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = C(F_{X_1}(x_1),\cdots,F_{X_n}(x_n)).$$

1.3.1 Copules archimédiennes

Une des familles de copules utilisée la plus fréquemment est la famille des copules archimédiennes. Ces copules passent par un certain générateur de copule ψ de sorte que

$$C(F_{X_1}(x_1),\cdots,F_{X_n}(x_n)) = \psi^{-1}(\psi(F_{X_1}(x_1)) + \cdots + \psi(F_{X_n}(x_n))).$$

Dans le cas où des variables sont indépendantes, il est connu que la distribution multivariée est simplement le produit des distributions marginales. Ainsi, la copule d'indépendance a un générateur $\psi(t) = e^{-t}$, de sorte que

$$C(F_{X_1}(x_1),\cdots,F_{X_n}(x_n)) = e^{-(-\ln(F_{X_1}(x_1)-\cdots-\ln(F_{X_n}(x_n)))} = F_{X_1}(x_1)\times\cdots\times F_{X_n}(x_n).$$

Ensuite, trois autres copules appartenant à cette famille sont la copule de Clayton, Frank, et Gumbel (Hofert et al., 2012). Elles ont des caractéristiques qui leur sont propres qui permettent de capturer différents types de dépendance, et sont reliées au tau de Kendall (1948), défini en tant que

$$\tau = \frac{C - D}{C + D},$$

où C est le nombre de paires dont le rang est concordant et D est le nombre de paires discordantes.

La copule de Clayton est caractérisée par un générateur $\psi_\alpha(t) = \frac{1}{\alpha}(t^{-\alpha} - 1)$ et présente une dépendance plus forte dans la queue inférieure que supérieure. Ainsi,

$$C_\alpha(F_{X_1}(x_1), F_{X_2}(x_2)) = \max([F_{X_1}(x_1)^{-\alpha} + F_{X_2}(x_2)^{-\alpha} - 1], 0),$$

où $\alpha = 2\tau/(1-\tau)$.

La copule de Frank présente quant à elle une dépendance symétrique des queues et a un générateur $\psi_\alpha(t) = -\ln\left(\frac{\exp(-\alpha t)-1}{\exp(-\alpha)-1}\right)$ tel que

$$C_\alpha(F_{X_1}(x_1), F_{X_2}(x_2)) = -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha F_{X_1}(x_1)} - 1)(e^{-\alpha F_{X_2}(x_2)} - 1)}{e^{-\alpha} - 1}\right),$$

où $\tau = 1 + 4/\alpha(D_1(\alpha) - 1)$ avec

$$D_1(\alpha) = \frac{1}{\alpha}\int_0^\alpha \frac{t}{e^t - 1}dt.$$

Finalement, la copule de Gumbel permet d'avoir une dépendance plus forte dans la queue supérieure que dans la queue inférieure. Son générateur est $\psi = (-\ln t)^\alpha$, où $\alpha = 1/(1-\tau)$, et donc

$$C_\alpha(F_{X_1}(x_1), F_{X_2}(x_2)) = \exp\left(-\left[(-\ln F_{X_1}(x_1))^\alpha + (-\ln F_{X_2}(x_2))^\alpha\right]^{1/\alpha}\right).$$

Il existe plusieurs autres copules permettant de modéliser différentes structures de dépendance, pour une discussion plus complète voir (Joe, 2014).

CHAPTER II

INTRODUCTION

Non-life insurance companies face high volatility due to the nature of losses they must provide coverage for. Regulation thus requires insurers to maintain funds under solvency constraints to ensure that up to a certain risk level, insureds' claims will not suffer from an insurer's solvency issues. These funds form actuarial reserves, and their risk requires accurate and reliable actuarial models. As such, a significant portion of actuarial literature is focused on modelling reserves and ensuring optimal capital allocation.

There exist very specific guidelines to determine the capital requirement to be maintained by an insurer varying from one country and even one state/province to another. For Property and Casualty (P&C) insurers, also known as non-life insurers, US regulation as defined by the National Association of Insurance Commissioners uses risk-based capital requirements (see (Feldblum, 1996) for more information), while the Canadian requirement is set forth as the conditional tail expectation (CTE) at a 99% level for insurance risk (Office of the Superintendent of Financial Institutions, 2018a). Different guidelines exist elsewhere in the world. Insurance risk can generally be further broken down into four parts: capital required for unpaid claim liabilities, capital required for premium liabilities, margin required for reinsurance ceded to unregistered reinsurers and catastrophe reserves.

An insured may or may not incur a loss over the duration of their contract. This contract is written on a certain effective date, and obligates the insurer to cover for losses defined in the contract for its specified length, which is usually but not always one year. In the event of a loss, there are often delays between occurrence and reporting dates, and depending on the nature of the claim, it may lead to multiple payments before the file is closed. Actuaries must establish the amount of capital to be maintained as reserves on a specific date called the evaluation date.

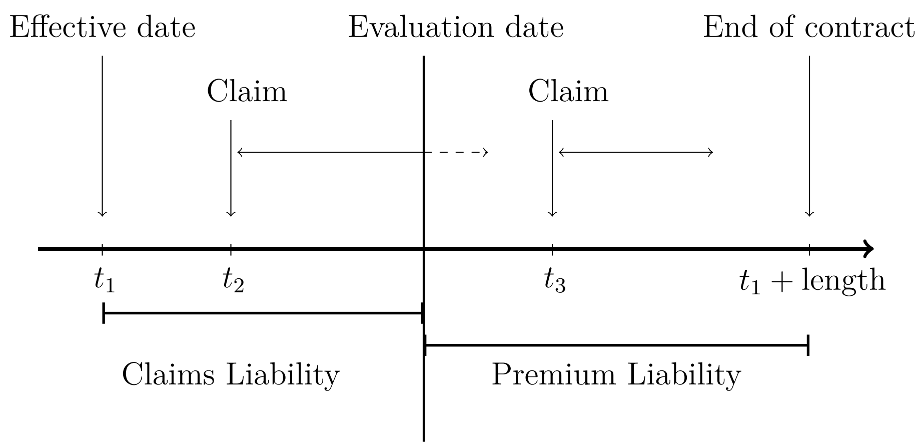


Figure 2.1: Split between claims liabilities and premium liabilities

Classical reserving methods usually focus on unpaid claim liabilities. This liability stems from the outstanding loss amount left to be paid from accidents incurred on or before the evaluation date. In figure 2.1, the outstanding loss from any claim occurring between the effective date and the evaluation date would fall into unpaid claim liabilities. This includes reported but not settled (RBNS) claims, and evaluation of incurred but not reported (IBNR) claims. There are many existing methods to evaluate such claims, see (Wüthrich & Merz, 2008) and (Friedland, 2010) for an extensive discussion of existing methods.

Premium liability is what drives the risk linked to unearned premium, stemming from potential future losses occurring after the evaluation date from contracts effective prior to the evaluation date. Figure 2.1 gives a visual representation of this separation. In other words, the risk linked to unearned premium is that the unearned premium reserve (UPR) will be insufficient to cover for premium liability.

This type of loss is the focus of our article. We use piece-wise constant risk exposure and cross-coverage dependence to evaluate unearned premium risk. Through this, we aim to identify the main drivers of risk relating to unearned premium and determine the role of cross-coverage dependence. Our hypothesis is that risk is driven by loss seasonality, loss distributions, the acquisition pattern of premium, contract subscription patterns, and cross-coverage dependence. In chapter 3, we explain what is unearned premium risk and what methods are available to build a model around it. In chapter 4, we present the models proposed to evaluate this risk by including dependence between coverages, and apply these models to actual data in chapter 5. Finally, we provide concluding remarks and potential adaptations in the conclusion.

CHAPTER III

UNEARNED PREMIUM RISK

To evaluate the risk linked to unearned premium, there are two main options: to use an aggregated framework or to use individual data. Aggregated methods essentially group all claims based on accident year and the time elapsed since the loss, generally referred to as development period, whereas individual methods keep all claim information on an individual level.

3.1 Traditional methods

Historically speaking, aggregated models have been favoured to calculate reserves due to their ease of use and non-intensive computing time. In this framework, one would use loss development triangles. This tool owes its name to its shape; to build a loss development triangle one must aggregate loss payments by accident year and development period, thus creating a triangle shape. Table 3.1 illustrates this aggregation. From paid amounts (P), one can project outstanding claim amounts (O) using methods such as the stochastic Chain-Ladder method (Mack, 1993) or Bornhuetter-Ferguson method (Bornhuetter & Ferguson, 1972).

Development triangles allowed for a convenient way of aggregating losses and calculate outstanding losses in non-computationally intensive ways. These methods are useful in valuing unpaid claim liabilities, but do not tell us much about the

Table 3.1: Loss development triangle with subsequent year

	Development period					
AY	**1**	**2**	**3**	**4**	**5**	**6**
2014	P	P	P	P	P	O
2015	P	P	P	P	O	O
2016	P	P	P	O	O	O
2017	P	P	O	O	O	O
2018	P	O	O	O	O	O
2019	F	F	F	F	F	F

Note: paid claims (P), outstanding claims (O), future claims (F)

last line of table 3.1, which consists of future claims (F). A part of these claims comes from contracts effective on the evaluation date: these make up most of premium liability, which is what we attempt to evaluate through our proposed model.

Loss analysis linked to unearned premium has only recently started receiving attention. As suggested by the Canadian Institute of Actuaries (2014), through triangles one would require a projection of the loss ratio for the following year and an estimation of the UPR to calculate premium liability. One can then evaluate ultimate losses for each year evaluated through methods such as (Mack, 1993), (Bornhuetter & Ferguson, 1972), and many others, then compare these losses to the premium written each year. This allows for determining the loss ratio to be multiplied by the UPR. The UPR is usually evaluated by supposing a uniform acquisition of premium, such that it is equal to the sum of written premium multiplied by the remaining fraction of contracts across all contracts.

The aggregated framework idea was reworked to determine the variance of such a method. Li (2010) proposes an extension of Mack's model using the next accident year's expected loss ratio to find an estimate of the prediction error of premium liability without assumptions concerning the underlying loss distribution. Priest (2012) instead proposes a model supposing that losses are driven by three factors, one specific to each accident year, one depending on both accident year and development period, and some random effect. Both Li and Priest rely on the hypothesis that the evaluation of premium liability is an extension of the evaluation of outstanding claims.

Some issues arise with these proposals. Barnett et al. (2005) demonstrates that the development method supposes independence between accident years, however Meyers (2013) suggests that this is in fact not the case as most contracts span more than one accident year. For example, a contract written on July 1^{st}, 2017 for a one year duration will end on June 30^{th}, 2018. Moreover, as most insureds keep the same insurer year after year, an insurer's portfolio will consist of mostly the same risks, which should induce between-year dependence.

Beyond between-year dependence, one should consider between-coverage dependence. Regulatory guidelines for Canadian insurers as provided by the Office of the Superintendent of Financial Institutions (2018b) require insurers to hold reserves for expected outstanding amount for each coverage. This is intuitively affected by dependence between coverages. Literature exists concerning dependence between lines of business when modelling loss reserve through aggregated loss triangles, however this is very limited when considering dependence between coverages, which is most likely due to only recently having the computational capacity to look at this level of loss.

Another problem is that current methods suppose losses occur uniformly throughout the year, which is not necessarily true as explained by Collins & Hu (2003). Factors affecting this assumption include but are not limited to:

- loss seasonality;
- loss trends (e.g. inflation);
- legal changes affecting claims;
- climate change and such.

3.2 Reserve for unearned premium risk

Any aggregated method is in fact highly dependent on the acquisition pattern of premium, given that losses linked to unearned premium are a subset of total losses, and that premium is the only way of measuring the size of this subset. Evaluation of premium liability would then entirely depend on having the right UPR. Bearing in mind that insurers usually use a uniform acquisition pattern for premium, the following example quickly illustrates how this first option can be problematic.

Exemple 3.2.1. Suppose two insureds, A and B, have contracts with an effective date of April 1^{st}. A has an equal chance of incurring a loss any time of the year. Meanwhile, B is twice as likely to have an accident between January and March than the rest of the year. Based on loss exposure, as of January 1^{st}, A will have 1/4 of her risk remaining. B will however have 2/5 of his risk remaining. Under uniform acquisition of premium, both insureds would have an unearned premium of $0.25P$, with P being their premium, but B's unearned premium should be $0.4P$, for a shortfall of $0.15P$.

With Example 3.2.1 in mind, it becomes clear that loss seasonality plays an important role in the evaluation of unearned premium risk. With the same seasonality, but different effective dates, we would however obtain a somewhat different scenario.

Exemple 3.2.2. Suppose two insureds, A and B, have contracts with an effective date of October 1^{st}, with the same seasonality as in Example 3.2.1. Based on loss exposure, as of January 1^{st}, A will have 3/4 of her risk remaining. B will however have 4/5 of his risk remaining. Under uniform acquisition of premium, both insureds would have an unearned premium of $0.75P$, with P being their premium, but B's unearned premium should be $0.80P$, for a shortfall of $0.05P$.

This example allows us to see that for the same loss seasonality, the effective date at which contracts are written affects the unearned premium risk, and with perfect recognition of seasonality in the acquisition of premium, then there would be no unearned premium risk. Evidently the behaviour of losses also needs to be taken into account to determine P, so these examples allow us to postulate that the main drivers of unearned premium risk are: loss distributions, loss seasonality, contract subscription patterns, and premium acquisition.

3.2.1 Individual loss reserving models

With advances in computational power, we have accessible information about each insured and claim, meaning that instead of only having data concerning the total amount paid for a given accident year and development period, we have data for every insured. This change in data availability allows us to use models such as (Antonio & Plat, 2014) and (Pigeon et al., 2013) for reserving purposes. An approach of this type allows for modelling losses linked to unearned premium directly from previous losses of the same type. This eliminates the issue of depending on premium to evaluate which portion of losses is associated to unearned premium.

We can take the individual model approach one step further by taking into account that insurance is generally separated into multiple coverages. For example, in car insurance in Ontario, accident benefits is a no-fault coverage that pays for an insured's injuries resulting from a car accident, while the third party liability coverage mainly pays for injuries caused by an insured driver to another person (such as the other driver). These coverages are usually valued separately in terms of costs and expected claims. It is however a fairly intuitive leap to see how if a car crash is severe enough to cause injuries, it is likely to do so for both drivers, which leads to dependence in both occurrence of losses between coverages and in loss amounts.

Using reserving approaches, dependence has mostly been studied from a line of business point of view. Côté et al. (2016) fit generalised linear models (GLM) to marginal lines of business and select copulas to capture dependence through rank-based methods. Cossette et al. (2013a) consider dependence between risks within an insurance portfolio and fit a Farlie-Gumbel-Morgenstern copula to model this dependence.

Although there exist many individual loss models in reserving, these models use past data to model claims that have already occurred; we are interested in future losses, which have not yet occurred, and so are more interested in a pricing approach.

3.2.2 Using a pricing approach

Under a pricing approach, Frees (2008) suggests that considering dependence within an insurance contract is important for pricing purposes. This dependence can take multiple forms, both in claim occurrence and in claim amounts. Abdallah et al. (2016) use the Sarmanov family of multivariate distributions to build a bivariate claim count model. Frees & Valdez (2008) instead model which coverages are affected when an accident occurs through a multinomial logit model, then use a t-copula to model dependence between losses. Recently, Pechon et al. (2019) have taken an approach of combining guarantees (coverages) and policyholders through a multivariate Poisson-mixture to capture dependence.

With this range of approaches to capturing dependence between coverages, we would then want to take into account dependence between coverages in an individual model with appropriate copulas, and model dependence between coverage occurrence. This idea stems from the intuition that if a claim involves multiple coverages, there is likely to be a link between those coverages, and it is worthwhile to investigate how this dependence behaves. We thus want a model capable of capturing loss seasonality, contract subscription patterns, while considering loss dependence between coverages.

CHAPTER IV

MODELLING APPROACH

We are interested in evaluating the risk linked to unearned premium, which is the potential excess of future losses over the unearned premium reserve. When a claim occurs, one or more coverages can be affected. To this end, as mentioned in the previous chapter, to model future losses we need to consider claim frequency, loss amount, as well as dependence between coverages.

In regard to frequency, we make a simplifying assumption for the model. A contract can in theory have more than one loss in a year. In actuarial databases, this is however rather infrequent. For example, first looking at a full year, in (Shi et al., 2018), table 1 presents some summary statistics of claim frequency, where we see that 0.35% of people will experience 2 or 3 losses in one year, or in table 2 of (Boucher et al., 2007) where 0.50% of contracts have more than one reported claim. These examples are for the year as a whole; future losses are only those losses that occur after the evaluation date. Multiple future losses in one year are thus very rare. As such, we suppose that a contract can only incur one future loss in a given year without losing much information through an indicator variable. In this way, we avoid having to consider multiple future losses for a single contract by supposing that if there is presence of loss, only one loss occurs. Another important assumption we will use is that we suppose that all contracts are written for a one year term, which is a standard assumption in actuarial models. Lastly, we assume that loss amount is independent from loss occurrence, meaning that the moment when a claim happens during the year has no incidence on loss amount, which is again a standard assumption. We are aware that this is a strong assumption with a potentially significant impact on the results. This could be explored in a further article by relaxing this limiting assumption.

In section 4.1, we define notation used throughout the rest of the paper. Section 4.2 presents the model in general form for a certain number C of coverages, as well as explaining how we consider dependence between coverages. Section 4.3 gives justification for coverage grouping in our model and section 4.4 explains the reserve for unearned premium risk.

4.1 Notation

The following are necessary definitions used in our model for some k^{th} contract. Figure 4.1 illustrates the various events related to unearned premium risk.

- N_k is a discrete random variable for the number of losses occurring after the evaluation date (future losses) that serves as a basis to form an indicator variable;
- J_k is an indicator random variable for the presence of a future loss as per our hypothesis such that

$$J_k = \begin{cases} 1 & \text{if } N_k > 0 \\ 0 & \text{otherwise;} \end{cases}$$

- $\mathbf{Y}_k = [Y_k^{(1)} \cdots Y_k^{(C)}]$ is a vector of continuous positive random variables for the paid future loss amount for each coverage c, $c = 1, \ldots, C$;

- $\mathbf{I}_k = [I_k^{(1)} \cdots I_k^{(C)}]$ is a vector of indicator variables for the presence of a loss for each coverage c, $c = 1, \ldots, C$;
- $t_k^{(E)}$ is the time since the last evaluation date on the effective date of a contract, where for example a contract written on November 1st with an evaluation date of December 31st would have $t_k^{(E)} = 0.8329$;
- t_k is the time of a claim;
- E_k is the exposure to future risk, which will be further defined in Section 4.3;
- P_k^{UE} is the unearned premium. Based on the acquisition pattern used by an insurer, this is not a random variable as the remaining portion of risk can easily be evaluated at the evaluation date. For example, if we assume a uniform acquisition of premium, then $P_k^{UE} = P_k(1 - t_k^{(E)})$, with P_k the written premium for a k^{th} contract.

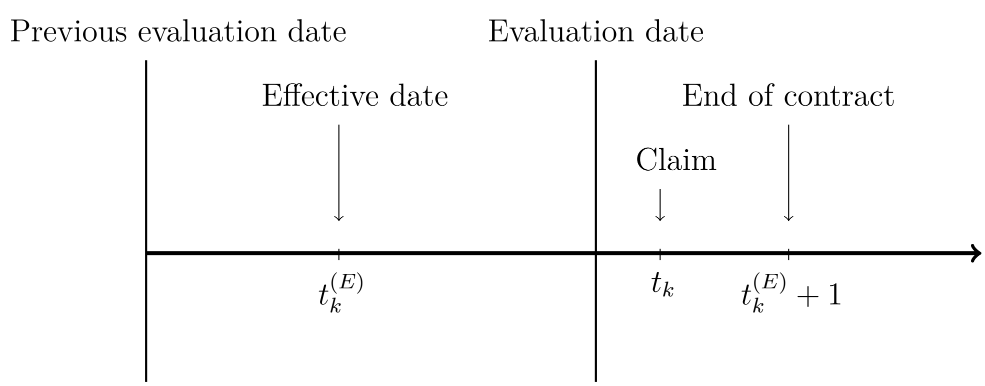


Figure 4.1: Illustration of time-related random variables for a one-year contract

4.2 General model in dimension C

Based on our previous definitions, the risk linked to unearned premium, which we call Z, depends on total future losses S^* and the total unearned premium across all n contracts, such that

$$Z = S^* - \sum_{k=1}^{n} P_k^{UE}. \tag{4.1}$$

As explained in subsection 4.1, the sum of unearned premium can be calculated at the evaluation date based on an insurer's acquisition pattern. In fact, for the rest of this paper, we will generally suppose that an insurer uses uniform acquisition of premium, such that P_k^{UE} is known at the evaluation date. The model is flexible enough to use other different acquisition patterns, but for illustration, and because it is almost always used in practice, uniform acquisition is supposed. Our interest is thus on modelling S^*. Recall our hypotheses, which are that

- A contract can only incur one future loss in a year;
- Contracts are written for a one-year period;
- Loss amount is independent from loss occurrence.

Grouping theses assumptions, we can then define future loss S_k in terms of $I_k^{(c)}$ and $Y_k^{(c)}$:

$$S_k = \begin{cases} \sum_{c=1}^{C} I_k^{(c)} Y_k^{(c)} & \text{if } J_k = 1, \\ 0 & \text{if } J_k = 0, \end{cases} \tag{4.2}$$

and subsequently

$$S^* = \sum_{k=1}^{n} S_k.$$

Supposing that the total unearned premium is known at the evaluation date, to model Z we thus need to model frequency N_k, coverage occurrence $\mathbf{I}_k$, and coverage loss amount $\mathbf{Y}_k$.

We include dependence in this model when considering $\mathbf{I}_k$ and $\mathbf{Y}_k$ and consider two cases: independence between coverages and presence of cross-coverage dependence. This choice of consideration follows from our research focus to determine the role of cross-coverage dependence.

For coverage occurrence $\mathbf{I}_k$, we can consider separate occurrence for each coverage or use joint occurrence probabilities. The problem with using an independent occurrence hypothesis is that claims involving more than two coverages seldom happen, and so independent occurrence can increase those probabilities. We choose to keep $\mathbf{I}_k$ fixed across both cases as the empirical distribution of coverage occurrence, we will elaborate on this idea in section 5.

Then, for both the independent and dependent scenarios, we model $\mathbf{Y}_k$ through copulas, where

$$F_{\mathbf{Y}_k}(\mathbf{y}) = C(F_{Y_k^{(1)}}(y^{(1)}), \cdots, F_{Y_k^{(C)}}(y^{(C)})).$$

in the perspective of considering between-coverage dependence in losses. See (Sklar, 1973) or a classical textbook on modelling dependence with copulas such as (Joe, 2014) for more details on copulas. In the independent case, we use the independence copula whereas when taking dependence between loss types we allow for other copulas.

4.3 Seasonality

As introduced in Example 3.2.1, risk is not always distributed uniformly throughout the year, despite the simplifying assumption used by most insurers, which is the basis of risk linked to unearned premium. In this paper, we use an idea proposed by Verrall & Wüthrich (2016), which is to suppose that the arrival rate at time s for the k^{th} contract $\lambda_0(s)$ is piece-wise constant. In other words, there exists a partition of time through the year $\mathcal{T} = \{t_m\}_{m=0,\ldots,M}$ such that we can define sets $A_m, m \in \{1, \ldots, M\}$ for which all events in A_m happen between t_{m-1} and t_m. In this way, for each A_m, we have a hazard rate $\lambda_0(s) = \lambda_m$, with $m = 1, \ldots, M$, where $\lambda_0(s)$ is a time-dependent hazard rate function.

Assume N_k follows a non-stationary Poisson process with multiplicative intensity function, similarly to (Zhao & Zhou, 2010). Let $A_m = [t_{m-1}, t_m)$. Then we can build an exposure function $E_0^{(C)}(t)$ s.t.

$$E_0^{(C)}(t) = \frac{\sum_{m=1}^{M} \frac{\int_{t_{m-1}}^{\min(t_m,t)} \mathbb{I}(t<t_m)\lambda_0(s)ds}{t_m - t_{m-1}}}{\sum_{m=1}^{M} \lambda_m}$$
$$= \frac{\sum_{m=1}^{M} \mathbb{I}(t < t_m)\lambda_m \frac{\min(t_m,t)-t_{m-1}}{t_m - t_{m-1}}}{\sum_{m=1}^{M} \lambda_m}, \tag{4.3}$$

where each set A_m has a cumulative hazard rate (or risk) of λ_m, with maximum risk for a full year being $\sum_{m=1}^{M} \lambda_m$. Using this exposure measure, we then have

$$\lambda_k(t|\mathbf{X}_k) = E_0^{(C)}(t)e^{\mathbf{X}_k^T \boldsymbol{\beta}},\ k = 1, \ldots, n, \tag{4.4}$$

with $\mathbf{X}_k$ a $(p \times 1)$ set of covariates and $\boldsymbol{\beta}$ a $(p \times 1)$ corresponding set of predictors obtained through maximum likelihood estimation (MLE), with p the number of covariates. Note that here we implicitly assume that the covariates have no impact on the exposition function.

In figure 4.1, consider the previous evaluation date as $t = 0$ and current evaluation date as $t = 1$. Under our previous assumption that a contract has a one year length, that contract's exposure to future risk is equivalent to $[1, t_k^{(E)}+1] \equiv [0, t_k^{(E)}]$, supposing that seasonality does not change from one year to the next. From there, using Equation (4.4) we trivially obtain

$$\mathbb{E}[N_k|\mathbf{X}_k] = E_0^{(C)}(t_k^{(E)})e^{\mathbf{X}_k^T \boldsymbol{\beta}}. \tag{4.5}$$

4.4 Reserve linked to unearned premium risk

Proposition 4.4.1. Let Z be a random variable as defined in Equation (4.1) for the risk linked to unearned premium. For a portfolio containing n contracts independent and C coverages, the expected value is given by

$$\mathbb{E}\left[Z|t_k^{(E)}\right] = \sum_{k=1}^{n} \left(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right) \sum_{c=1}^{C} \Pr(\mathbb{I}_k^{(c)} = 1)\mathbb{E}\left[Y_k^{(c)}\right] - \sum_{k=1}^{n} P_k^{(UE)}. \tag{4.6}$$

and the variance is given by

$$\begin{aligned}
\mathrm{Var}\Big[Z|t_k^{(E)}\Big] = \Big(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\Big) \Bigg[\sum_{c=1}^{C} \mathbb{E}\Big[(I_k^{(c)})^2\Big]\,\mathbb{E}\Big[(Y_k^{(c)})^2\Big] - \mathbb{E}\Big[I_k^{(c)}\Big]^2\,\mathbb{E}\Big[Y_k^{(c)}\Big]^2 \\
+2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\Big(\mathbb{E}\Big[Y_k^{(i)}Y_k^{(j)}\Big]\,\Pr(I_k^{(i)} = 1, I_k^{(j)} = 1) \\
-\mathbb{E}\Big[I_k^{(i)}\Big]\,\mathbb{E}\Big[I_k^{(j)}\Big]\,\mathbb{E}\Big[Y_k^{(i)}\Big]\,\mathbb{E}\Big[Y_k^{(j)}\Big]\Big)\Bigg] \\
+ \Big(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\Big)\, e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)} \Bigg[\sum_{c=1}^{C}\mathbb{E}\Big[I_k^{(c)}\Big]\,\mathbb{E}\Big[Y_k^{(c)}\Big]\Bigg]^2
\end{aligned} \tag{4.7}$$

Proof. The proof is in Appendix A. □

Moreover, given the probability density function $f_{T^{(E)}}(s)$ of $T^{(E)} \in [0,1]$, a continuous random variable for the effective date of contracts,

$$\mathbb{E}[Z] = \sum_{k=1}^{n}\Bigg[\sum_{c=1}^{C}\Pr(\mathbb{I}_k^{(c)} = 1)\mathbb{E}\Big[Y_k^{(c)}\Big]\int_0^1 \Big(1 - e^{-\lambda_k(s|\mathbf{X}_k)}\Big) f_{T^{(E)}}(s)ds\Bigg] - \sum_{k=1}^{n} P_k^{(UE)}.$$

Then, given the distribution of $T^{(E)}$,

$$\begin{aligned}
\mathrm{Var}[Z] = \sum_{k=1}^{n}\Bigg(\Bigg[\sum_{c=1}^{C} \mathbb{E}\Big[(I_k^{(c)})^2\Big]\,\mathbb{E}\Big[(Y_k^{(c)})^2\Big] - \mathbb{E}\Big[I_k^{(c)}\Big]^2\,\mathbb{E}\Big[Y_k^{(c)}\Big]^2 \\
+ 2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\Big(\mathbb{E}\Big[Y_k^{(i)}Y_k^{(j)}\Big]\,\Pr(I_k^{(i)} = 1, I_k^{(j)} = 1) \\
-\mathbb{E}\Big[I_k^{(i)}\Big]\,\mathbb{E}\Big[I_k^{(j)}\Big]\,\mathbb{E}\Big[Y_k^{(i)}\Big]\,\mathbb{E}\Big[Y_k^{(j)}\Big]\Big)\Bigg] \\
\times \int_0^1 \Big(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\Big) f_{T^{(E)}}(s)ds \\
+ \Bigg[\sum_{c=1}^{C}\mathbb{E}\Big[I_k^{(c)}\Big]\,\mathbb{E}\Big[Y_k^{(c)}\Big]\Bigg]^2 \int_0^1 \Big(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\Big)\, e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)} f_{T^{(E)}}(s)ds\Bigg).
\end{aligned}$$

Knowing the expected value and variance give us some information about the distribution of Z, however as suggested by Kaye (2005), there are multiple ways of measuring risk and allocating capital in general insurance with different risk measures for which we need to know the full distribution of Z. For example, Cossette et al. (2013b) give possible applications of the Value-at-Risk as the most widely used risk measure in both insurance and finance for capital allocation. As such, we can define the reserve, or allocated capital, R for Z as

$$R = \rho_\alpha(Z),$$

where $\rho_\alpha : \mathcal{L} \to \mathbb{R}$ is a risk measure based on the distribution of $Z \in \mathcal{L}$ such as the Value-at-Risk or the conditional tail expectation at a certain risk level α.

4.4.1 Algorithm to find the reserve for unearned premium risk

The algorithm to simulate the reserve for unearned premium risk is thus as follows:

1. For all contracts, generate a realisation of J_k through N_k and the contract's exposure $E_0^{(C)}(t_k)$, where $J_k = 1$ if $N_k > 0$ and 0 otherwise.
2. If $J_k = 1$, then there is a loss. Generate a realisation of $\mathbf{I}_k$ from the possible loss scenarios based on empirical observations.
3. Generate a realisation of losses $y^{(c)} \sim Y^{(c)}$ for affected coverages. In the independent case, use the independence copula; in the dependent case, use an appropriate copula to capture dependence.
4. Sum losses across all coverages.
5. Calculate the reserve for unearned premium risk as the sum of losses less the total unearned premium.
6. Repeat this procedure a large number of times.
7. Use these results to obtain the predictive distribution of Z.
8. Use appropriate risk measures to determine the reserve for unearned premium risk.

CHAPTER V

ANALYSIS

5.1 Data

We analyse data from a Canadian insurer in Ontario consisting of 132,093 auto contracts with information concerning 45 different coverages with effective dates ranging from December 15, 2015 to December 31, 2018. We choose to focus on five main coverages (Financial Services Commission of Ontario, 2016):

- Accident Benefits (AB), which is a no-fault coverage for benefits that the driver or another insured person may receive if injured or killed in an auto accident (income replacement, medical fees, rehabilitation, etc.);
- Collision (Coll), which covers for material damage to an insured's vehicle when involved in a collision with another object or the insured's vehicle rolls over and for which they are at least partially responsible for;
- Comprehensive (Comp), which covers for damage to an insured's vehicle from perils not linked to a collision with another vehicle, such as hail, theft, vandalism, or hitting a wild animal;
- Direct Compensation Property Damage (DCPD), which is direct compensation covering damage to an insured's vehicle when they are not at fault in

an accident, with at least one of the other drivers being insured. It can be seen as the flip side of Collision;

- Third Party Liability (TPL), which can cover for injuries caused to another person or damage to someone's property, as well as protect the insured in the event of lawsuits.

There are 10,423 claims for these five coverages, of which 87.4% involve only one coverage, 11.4% involve two coverages, and 1.2% involve three or more coverages. Due to the limited sample size and out of parsimony, we limit our analysis to two coverages at a time in our analysis of dependence.

The database is built on a transactional basis, where each intervention between the insured and the insurer creates a new line, even if there are no changes. Claim information consists of the date of occurrence, coverages affected, and incurred amount for each coverage.

Using this data, we attempt to address the following research questions:

1. How significant is unearned premium risk?

2. What factors actually drive unearned premium risk?

3. What is the impact of cross-coverage dependence on this risk?

5.2 Fitting the model

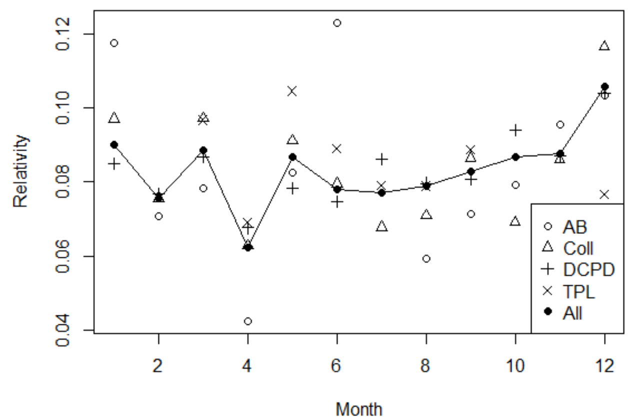
We use December 31$^{\text{st}}$ as our evaluation date. Due to our focus on the risk linked to unearned premium, this means we only have two years of losses to work with, per se losses occurring in 2017 from contracts effective in 2016, and losses occurring in 2018 from 2017 policies. We choose to fit our model on 2017 losses then observe its accuracy on 2018 losses. This is a limiting factor in our analysis due to the very small number of years available for analysis and the volatile nature of losses from one year to the next.

To get total future loss S^*, we use two models, $S^{(comp)}$ with dimension $C = 1$, and $S^{(crash)}$ with dimension $C = 4$, consisting of respectively only the comprehensive coverage, and the four other coverages, such that

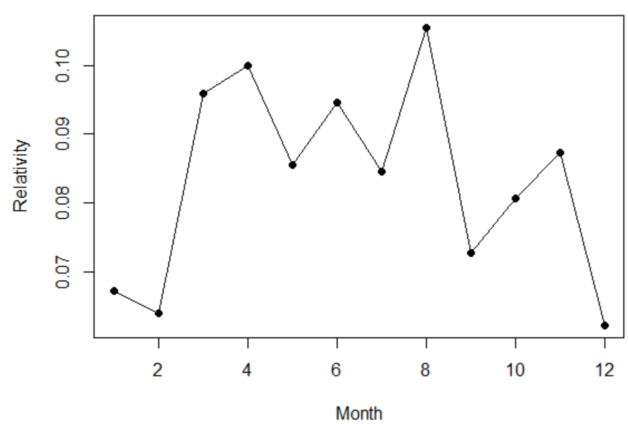
$$S_k^{(comp)} = I^{(comp)} * Y^{(comp)}, \text{ and}$$
$$S_k^{(crash)} = I^{(AB)} * Y^{(AB)} + I^{(coll)} * Y^{(coll)} + I^{(DCPD)} * Y^{(DCPD)} + I^{(TPL)} * Y^{(TPL)},$$

provided $J_k = 1$. The reason behind this choice is the distribution of $\lambda_0(t)$. Figure 5.1 shows the observed frequency by month for AB, Coll, DCPD and TPL (top) as compared to the frequency for Comp (bottom). The seasonality for the group of four coverages is very similar, while Comp has a very different seasonality. This is mostly due to the fact that Comp provides coverage for accidents that do not occur while driving (vandalism, hail, theft), whereas the other coverages require accidents occurring while on the road, justifying keeping it separate. Note however that we observe weak seasonality, which is a limiting factor in our analysis of its impact on unearned premium risk.

We explored the potential impact of stronger seasonality using a toy example of Gamma distributed losses with different levels of seasonality and inflation comparing an individual model to classical triangles. Scenarios included a 30% seasonality in the first three months of the year (GAA, 2015), the same seasonality in the last three months of the year, and different levels of inflation in premiums and losses. We found that under agregate methods (e.g. triangles), recognising seasonality in premium acquisition was essential to obtaining an accurate reserve. An individual model was more robust to seasonality, provided that subscription patterns and seasonality had the same distribution from year to year, but under changing subscription/seasonality an accurate forecast was necessary. This idea was however not applied directly to our model and would be worth exploring in a future analysis.



(a) Relativity - car crashes from Jan to Dec



(b) Relativity - comprehensive from Jan to Dec

Figure 5.1: Relativity of claim occurrence by month for claims linked to car crashes (top) and comprehensive (bottom)

For both models, we need $\lambda_0(t)$, N_k, $\mathbf{I}_k$ and $\mathbf{Y}_k$. As mentioned in subsection 4.2, we use two approaches to investigate the impact of cross-coverage dependence on unearned premium risk; one where losses are assumed independent between coverages and one where we consider dependence between coverages.

In both cases, for each model $\lambda_0(t)$ is fitted using the empirical distribution observed in figure 5.1. That is, to determine $E_0^{(C)}(t)$ to be used as an exposure measure, we use an empirical approach by using the hypothesis that the partitions in Equation (4.3) are monthly, meaning that we suppose the risk level is constant through January, but different from February, and so on. This allows us to find $\mathbb{E}[N_k]$ for each contract using Equation (4.5).

We use the empirical distribution of observed groupings of coverages to simulate the behaviour of coverage occurrence $\mathbf{I}_k$. In the four-coverage case, this does induce some dependence in our independent scenario, but as mentioned previously, the probability of multiple coverages occurring would be over-evaluated if we used independent occurrence of coverages. Using three coverages as an example, in reality we observed 121 out of 10 423 claims involving three or more coverages, or a 1.2% probability, but in an independent model this probability increases to 3.9%. Given that 121 data points is insufficient to model dependence, we restrict our analysis to two coverages per claim, and so imposing a joint empirical distribution allows us to prevent more than two coverages occurring simultaneously. Table 5.1 lists scenario probabilities across collision-linked events adjusted for the occurrence of only two coverages.

Table 5.1: Scenario probabilities of coverages for an accident

	AB	**Coll**	**DCPD**	**TPL**	**Probability**
1 coverage	1	0	0	0	2.77%
	0	1	0	0	35.30%
	0	0	1	0	40.78%
	0	0	0	1	1.29%
2 coverages	1	1	0	0	3.54%
	1	0	1	0	7.23%
	1	0	0	1	1.89%
	0	1	1	0	3.01%
	0	1	0	1	3.27%
	0	0	1	1	0.94%

Note 1: Probabilities are adjusted to sum to 100%

Note 2: a 1 indicates presence in an accident.

Next, in order to model $\mathbf{Y}_k$, we need to select marginal claim distributions for each coverage, and appropriate copulas based on the observed dependence structure. To evaluate the impact of loss distributions on unearned premium risk, we consider two cases for our marginal distributions. First, we select a homogeneous distribution for all losses of a particular coverage, and then we use a generalised linear models (GLM) approach to have a heterogeneous distribution of losses based on individual characteristics.

Although normally we would select among common loss distributions in actuarial literature (see (Klugman et al., 2012)) for potential marginals, we choose to restrict ourselves to the Gamma distribution, optimised using the *actuar* (Dutang et al., 2008) package in *R*. The motivation behind this choice lies in comparing similar distributions, in this case a Gamma distribution with another Gamma distribution, where in the heterogeneous model

$$\mathbb{E}\left[Y_k^{(c)}\right] = e^{\mathbf{D}_k^T \boldsymbol{\gamma}},\ k = 1, \ldots, \text{n} \tag{5.1}$$

with $\mathbf{D}_k$ a $(q \times 1)$ set of covariates which can be different from the one used for frequency, and $\boldsymbol{\gamma}_k$ a $(q \times 1)$ set of predictors optimised through maximum likelihood estimation. The parameters obtained for the homogeneous case are in Appendix C while the predictors obtained for the heterogeneous model are in Appendix D.

Then, to determine which pairs of coverages may have dependent loss amounts, we look at Kendall's tau and Spearman's rho (see (Kendall, 1948)), listed in table 5.2, and calculated with the help of the *VGAM* (Yee et al., 2010) package in *R*. We determine that there is weak dependence between Accident Benefits and DCPD, tail dependence between Accident Benefits and Liability, and strong dependence between Collision and DCPD, as we can observe in figures 5.2 to 5.4. The line observed near 0.44 for accident benefits in figure 5.2 stems from policy limits and is not abnormal.

Table 5.2: Kendall's Tau and Spearman's Rho for potentially dependent coverages

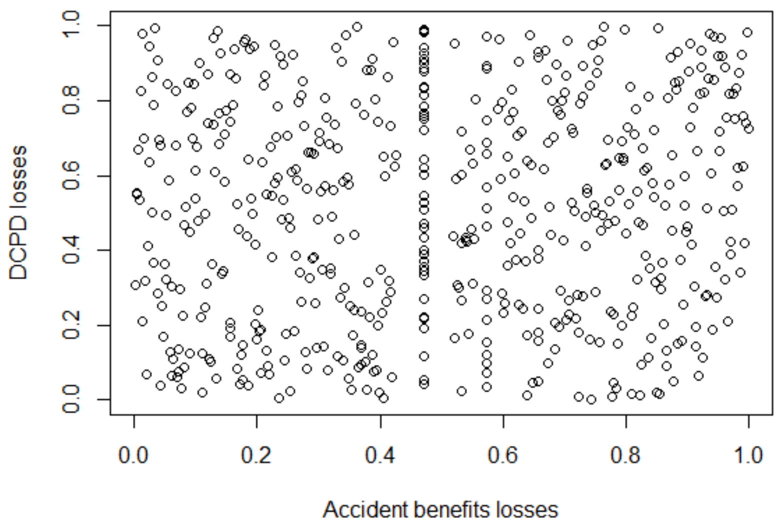
Coverages	**Kendall's Tau**	**Spearman's Rho**
AB, Coll	0.011	0.077
AB, DCPD	**0.053**	**0.097**
AB, Liab	**0.280**	**0.397**
Coll, DCPD	**0.429**	**0.435**
Coll, Liab	0.027	0.043
DCPD, Liab	-0.005	-0.000

Note: lines in bold indicate significant correlation

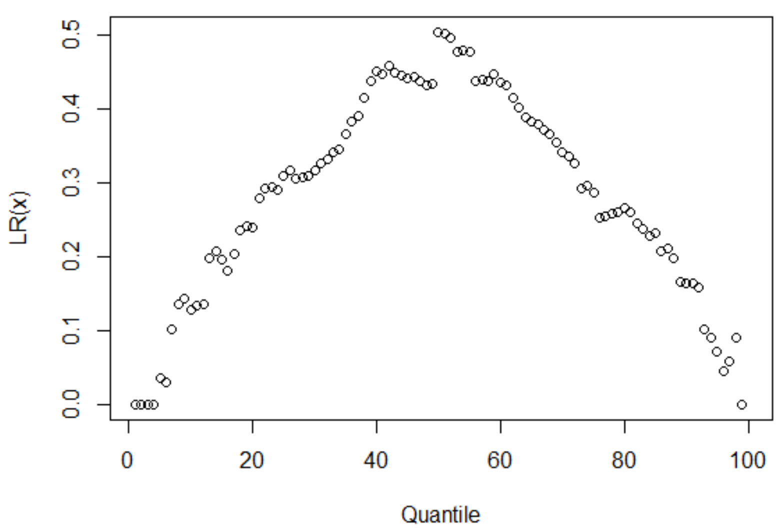
For copulas, we use rank plots and a left-right tail function defined as

$$LR(z) = \begin{cases} \Pr(F_X(X) < z | F_Y(Y) < z), & \text{if } 0 \leq z < 0.5 \\ \Pr(F_X(X) > z | F_Y(Y) > z), & \text{if } 0.5 \leq z < 1, \end{cases}$$

which allows for comparing the curve obtained through the function with a theoretical curve and choosing the closest fit. More information about this function can be found in (Venter, 2002) and (Boucher et al., 2008). Supporting theoretical graphics corresponding to the fitted copulas are in Appendix F.

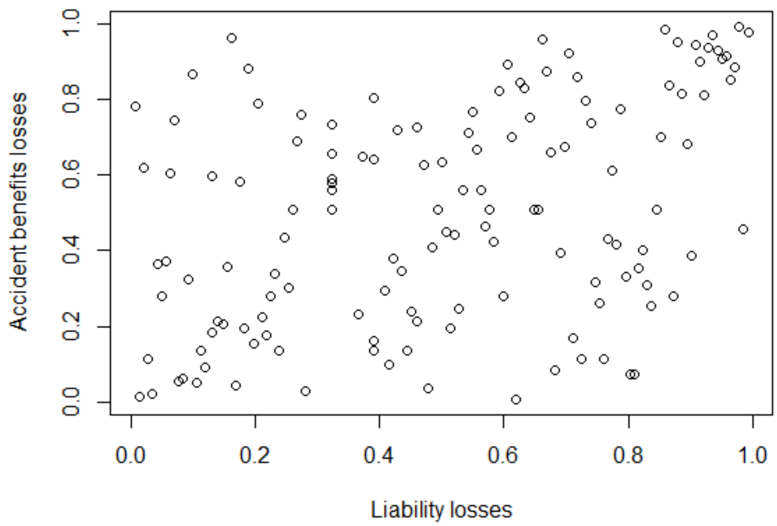


(a) Rank plot AB-DCPD

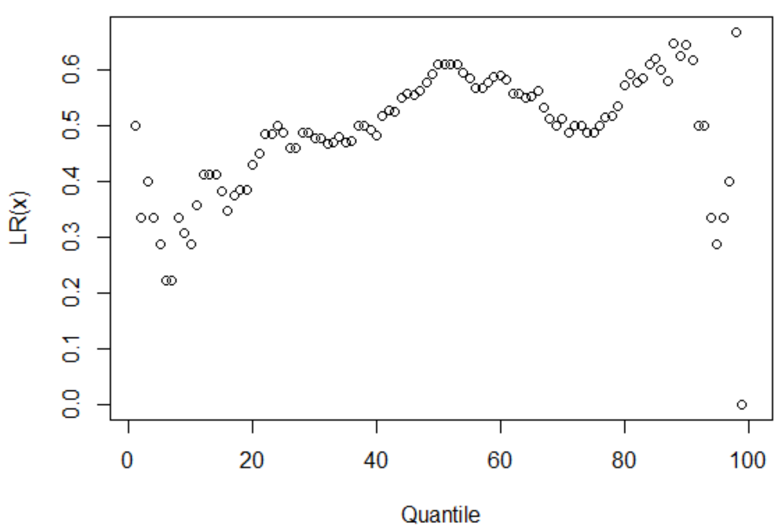


(b) Left-right tail function AB-DCPD

Figure 5.2: Rank plot (top) and left-right tail function (bottom) for Accident Benefits and DCPD



(a) Rank plot AB-TPL



(b) Left-right tail function AB-TPL

Figure 5.3: Rank plot (top) and left-right tail function (bottom) for Accident Benefits and Third Party Liability

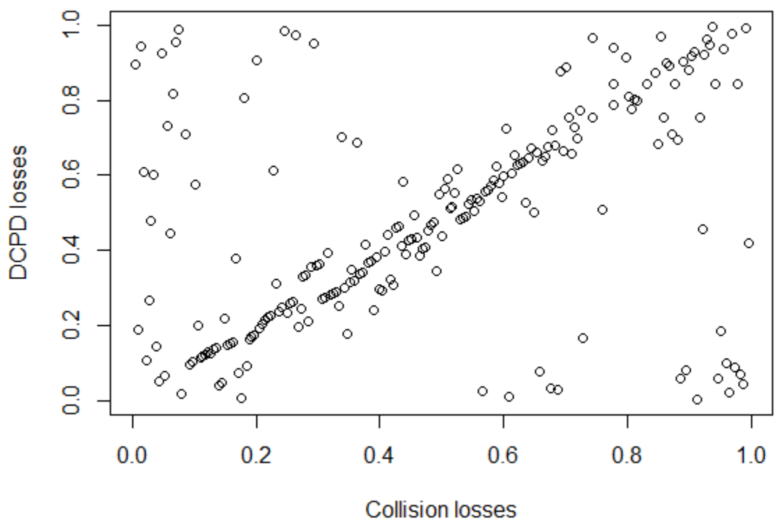


Figure 5.4: Rank plot for Collision and DCPD

These figures allow us to select a Frank copula between AB and DCPD, a Gumbel copula between AB and Liability, and we happen to know the exact link between Collision and DCPD. These coverages are in fact two sides of the same coin, where claim cost is split based on percentage of responsibility. We thus model this link through observed frequency of percentage breakdown between 0/100, 25/75, 40/60, and 50/50, with proportions found in table 5.3. Note that in 100% at fault or not at fault cases, the accident will be covered fully respectively by Coll (at fault) or DCPD (not at fault).

Table 5.3: Collision/DCPD percentage split

	Coll/DCPD split						
	0/100	**25/75**	**40/60**	**50/50**	**60/40**	**75/25**	**100/0**
Proportion	9.25%	3.96%	7.49%	65.20%	4.85%	1.76%	7.49%

Note: 0/100 or 100/0 implies respectively fully collision or DCPD.

5.3 Results

With our previous selections, we can obtain reserve amounts for the unearned premium risk, assuming the insurer uses uniform acquisition of premium. We run 20,000 simulations following the algorithm described in Section 4.4 with the *copula* package in R, allowing us to obtain an unearned premium risk distribution and evaluate different risk measures. The choice of 20,000 is motivated by a balance between sufficient data and computing time. Given that we have an Ontario database, we consider the Financial Services Commission of Ontario (FSCO) guidelines, which follow Canadian federal guidelines, and so we look at the 99$^{\text{th}}$ Conditional Tail Expected (CTE) as well as the 99.5$^{\text{th}}$ Value-at-Risk (VaR), presented in table 5.4 under both homogeneous and heterogeneous distributions. We present what would happen under independent occurrence (that is, independent $I_k^{(c)}$) in table 5.5.

Table 5.4: Reserve amounts using multiple loss approaches

Approach		Fitted set (000s)		Predictive set (000s)	
		Ind.	**Dep.**	**Ind.**	**Dep.**
Homogeneous	**Mean**	−2 485	−2 662*	−2 284	−2 513*
	VaR (99.5%)	2 417	2 175	2 949	2 699
	CTE (99%)	2 622	2 286	3 105	2 861
Heterogeneous	**Mean**	−4 009	−4 389*	−4 302	−4 770*
	VaR (99.5%)	3 167	2 901	3 189	2 835
	CTE (99%)	3 548	3 299	3 617	3 248
	St. Dev.	2 294		2 416	
Formula (Prop. 4.4.1)	**Mean**	−4 013	−4 285*	−4 304	−4 634*
	St. Dev.	2 292		2 463	
Observed		−2 952		−8 247	

* The mean is lower for the dependent case because of the added joint distribution of (Coll, DCPD).

Table 5.5: Reserve amounts in a purely independent heterogeneous model

	Fitted set (000s)	**Predictive set (000s)**
Mean	−3 990	−4 271
VaR (99.5%)	1 100	1 119
CTE (99%)	1 224	1 285

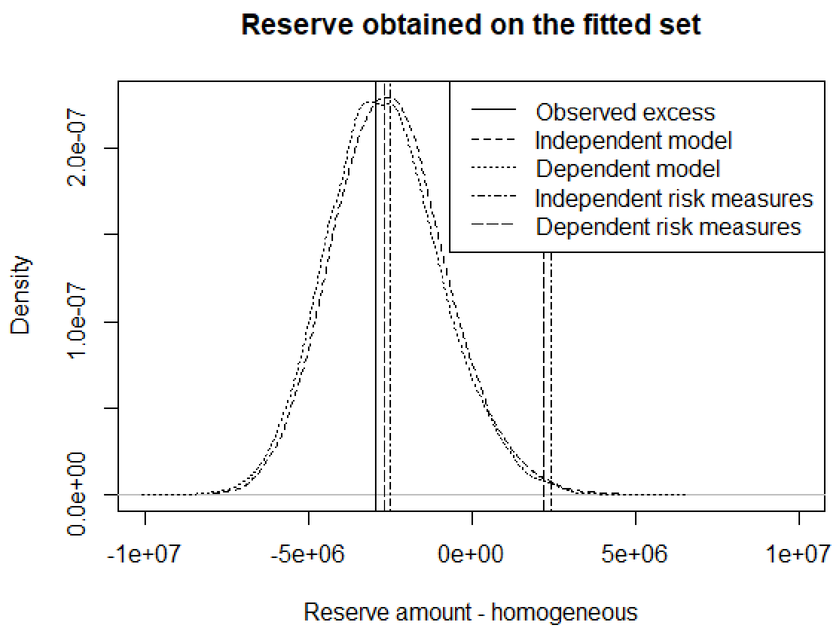
It is worth noting that written premium for the training and validation datasets are respectively 42 583 509 and 57 541 346 so the unearned premium risk is a relatively small percentage of written premium. This is potentially due to the weak seasonality observed in the data and should be investigated under stronger seasonality.

We note that the dependent model always returns a lower reserve than the heterogeneous case. While seemingly counter-intuitive due to the positive dependence, this is expected. In the particular case of the Coll-DCPD relationship, modelling claims involving both these coverages leads to lower amounts than modelling each coverage separately, then adding them up. AB and TPL claims can incur large amounts, but given our choice of the Gamma distribution which does not have a heavy tail, in the absence of large AB-Liab claims, the Coll-DCPD relationship leads to lower simulated reserve amounts.

Bearing in mind the research questions of drivers of unearned premium risk and impact of cross-coverage dependence, these results, despite not being directly intuitive, highlight the importance of properly recognising the relationship between coverages. While one might expect positive dependence to lead to higher loss reserves, it is important to take into account what actually occurs when multiple coverages are implied, as joint losses might behave differently than independent losses. Looking at means as a simple example of this idea, the mean of joint Coll-DCPD claims is much lower than the mean of collision claims added to the mean of DCPD claims, which creates bias in an independent model that can be corrected in an individual model taking dependence into account. We can therefore see that cross-coverage dependence plays a central role in proper evaluation of reserve amounts.

Next, we note that the means are negative; this indicates a surplus. This is to be expected, as an insurer does not only pay losses, but also pays expenses and generally keeps a certain profit margin. One would thus expect an average loss ratio around 70%. While on the fitted data both the homogeneous and heterogeneous models capture this well enough, on the predictive set we see that the homogeneous model has a distribution almost entirely to the right of the observed reserve amount while the heterogeneous model presents a better fit. We thus see that for the same loss distribution (Gamma in this case), using an individual model enables more predictive power, and so the choice of loss model is an important driver of risk.

Finally, we see by comparing tables 5.4 and 5.5 that using independent occurrence of coverages increases Z. This is expected, as explained in subsection 5.2, as independent occurrence leads to a disproportionate percentage of claims with multiple coverages, which in turn leads to potentially higher total loss. This highlights the importance of respecting the dependence structure in coverage occurrence and not limiting ourselves to dependence in loss distributions. So we can conclude that cross-coverage dependence, both in terms of occurrence and loss amount, is an important element of modelling unearned premium risk.



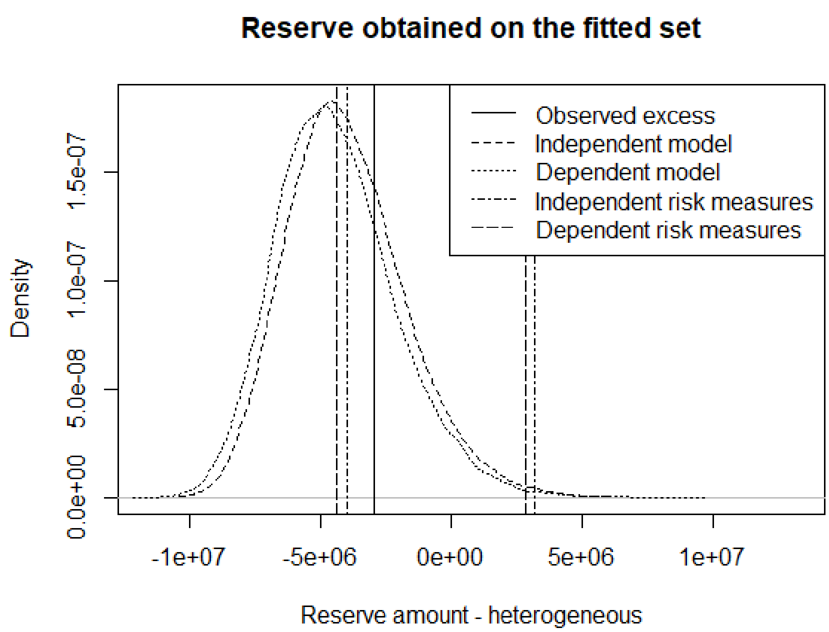


Figure 5.5: Reserve amounts for unearned premium risk for fitted data using a homogeneous model (top) and heterogeneous model (bottom)

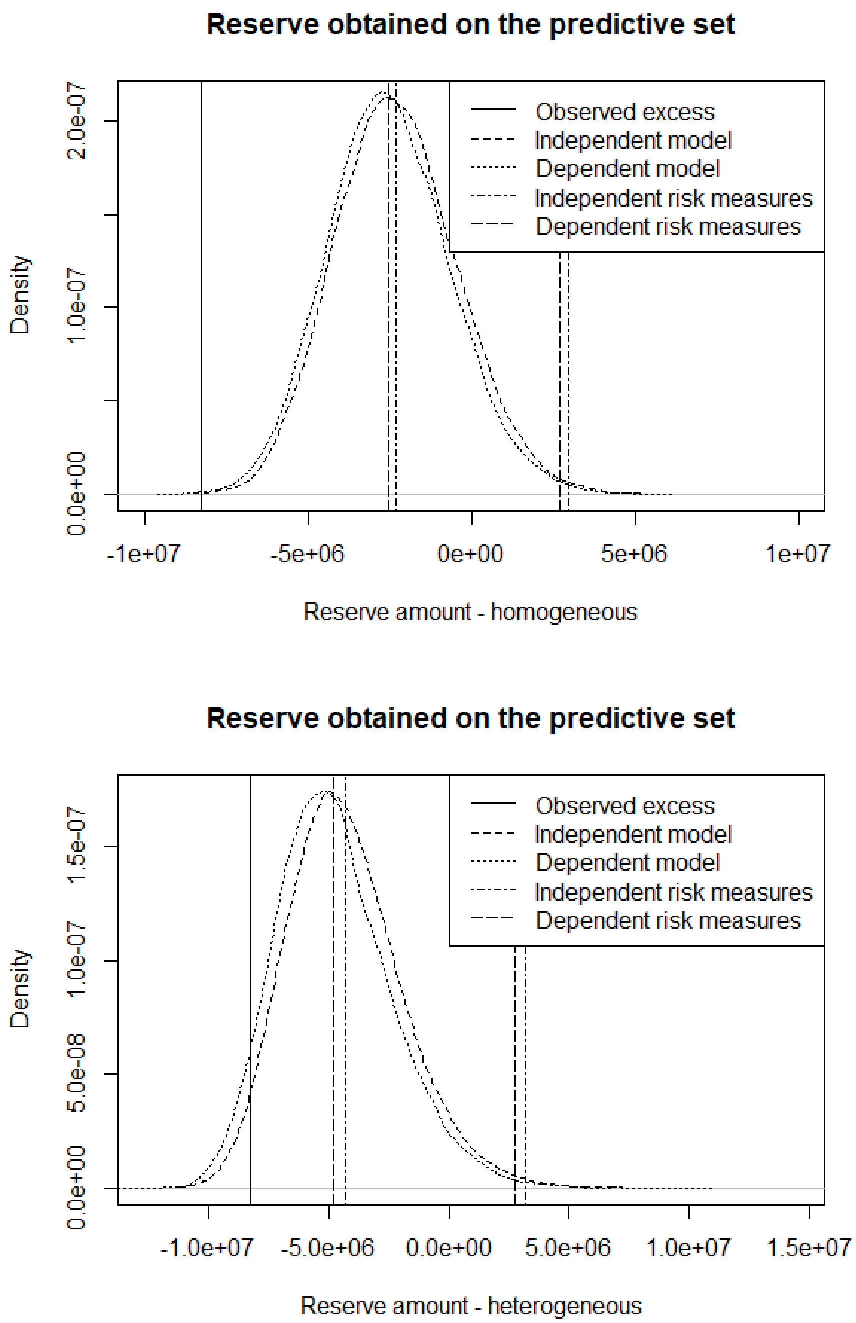


Figure 5.6: Reserve amounts for unearned premium risk for predictive set using a homogeneous model (top) and heterogeneous model (bottom)

In figures 5.5 and 5.6, we compare the distribution of unearned premium risk with and without dependence for both a homogeneous loss model and a heterogeneous one. We only present the models using seasonal exposure instead of uniform exposure because both curves are nearly identical; this is due to weak seasonality in our data and would likely create a larger difference with stronger seasonality. Our model implies that seasonality is in fact one of the main drivers of unearned premium risk, and so in a situation with strong seasonality, acknowledging seasonality in our exposure measure and premium acquisition would become important. The vertical full line is the actual observed amount of future losses minus the actual unearned premium reserve supposing a uniform acquisition pattern. The dotted vertical lines are the mean and 99.5% quantile for the independent case while the long dashes are the same measures for the dependent case. Notice that the dependent model is lower than the independent model in both the fitted dataset and the predictive dataset, as explained earlier in this subsection.

Relating this back to our supposition that loss distribution is one of the main drivers of unearned premium risk, we can deduce from figures 5.5 and 5.6 as well as table 5.4 that the homogeneous models vastly over-evaluate the necessary reserves while the heterogeneous models provide an accurate reserve, thus suggesting that using an individual GLM approach is better than using a homogeneous model.

5.3.1 Fully optimised homogeneous model

For the sake of completeness, we present the reserve distribution for the homogeneous model obtained by selecting from the Gamma, Pareto, lognormal, and Burr distributions. They are listed in table 5.6 along with their goodness of fit statistics calculated using the *actuar* (Dutang et al., 2008) package in *R*. Supporting graphs with cumulative distribution functions and Q-Q plots are in Appendix. In cases where the BIC criterion and the KS distance measure provided contradictory results, we favoured the KS distance measure, except for Collision+DCPD where we picked from the underlying distributions. The reason for favouring KS distance is that it is an actual distance measure and represented a proportionally larger difference than BIC. To avoid unnecessarily burdening the text, fitted parameters are in Appendix as well. We need 6 marginals despite having 5 coverages because in the dependent model, splitting Collision and DCPD by percentage of responsibility requires us to have a distribution for the full amount of loss, which we then split by percentage, again with the proportions based on empirical observation.

Table 5.6: Distribution selection for each marginal distribution

Coverage	**Marginal**	**BIC**	**KS**
Accident benefits	Gamma	4 964	0.2280
	Pareto	4 875	0.1574
	lognormal	4 867	0.1822
	Burr	4 847	0.1161
Collision	Gamma	11 855	0.0478
	Pareto	11 893	0.0940
	lognormal	11 874	0.0516
	Burr	11 836	0.0269
Comprehensive	Gamma	10 269	0.2220
	Pareto	10 075	0.2098
	lognormal	10 032	0.2462
	Burr	9 732	0.2897
DCPD	Gamma	15 236	0.0882
	Pareto	15 286	0.1283
	lognormal	15 150	0.0315
	Burr	15 148	0.0496
Collision + DCPD	Gamma	4 458	0.0711
	Pareto	4 471	0.1154
	lognormal	4 427	0.0616
	Burr	4 439	0.0605
TPL	Gamma	2 395	0.2461
	Pareto	2 385	0.1967
	lognormal	2 379	0.1895
	Burr	2 389	0.2019

Note: Selected marginals are in **bold**.

With these selections in mind, we thus obtain the reserve values in table 5.7.

We see that using heavier tailed distributions leads to a large over-evaluation of the expected reserve, and so emphasises the importance of selecting proper loss distributions that give the right amount of weight to extreme value scenarios. This would however offer the advantage of a more conservative estimate than the Gamma distribution, which is not heavy tailed. It may be worthwhile to investigate using different distributions around the tail values, where we would for example have $f(x) = \alpha f_{X_1}(x) + (1-\alpha) f_{X_2}(x)$, with $\alpha \in [0,1]$. We leave this question open for further exploration.

Table 5.7: Reserve amounts using homogeneous loss distributions

	Fitted set (000s)		**Predictive set (000s)**	
	Ind.	**Dep.**	**Ind.**	**Dep.**
Mean	1 689	2 222	2 473	2 317
VaR (99.5%)	10 872	11 153	12 240	12 148
CTE (99%)	11 226	11 538	12 683	12 356

CONCLUSION

In this paper, we analysed the risk linked to unearned premium in Property & Casualty insurance through a non-homogeneous Poisson process and an individual model including between-coverage dependence. That is, we used piece-wise constant seasonality of losses as an exposure base to build a generalised linear model for claim occurrence, and modelled loss amounts through copulas.

Recalling the main questions at the start of chapter 5, we can conclude that there are four main drivers to unearned premium risk. Given that exposure to risk in our model is driven by loss seasonality and the effective date of contracts, we see that the timing of losses during the year and when contracts are written, or subscription patterns, are two important drivers of risk.

Next, unearned premium risk being the risk that future losses exceed the unearned premium reserve, loss distribution and premium acquisition are necessarily other main drivers of risk. In terms of loss distribution, our data study clearly indicates that recognising cross-coverage dependence has an impact on projecting losses and so for a more accurate model, it is preferable to include dependence. In fact, we showed that some coverages, such as Collision and DCPD, have strong dependence and should not be considered independently. As such, current models for Claims Liability would also gain in using between-coverage dependence, allowing for a more realistic claims projection. Then, we suggested that the UPR can easily be established at the evaluation date. It is easy to see how recognising seasonality in the acquisition of premium would lead to a better cashflow matching between losses and premium, thus potentially decreasing unearned premium risk.

There are therefore four main drivers to unearned premium risk: seasonality, loss distributions, the acquisition pattern of premium, and the subscription pattern of insureds. Our two other questions are addressed within these four drivers of risk. All things considered, we can conclude that given the advances in computational power, it would be encouraged to stop working with loss development triangles and to use stronger statistical tools to evaluate reserves while including dependence.

Given the high variability of losses in P&C insurance, it would be interesting to have more accident years to work with, as we only had data to work with for two accident years. Having multiple years of data would also enable us to observe the impact of trends on unearned premium risk. Our database unfortunately had fairly weak seasonality and so we could not fully observe its impact on unearned premium risk. It would be interesting to see how risk shifts with stronger presence of seasonality, for example in a heavily snow-prone region where winter generally implies more accidents. Moreover, in the presence of strong seasonality one may suppose that a particular type of accident would be more prevalent (e.g. sliding on ice, storm surges, hail storms), leading to a relation between time of loss and type/amount of loss. It would thus also be interesting to relax the hypothesis of independence between time of loss and loss amount. We may also want to generalise our model to more than one claim, where we use N_k instead of transforming it into a binary variable.

Finally, the model was built around automobile insurance, but it follows that it could be applied to homeowners insurance, which faces high seasonality and evolving climate-related perils. It would thus be interesting to see how the model behaves when faced with this different type of insurance.

APPENDIX

Annexe A - Famille exponentielle linéaire

Afin d'avoir une fonction de lien, Y doit faire partie d'une famille de distribution appelée la famille exponentielle linéaire. Pour appartenir à cette famille, la distribution d'une variable doit suivre

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

avec θ et ϕ des paramètres et $c(y,\phi)$ et $a(\theta)$ des fonctions telles que $a(\theta)$ est convexe. θ et ϕ sont respectivement appelés les paramètres canoniques et de dispersion, et sont liés aux deux premiers moments de Y_i :

$$\begin{aligned} \mathbb{E}[Y_i] &= \mu_i = \exp(\mathbf{X}_i\boldsymbol{\beta}) = a'(\theta_i) \text{ et} \\ \mathrm{Var}[Y_i] &= \phi_i a''(\theta_i) = \phi_i V(\mu_i), \end{aligned}$$

où $a'(\theta_i)$ et $a''(\theta_i)$ sont les deux premières dérivées de $a(\theta)$ par rapport à θ, et V est une fonction de variance. Certaines des distributions les plus connues appartenant à cette famille sont les distributions normales, binomiales, Poisson et gamma.

Pour déterminer $a()$ et $c()$, il est généralement mieux de passer par

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

En fait avec cette forme il est possible de maximiser la logvraisemblance

$$\begin{aligned} l(\mathbf{y};\theta,\phi) &= \sum_{i=1}^{n}\left(\ln\left(c(y_i,\phi)\right) + \left(\frac{y_i\theta - a(\theta)}{\phi}\right)\right) \\ &= \left(\frac{n(\bar{y}\theta - a(\theta))}{\phi}\right) + \sum_{i=1}^{n}\ln\left(c(y_i,\phi)\right) \end{aligned}$$

de sorte que

$$\begin{aligned} \frac{dl}{d\theta} &= \left(\frac{n(\bar{y} - a'(\theta))}{\phi}\right) = 0 \\ &\Rightarrow a'(\theta) = \hat{\theta} = \bar{y} \end{aligned}$$

pour toute distribution faisant partie de la distribution exponentielle.

En répétant l'exercice mais en différentiant par rapport à β_i, on trouve que chaque β_i doit respecter la condition de premier ordre

$$\sum_{i=1}^{n}\frac{y_i - \mu_i}{g'(\mu_i)a''(\theta_i)} = 0,$$

où $g'(\mu_i) = \frac{dg(\mu_i)}{d\mu_i}$, avec $g(\mu_i)$ une fonction de lien telle que $g(\mu_i) = \boldsymbol{X}'\boldsymbol{\beta}$. Qui plus est, en prenant le 2^e dérivé en fonction de β, on obtient la matrice d'information de Fisher (Lehmann & Casella, 2006)

$$\mathbf{I}(\boldsymbol{\beta})_{t,j} = -\mathbb{E}\left[\left(\frac{\delta^2 l(\mathbf{Y};\boldsymbol{\beta},\phi)}{\delta\beta_t\delta\beta_j}\right)\middle|\boldsymbol{\beta}\right],$$

à partir de quoi on peut trouver sous certaines conditions

$$\mathbf{I}(\boldsymbol{\beta})_{t,j} = \sum_{i=1}^{n}\frac{1}{g'(\mu_i)^2\mathrm{Var}(Y_i)}x_{i,t}x_{i,j},$$

avec

$$\mathrm{Var}[Y] = \phi a''(\theta).$$

Ceci nous permet finalement de trouver la matrice de variance-covariance des estimateurs, soit

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{I}(\boldsymbol{\beta})^{-1}.$$

On en déduit donc par la version faible de la loi des grands nombres et le théorème de Slutsky (Van der Vaart, 2000) que

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0,I^{-1}(\theta_0)).$$

Finalement, pour certaines distributions le paramètre ϕ devient utile pour faire des tests statistiques. L'utilisation d'une méthode de maximum de vraisemblance ne permettant pas de dériver une formule explicite pour ϕ_i, on utilise plutôt la statistique du χ^2 de Pearson pour approximer un paramètre global,

$$\hat{\phi} = \left(\frac{1}{n-p}\right)\sum_{i=1}^{n}\frac{(y_i - \hat{\mu_i})^2}{V(\mu_i)},$$

avec n le nombre d'observations et p le nombre de paramètres estimés.

Appendix B - Proof of Proposition 2.1

We seek to evaluate the sum of losses linked to unearned premium. For a loss to fall into this category, it must arrive after the evaluation date but before the end of the contract. Recall that N_k is a discrete random variable for the number of future losses for the k^{th} contract, $k = 1, \ldots, n$. Then S_k the future loss across C coverages is

$$S_k = \begin{cases} \sum_{c=1}^{C} \mathbb{I}_k^{(c)} Y_k^{(c)} & \text{if } N_k > 0 \\ 0 & \text{if } N_k = 0. \end{cases}$$

It then follows that

$$\mathbb{E}\left[S_k | t_k^{(E)}\right] = \mathbb{E}\left[\mathbb{E}\left[S_k | N_k, t_k^{(E)}\right] | t_k^{(E)}\right] = \Pr(N_k > 0)\sum_{c=1}^{C}\Pr(\mathbb{I}_k^{(c)} = 1)\mathbb{E}\left[Y_k^{(c)}\right],$$

which makes sense intuitively: the expected future loss is the sum of weighted expected losses by coverage, weighted by the probability of a loss occurring. Now, recalling equations (4.4) and (4.5), we can rewrite the expected value as

$$\mathbb{E}\left[S_k | t_k^{(E)}\right] = \left(1 - e^{-\lambda_k(t_k^{(E)} | \mathbf{X}_k)}\right)\sum_{c=1}^{C}\Pr(\mathbb{I}_k^{(c)} = 1)\mathbb{E}\left[Y_k^{(c)}\right].$$

Straightforwardly, we obtain

$$\mathbb{E}\left[Z|t_k^{(E)}\right] = \sum_{k=1}^{n}\left(1-e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right)\sum_{c=1}^{C}\Pr(\mathbb{I}_k^{(c)}=1)\mathbb{E}\left[Y_k^{(c)}\right]-\sum_{k=1}^{n}P_k^{(UE)}.$$

The proof for the variance is similar to the one for the expected value. Consider the indicator function

$$J_k = \begin{cases} 1 & \text{if } N_k > 0 \\ 0 & \text{if } N_k = 0, \end{cases}$$

defined in subsection 4.1. We have

$$\begin{aligned}
\mathrm{Var}[S_k] =& \mathbb{E}[\mathrm{Var}[S_k|J_k]] + \mathrm{Var}[\mathbb{E}[S_k|J_k]] \\
\mathrm{Var}[S_k] =& \Pr(N_k>0)\mathrm{Var}\left[\mathbf{I}_k^T\mathbf{Y}_k\right] + \Pr(N_k>0)\Pr(N_k=0)\left[\mathbb{E}\left[\mathbf{I}_k^T\mathbf{Y}_k\right]\right]^2 \\
=& \Pr(N_k>0)\left[\sum_{c=1}^{C}\mathrm{Var}\left[I_k^{(c)}Y_k^{(c)}\right] + 2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\mathrm{Cov}\left[I_k^{(i)}Y_k^{(i)}, I_k^{(j)}Y_k^{(j)}\right]\right] \\
&+ \Pr(N_k>0)\Pr(N_k=0)\left[\sum_{c=1}^{C}\mathbb{E}\left[I_k^{(c)}\right]\mathbb{E}\left[Y_k^{(c)}\right]\right]^2 \\
=& \Pr(N_k>0)\left[\sum_{c=1}^{C}\mathbb{E}\left[(I_k^{(c)})^2\right]\mathbb{E}\left[(Y_k^{(c)})^2\right] - \mathbb{E}\left[I_k^{(c)}\right]^2\mathbb{E}\left[Y_k^{(c)}\right]^2\right. \\
&\left. +2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\mathbb{E}\left[I_k^{(i)}Y_k^{(i)}I_k^{(j)}Y_k^{(j)}\right] - \mathbb{E}\left[I_k^{(i)}Y_k^{(i)}\right]\mathbb{E}\left[I_k^{(j)}Y_k^{(j)}\right]\right] \\
&+ \Pr(N_k>0)\Pr(N_k=0)\left[\sum_{c=1}^{C}\mathbb{E}\left[I_k^{(c)}\right]\mathbb{E}\left[Y_k^{(c)}\right]\right]^2 \\
=& \left(1-e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right)\left[\sum_{c=1}^{C}\mathbb{E}\left[(I_k^{(c)})^2\right]\mathbb{E}\left[(Y_k^{(c)})^2\right] - \mathbb{E}\left[I_k^{(c)}\right]^2\mathbb{E}\left[Y_k^{(c)}\right]^2\right. \\
&+2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\left(\mathbb{E}\left[Y_k^{(i)}Y_k^{(j)}\right]\Pr(I_k^{(i)}=1, I_k^{(j)}=1)\right. \\
&\left.\left. -\mathbb{E}\left[I_K^{(i)}\right]\mathbb{E}\left[I_k^{(j)}\right]\mathbb{E}\left[Y_k^{(i)}\right]\mathbb{E}\left[Y_k^{(j)}\right]\right)\right] \\
&+ \left(1-e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right)e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\left[\sum_{c=1}^{C}\mathbb{E}\left[I_k^{(c)}\right]\mathbb{E}\left[Y_k^{(c)}\right]\right]^2
\end{aligned}$$

Appendix C - Homogeneous Gamma distribution parameters

Table 5.8: MLE parameters for all coverages for the Gamma distribution

	AB	**Coll**	**Comp**	**DCPD**	**Coll+DCPD**	**TPL**
Shape	0.50 (0.00)	1.40 (0.07)	0.47 (0.02)	1.41 (0.04)	1.41 (0.12)	0.34 (0.04)
Scale	63 747.38 (7 838.25)	4 460.40 (277.31)	4 254.52 (319.18)	6 043.20 (349.27)	4 838.24 (492.46)	203 320 (41 129.67)

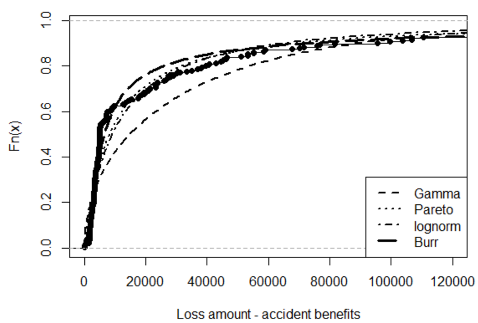
Appendix D - Heterogeneous Gamma distribution parameters

Table 5.9: GLM loss predictors obtained by MLE by coverage

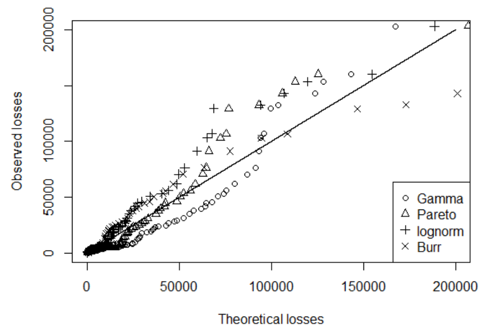
	AB		**Coll**		**Comp**		**DCPD**		**TPL**	
Dispersion	3.44	NA	0.76	NA	3.46	NA	0.96	NA	3.67	NA
Intercept	9.29	(0.59)***	8.35	(0.14)***	6.94	(0.28)***	8.38	(0.14)***	11.03	(1.29)***
Var1	0.26	(0.26)	0.11	(0.07)	0.44	(0.15)**	0.18	(0.07)*	NA	NA
Var2	0.07	(0.34)	0.14	0.10	0.06	(0.24)	-0.00	(0.09)	NA	NA
Var3	0.64	(0.27)*	-0.03	(0.07)	-0.05	(0.16)	-0.10	(0.07)	NA	NA
Var4	-0.56	(0.41)	-0.17	(0.11)	-0.43	(0.21)*	-0.31	(0.11)**	-0.75	0.60
Var5	-0.98	(0.44)*	-0.08	(0.12)	0.08	(0.28)	-0.17	(0.11)	-0.48	0.66
Var6	1.24	(0.85)	0.40	(0.17)*	0.32	(0.37)	0.35	(0.20)+	-0.66	(0.41)
Var7	0.44	(0.46)	-0.04	(0.13)	-0.20	(0.24)	0.08	(0.11)	-0.62	(0.90)
Var8	0.74	(0.50)	0.19	(0.12)	0.56	(0.24)*	0.01	0.11	0.59	(0.71)
Var9	0.13	(0.69)	0.34	(0.17)+	0.58	(0.34)+	-0.03	(0.17)	0.69	(1.19)
Var10	-0.26	(0.27)	0.11	(0.08)	-0.16	(0.16)	0.27	(0.07)***	1.03	(1.37)
Var11	1.44	(0.63)*	-0.10	(0.34)	0.10	0.45	0.05	0.21	-1.07	(0.42)*
Var12	-0.27	(0.43)	0.34	(0.13)	0.74	(0.31)*	0.05	(0.14)	-0.39	(0.79)

Note: Variable names removed for confidentiality purposes. TPL variables limited due to lower data than the other coverages. ***: significant at 0.1%, **: significant at 1%, *: significant at 5%, +: significant at 10%.

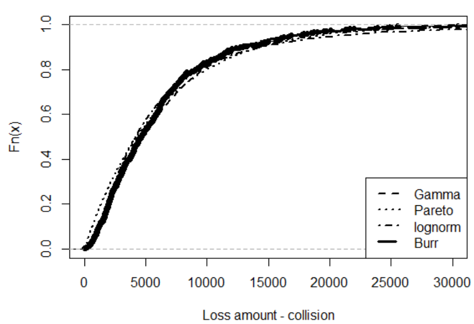
Appendix E - CDF and Q-Q plots by coverage



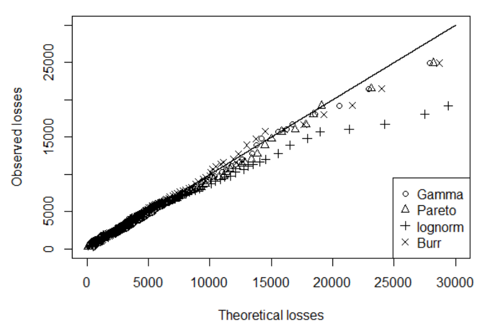
(a) CDF - AB



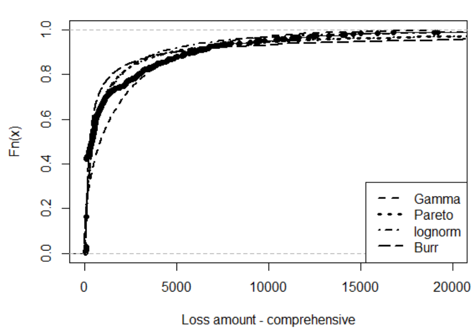
(b) QQ Plot - AB



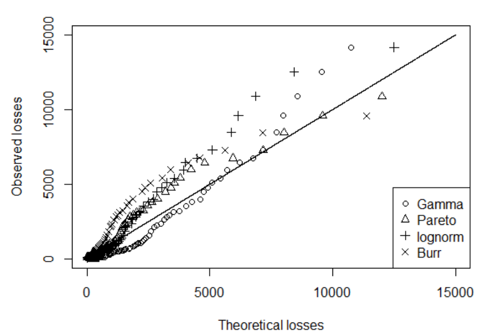
(c) CDF - Coll



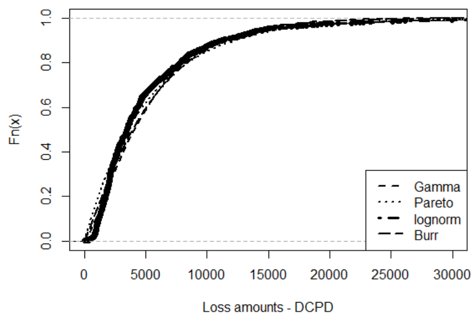
(d) QQ Plot - Coll



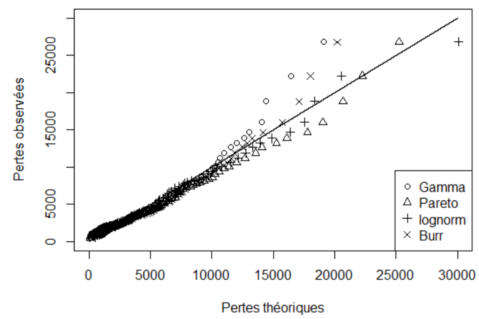
(e) CDF - Comp



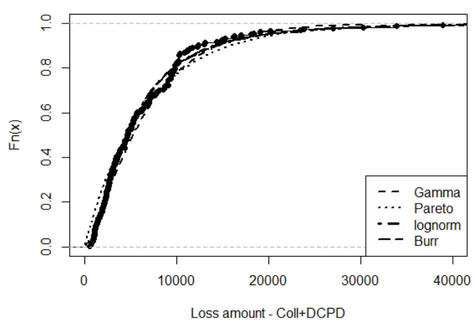
(f) QQ Plot - Comp



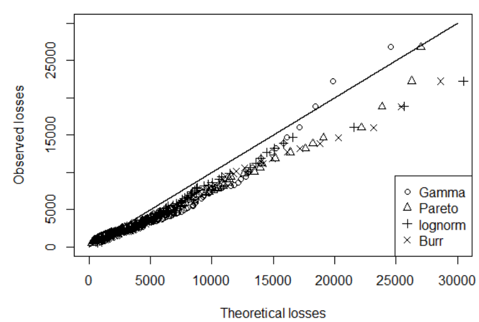
(g) CDF - DCPD



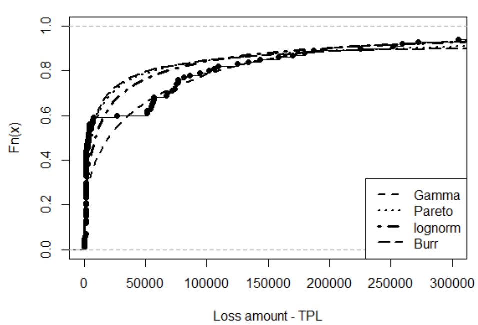
(h) QQ Plot - DCPD



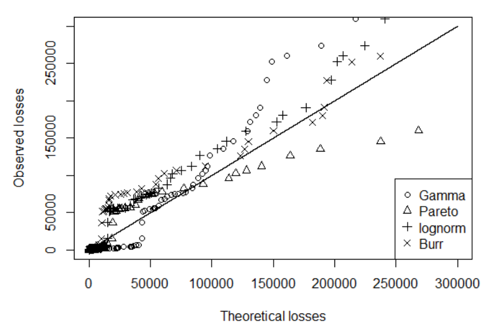
(i) CDF - Coll+DCPD



(j) QQ Plot - Coll+DCPD



(k) CDF - TPL



(l) QQ Plot - TPL

Figure 5.7: Empirical cumulative distribution (left) and quantile-quantile plot (right)

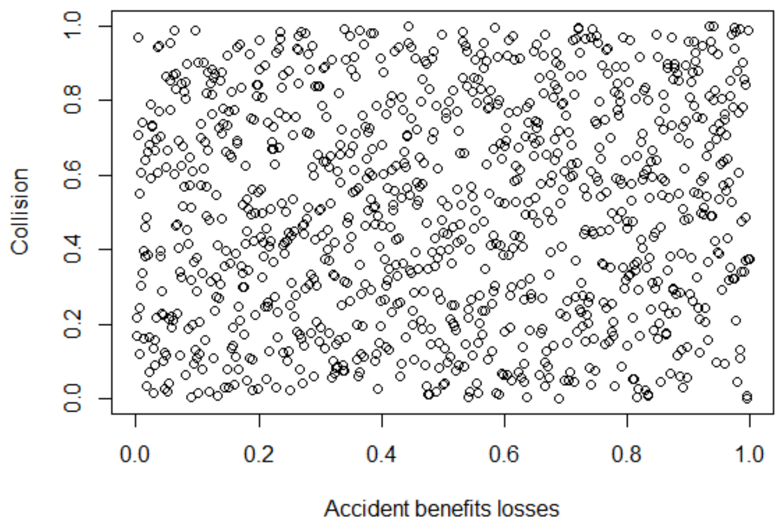
Appendix F - Homogeneous distribution parameters by distribution and coverage

Table 5.10: MLE parameters for all coverages by distribution

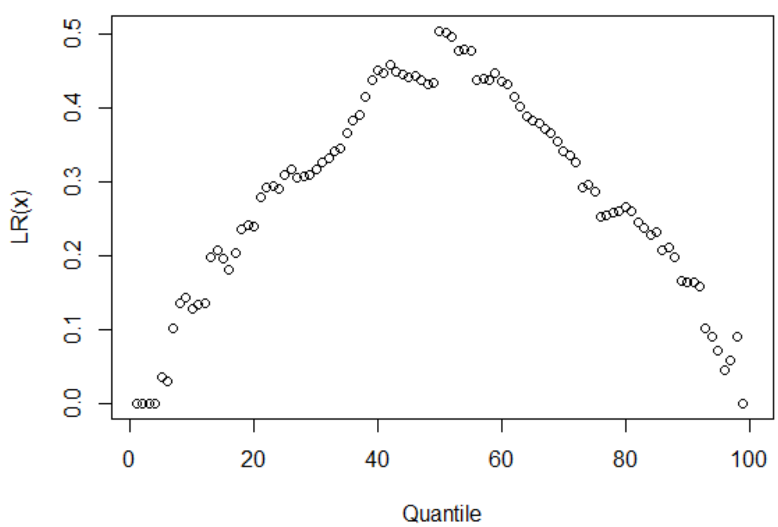
	Gamma		Pareto		lognormal		Burr		
	shape	**scale**	**shape**	**scale**	**meanlog**	**sdlog**	**shape1**	**shape2**	**scale**
AB	0.50	63 747.38	1.11	9 429.53	9.10	1.52	**0.27**	**2.57**	**2 431.02**
	(0.00)	(7 838.25)	(0.20)	(3 046.78)	(0.10)	(0.07)	**(0.05)**	**(0.36)**	**(274.19)**
Coll	1.40	4 460.40	52.74	322 465.52	8.34	0.96	**2.65**	**1.50**	**10 287.02**
	(0.07)	(277.31)	(2.16)	(2 123.94)	(0.04)	(0.028)	**(0.67)**	**(0.08)**	**(2 477.47)**
Comp	0.47	4 254.52	**0.87**	**366.12**	6.26	1.62	0.05	12.57	93.14
	(0.02)	(319.18)	**(0.06)**	**(49.60)**	(0.07)	(0.05)	(0.00)	(1.21)	(1.99)
DCPD	1.41	6 043.20	22.62	112 603.35	**8.16**	**0.89**	0.97	1.99	3 368.82
	(0.04)	(349.27)	(14.70)	(76 277.88)	**(0.03)**	**(0.02)**	(0.15)	(0.12)	(410.11)
Coll+DCPD	1.41	4 838.24	18.35	118 140.57	**8.43**	**0.88**	1.55	1.67	6 742.60
	(0.12)	(492.46)	(17.18)	(116 321.40)	**(0.06)**	**(0.04)**	(0.47)	(0.16)	(1 865.88)
TPL	0.34	203 320	0.48	1 951.65	**9.11**	**2.38**	0.33	1.31	1 272.89
	(0.04)	(41 129.67)	(0.07)	(611.05)	**(0.23)**	**(0.16)**	(0.10)	(0.31)	(440.94)

Note: selected marginal parameters are in **bold**

Appendix G - Theoretical rank plots and left-right tail functions for the proposed copulas



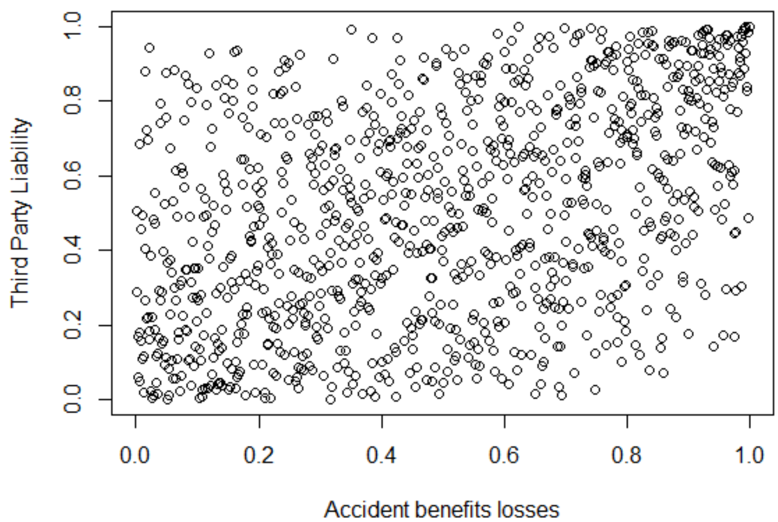
(a) Theoretical rank plot AB-DCPD



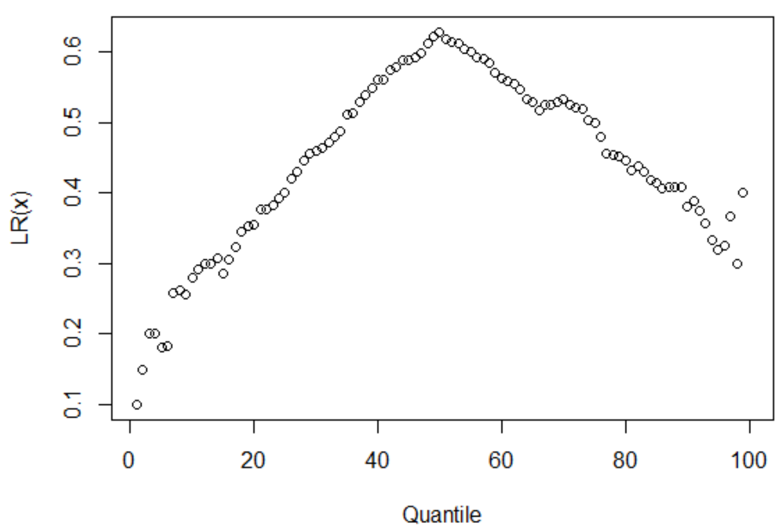
(b) Left-right tail function AB-DCPD

Figure 5.8: Rank plot (top) and left-right tail function (bottom)

for Accident Benefits and DCPD



(a) Theoretical rank plot AB-TPL



(b) Left-right tail function AB-TPL

Figure 5.9: Rank plot (top) and left-right tail function (bottom) for Accident Benefits and Third Party Liability

BIBLIOGRAPHY

Abdallah, A. (2016). Modèles de dépendance hiérarchique pour l'évaluation des passifs et la tarification en actuariat.

Abdallah, A., Boucher, J.-P. & Cossette, H. (2016). Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance: Mathematics and Economics*, *68*, 120–133.

Antonio, K. & Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, *2014*(7), 649–669.

Barnett, G., Zehnwirth, B. & Dubbossarsky, E. (2005). When can accident years be regarded as development years? *Proceedings of the Casualty Actuarial Society*, *92*(177), 239–256.

Bornhuetter, R. L. & Ferguson, R. E. (1972). The actuary and IBNR. *Proceedings of the casualty actuarial society*, *59*(112), 181–195.

Boucher, J.-P., Denuit, M. & Guillén, M. (2007). Risk classification for claim counts: a comparative analysis of various zeroinflated mixed poisson and hurdle models. *North American Actuarial Journal*, *11*(4), 110–131.

Boucher, J.-P., Denuit, M. & Guillen, M. (2008). Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, *2*(1), 135–162.

Canadian Institute of Actuaries (2014). Premium liabilities. *Educational Notes*.

Charpentier, A. (2014). *Computational actuarial science with R*. CRC Press.

Collins, E. & Hu, S. (2003). Practical considerations in valuing premium liabilities. *Institute of Actuaries of Australia 14th General Insurance Seminar*.

Cossette, H., Côté, M.-P., Marceau, E. & Moutanabbir, K. (2013a). Multivariate distribution defined with farlie–gumbel–morgenstern copula and mixed erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics*, *52*(3), 560–572.

Cossette, H., Mailhot, M., Marceau, É. & Mesfioui, M. (2013b). Bivariate lower and upper orthant value-at-risk. *European actuarial journal*, *3*(2), 321–357.

Côté, M.-P., Genest, C. & Abdallah, A. (2016). Rank-based methods for modeling dependence between loss triangles. *European Actuarial Journal*, *6*(2), 377–408.

Dutang, C., Goulet, V., Pigeon, M. et al. (2008). actuar: An r package for actuarial science. *Journal of Statistical software*, *25*(7), 1–37.

England, P. D. & Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, *8*(3), 443–518.

Feldblum, S. (1996). Naic property/casualty insurance company risk-based capital requirements. *Proceedings of the Casualty Actuarial Society*, *83*(159), 300–389.

Financial Services Commission of Ontario (2016). What do the coverages mean. url http://www.fsco.gov.on.ca/fr/auto/brochures/pages/what-the-coverages-mean-glossary.aspx. Last visited 2019-05-18.

Frees, E. W. (2008). Dependent insurance risks. *Encyclopedia of Quantitative Risk Analysis and Assessment*, *2*.

Frees, E. W. & Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, *103*(484), 1457–1469.

Friedland, J. (2010). Estimating unpaid claims using basic techniques. In *Casualty Actuarial Society*, volume 201.

GAA (2015). Les collisions augmentent de 30% en hiver sur les routes du québec. *Communiqué*.

Hofert, M., Mächler, M. & Mcneil, A. J. (2012). Likelihood inference for archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, *110*, 133–150.

Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC.

Kaye, P. (2005). Risk measurement in insurance a guide to risk measurement, capital allocation and related decision support issues. *Casualty Actuarial Society Discussion Paper*.

Kendall, M. G. (1948). Rank correlation methods.

Klugman, S. A., Panjer, H. H. & Willmot, G. E. (2012). *Loss models: from data to decisions*. John Wiley & Sons.

Kremer, E. (1982). IBNR-claims and the two-way model of anova. *Scandinavian Actuarial Journal*, *1982*(1), 47–55.

Lehmann, E. L. & Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.

Li, J. (2010). Prediction error of the future claims component of premium liabilities under the loss ratio approach. *Variance*, *4*(2), 155–169.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, *23*(2), 213–225.

Meyers, G. (2013). Correlations between accident years in loss reserve models. *The Actuarial Review*, *40*(1), 27–28.

Office of the Superintendent of Financial Institutions (2018a). Minimum capital test for federally regulated property and casualty insurance companies. url http://www.osfi-bsif.gc.ca/eng/fi-if/rg-ro/gdn-ort/gl-ld/Pages/mct2018.aspx. Last visited 2019-06-27.

Office of the Superintendent of Financial Institutions (2018b). Own risk and solvency assessment. url http://www.osfi-bsif.gc.ca/eng/fi-if/rg-ro/gdn-ort/gl-ld/Pages/e1918.aspx. Last visited on 20019-08-07.

Pechon, F., Denuit, M. & Trufin, J. (2019). Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal*, pp. 1–28.

Pigeon, M., Antonio, K. & Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin: The Journal of the IAA*, *43*(3), 399–428.

Priest, C. (2012). Premium liability correlations and systemic risk. *Australian Actuarial Journal*, *18*(1), 1–66.

Shi, P., Zhang, W. & Boucher, J.-P. (2018). Dynamic moral hazard: A longitudinal examination of automobile insurance in canada. *Journal of Risk and Insurance*, *85*(4), 939–958.

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, *9*(6), 449–460.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229–231.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Venter, G. G. (2002). Tails of copulas. *Proceedings of the Casualty Actuarial Society*, *89*(171), 68–113.

Verrall, R. & Wüthrich, M. (2016). Understanding reporting delay in general insurance. *Risks*, *4*(3), 25.

Verrall, R. J. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: mathematics and economics*, *26*(1), 91–99.

Wright, T. S. (1990). A stochastic method for claims reserving in general insurance. *Journal of the Institute of Actuaries*, *117*(3), 677–731.

Wüthrich, M. V. & Merz, M. (2008). *Stochastic claims reserving methods in insurance*. John Wiley & Sons.

Yee, T. W. et al. (2010). The vgam package for categorical data analysis. *Journal of Statistical Software*, *32*(10), 1–34.

Zhao, X. & Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, *46*(2), 290–299.