



HAL
open science

Modélisation de la dépendance temporelle des sinistres en assurance non vie et enjeux de l'évaluation du Passif

Sawssen Araichi

► **To cite this version:**

Sawssen Araichi. Modélisation de la dépendance temporelle des sinistres en assurance non vie et enjeux de l'évaluation du Passif. Gestion et management. Université Claude Bernard - Lyon I, 2015. Français. NNT : 2015LYO10170 . tel-01369002

HAL Id: tel-01369002

<https://theses.hal.science/tel-01369002>

Submitted on 20 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modélisation de la dépendance temporelle des sinistres en assurance non vie et enjeux de l'évaluation du Passif

THÈSE

présentée et soutenue publiquement le 29 Septembre 2015

pour l'obtention du

Doctorat de l'Université Claude Bernard Lyon I et de l'Université
de Sousse

Spécialité Gestion

par

Sawssen ARAICHI HAMROUNI

Composition du jury

Directeurs de thèse : Christian DE PERETTI, Maître de conférences à l'École Centrale de Lyon
Lotfi BELKACEM, Professeur à l'Université de Sousse.

Rapporteurs : Christian WALTER, Titulaire de la chaire Ethique et finance, collègue d'études
mondiales, Fondation Maison des sciences de l'homme, chercheur associé
au CEFRA de l'EM Lyon
Mokhtar KOUKI, Professeur à l'École Supérieure de la Statistique et de l'analyse
de l'information (ESSAI), Université de Carthage, Tunis, Tunisie.

Examineurs : Jean Louis RULLIÈRE, Professeur à l'Université Claude Bernard Lyon 1.
Faysal MANSOURI, Professeur à l'Université de Sousse.

Remerciements

Arrive enfin ce moment tant attendu, qui marque la fois la fin de la thèse et en même temps la gratitude que j'éprouve à toutes les personnes qui m'ont soutenue pour que ce travail arrive à terme.

Je tiens à remercier tout d'abord mes deux directeurs de thèse Christian de Peretti et Lotfi Belkacem pour tout ce qu'ils m'ont apporté: pour leur aide scientifique, et pour leur soutien humain. Leurs conseils, leurs remarques constructives, leur expérience et leur confiance ont été une grande source de motivation durant ces années.

Je remercie le Professeur Lotfi Belkacem pour m'avoir accompagné et dirigé dans mes recherches depuis le mastère et pendant ces années de thèse. Pour la formation solide dont il m'a fait bénéficier durant mes années à l'HEC. Je lui adresse mon infinie gratitude pour sa disponibilité, sa rigueur scientifique, son expérience en tant que directeur de thèse depuis plusieurs années, m'a beaucoup aidé. Je lui remercie pour son dévouement considérable et son aide si précieuse dans la publication. Je lui exprime mon profond respect.

Je remercie également le Professeur Christian de Peretti pour ses grandes connaissances en économétrie et en mathématiques. Je le remercie pour l'opportunité qu'il m'a accordé, pour le temps qu'il m'a consacré. Je le remercie tout particulièrement pour les travaux que nous avons effectués ensemble, pour nos discussions sans fin sur l'actuariat non-vie et pour le soutien qu'il m'a apporté tout au long de cette collaboration. Je n'oublie jamais qu'il a tenu son rendez-vous, malgré que sa fille s'est blessée à la jambe. Ses qualités scientifiques ainsi que ses qualités humaines m'impressionnent vraiment.

Je suis très honorée que Christian Walter et Mokhtar Kouki aient accepté d'être rapporteurs de ma thèse. Merci à eux pour le temps consacré, je leur en suis très reconnaissante.

Je tiens à remercier chaleureusement Faysal Mansouri et Jean Louis Rullière, de m'avoir fait l'honneur d'accepter de faire partie du jury de cette thèse.

Je tiens aussi à remercier du fond du coeur Frédéric Planchet qui a joué un rôle central dans l'élaboration de cette thèse. Je le remercie pour la base de donnée qui nous a fourni et à ses conseils et remarques si précieuses.

Un grand merci à toutes les personnes qui composent le Laboratoire de Sciences Actuarielle et Financière (SAF). Je remercie Christian Robert pour son aide et pour son accueil chaleureux. Merci à Esterina Masiello, Pierre Ribereau, Areski Cousin, Didier Rullière, Denis Clot et Stéphane Loisel pour leur accueil toujours chaleureux et leur gentillesse.

Je suis très reconnaissante à mes amis qui m'ont facilité mes séjours à Lyon. Un grand merci à Manel, pour tes conseils, pour ton encouragement, pour tes invitations et les sorties ensemble. également, merci à Abdou et Elmira pour leurs aides immenses. Merci aussi à Mehdi, Yeza et Layal qui m'ont beaucoup aidé à bien s'installer à Lyon. Merci à Samy et Soffana pour leur accueils, pour leur amitiés et pour tous les repas de midi partagés au RU ! Merci à Khouloud et Hanen et les bons moments passés à Paris. Je tiens également à souligner toute mon amitié à mes collègues doctorants du laboratoire SAF: Anisa, Rudy, Elias, Julien A. , Julien V., Inés, Yahia, Romain, Jean-Charles, Andrés, Khader, Ibrahima, Huy, hanen et à ceux que j'aurais oubliés.

Je remercie également mes amis du laboratoire LAREMFIQ de l'IHEC de Sousse: Merci à Souad, Hela, Meryem, Manel.A, Naouel, Manel.Y, Rihab, Sana, Zied pour leur sympathie et pour tous les bons moments passés ensemble.

Merci à tous mes collègues enseignants de l'ISG de Sousse, en particulier: Fethi, Chokri, Mehdi, Bochra, Ibtissem, Asma, Leyla, Mouna, Najla, Nouhé, Fafani, Imen, Aicha, Ines et Amani, pour tous les bons moments passés ensemble.

Là j'arrive à la partie qui m'émeut le plus, dédiée aux êtres les plus chers à mon existence.

Du fond du coeur, je remercie mes chers parents. Je suis grandement reconnaissante pour tout ce qu'ils m'ont apporté ! Merci d'être toujours à mes côtés pour m'encourager et me soutenir durant mes nombreuses années d'études. Je ne les remercierai jamais assez pour la vie équilibrée qu'ils m'ont donnée et pour l'amour inconditionnel qu'ils m'ont apporté.

Merci à mes frères Neyer et Oussama et mon adorable petite soeur Yesmine. Vous étiez toujours là pour m'encourager !

Un immense merci à mon bébé chérie; ma fille Salma qui me comble au quotidien du bonheur et de joie.

Un grand merci à ma belle mère, pour son aide immense au quotidien. Merci pour tes relectures et tes conseils.

Je ne crois pas qu'un simple merci suffirait pour remercier mon mari. Tu étais toujours là pour m'écouter et m'encourager. Merci à ta PATIENCE et ton grand AMOUR.

Je suis très reconnaissante à ma magnifique cousine Inés, elle m'a énormément aidé durant mes voyages en France. Elle m'a accueilli de nombreuses fois dans son appartement à Paris. Elle était toujours chaleureuse, encourageante et très aimable. Merci également à son mari Karim et leur bout de choux Adam. Merci Aussi à mon adorable tante Souad, son mari Hamda et ses fils Walid et Jemil.

Mes remerciements s'adressent enfin, à toute ma grande famille et en particulier à mon cher oncle Nouri, qui m'a énormément encouragé durant toutes mes années d'études. Merci également à sa femme Wafa et leurs filles.

Je remercie également mes soeurs de coeur Menyar et son adorable fille Manel, Leyla, Soukaina et Mourad, pour leur aide et leur encouragement.

Merci à mes très chères amies Samia, Wafa, Amira.A, Amira.S et Mounira pour m'avoir soutenue et pour avoir été là pour moi.

Modélisation de la dépendance temporelle des sinistres en assurance non vie et enjeux de l'évaluation du passif

Résumé

Initialement, la modélisation des risques en assurance non vie, supposait l'indépendance entre les différentes variables des modèles actuariels. De nos jours, cette hypothèse d'indépendance est souvent relâchée afin de tenir compte de possibles interactions entre les différents éléments.

Cette thèse a pour but de contribuer à la littérature existante de la modélisation de la dépendance en assurance non vie. Concrètement, nous introduisons une nouvelle méthodologie d'analyse des risques en assurance à travers le développement des modèles de dépendance, principalement dans un cadre dynamique.

Dans le premier chapitre de la thèse nous introduisons le contexte actuel de solvabilité, ainsi que la modélisation de la dépendance en assurance, avec une présentation des principaux résultats.

Le deuxième chapitre est essentiellement constitué d'un article coécrit avec Christian de Peretti et Lotfi Belkacem, intitulé "*Modelling Temporal Dependence of Claims In Insurance Using Autoregressive Conditional Amount Models*" (voir Araïchi et al. (2013)). Dans ce chapitre nous montrons l'existence d'une forme de dépendance temporelle (dynamique) entre les montants de sinistres d'une même branche d'assurance.

Nous proposons un nouveau modèle nommé *Autoregressive Conditional Amount Model (ACA)*, qui permet de capturer le comportement dynamique des sinistres.

également, nous développons un nouveau modèle nommé *Generalized Extreme Value ACA model (GEVACA)*, afin d'analyser la dépendance dynamique des montants élevés, au niveau des queues de distribution.

Enfin, nous donnons une nouvelle expression pour la *Value at Risk (VaR)* paramétrique adaptée pour des risques à dépendance temporelle.

Des applications sur des données réelles et des techniques de *backtesting* sont ensuite effectuées afin de montrer la pertinence des modèles proposés.

Le troisième chapitre est constitué d'un article coécrit avec Christian de Peretti et Lotfi Belkacem, intitulé "*Generalized Autoregressive Conditional Sinistrality Model : A novel model for claims reserving in Non life insurance*", (voir Araïchi et al. (2015)).

Dans ce chapitre, nous abordons d'abord le problème de l'évaluation des réserves dans un cadre dynamique. Nous montrons l'existence d'une forme de dépendance dynamique dans un triangle de liquidation. En particulier, nous nous intéressons à l'analyse de la dépendance temporelle entre les sinistres, ainsi qu'entre les années de développement.

Nous proposons un nouveau modèle nommé "*Generalized Autoregressive Conditional Sinistrality Model (GACSM)*", qui constitue une extension du modèle linéaire généralisé classique. Ensuite, nous fournissons une méthode de simulation bootstrap basée sur le modèle GACSM, qui permet d'évaluer les réserves en tenant compte du caractère dynamique des sinistres.

Enfin, afin de montrer l'impact du modèle proposé sur l'évaluation des réserves et du capital, nous effectuons une comparaison des résultats obtenus avec ceux obtenus des modèles classiques (Chain Ladder et modèle linéaire généralisé).

Dans le quatrième chapitre de la thèse, qui est constitué d'un article, coécrit avec Christian de Peretti et Lotfi Belkacem, intitulé "*Time Varying Copula Model for claims reserving in Non life insurance*". Nous intéressons à évaluer le montant agrégé des sinistres, en analysant conjointement la dépendance dynamique inter-sinistres ainsi qu'entre les sinistres de deux branches. Nous proposons un modèle basé sur le modèle GACSM et les copules conditionnelles, qui permettent de suivre l'évolution de la dépendance au cours du temps. Enfin, nous effectuons des applications sur des données réelles, ainsi que des méthodes de simulation sont considérées. En comparant les résultats obtenus, nous avons pu illustrer l'impact de la dépendance dynamique sur les réserves et le besoin en capital.

Modelling temporal dependence of claims in non life insurance

Abstract

In this thesis a different aspects of dependence modelling are considered. Indeed, temporal dependence structures between claims amounts and between lines of business are analysed.

In the first chapter, a general introduction on modelling dependence in insurance is provided.

The second chapter is essentially constituted by the article "*Modelling Temporal Dependence of Claims In Insurance Using Autoregressive Conditional Amount Models*", written with Christian de Peretti and Lotfi Belkacem, (see Araichi et al. (2013)) It deals with the problem of existing a temporal dependence structure between claims amounts of one line of business. To this end, we propose a new model for handling the dynamic behaviour of claims amounts in insurance companies using an Autoregressive Conditional Amount (ACA) framework. This model will be named Autoregressive Conditional Amount Model (ACA). A Gamma ACA model and a Generalized Extreme Value ACA model are proposed. It is shown that these models are more appropriate to describe and to forecast the process of claims of the lines Auto Damage and Auto Liability than traditional models. Furthermore, a parametric Value at Risk based on ACA framework (VaR ACA) is proposed for evaluating a coverage amount of these claims. Using backtesting techniques, the VaR ACA provides an accurate estimation of risk.

The third chapter of this thesis is based on the article "*Generalized Autoregressive Conditional Sinistrality Model: A novel model for claims reserving in Non life insurance*", written with Christian de Peretti and Lotfi Belkacem, (see Araichi et al. (2015)). In this chapter, a Generalized Autoregressive Conditional Sinistrality Model (GACSM) for claims is proposed. We extend the Generalized Linear Model (GLM) by incorporating temporal dependence between claims amounts of one triangle. The GACSM is used for model parameter estimation and consistency of such estimate is proved. Bootstrap procedure is implemented for prediction reserves and prediction errors. Results show that taking into account the temporal dependence between losses improves the precision of the reserve distribution estimate, and thus evaluates an accurate SCR.

Finally the fourth chapter is based on the article "*Time Varying Copula Model for claims reserving in Non life insurance*", written with Christian de Peretti and Lotfi Belkacem. In this chapter, a time varying copula models to understand the behaviour of claims amounts of two lines of business. Time varying copula functions with a Generalized Autoregressive Conditional Sinistrality model are used to analyse the evolution in time of dependence between two lines and the temporal dependence between claims of each line.

Simulation study is performed to highlight the impact on reserves and Solvency Capital Requirement. Results show that our approach provides a diversification effect between claims amounts.

Table des matières

Table des figures	xi
-------------------	----

Introduction générale

Chapitre 1

Modélisation de la dépendance en assurance

1.1	Contexte actuel de solvabilité	5
1.1.1	Contexte des règles quantitatives : Fonds propres et provisions techniques	7
1.1.2	Évaluation du SCR : Modèle Standard, Modèle Interne	8
1.1.3	Les Mesures de risque	9
1.2	Modèle d'évaluation du Best Estimate	11
1.2.1	Triangle de liquidation : base de calcul des provisions	11
1.3	Modélisation de la dépendance en assurance	13
1.3.1	Dépendance dynamique inter-sinistres	14
1.3.2	Dépendance dynamique dans un triangle de liquidation	15
1.4	Modélisation de la dépendance inter-branches	16
1.5	Contribution originale de la thèse	19
1.5.1	Modèle de sinistralité conditionnelle autoregressif généralisé	22
1.5.2	Modélisation de la dépendance inter-branches	24

Chapitre 2

Modélisation de la dépendance temporelle des sinistres

Résumé	28	
2.1	Introduction	29
2.2	Autoregressive Conditional Amount Models (ACA)	30
2.2.1	Autoregressive Conditional Amount Model (ACA)	31

2.2.2	Autoregressive Conditional Amount Model (ACA)	32
2.2.3	Generalized Extreme Value ACA Model	32
2.2.4	Implementing the mean squared errors in the case of time series	33
2.3	VaR ACA measure	34
2.4	Empirical Analysis	35
2.4.1	Data description	35
2.4.2	Estimation Results with ACA model	39
2.4.3	Comparison with ARMA model	43
2.5	Conclusion	45
2.6	Appendix	47
2.7	Model specifications	47
2.7.1	Exponential ACA model (EACA)	47
2.7.2	Weibull ACA model (WACA)	47
2.7.3	Gamma ACA model (GACA)	47
2.8	Proof	48
2.9	QQ plots	49
	Bibliography	51

Chapitre 3

Évaluation des réserves et du Capital de Solvabilité par le modèle (GACSM)

Résumé	54
3.1 Generalized Autoregressive Conditional Sinistrality Model	56
3.1.1 Claims reserving notation	56
3.1.2 Autoregressive Conditional Amount model (ACAM)	56
3.1.3 Generalized Autoregressive Conditional Sinistrality Model	57
3.2 Bootstrap method and prediction errors	59
3.2.1 Bootstrap method	59
3.2.2 Mean squared error of prediction MSE	60
3.3 Empirical results	61
3.4 Conclusion	72
3.5 Appendix	74
3.6 Classical reserving models	74
3.6.1 Mack Chain Ladder model	74
3.6.2 Generalized Linear Model (GLM)	75

Bibliography	76
------------------------	----

Chapitre 4

Évaluation du Capital de Solvabilité : approche reposant sur les copules dynamiques
--

Résumé	80
4.1 Marginal models	82
4.2 Time varying copula model	84
4.2.1 Motivation for analysing the time varying copula	84
4.2.2 Conditional copulas models	84
4.2.3 Copula estimation	87
4.2.4 Simulation method	87
4.3 Empirical results	88
4.3.1 Marginal modelling	88
4.3.2 Modelling copulas	89
4.3.3 Impact of time varying dependence on total loss	92
4.4 Conclusion	96
4.5 Appendix	97
4.6 Autocorrelations of the lines Auto Damage and Auto liability	97
4.7 QQ plots	97
4.8 Copulas	99
Bibliography	101

Conclusion et perspectives

Bibliographie

Table des figures

1.1	Les trois piliers du projet Solvabilité 2	6
1.2	Actifs et Passifs du Bilan d'une société d'assurance	7
1.3	Triangle de liquidation	12
2.1	plots of the lines Auto Damage and Auto Liability	35
2.2	Histogram and Descriptive Statistics of the line Auto Damage	36
2.3	Histogram and Descriptive Statistics of the line Auto Liability	36
2.4	Autocorrelations of the lines Auto Damage and Auto Liability	36
2.5	Partial Autocorrelations of the lines Auto Damage and Auto Liability	37
2.6	VaR _{99.5%} GACA(1,1) of the line Auto Damage	44
2.7	VaR _{99.5%} GEVACA(1,1) of the line Auto Liability	45
2.8	QQplot of standardized claims amounts of the line Auto Damage versus Gamma(1,1.7)	49
2.9	QQplot of standardized claims amounts of the line Auto Liability versus GEV(0.3,1,0.6)	49
2.10	QQplot of Residuals from ARMA(1,1) of the Auto Damage	50
2.11	QQplot of Residuals from ARMA(1,1) of the Auto Liability	50
3.1	Run- off triangle for incremental claims amounts	56
3.2	Autocorrelations of the line Auto Damage and Auto liability	62
3.3	Reserves estimates of each quarter (in thousands €) for Auto Damage.	67
3.4	Reserves estimates of each quarter (in thousands €) for Auto Liability.	67
3.5	RMSE of prediction as % of reserve estimate of each quarter for Auto Damage.	70
3.6	RMSE of prediction as % of reserve estimate of each quarter for Auto Liability.	70
4.1	Time varying parameters estimates corresponding to Normal, Clayton, SJC and Gumbel copulas	91
4.2	Scatter plots of the aggregate total reserves	94
4.3	Total aggregate reserves distributions from different models	94
4.4	Total aggregate reserves distributions	95
4.5	Autocorrelations of the line Auto Damage and Auto liability	97
4.6	QQplot Auto damage residuals versus Gamma (0.367,2.75)	98
4.7	QQplot Auto liability residuals versus Gamma (0.527,1.9)	98

Introduction générale

L'actuariat non-vie consiste en la modélisation mathématique et en l'étude des différents aspects de l'activité d'assurance. La problématique la plus souvent posée dans ce domaine est l'évaluation des réserves et du capital de solvabilité.

En effet, la société d'assurance se doit d'être solvable c'est à dire, elle doit adopter un mode de gestion optimal de ses fonds propres qui garantit le respect des engagements qu'elle a contracté auprès de ses assurés. En souscrivant un contrat d'assurance, l'assureur s'engage à indemniser l'assuré moyennant une prime, de certains risques ou sinistres (accidents, incendies...) touchant ses biens (voiture, logement...) ou sa personne (responsabilité civile).

Afin de respecter ses engagements, la réforme réglementaire européenne solvabilité 2 impose aux sociétés d'assurance de disposer d'un montant de couverture adéquat aux risques souscrits.

En effet, en 2001, la commission européenne a lancé le projet solvabilité 2, destiné à évaluer le besoin d'un changement plus radical du régime de solvabilité. Il vise à moderniser et à harmoniser les règles de solvabilité applicables à l'ensemble des entreprises d'assurances afin de renforcer la protection des assurés, d'inciter les entreprises à améliorer la connaissance et la gestion des risques auxquels ils sont confrontés et d'assurer une application harmonisée entre les pays européens.

le nouveau référentiel se focalise principalement sur trois valeurs de références qui doivent être mesurées. La première fixe le capital minimum à détenir, la seconde estime le niveau des provisions techniques représentant les engagements de la société d'assurance vis-à-vis des assurés et la troisième concerne le niveau du capital de solvabilité à constituer au passif et qui doit contrôler le risque global de la société.

C'est ainsi qu'elle propose deux approches d'évaluation du capital de solvabilité, à savoir le modèle standard et le modèle interne.

La première approche est commune à toutes les compagnies d'assurance européennes, détermine un niveau de capital prudent à constituer, alors que la deuxième est basée sur la mise en place d'un modèle interne à la société, qui reflète la spécificité du profil de risque de la compagnie. La construction d'un tel modèle est une problématique majeure pour les sociétés d'assurance : elle nécessite une analyse fine des risques et notamment de leurs interactions, en d'autre terme : la dépendance qui peut exister entre eux.

En effet, l'hypothèse d'indépendance entre les risques en assurance non vie peut engendrer des conséquences très importantes. Elle peut entraîner une sous tarification des produits, des réserves insuffisantes, une mauvaise allocation d'actifs...

Ceci peuvent provoquer des conséquences néfastes à la compagnie d'assurance et donc le non-respect des engagements de l'assureur envers ses assurés ou bénéficiaires de contrats. Par conséquent, La prise en compte de la dépendance entre les risques constitue un enjeu primordial pour une compagnie d'assurance.

les risques en assurance peuvent présenter deux formes de dépendance. Une forme de dépendance inter-sinistres et une forme de dépendance inter-branches.

La première peut se présenter dans une même branche d'assurance engendrant des sinistres corrélés. Par exemple, les états du climat (sec, pluvieux, neigeux...) peuvent influencer le nombre de sinistres ainsi que leurs montants au sein de la même branche d'assurance. Tandis que La deuxième forme peut être observée entre deux ou plusieurs branches d'assurance. Par exemple, un accident de voiture peut toucher à la fois la branche auto dommage ainsi que la branche responsabilité civile.

Un autre phénomène observé dans les compagnies d'assurance est que la structure de dépendance n'est pas constante au cours du temps et qu'elle suit des processus différents et dynamiques. En effet, les sinistres peuvent être faiblement corrélés (respectivement fortement corrélés), conditionnellement à l'intensité de l'événement qui survient. Vers ce contexte, le changement dans la structure de la dépendance peut être une des causes de la sous tarification ou des provisions insuffisantes pour couvrir la charge des sinistres.

S'inscrivant dans ce contexte, cette thèse consiste essentiellement à étudier différents aspects de modélisation de la dépendance entre les risques.

Notamment, nous effectuons une analyse de la relation entre les sinistres de la même branche d'assurance ainsi qu'entre deux branches.

Notre étude est axée principalement sur la modélisation de la dépendance temporelle des sinistres, où la dépendance est considérée comme une variable qui change au fil du temps.

Des questions fondamentales qui se posent : Comment peut-on définir la dépendance en assurance ? Est ce que cette dépendance sera stable au cours du temps ? La modélisation de cette dépendance permettra t-elle d'évaluer un capital qui permet de pérenniser la compagnie d'assurance ? Les compagnies d'assurance ont elles intérêt à y procéder ? Autrement dit, quel sera l'impact de cette modélisation sur l'évaluation des réserves et du capital de solvabilité ?

Nous nous proposons dans cette thèse de répondre à ces questions en analysant différentes structures de dépendance entre les risques d'assurance non vie, principalement dans un cadre dynamique, où la structure de dépendance évolue au cours du temps.

En effet, l'objectif de ce travail est d'analyser deux formes de dépendance entre les risques ; Une dépendance entre les montants des sinistres eux même, ainsi qu'entre les différentes branches.

Une autre voie explorée dans ce travail est l'étude de l'effet de la dépendance dynamique sur l'évaluation des réserves et du capital de solvabilité.

En effet, la présente thèse contribue à la modélisation de la dépendance au-delà de la corrélation linéaire, en ajoutant une analyse de la dépendance conditionnelle dans un portefeuille non vie et explore également l'existence des changements dans la structure de dépendance. Concrètement, nous introduisons une nouvelle méthodologie d'analyse des risques en assurance à travers l'utilisation des modèles dynamiques, en particulier les modèles autoregressifs conditionnels et les modèles de copules dynamiques, et voir l'impact de la prise en compte de ces phénomènes sur l'évaluation du passif non vie.

Trois modèles sont proposés dans cette thèse pour analyser les risques : Modèle des montants conditionnels autoregressifs, modèle de sinistralité conditionnelle généralisée et le modèle des copules conditionnelles.

Nous nous intéressons donc à une compagnie d'assurance composée de deux branches d'affaire. Ces deux branches peuvent présenter des formes de dépendance entre les risques de chacune, ainsi qu'entre les deux branches.

A notre connaissance, ces questions n'ont pas été abordées dans la littérature actuarielle.

La thèse est composée de quatre chapitres qui font l'objet d'analyse de la dépendance entre les risques.

Dans le premier chapitre de la thèse nous introduisons le contexte actuel de solvabilité, ainsi que la modélisation de la dépendance en assurance, avec une présentation des principaux résultats.

Dans le deuxième chapitre, nous montrons l'existence d'une forme de dépendance temporelle (dynamique) entre les montants de sinistres d'une même branche d'assurance. En effet, de nombreux facteurs économiques, environnementaux ou juridiques peuvent créer de la corrélation entre les risques. Ce qui prouve que les sinistres sont dépendants. Par ailleurs, cette corrélation peut être dynamique. Autrement dit, la dépendance inter-sinistres peut évoluer au fil du temps.

Dans ce chapitre, nous proposons un nouveau modèle nommé *Autoregressive Conditional Amount Model (ACA)*, qui permet de capturer le comportement dynamique des sinistres.

Également, nous développons un nouveau modèle nommé *Generalized Extreme Value ACA model (GEVACA)*, afin d'analyser la dépendance dynamique des montants élevés, au niveau des queues de distribution.

Enfin, nous donnons une nouvelle expression pour la *Value at Risk* (VaR) paramétrique adaptée pour des risques à dépendance temporelle.

Des applications sur des données réelles et des techniques de *backtesting* sont ensuite effectuées afin de montrer la pertinence des modèles proposés.

La problématique évoquée dans ce chapitre, considère les montants des sinistres comme des données longitudinales. Cependant, en assurance les données sont souvent classées dans un triangle de liquidation, permettant pour un actuaire non vie d'évaluer les réserves et le capital de solvabilité. Ce ci fait l'objet du troisième chapitre de la thèse.

Dans le troisième chapitre, nous abordons d'abord le problème de l'évaluation des réserves dans un cadre dynamique. Nous montrons l'existence d'une forme de dépendance dynamique dans un triangle de liquidation. En particulier, nous nous intéressons à l'analyse de la dépendance temporelle entre les sinistres, ainsi qu'entre les années de développement. Nous proposons un nouveau modèle nommé "*Generalized Autoregressive Conditional Sinistrality Model (GACSM)*", qui constitue une extension du modèle linéaire généralisé classique. Ensuite, nous fournissons une méthode de simulation bootstrap basée sur le modèle GACSM, qui permet d'évaluer les réserves en tenant compte du caractère dynamique des sinistres.

Enfin, afin de montrer l'impact du modèle proposé sur l'évaluation des réserves et du capital, nous effectuons une comparaison des résultats obtenus avec ceux obtenus des modèles classiques (Chain Ladder et modèle linéaire généralisé (GLM)).

Ce chapitre s'intéresse principalement à étudier la dépendance dynamique et son impact sur l'évaluation des réserves et du capital dans un cadre unidimensionnelle (au sein d'une seule branche d'assurance). Le chapitre suivant donne des contributions supplémentaires de l'analyse dynamique des risque à une dimension plus grande que un.

Dans le quatrième chapitre de la thèse, nous intéressons à évaluer le montant agrégé des sinistres, en analysant conjointement la dépendance dynamique inter-sinistres ainsi qu'entre les sinistres de deux branches. Nous proposons un modèle basé sur le modèle GACSM et les copules conditionnelles, qui permettent de suivre l'évolution de la dépendance au cours du temps. Enfin, nous effectuons des applications sur des données réelles, ainsi que des méthodes de simulation sont considérées. En comparant les résultats obtenus, nous avons pu illustrer l'impact de la dépendance dynamique sur les réserves agrégées et le besoin en capital.

Cette thèse est organisée comme suit : le premier chapitre est un chapitre introductif sur la modélisation de la dépendance en assurance. Le chapitre 2 est dédié à l'analyse de la dépendance temporelle inter-sinistres. Le chapitre 3 étudie l'impact de la modélisation de la dépendance inter-sinistre sur l'évaluation des réserves et du capital de solvabilité. Le chapitre 4 étend l'analyse à un cadre bivarié en considérant deux branches d'assurance.

Chapitre 1

Modélisation de la dépendance en assurance

Ce chapitre est un chapitre introductif à la thèse qui va développer la problématique et les objectifs de chaque chapitre. La première partie de ce chapitre est basée sur un article coécrit avec Lotfi Belkacem, intitulé *Solvency capital for non life insurance : modelling dependence using copulas*, publié dans *Journal of Applied Economic Sciences* (voir Araichi et Belkacem (2014)).

De prime abord, nous introduisons le contexte actuel de solvabilité. Ensuite, nous présentons les formes de dépendances modélisées dans la thèse. Enfin, nous donnons des résultats utiles pour chaque forme étudiée.

1.1 Contexte actuel de solvabilité

Les compagnies d'assurance évoluent dans un monde de plus en plus incertain (concurrence accrue, internationalisation des compagnies, nouveaux risques,...) ce qui mène le Comité Européen des Contrôleurs d'Assureurs et de Pensions Privées CEIOPS (*Committee of European Insurance and Occupational Pensions Supervisors*)¹, à lancer pour le secteur de l'assurance une réforme réglementaire, dite directive "solvabilité 2". Solvabilité 2, qui vient remplacer le projet solvabilité 1 a pour objectifs d'assurer un bon niveau de protection des assurés et d'améliorer le système de solvabilité actuel en procédant à une révision en profondeur des exigences s'appliquant aux organismes d'assurance, afin de garantir que ces derniers soient financièrement solides et capables de faire face à des circonstances défavorables. En outre, il vise à établir un calcul de solvabilité intégrant tous les risques qui pèsent de manière effective sur les assureurs et qui satisfait les normes comptables IAS-IFRS ².

¹Remplacé le 24 novembre 2010 par l'autorité européenne des assurances et des pensions professionnelles (AEAPP), *European Insurance and Occupational Pensions Authority (EIOPA)*

²(*International Accounting Standards*) : Norme comptable International élaboré en 1991 par l'IASB (*International Accounting Standards Board*). En 2001 certains de ces normes ont été remplacé par des IFRS (*International Financial Reporting Standards*), d'autres sont toujours en vigueur. L'ensemble IAS/IFRS a été imposé comme cadre comptable par l'union européenne pour les entreprises cotées, à

Par ailleurs, ce projet incite les compagnies d'assurance à recourir à un modèle interne pour déterminer le capital réglementaire. Par conséquent, ce nouveau système permet de renforcer la compétitivité internationale des assureurs européens.

Solvabilité 2 s'articule autour de trois piliers (voir figure 1.1) qui constituent les objectifs principaux de la réforme : Le premier pilier concerne les exigences quantitatives, prévoit essentiellement les règles se rapportant à l'évaluation du passif, (à savoir le niveau des provisions techniques et le niveau des fonds propres). A cet effet, deux niveaux de fonds propres sont définis, le capital minimum requis (MCR) et le capital de solvabilité requis (SCR).

Le deuxième pilier concerne les exigences quantitatives de l'assureur, il prévoit les règles de suivi des risques en interne, par la mise en place d'un système de gouvernance adéquat, qui garantit une gestion prudente de l'activité, et dans la supervision faite par les autorités de contrôle. Enfin, le troisième pilier comprend des exigences en matière d'information prudentielle et de publication en vue d'améliorer la transparence de l'information et de promouvoir une meilleure discipline de marché. Ce travail est centré uniquement sur les exigences quantitatives et concerne donc exclusivement le pilier 1.

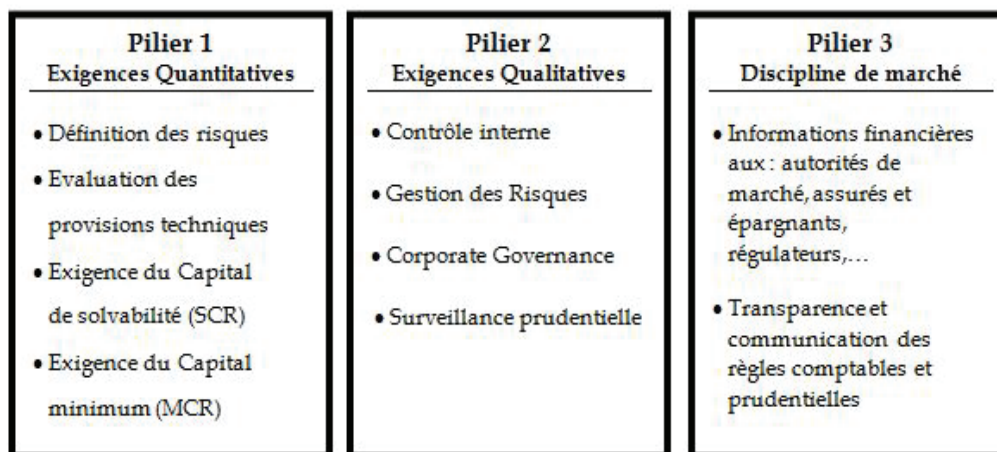


FIGURE 1.1 – Les trois piliers du projet Solvabilité 2

partir du premier janvier 2005.

1.1.1 Contexte des règles quantitatives : Fonds propres et provisions techniques

Les Fonds propres

Les exigences quantitatives (pilier 1) sont basées sur une vision "*market consistent*" où les actifs couvrent adéquatement les passifs (voir figure 1.2).

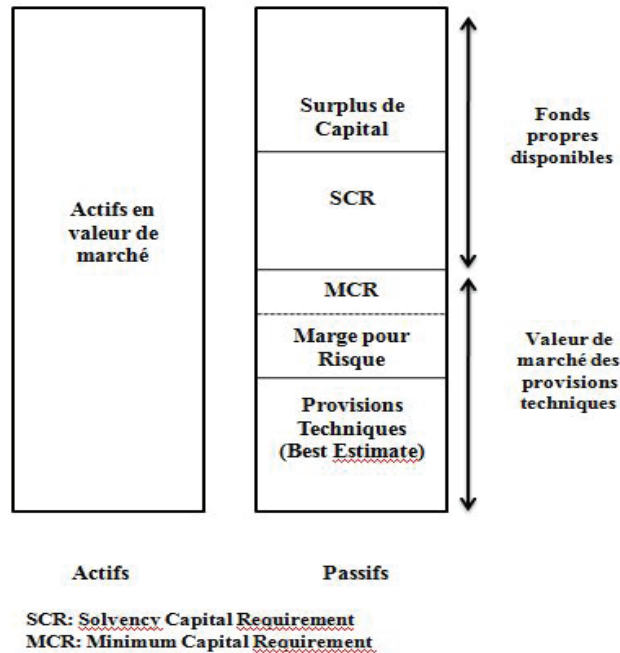


FIGURE 1.2 – Actifs et Passifs du Bilan d'une société d'assurance

En termes de ces exigences quantitatives, solvabilité 2 fixe deux niveaux de fonds propres : le Capital Minimum Requis (*Minimum Capital Requirement MCR*) et le Capital de Solvabilité Requis (*Solvency Capital Requirement SCR*).

Capital Minimum Requis (MCR)

C'est le capital, au dessous duquel les autorités de contrôle interviennent, dans la mesure où son dépassement déclenchera une intervention juridique pouvant aller jusqu'au retrait d'agrément. Son calcul est simple et robuste.

En effet, il est calibré selon la VaR avec un niveau de confiance de 80% à 90% à l'horizon d'un an. Il a un seuil plancher absolu d'un million d'euros pour les entreprises d'assurance et de réassurance non-vie.

Capital de Solvabilité Requis (SCR)

On l'appelle aussi Capital Ajusté au Risque (*Risk Adjusted Capital*) ou Capital Cible, ce dernier désigne le niveau de capital permettant à une société d'assurance d'absorber les sinistres imprévus significatifs et de donner aux preneurs d'assurance l'assurance raisonnable que les versements seront effectués à leur échéance (voir Commission (2003)). Il s'agit donc du capital nécessaire à la société pour qu'elle soit certaine de demeurer solvable et de pouvoir poursuivre son activité dans les années futures.

Le SCR pourra être obtenu, soit par l'application d'un modèle standard commun à tous les assureurs européens, soit par un modèle interne plus adapté aux risques réellement supportés par la compagnie. Par conséquent, le SCR correspond au fond propre requis qui doit remplir les engagements futurs à un horizon temporel donné, et en fonction d'un niveau de confiance (équivalent à une probabilité de 99.5%).

Les provisions techniques et la marge pour risque

Les provisions techniques représentent les engagements de la société d'assurance vis-à-vis des assurés et des bénéficiaires du contrat. Elles doivent être calculées de manière prudente, fiable et objective. Le montant des provisions techniques est égal à la somme du Best Estimate et d'une marge pour risque.

En effet, le Best Estimate correspond à la valeur actuelle, attendue des flux de trésorerie futurs (voir Commission (2010)), basé non seulement sur des informations crédibles et des hypothèses réalistes, mais aussi sur des méthodes actuarielles et techniques statistiques qui tiennent compte des caractéristiques du portefeuille d'assurance et de tous les risques liés à la sinistralité future.

La marge pour risque correspond au coût que représente la mobilisation d'un montant de fonds propres éligibles (égal au SCR) nécessaire pour assumer les engagements futurs d'assurance et de réassurance sur toute leur durée de vie.

1.1.2 Évaluation du SCR : Modèle Standard, Modèle Interne

Modèle Standard

Le CEIOPS définit une formule standard commune pour toutes les compagnies d'assurance afin de gérer leurs risques. En effet, le chargement en capital (SCR) au titre du risque de provisionnement est calculé comme suit

$$SCR = \rho(\sigma_{totale})V_{totale} \tag{1.1}$$

avec

$$\rho(\sigma_{total}) = \frac{\exp(N_{0.995} * \sqrt{\log(\sigma_{total}^2 + 1)})}{\sqrt{\sigma_{total}^2 + 1}} - 1$$

$N_{0.995}$ Quantile à 99,5% de la distribution normale standard.

$$\sigma_{totale} = \sqrt{\frac{1}{V_{total}^2} \cdot \sum_{k,l=1}^n Corr_{k,l} \cdot \sigma_k \cdot \sigma_l \cdot V_k \cdot V_l},$$

σ_k, σ_l sont les écarts types qui correspondent respectivement aux branches d'assurance k et l , ces écarts types sont fixés par l'EIOPA voir table Commission (2010).

$V_{total} = V_k + V_l$ est le montant total des réserves, évalué comme la somme des réserves de la branche k et la branche l .

Cependant, la formule standard proposée par le CEIOPS présente des limites à la fois quant aux exigences en capital qui ne tiennent pas compte de leur adéquation aux passifs et quant aux données statistiques qu'elle exige.

Modèle Interne

Le CEIOPS recommande l'utilisation d'un modèle interne adapté aux risques réellement supportés par l'assureur. Ce modèle vise à moderniser les règles de solvabilité applicables, d'améliorer la connaissance et la gestion des risques internes. Par conséquent, le modèle interne est un véritable outil pour la gestion d'une compagnie d'assurance, influençant notamment les politiques de prise et de cession des risques que la formule standard.

L'unique exigence imposée par le CEIOPS, pour le modèle interne est que ce dernier doit être conforme aux exigences de solvabilité 2 à savoir un horizon d'un an et une mesure de risque correspondant à la Value at Risk au seuil de 99,5%.

La formule du chargement en capital au titre du risque de provisionnement est donnée par :

$$SCR_k = VaR_{99,5\%}^k - BE_k \quad (1.2)$$

avec BE_k est le montant des réserves de la branche k .

Après avoir calculé un SCR pour chaque branche d'assurance, on fait l'agrégation des risques qui se gère simplement par la formule :

$$SCR_{aggreg} = \sqrt{\sum_{k,l} Corr_{k,l} SCR_k SCR_l} \quad (1.3)$$

$Corr_{k,l}$ est la matrice de corrélation des deux branches k et l .

SCR_k est le capital de solvabilité de la branche k et SCR_l est le capital de solvabilité de la branche l .

1.1.3 Les Mesures de risque

Les mesures de risque sont des outils de quantification de risque, qui permettent d'évaluer un niveau de dangerosité d'un risque et de comparer différents risques entre eux. La définition d'une mesure de risque formalisé dans Denuit et Charpentier (2004) est la suivante :

Définition 1.1.1 (mesure de risque) On appelle mesure de risque toute application ρ associant un risque X à un réel $\rho(X) \in \mathbb{R}_+ \cup \{+\infty\}$.

La quantification et la comparaison des risques peuvent ensuite être utilisées à plusieurs fins telles que l'évaluation de primes, l'allocation de capital, la détermination de risques d'un portefeuille d'assurance ou de réassurance.

La définition d'une mesure de risque est très générale puisque toute fonctionnelle réelle positive d'une variable aléatoire peut être considérée comme étant comme une mesure de risque. En pratique, on exige qu'une mesure de risque ρ possède certaines propriétés afin d'être "cohérente".

Définition 1.1.2 (Mesure de risque cohérente) Une mesure de risque ρ est dite cohérente au sens de Artzner et al. (1999) si, pour deux variables aléatoires X et Y elle satisfait les propriétés suivantes :

- *Monotonie* : si $\mathbb{P}(X \leq Y) = 1$, alors $\rho(X) \leq \rho(Y)$;
- *Sous-additivité* : $\rho(X + Y) \leq \rho(X) + \rho(Y)$;
- *Homogénéité positive* : $\rho(a X) = a \rho(X)$;
- *Invariance par translation* : $\rho(X + b) = \rho(X) + b$.

Exemple de mesures de risque

Value at Risk (VaR)

Développée par Jorion [1997], la *Value-at-Risk* (VaR) est une mesure de risque commune aux organismes financiers qui est totalement équivalente au concept de sinistre maximum probable, très familier des assureurs non vie. Elle est égale à la perte maximale que peut subir une organisation, dans des conditions normales de marché, sur une période de temps donnée (H) pour un certain niveau de probabilité α . En terme statistique, elle correspond à la notion de quantile :

$$VaR_\alpha(X) = F_X^{-1}(\alpha), \quad \forall \alpha \in (0, 1), \quad (1.4)$$

La VaR exhibe l'avantage d'être un concept simple et facile à calculer. Cependant elle n'est pas sous additive selon Artzner [1999], ce qui constitue la critique la plus souvent formulée à son égard. De plus la VaR ne donne qu'une information ponctuelle au quantile α de la distribution du sinistre et aucune information au delà de ce point. C'est pourquoi des mesures de risque alternatives ont été développées.

Conditional Value at Risk (CVaR)

La *Conditional Value at Risk* (CVaR) de niveau α est une mesure de risque définie comme l'espérance de la variable aléatoire X , sachant que celle ci dépasse la VaR au niveau de confiance α :

$$CVaR_\alpha(X) = E[X | X \geq VaR_\alpha(X)], \quad (1.5)$$

Cette mesure de risque, contrairement à la VaR , donne des informations sur la moyenne des pertes au delà de la VaR et donc sur l'épaisseur de la queue de distribution. Ceci permet de déterminer un capital de manière à surmonter en moyenne les pires sinistres.

Tail-Value-at-Risk (TVaR)

La *Tail-Value-at-Risk* de niveau α associée au risque X est définie comme la moyenne arithmétique des valeurs de la VaR pour les niveaux de confiance au-delà de α :

$$\text{TVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(X) du. \quad (1.6)$$

Cette mesure coïncide avec la CTE lorsque la fonction de répartition du risque X est continue.

La CTE ne respecte pas toutes les propriétés d'une mesure de risque cohérente alors que la TVaR sera cohérente aussi bien pour des variables aléatoires continues que discrètes.

1.2 Modèle d'évaluation du Best Estimate

En assurance Non vie, il existe différentes méthodes du calcul du Best Estimate des provisions pour sinistre à payer. Nous présenterons dans ce qui suit les modèles les plus usités dans les compagnies d'assurances, à savoir les modèles de type Chain Ladder et les modèles linéaires généralisés, basés sur les triangles de liquidation.

1.2.1 Triangle de liquidation : base de calcul des provisions

Du fait du décalage entre l'exercice de survenance d'un sinistre et l'exercice de règlement, la sinistralité d'une branche peut être représentée d'une manière synthétique dans un triangle de liquidation reflétant la dynamique des sinistres et permettant de suivre le déroulement temporel de paiement. En effet, les prestations à payer par une compagnie d'assurance relatives à une année de survenance sont étalées sur plusieurs années de développement. Dans la suite, les notations suivantes seront utilisées :

i : année de survenance du sinistre $1 \leq i \leq n$.

j : année de développement, comptée à partir de l'année de survenance, $1 \leq j \leq n$.

$y_{i,j}$: représente les règlements relatifs aux sinistres survenus à l'année de survenance i et effectués j^{me} année de développement pour $1 \leq j \leq n$.

Le modèle Chain Ladder

Le modèle Chain Ladder s'applique à des triangles de paiement cumulés définis par

$$C_{i,j} = y_{i,j} + \sum_{k=1}^{j-1} y_{i,k} \quad (1.7)$$

Année de survenance	Année de développement						
	1	2j	n
1							
2							
...							
i							
...							
...							
...							
n							

FIGURE 1.3 – Triangle de liquidation

Ce modèle est fondé sur l'utilisation de link-ratios, encore appelés coefficients de passage ou facteur de développement qui serviront à estimer la partie inférieure du triangle de liquidation définis par

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}} \quad 1 \leq j \leq n-1 \quad (1.8)$$

Les paiements futurs sont estimés par

$$\hat{C}_{i,n} = \hat{C}_{i,n-1} \prod_{k=n-i}^{n-1} \lambda_k \quad (1.9)$$

Ayant estimé les paiements futurs, le montant de réserve pour l'année de survenance i est donné par

$$\hat{R}_i = \hat{C}_{i,n} - \hat{C}_{i,n-i+1} \quad (1.10)$$

Le montant de réserve totale est estimé par

$$\hat{R} = \sum_{i=1}^n \hat{R}_i \quad (1.11)$$

Malgré la simplicité et la clarté de cette méthode, elle présente plusieurs inconvénients : L'estimation de la charge finale se fait à l'aide des coefficients multiplicatifs (link ratios), le risque d'accumulation d'erreur est donc important et d'autant plus important que l'exercice de survenance du sinistre est récent.

De plus, l'approche de Chain Ladder repose sur l'hypothèse d'une cadence des paiements constatée par le passé qui va se reproduire dans le futur. Ainsi, elle ne prend pas en considération les différents changements susceptibles de se produire dans le rythme des paiements, ainsi que dans les tendances du marché (fluctuations de l'inflation, évolution de jurisprudence).

Une autre approche est proposée dans la littérature actuarielle qui est la méthode linéaire généralisée. Nous présentons ci dessous une brève présentation de ce modèle.

Le modèle linéaire Généralisé (GLM)

Contrairement au modèle Chain Ladder, le modèle GLM s'applique à des triangles de paiements incrémentaux. C'est un modèle statistique du type variable à expliquer/variables explicatives.

Les incréments du triangle supérieur $y_{i,j}$ sont supposés être des variables indépendantes dont la loi de probabilité est de type exponentiel, de densité de probabilité

$$f(y_{i,j}, \theta_{i,j}, \phi) = \exp\left\{\frac{y_{i,j}\theta_{i,j} - b(\theta_{i,j})}{a_{i,j}(\phi)} + c(y_{i,j}, \phi)\right\} \quad (1.12)$$

avec $a_{i,j}(\phi)$, $b(\theta_{i,j})$ et $c(y_{i,j}, \phi)$ sont des fonctions spécifiques à la distribution. $\theta_{i,j}$ est un paramètre relié à la moyenne, et ϕ est un paramètre d'échelle relié à la variance. Pour plus de détails à propos du modèle GLM, nous renvoyons le lecteur à

Le montant de paiement futur est estimé par $\hat{\mu}_{i,j} = E(y_{i,j}) = b^{-1}(\theta_{i,j})$ et de variance égale à $v(y_{i,j}) = \phi v(\mu_{i,j}) = \phi b'(\theta_{i,j})$.

La moyenne $\mu_{i,j}$, est relié aux covariables via une fonction lien g . Cette fonction lien peut être : identité, puissance, logarithmique.

On a $g(\mu_{i,j}) = \eta_{i,j}$, avec $\eta_{i,j} = X_{i,j}\beta$, où $X_{i,j}$ est la matrice des covariables, et β est le vecteur des paramètres.

Le montant des réserves pour chaque année de survenance est estimé par

$$R_i = \sum_{j>n-i+1} \hat{\mu}_{i,j} \quad (1.13)$$

Le montant de réserve totale est

$$R = \sum_{i=1}^n R_i \quad (1.14)$$

Ces modèles sont basés sur l'hypothèse d'indépendance entre les montants des sinistres, jugée trop restrictive pour décrire la dynamique des risques. Nous évoquons dans ce qui suit la problématique de la dépendance entre les risques et nous proposons une autre alternative des modèles qui introduisent la dépendance entre les montants des sinistres. On note que dans toute la thèse, l'évaluation des réserves et du capital de solvabilité se fait sur l'horizon ultime de développement.

1.3 Modélisation de la dépendance en assurance

Historiquement, les risques auxquels sont soumises les compagnies d'assurance sont considérés comme indépendants. En effet, l'hypothèse d'indépendance est l'une des fondements de base de l'actuariat classique comme rappelait Hans Buhlmann en 1963 :

The independence hypothesis is so common to be made that many authors forget to mention it.

Cependant, intuitivement de nombreux exemples peuvent laisser penser que ce n'est pas le cas en pratique.

Notamment en assurance non vie, on peut relever différentes formes de dépendance. En effet, il existe des dépendances intervenant entre les contrats d'une même branche d'assurance, se traduisant par une dépendance entre les montants des sinistres. Par exemple, les états de climats (sec, pluvieux, neigeux,...) peuvent influencer considérablement le nombre de sinistres au sein de la branche automobile. De même, un incendie frappant un bâtiment peut se propager dans d'autres bâtiments engendrant des risques de perte au sein de la même branche. On peut également, constater des formes de dépendances non plus entre les sinistres d'une même branche d'assurance, mais entre différentes branches d'assurance. En effet, un accident automobile peut toucher à la fois la branche automobile ainsi que la branche responsabilité civile, vu qu'il engendre des dommages matériels, et il peut fréquemment causer des dommages corporels liés à la garantie responsabilité civile.

C'est pourquoi nous nous sommes intéressés dans ce travail à étudier ces deux formes de dépendance qui peuvent impacter le résultat d'une société d'assurance et évaluer ses effets. Nous nous focalisons dans cette thèse sur l'analyse d'une dépendance temporelle en d'autre terme, dynamique qui évolue au fil du temps

1.3.1 Dépendance dynamique inter-sinistres

L'assurance est considérée comme un transfert de risque de l'assuré vers l'assureur moyennant une prime.

Dans ce contexte, la société d'assurance doit recourir à une modélisation adéquate de ces risques afin d'être solvable dans les années futures.

Une grande partie de la littérature actuarielle, notamment en assurance non vie, gravite autour des sujets fondamentaux suivants : modélisation des montants des sinistres et évaluation des réserves et du capital de solvabilité. En effet, dans la modélisation des montants des sinistres d'assurance, l'hypothèse d'indépendance est l'un des fondements de base des modèles proposés. On peut citer le modèle lognormale de Kremer (1982), le modèle binomial négatif de Richard (2000), le modèle Generalized Beta of the Second Kind (GB2) Cummins et al. (1990). Par ailleurs Mikosch (2006) a montré que les montants des sinistres en assurance suivent généralement des distributions de la famille exponentielle.

Ces modèles supposent que les montants des sinistres sont indépendants et identiquement distribués. Cependant, de nombreux facteurs économiques, environnementaux ou juridiques peuvent créer de la corrélation entre les risques. Par exemple, les états du climat, peuvent influencer le nombre et le montant des sinistres au sein de la même branche d'assurance. De même, un incendie frappant un bâtiment peut se propager dans d'autres bâtiments engendrant des risques de perte au sein de la même branche

d'assurance incendie. Ces exemples montrent bien qu'un événement extérieur peut entraîner des sinistres non envisagés, et peut sous-estimer les risques au sein de la même branche d'assurance.

Ainsi la problématique de dépendance entre les sinistres de la même branche est cruciale pour la compagnie d'assurance. Dans ce contexte, les modèles d'indépendance inter-sinistres apparaissent donc inappropriés.

A ce problème de dépendance, s'ajoute la problématique de la dynamique des corrélations. Autrement dit, la dépendance inter-sinistres peut évoluer au fil du temps. En effet, les sinistres peuvent être faiblement corrélés (respectivement fortement corrélés), conditionnellement à l'intensité de l'événement qui survient.

Dans la littérature, les modèles de séries temporelles sont essentiellement établis pour répondre à ce genre de questions. En effet, les séries temporelles permettent d'étudier la dépendance temporelle et estimer les processus dynamiques.

En science actuarielle, il n'existe qu'assez peu de recherche consacrées à l'étude de la dépendance temporelle des montants des sinistres. Nous pouvons citer, Gerber (1982), Cummins (1985), El-Bassiouni et El-Habashi (1991) et Promislow (1991) qui ont utilisé les séries temporelles dans la modélisation de la dépendance temporelle entre les montants d'assurance. Les modèles proposés considèrent que l'intervalle temporelle des événements est fixe, c'est à dire que la survenance des sinistres est régulière dans le temps (chaque année, chaque mois,...). Cependant, en assurance on observe bien que les sinistres surviennent dans un intervalle de temps irrégulier et qui sont espacés dans le temps, et par conséquent l'indemnisation des sinistres se fait dans un intervalle de temps irrégulier.

Donc la question que nous nous proposons porte sur le développement d'un modèle qui permet de modéliser la dépendance temporelle des montants des sinistres et prendre en compte le caractère irrégulier des paiements.

1.3.2 Dépendance dynamique dans un triangle de liquidation

La problématique évoquée dans la section 1.3.1, considère les montants des sinistres comme des données longitudinales. Cependant, en assurance les données sont souvent classées dans un triangle de liquidation permettant de suivre l'évolution temporelle des paiements. En effet, la sinistralité d'une branche relative à une année de survenance donnée est étalée sur plusieurs années de développement. Donc, un modèle de dépendance temporelle peut-il permettre de déterminer correctement les réserves futures ?

L'évaluation des réserves constitue une problématique majeure pour un actuaire non vie. En effet, l'actuaire a besoin de déterminer le montant des réserves afin de faire face à ses engagements futurs.

La littérature actuarielle étant abondante sur ce sujet, nous pouvons citer Mack (1993), Mack (1994), England et Verrall (2002), Wüthrich et Merz (2008).

Parmi les approches classiques d'évaluation des réserves, Le modèle linéaire généralisé et le modèle Chain Ladder sont les plus usités en assurance. Ces modèles sont basés essentiellement sur l'hypothèse d'indépendance entre les montants des sinistres ainsi qu'entre les années de survenance et de développement.

Cette hypothèse est souvent éloignée de la réalité et mène à une sous estimation des montants de réserves, voire même à des conséquences néfastes au sein de la compagnie d'assurance, qui se trouve incapable de gérer ses risques. En effet, un accident de voiture peut toucher simultanément plusieurs contrats au sein de la branche automobile, ce qui rend l'hypothèse d'indépendance discutable voire même irréaliste .

Au cours des dernières années, une partie de la littérature s'est intéressée au problème de la dépendance au sein du triangle de liquidation suggéré par Antonio et Beirlant (2007), qui propose le modèle linéaire généralisé mixte, pour modéliser la dépendance des montants des sinistres entre les années de développement. Cependant, ce modèle nécessite des informations externes telles que des informations sur le marché, qui sont parfois très coûteuses. De plus, ce modèle exige des données très détaillées (sinistre par sinistre).

Récemment, Hudcová et Pešta (2013) proposent d'appliquer le modèle (*generalized estimating equations (GEE)*) pour estimer les réserves futures. La dépendance entre les années de développement est modélisée à travers une matrice de corrélation.

Par ailleurs, Pešta et Okhrin (2014) ont développé une autre attitude de modéliser la dépendance au sein d'un triangle de liquidation, et ce à travers un modèle de moyenne-variance conditionnelle ; en supposant une structure de dépendance par copule entre les montants des sinistres.

Comme nous pouvons observer, la littérature est restreinte au sujet de la dépendance au sein d'un seul triangle de liquidation. C'est pourquoi, nous nous sommes intéressés à développer une nouvelle alternative de modélisation de la dépendance nommée *Generalized Autoregressive Conditional Sinistrality model (GACSM)*, (voir chapitre 4).

Dans ce qui suit, une deuxième forme de dépendance est étudiée. En effet, nous évoquons la problématique de la dépendance entre les branches d'assurance.

1.4 Modélisation de la dépendance inter-branches

Une compagnie d'assurance est généralement formée de multiple branche d'activité. La réforme réglementaire solvabilité 2 propose d'étudier chaque branche indépendamment des autres et ensuite agréger les risques en utilisant une matrice de corrélation. En effet, les coefficients de corrélation sont fixés par la directive et ils ne dépendent que de l'appartenance du risque à une certaine catégorie (catastrophe naturelle, marché, mortalité, longévité...), voir Commission (2010).

Ce coefficient présente certaines limites voir Rolski et al. (1999). En effet, il constitue une mesure très imparfaite de la dépendance dans le sens où il n'intègre que la composante linéaire et il n'est pas défini si l'une des variables a une variance infinie.

Aussi ce coefficient devient insatisfaisant, si on quitte l'univers gaussien. Par ailleurs, deux variables peuvent avoir un coefficient de corrélation nul alors qu'elles sont fortement dépendants.

Ainsi, le coefficient de corrélation ne permet pas d'étudier de manière correcte la structure de dépendance entre les branches d'assurance.

Bien que le régulateur impose une matrice de corrélation entre les risques d'assurance, il n'empêche pas les sociétés d'opter pour un modèle interne et de développer des outils pour modéliser leurs structures de dépendance.

Différents modèles de risque en assurance se sont concentré sur l'étude de situations dans lesquelles les risques sont dépendants et ce à travers la théorie des copules. En effet, cette théorie est essentiellement née pour répondre aux problèmes de dépendance entre les variables. Pour une ample présentation des fondements de la théorie de copule, nous renvoyons le lecteur à Joe (1997) et Nelsen (2006). Des applications en actuariat et gestion des risques sont présentées dans le livre de Denuit et al. (2005), McNeil et al. (2005) et Jaworski et al. (2010).

Le concept de copule dont la signification est une fonction qui permet de constituer un véritable lien entre la distribution multivariée et les lois marginales, a été introduit pour la première fois par Abe Skalar en 1959, en réponse à un problème concernant les espaces métriques aléatoires énoncé par Maurice Fréchet.

Outre une grande souplesse dans la mise en oeuvre de la modélisation multivariée, car elles permettent d'alléger le problème des modèles multivariés en le décomposant en deux étapes :

La première étape consiste à identifier les distributions marginales, ce qui permet d'utiliser les techniques développées dans le cas unidimensionnel.

La deuxième étape consiste à caractériser la structure de dépendance via le choix d'une copule appropriée. Elles permettent une représentation moins naïve de la dépendance statistique fondée sur la mesure traditionnelle de corrélation voir Embrechts et al. (1999a). De plus, elles autorisent des distributions de probabilités jointes moins restrictives, prenant mieux en compte les faits stylisés (leptokurticité, asymétrie et dépendance de queue). Par ailleurs, l'avantage indéniable présenté par les copules, c'est qu'elles permettent de lier des marginales de lois différentes.

Nous introduisons ici la théorie des copules dans un cadre bivarié. Cette théorie s'étend évidemment à des modèles multivariés.

Définition 1.4.1 (Copule) *Une copule C bivariée est une fonction de répartition bidimensionnelle sur $[0, 1]^2$, dont les lois marginales sont uniformes sur $[0, 1]$, satisfaisant les propriétés suivantes.*

i. $\lim_{u_i \rightarrow 0} C(u_1, u_2) = 0$ pour $i = 1, 2$

ii. $\lim_{u_i \rightarrow 1} C(u_1, u_2) = u_j$ pour $i = 1, 2, j = 1, 2, i \neq j$

iii. C est 2 increasing (croissante) :

$$C(v_1, v_1) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_1) \geq 0 \quad \forall u_1 \leq v_1 \text{ et } u_2 \leq v_2,$$

Pour un examen exhaustif des propriétés de copules voir par exemple Nelsen (2006).

A partir de cette définition, C peut représenter la fonction de répartition jointe de variables aléatoires u_1, u_2 . Par conséquent, les copules sont des outils parfaitement adaptés à la construction des distributions bidimensionnelle, c'est justement ce qui stipule le théorème de Sklar.

Théorème 1.4.2 (Théorème de Sklar) *Soit F une fonction de distribution bidimensionnelle dont les marges sont F_1 et F_2 . Alors F admet une représentation copule $C : [0, 1]^2 \rightarrow [0, 1]$ telle que, pour tout x_1 et $x_2 \in [-\infty, +\infty]$,*

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \tag{1.15}$$

Si F_1 et F_2 sont continues alors C est unique.

Inversement, si C est une copule et F_1 et F_2 sont des fonctions de répartition univariées, alors la fonction F est une fonction de répartition bivariée avec F_1 et F_2 sont les fonctions de répartitions marginales.

Si la distribution bivariée est continue, alors elle admet une densité

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2))f(x_1)f(x_2) \tag{1.16}$$

où c est la densité de la copule définie, pour u_1 et u_2 dans $[0, 1]$, par

$$c(u_1, u_2) = \frac{\partial \partial C(u_1, u_2)}{\partial u_1 \partial u_2} \tag{1.17}$$

En assurance et essentiellement en assurance non vie, la notion de copule a été largement étudiée afin de modéliser les différentes structures de dépendance entre les risques. En effet, le concept de copule a été introduit pour la première fois en actuariat par Frees et Valdez (1998), qui ont modélisé la dépendance entre les branches d'assurance. Depuis, l'introduction des copules dans l'analyse des risques d'assurance est si fortement développée, voir Frees et Wang (2006) dans la théorie de la crédibilité, Kaishev et Dimitrova (2006) en réassurance, Antonin et Benjamin (2001), Belguise et Levi (2002), Faivre (2002), Cadoux et Loizeau (2004), Diers et al. (2012), Zhang et Dukic (2013) dans le problème d'évaluation des réserves.

Récemment, une partie encore peu développée de la littérature s'est intéressé au problème de la dépendance entre les risques à l'aide d'un modèle de copule de régression. Des applications en actuariat présentées par Frees et Valdez (2008), Frees et al. (2009) pour modéliser les risques en assurance, Zhao et Zhou (2010), Shi et Frees (2011) pour l'évaluation des réserves, Shi et Valdez (2011), Shi et Valdez (2012) pour examiner l'asymétrie d'information, Shi et Frees (2010), Krämer et al. (2013) pour analyser les risques d'assurance.

Malgré la popularité et la pertinence de ces modèles, l'application des copules avec des modèles de séries temporelles en actuariat, n'est que rarement observée. En effet, le problème de la dépendance temporelle, qui prend en compte l'évolution de la structure de dépendance à travers le temps est très peu développée. Très récemment, Pešta and Okhrin (2014) ont évoqué le problème de dépendance temporelle des sinistres.

Ils considèrent les montants des sinistres comme des séries temporelles présentant des structures de dépendance temporelles. Ils modélisent les montants des sinistres à l'aide d'un modèle moyenne-variance conditionnelle et en supposant une structure de dépendance par copule. Cette étude a donné des résultats consistants et pertinents. Cependant, elle reste inefficace au sujet de la dépendance temporelle entre plusieurs branches d'assurance.

C'est pourquoi, nous nous sommes intéressés dans le chapitre 4 à développer un modèle qui tient en compte dépendance temporelle entre les branches d'assurance. En effet, nous nous proposons des copules dynamiques qui prennent également en compte la dépendance temporelle inter-sinistres. Nous nous sommes focalisés principalement sur deux branches d'assurance, notre étude peut s'étendre évidemment à plusieurs branches d'assurance.

1.5 Contribution originale de la thèse

Nous nous proposons dans cette thèse d'analyser les montants des sinistres d'une compagnie d'assurance dans un contexte dynamique et ce via 3 chapitres. Les chapitres 2 et 3 s'insèrent dans l'analyse de la dépendance dynamique dans un cadre univarié (une seule branche d'assurance). Tandis que le chapitre 4 intègre la dimension bivariée (deux branches d'assurance).

Le deuxième chapitre de la thèse, intitulé **Modélisation de la dépendance temporelle des sinistres : approche fondée sur les modèles des montants Conditionnels Autoregressifs**.

Ce chapitre est essentiellement constitué d'un article **Modeling Temporal Dependence of Claims In Insurance Using Autoregressive Conditional Amount Models**. Ce chapitre porte essentiellement, sur la modélisation de la dépendance temporelle des montants des sinistres de deux branches d'affaires non vie, et sur l'étude de l'effet de cette dépendance sur la mesure de risque Value at Risk. Nous déterminons les processus de survenance de sinistres afin d'estimer ces montants dans le futur.

Nous proposons un nouveau modèle basé sur la théorie des valeurs extrêmes, qui permet de modéliser les montants élevés des sinistres. Notre modèle se fonde aussi sur une extension de la mesure Value at Risk. Dans notre procédure d'estimation, nous calculons une VaR paramétrique qui tient compte de la dépendance temporelle des montants observés. Des techniques de backtesting sont appliquées afin de montrer la validité de la mesure calculée.

Le troisième chapitre, intitulé **Évaluation des réserves et du Capital de Solvabilité par le modèle de Sinistralité Conditionnelle Autoregressif Généralisé (GACSM)**, est fondé sur un article **Generalized Autoregressive Conditional Sinistrality Model : a novel model for claims reserving in Non life insurance**. Il n'existe qu'assez peu de recherche consacrée à la dynamique des réserves et son évolution à travers le temps.

Ce chapitre est consacré à étudier des triangles de liquidation de deux branches d'affaires. Nous proposons un nouveau modèle d'évaluation des réserves et du capital de solvabilité qui permet de tenir compte de la dépendance temporelle des sinistres. En ce qui concerne les réserves, nous mettons l'accent sur les modèles existants qui se basent sur l'indépendance des sinistres, et qui peuvent mener à une sous tarification des risques ou à des réserves insuffisantes pour couvrir la charge des sinistres. Le modèle proposé est une combinaison du modèle des montants Conditionnels Autoregressifs et du modèle autorégressif et moyenne mobile Généralisé. Ce modèle permet l'analyse du comportement dynamique des réserves et son évolution au fil du temps. Nous nous intéressons à montrer l'importance de ce modèle pour un actuaire non vie qui lui permet d'avoir une appréhension des risques souscrits sur un horizon de temps donné.

Le quatrième chapitre, intitulé **Évaluation du Capital de Solvabilité : approche reposant sur les copules dynamiques**, fait l'objet d'un article **Time Varying Copula Model for claims reserving in Non life insurance**. Dans ce chapitre, nous abordons d'abord le problème de dépendance bidimensionnelle, c'est à dire, la dépendance entre deux branches d'assurance. Grâce à des modèles de copules dynamiques, nous déterminons la structure de dépendance qui évolue à travers le temps de ces deux branches. Ensuite, nous procédons à évaluer les réserves et le capital de solvabilité en tenant compte de cette dépendance dynamique.

Chaque chapitre fait l'objet d'une application empirique sur des données réelles. Et compte tenu des résultats obtenus et des analyses effectuées, nous allons conclure notre travail par des recommandations qui permettent à un actuaire non vie de gérer d'une manière efficace ses fonds propres (réserves et capital de solvabilité).

Nous présentons dans la suite un aperçu des hypothèses et des différents résultats de ces travaux.

Une grande partie de la littérature en sciences actuarielles gravite autour des sujets : l'étude de la dépendance entre les risques et l'étude des queues de distribution.

En effet, les risques en assurance prouvent qu'ils sont dépendants avec des distributions à queues épaisses. On peut relever une dépendance entre les contrats de la même branche. Comme par exemple, un incendie frappant un bâtiment peut se propager dans d'autres bâtiments engendrant des risques de perte au sein de la même branche d'assurance incendie. Aussi, un accident de voiture peut toucher plusieurs contrats au sein de la branche automobile. De plus, la survenance de ces sinistres est espacée dans le temps. En d'autres termes, elle n'est pas régulière : on peut prévoir des sinistres après un jour, deux jours ou même plus.

L'analyse des montants payés par les compagnies d'assurance montre que les montants des sinistres présentent des dépendances temporelles et également des montants élevés. Nous nous proposons dans ce chapitre d'étudier la dépendance temporelle entre les montants des sinistres et d'analyser ses effets sur la mesure de risque Value at Risk.

Nous abordons d'abord le problème de la dépendance temporelle entre les sinistres et nous proposons un modèle nommé : modèle des montants conditionnels autoregressifs (ACA).

Notre travail se démarque des papiers précédents ayant analysés la dépendance entre les sinistres par l'introduction d'une structure de dépendance temporelle et ce ci à travers le processus

$$y_t = \psi_t \epsilon_t, \quad (1.18)$$

$$\epsilon_t \sim \text{i.i.d} \quad \text{avec } E(\epsilon_t) = 1. \quad (1.19)$$

ψ_t correspond à l'espérance conditionnelle des sinistres, $\psi_t = E(y_t | \Omega_{t-1})$, avec Ω_{t-1} représente toute l'information passée jusqu'à la date $t - 1$.

Nous supposons que ψ_t suit un processus autoregressif

$$\psi_t = \omega + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{j=1}^q \beta_j \psi_{t-j}, \quad (1.20)$$

avec $\omega, \alpha_i, \beta_j > 0$ et $\sum(\alpha_i + \beta_j) < 1$.

Ce processus permet de prendre en compte de dépendance temporelle entre le montant présent et le montant passé. De plus, ce processus n'exige pas que le paiement des montants soit régulier dans le temps.

Par ailleurs, un des objectif poursuivi dans notre article est d'analyser la dépendance temporelle au niveau des queues de distribution, c'est à dire entre les montants élevés des sinistres. Nous développons un modèle basé sur la loi généralisée des valeurs extrêmes (GEV), et qui tient compte de la dépendance temporelle entre les montants élevés. Notre modèle nommé modèle d'extremum généralisé pour montant de sinistre (GEVACA), diffère de celui de GEV qui considère que les variables sont iid. Il inclut ψ_t qui modélise la dépendance temporelle des variables.

La fonction de distribution d'une variable suivant un GEVACA se présente ainsi

$$G_\theta(y_t | \Omega_{t-1}) = \exp \left\{ - \left[1 + \frac{(\frac{y_t}{\psi_t} - \mu) [1 - \Gamma(1 - \xi)]}{(\mu - 1)} \right]^{-1/\xi} \right\}, \quad (1.21)$$

avec μ est un paramètre de localisation et ξ est un paramètre d'échelle.

Sa fonction de densité est

$$g_\theta(y_t | \Omega_{t-1}) = \frac{1 - \Gamma(1 - \xi)}{\xi(\mu - 1)\psi_t} \left[1 + \frac{(\frac{y_t}{\psi_t} - \mu)(1 - \Gamma(1 - \xi))}{(\mu - 1)} \right]^{-\frac{1}{\xi}-1} \exp \left\{ - \left[1 + \frac{(\frac{y_t}{\psi_t} - \mu)[1 - \Gamma(1 - \xi)]}{(\mu - 1)} \right]^{-1/\xi} \right\} \quad (1.22)$$

Les paramètres du modèle sont estimés par maximum de vraisemblance.

Ensuite, nous nous attachons dans notre travail à évaluer l'impact de la dépendance temporelle sur la mesure de risque Value at Risk (VaR). Nous proposons une VaR paramétrique nommée VAR ACA. En effet, nous dérivons l'expression d'une VaR dynamique qui évolue au fil du temps, définie comme

$$\widehat{VaR}_{t+1}(\alpha) = \hat{\psi}_{t+1} F_{\hat{\theta}_\epsilon}^{-1}(\alpha), \quad (1.23)$$

avec F est la distribution paramétrique de ϵ_t et α est le niveau de confiance.

Nous déduisons l'expression de la VaR pour différentes distributions : VaR exponentielle, VaR weibull, VaR Gamma et VaR GEV.

En particulier, la considération d'une VaR dynamique met en évidence la nécessité d'utiliser des mesures adéquates pour l'étude de la dépendance temporelle au niveau des queues de distribution des sinistres sans recourir à une VaR ponctuelle basée sur la notion de quantile à un niveau de confiance α , qui essentiellement ne prend pas compte de la dépendance au niveau des queues.

Cette méthode fournit une description de l'évolution du montant de couverture à travers le temps et permet à un actuaire non vie de se prémunir contre les pertes extrêmes.

Dans notre article, nous nous proposons de tester empiriquement ces modèles et ce, en considérant deux branches d'assurance non vie Auto Dommage et Auto Responsabilité Civile, constituées des montants de sinistres journaliers.

L'analyse des montants montre la présence des autocorrélations positives entre les sinistres. Nous estimons nos données avec différents modèles ACA.

Les résultats révèlent des paramètres estimés significatifs et des résidus iid. Ce ci mène à conclure que les modèles ACA sont adéquats pour modéliser les montants de sinistres. Un deuxième résultat est dérivé du calcul du montant de couverture et concerne donc la VaR ACA. Nous prouvons à l'aide d'une procédure de backtesting que la VaR ACA est une méthode adéquate pour évaluer le montant de couverture.

1.5.1 Modèle de sinistralité conditionnelle autoregressif généralisé

L'une des problématiques majeures d'un actuaire non vie est d'évaluer le montant futur des réserves qui permet de faire face aux engagements de la société d'assurance.

En science actuarielle, les méthodes proposées pour estimer les réserves sont basées sur les triangles de liquidation (Chain Ladder, GLM...). En effet, les montants de sinistres sont classés dans des triangles de liquidation. Ce classement se fait par année de survenance et années de développement. Les méthodes traditionnelles supposent que les montants sont indépendants ainsi que les années de développements et de survenances. La première question qui en découle est de connaître la relation entre ses montants, en d'autres termes, est ce que ces montants sont réellement indépendants ?

Existe-t-il une structure de dépendance entre les sinistres ? Si oui, une deuxième question qui se pose et qui est relative à l'effet de cette dépendance sur la prédiction des réserves.

Nous nous proposons dans ce travail d'analyser les montants de sinistres qui sont classés dans un triangle de liquidation. En effet, la forme de dépendance étudiée dans ce papier est une dépendance au sein de la même branche d'assurance. Pour ce faire, nous proposons un modèle de sinistralité conditionnel autoregressif généralisé (GACSM). Ce modèle qui est basé sur le modèle ACA du chapitre 2, suppose que les montants de sinistres sont dépendants ainsi que leurs années de développements. Cependant, il suppose que les années de survenances sont indépendants.

Notre modèle s'insère dans une branche de la littérature dédiée à la prédiction des réserves en assurance non vie, et s'inscrit plus précisément dans les papiers introduisant la dépendance au sein du triangle de liquidation Pešta et Okhrin (2014). Ces auteurs étudient principalement la dépendance entre les montants des sinistres de la même année de survenance, en ignorant la dépendance entre les covariables (années de survenances et années de développements). Nous nous proposons de pallier à cette insuffisance et d'apporter une analyse exhaustive de la dépendance.

Nous partons de même hypothèse que le modèle ACA, où nous proposons une structure dynamique aux montants de sinistres définie par

$$y_{i,t} = \psi_{i,t}\epsilon_{i,t}, \quad (1.24)$$

$$\epsilon_{i,t} \sim \text{i.i.d avec } E(\epsilon_{i,t}) = 1. \quad (1.25)$$

avec $\epsilon_{i,t}$ représentent les montants standardisés (résidus) relatives à l'année de survenance i et année de développement t .

On suppose que $\epsilon_{i,t}$ suivent une loi appartenant à la famille exponentielle.

$\psi_{i,t}$ est définie comme l'espérance conditionnelle des sinistres avec $E(y_{i,t} | \Omega_{i,t-1}) = \psi_{i,t} = [\psi_{i,1}, \dots, \psi_{i,T_i}]$, où $\Omega_{i,t-1} = \{x_{i,1}, x_{i,2}, \dots, x_{i,t-1}, y_{i,1}, y_{i,2}, \dots, y_{i,t-1}\}$ est l'information passée jusqu'à la date $t - 1$.

L'espérance conditionnelle des sinistres est reliée au prédicteur $\eta_{i,t}$ par une fonction analogue à celle du modèle GLM appelée fonction lien : $g(\psi_{i,t}) = \eta_{i,t}$.

Cette fonction lien peut être : identité, puissance, logarithmique...

Nous supposons que $\eta_{i,t}$ considère une structure de dépendance temporelle entre les montants $y_{i,t}$ ainsi qu'entre les années de développements, définie par

$$\eta_{i,t} = X'_{i,t}\beta + \sum_{j=1}^p \phi_j \{g(y_{i,t-j}) - X'_{i,t-j}\beta\} + \sum_{j=1}^q \theta_j \{g(y_{i,t-j}) - \eta_{i,t-j}\} \quad (1.26)$$

avec $X'_{i,t}$ est une matrice de $z \times p$ covariables (années de survenances et années de développements), et $\beta \in \mathbb{R}^{p \times 1} = [\gamma, \alpha_2, \dots, \alpha_n, \delta_2, \dots, \delta_{T_i}]$, est le vecteur des paramètres à estimer.

Nous remarquons que $\eta_{i,t}$ généralise une structure autoregressive ainsi qu'une structure moyenne mobile qui sont introduites d'une manière additionnelle dans le prédicteur. Notre modèle étend le modèle traditionnel GLM qui est basé sur l'hypothèse d'indépendance entre $y_{i,t}$ ainsi qu'entre les années de développement.

En effet, notre modèle permet de mitiger cette indépendance en montrant que la relation entre $y_{i,t}$ n'est pas indépendante et peut contenir certaines corrélations. De plus, ces corrélations sont considérées comme dynamiques et changent à travers le temps.

Dans notre papier, nous supposons que les montants standardisés suivent une loi Gamma avec $E(\epsilon_{i,t}) = 1$. Donc la fonction de densité se présente ainsi

$$f(y_{i,t} | \Omega_{i,t-1}) = \frac{y_{i,t}^{k-1} k^k}{\psi_{i,t}^k \Gamma(k)} \exp\left(-\frac{y_{i,t}^k}{\psi_{i,t}}\right) \quad (1.27)$$

où k est un paramètre de forme et $\psi_{i,t} = \exp(\eta_{i,t})$.

Les paramètres du modèle sont estimés par maximisation de la log vraisemblance de la fonction de densité.

Un premier résultat émane de ce calcul, est que tous les paramètres sont consistants et asymptotiquement distribués selon la loi normale.

Nous procédons ensuite à l'analyse empirique. Nous considérons deux branches d'assurance ayant chacune des montants de sinistres classés dans un triangle de liquidation. Nous montrons que chaque branche d'assurance présente des autocorrélations importantes, ce qui justifie la modélisation avec un modèle GACSM.

L'estimation de notre modèle débouche sur des résultats significatifs.

Nous nous intéressons ensuite d'étudier l'effet de cette dépendance sur l'évaluation des réserves de chaque branche d'assurance et sur le capital de solvabilité (SCR). Pour ce faire, nous nous proposons dans notre volet empirique une méthode bootstrap basée sur le modèle GACSM. Cette méthode permet de fournir la distribution des réserves, en tenant compte de la dépendance temporelle des sinistres.

Afin de prouver la pertinence du modèle proposé, une comparaison des résultats est faite avec le modèle traditionnel Chain Ladder et le modèle standard GLM.

Les résultats montrent que les réserves ainsi que le capital calculé par le modèle GACSM (faisant l'objet d'une analyse adéquate de la dépendance temporelle) donnent un montant de réserve et du SCR plus élevés qu'à ceux évalués par les modèles Chain Ladder et GLM.

Un deuxième résultat prouve que le modèle proposé donne des erreurs quadratiques moyennes plus faibles, donc il est plus efficace que les modèles Chain Ladder et GLM.

1.5.2 Modélisation de la dépendance inter-branches

En science actuarielle, l'étude de la dépendance entre les branches d'assurance s'est concentrée essentiellement sur une analyse basée sur les fonctions copules.

Récemment, une partie encore peu développée de la littérature s'est intéressée aux copules de régression.

Nous proposons dans cet article de modéliser la dépendance par des copules conditionnelles.

Notre travail se démarque des papiers empiriques ayant analysé la dépendance entre les branches d'assurance par l'introduction d'une structure de dépendance qui évolue au fil du temps. Cette dépendance dynamique permet de pallier à l'insuffisance de la dépendance considérée comme constante.

En se basant sur les travaux de Patton (2001), qui a introduit la notion de copule dynamique, nous prouvons que la dépendance entre les branches d'assurance est une dépendance qui change à travers le temps, et qui peut prendre des formes différentes. Dans ce travail, nous considérons plusieurs fonctions copules : Copule conditionnelle Normale, copule conditionnelle de Clayton, copule conditionnelle de Symmetrized Joe Clayton et copule conditionnelle de Gumbel.

Pour les marges, nous nous basons sur le modèle GACSM du chapitre 3.

Notre article diffère sur pas mal de points, par rapport à ceux de la littérature étudiant la dépendance entre les branches. En effet, notre papier considère que la dépendance est dynamique à la fois au niveau de chaque triangle de liquidation ainsi qu'entre les branches.

Dans notre volet empirique, nous proposons de tester ces copules dynamiques pour deux branches d'assurance non vie (Auto Dommage et Auto Responsabilité Civile). Nous obtenons des résultats satisfaisants quant à l'adéquation de ces modèles aux montants des sinistres.

Nous procédons ensuite à étudier l'effet des copules dynamiques sur l'évaluation des réserves agrégées et du capital de solvabilité. Pour ce faire, une comparaison est effectuée du point de vue dépendance entre les risques.

Nous considérons la copule conditionnelle de Gumbel, la copule statique Symmetrized Joe Clayton, le modèle GACSM et le modèle Chain Ladder. En effet, nous proposons en premier lieu, une méthode de simulation afin d'obtenir le montant agrégé des réserves. En second lieu, nous calculons des quantités numériques telles que VaR, CVaR, SCR.

Le premier résultat qui émane de l'évaluation des réserves, montre que le modèle proposé fournit un montant agrégé des réserves inférieur à celui du modèle GACSM, et donc un effet de diversification entre les risques peut être observé et qui permet de réduire le risque agrégé des réserves.

Cependant, les résultats du calcul de la VaR et de CVaR prouvent que notre modèle favorise la dépendance dynamique au niveau des queues de distribution et ce à travers l'évaluation d'un capital plus élevé qu'au modèle Symmetrized Joe Clayton et Chain Ladder.

Par ailleurs, d'après l'analyse effectuée, les résultats obtenus du modèle Chain Ladder aboutissent à affirmer que ce dernier sous estime le risque d'une compagnie d'assurance.

Chapitre 2

Modélisation de la dépendance
temporelle des sinistres : approche
fondée sur les modèles des montants
conditionnels autoregressifs

Résumé

Ce chapitre est composé d'un article, intitulé "Modelling Temporal Dependence of Claims In Insurance Using Autoregressive Conditional Amount Models" coécrit avec Christian de Peretti et Lotfi Belkacem.

L'objet de ce chapitre est de modéliser la dépendance dynamique des sinistres. Pour ce faire, nous proposons en premier lieu, un nouveau modèle nommé *Autoregressive Conditional Amount Model (ACA)*, qui permet de décrire une certaine forme de dépendance temporelle entre les montants.

En second lieu, nous évoquons le problème de dépendance temporelle entre les montants élevés au niveau des queues de distribution. Nous développons un nouveau modèle nommé *Generalized Extrem Value Autoregressive Conditional Amount Model (GE-VACA)*.

En dernier lieu, nous donnons une nouvelle expression pour la *Value at Risk (VaR)* paramétrique adaptée pour des risques à dépendance temporelle.

Des applications sur des données réelles sont ensuite effectuées afin de montrer la pertinence des modèles proposés.

2.1 Introduction

Insurance is considered as a transfer of risks from the insured to the insurer. It must have a level of liability, which enables it to absorb unfavourable risks and to be solvent in future years.

Much actuarial research has focused on how to find the best parametrized model which could fit claims amounts, over a given time period.

Several distributions are used to model insurance losses process: the Lognormal model Kremer (1982), the Negative binomial model Richard (2000), and Generalized Beta of the Second Kind (GB2) Cummins et al. (1990). Mikosch (2006) has shown that the exponential family distributions are adequate to model losses. Cooray and Ananda (2005) have modeled actuarial data with a composite Lognormal Pareto Distribution.

The common shortcoming of these models is that they ignore the temporal dependence that may exist between claims amounts. In the actuarial science, these claims amounts are generally considered as a sequence of independent and identically distributed (iid) variables see Bühlmann (1996) and Rolski et al. (1999). However, in most situations, this assumption may be unrealistic. For instance, a car accident may involve several policyholders at once in a collision, and can simultaneously hit several policies. Also, a bad weather (rainy, snowy) may induce a large number of claims in the same period. For that, it has been pointed out that methods enabling modelling of temporal dependencies among claims amounts, are required.

In actuarial science, actuaries seem to believe that identifying the temporal dependence of claims amounts improves their prediction in the future. One approach is the use of time series to model the dynamic effect of claims over time. Little attention has been granted to this area. The early papers by Gerber (1982) and Cummins (1985), and the more recent ones by El-Bassiouni and El-Habashi (1991) are examples of introducing time series in modelling temporal dependence of claims in insurance. In addition, Promislow (1991), Zhang (2005) have determined the ruin probability expression of the claims, under a linear time series model, they have modeled claims using the moving average and the autoregressive models.

In actuarial literature, modelling temporal dependence is a more common practice when developing credibility models. Frees et al. (1999) have introduced a longitudinal data model to analyze and forecast future claims. Bolancé et al. (2003) tested, and estimated an autoregressive specifications for time dependent random effects in a frequency claims models. They have shown that this statistics tool is relevant to determine the dynamics of the data. Boucher et al. (2008) have presented an overview of models where time dependence between contracts of the same insured is modeled.

This paper aims: to investigate the temporal dependence of claims amounts in order to improve their prediction in the future, and to evaluate the impact of this temporal dependence on the evaluation of the Solvency capital requirement.

In this paper, a new model handling temporal dependence between daily claims amounts is proposed. For that purpose, an Autoregressive Conditional Amount (ACA) framework is used. This model takes into account the dependence between claims amounts. Moreover, an alternative model called Generalized extreme value Autoregressive Conditional Amount Model (GEVACA) is introduced to forecast the large losses. Furthermore, a parametric Value at Risk ACA measure is derived to assess the coverage amounts of claims.

This paper is organized as follows: Section 2.2 presents the Autoregressive Conditional Amount Model. Section 2.3 presents the VaR ACA measure. Section 2.4 reports the empirical results, followed by concluding remarks in Section 2.5.

2.2 Autoregressive Conditional Amount Models (ACA)

Engle and Russell (1998) introduced the Autoregressive Conditional Duration (ACD) Model, which presents a new way for modeling irregularly spaced financial transaction data. They used the similar idea of the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model to develop the ACD model based on Exponential (EACD) and Weibull (WACD) distributions rather than symmetric distributions such as the Gaussian one. Their explicit objective is to model the intertemporally correlated durations. At present, this model is considered as the most powerful tools, provided by several applications to real financial data (Engle and Russell (1998), Engle (2000) and Engle (2002)).

Bauwens and Giot (2000) introduced the Log-ACD model, which implies a non linear relation between the durations and their lags. Grammig and Maurer (2000) developed an ACD model based on the Burr distribution.³ They have shown that their model provides a greater flexibility than the EACD and WACD models.

Zhang et al. (2001) proposed the Threshold Autoregressive Conditional Duration (TACD) model to allow the expected duration to nonlinearly depend on past information. Lunde (1999) extended the ACD model to Generalized Gamma duration (GACD), that allows a non monotonic hazard function taking for instance both tub shaped or inverted U-shaped forms. Ghysels et al. (2004) presented the stochastic volatility duration (SVD) model for processes that involve time varying uncertainty.

Moreover, ? developed a methodology of parametric modelling of time series dynamics, when the underlying loss function is linear-exponential. This model is principally based on ACD model of Engle and Russell (1998).

All these studies have shown that ACD models are relevant tools for detecting temporal dependence between durations. Financial durations share common characteristics with insurance claims amounts: they are positive variables and present a temporal de-

³A random variable t follows a Burr distribution if it has the following probability density function:

$$f(t) = \frac{\mu\chi t^{\chi-1}}{(1 + \sigma^2\mu t^\chi)^{\frac{1}{\sigma^2}+1}} \quad t > 0, \mu > 0, \chi > 0, \quad (2.1)$$

where μ and χ are the shape parameters and σ is the scale parameter.

pendence structure. Also claims amounts and financial durations follow principally the exponential family distributions, as proved by Engle and Russell (1998) for financial durations, and by Mikosch (2006) for claims amounts. Motivated by these characteristics, we adapted the ACD models to insurance claims amounts data. To avoid confusion, the model is named Autoregressive Conditional Amount (ACA) model in the context of insurance. In the following, the standard model is presented as well as a new model called GEVACA model.

2.2.1 Autoregressive Conditional Amount Model (ACA)

Let y_t represents the claims amount (loss) paid at the day t . Following Engle's ACD(p,q) model idea, the ACA model assumes that y_t follows a multiplicative structure:

$$y_t = \psi_t \epsilon_t, \quad (2.2)$$

$$\epsilon_t \sim \text{i.i.d non negative variable with } E(\epsilon_t) = 1. \quad (2.3)$$

The conditional expected loss is:

$$\psi_t = E(y_t | \Omega_{t-1}),$$

where Ω_{t-1} is the set of the past observations until the time $t - 1$. Engle and Russell (1998) specified an autoregressive structure for this conditional mean:

$$\psi_t = \omega + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{j=1}^q \beta_j \psi_{t-j}, \quad (2.4)$$

where $\omega, \alpha_i, \beta_j > 0$ to insure that the amount is positive and $\sum(\alpha_i + \beta_j) < 1$ for stability purpose. Consequently, ψ_t depends on the p past observations and q past expected losses.

The parameters of the ACA model are estimated by maximizing the log-likelihood function. Let $f_{\theta_\epsilon}(\epsilon_t)$ be the density function of ϵ_t with parameters θ_ϵ . Let $g_\theta(y_t | \Omega_{t-1})$ be the probability density function of y_t knowing the data observations available up to time $t - 1$. Since the standardized claims amounts $\frac{y_t}{\psi_t} = \epsilon_t \sim \text{i.i.d.} f_{\theta_\epsilon}, t = 1, \dots, T$, this implies that:

$$g_\theta(y_t | \Omega_{t-1}) = \frac{1}{\psi_t} f_{\theta_\epsilon} \left(\frac{y_t}{\psi_t} \right),$$

where θ is the vector of all the unknown parameters including θ_ϵ as well as the parameters in ψ_t . The log-likelihood function is then given by:

$$l(\theta; y_1, \dots, y_T) = \sum_{t=1}^T \log [g_\theta(y_t | \Omega_{t-1})] = \sum_{t=1}^T \left[\log f_{\theta_\epsilon} \left(\frac{y_t}{\psi_t} \right) - \log(\psi_t) \right]. \quad (2.5)$$

ACA model can be estimated with Exponential, Weibull and Gamma distributions for ϵ (see Appendix 2.7). In the following subsection, an alternative model is presented.

2.2.2 Autoregressive Conditional Amount Model (ACA)

Let y_t represents the claims amount (loss) paid at the day t . Following Engle's ACD(p,q) model idea, the ACA model assumes that y_t follows a multiplicative structure:

$$y_t = \psi_t \epsilon_t, \quad (2.6)$$

$$\epsilon_t \sim \text{i.i.d non negative variable with } E(\epsilon_t) = 1. \quad (2.7)$$

The conditional expected loss is:

$$\psi_t = E(y_t | \Omega_{t-1}),$$

where Ω_{t-1} is the set of the past observations until the time $t - 1$. Engle and Russell (1998) specified an autoregressive structure for this conditional mean:

$$\psi_t = \omega + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{j=1}^q \beta_j \psi_{t-j}, \quad (2.8)$$

where $\omega, \alpha_i, \beta_j > 0$ to insure that the amount is positive and $\sum(\alpha_i + \beta_j) < 1$ for stability purpose. Consequently, ψ_t depends on the p past observations and q past expected losses.

The parameters of the ACA model are estimated by maximizing the log-likelihood function. Let $f_{\theta_\epsilon}(\epsilon)$ be the density function of ϵ with parameters θ_ϵ . Let $g_\theta(y_t | \Omega_{t-1})$ be the probability density function of y_t knowing the data observations available up to time $t - 1$. Since the standardized claims amounts $\frac{y_t}{\psi_t} = \epsilon_t \sim \text{i.i.d. } f_{\theta_\epsilon}, t = 1, \dots, T$, this implies that:

$$g_\theta(y_t | \Omega_{t-1}) = \frac{1}{\psi_t} f_{\theta_\epsilon} \left(\frac{y_t}{\psi_t} \right),$$

where θ is the vector of all the unknown parameters including θ_ϵ as well as the parameters in ψ_t . The log-likelihood function is then given by:

$$l(\theta; y_1, \dots, y_T) = \sum_{t=1}^T \log [g_\theta(y_t | \Omega_{t-1})] = \sum_{t=1}^T \left[\log f_{\theta_\epsilon} \left(\frac{y_t}{\psi_t} \right) - \log(\psi_t) \right]. \quad (2.9)$$

ACA model can be estimated with Exponential, Weibull and Gamma distributions for ϵ (see Appendix 2.7). In the following subsection, an alternative model is presented.

2.2.3 Generalized Extreme Value ACA Model

Insurance companies are undergoing major changes and are increasingly exposed to catastrophic events that induce large losses. Extreme value theory yields methods for quantifying such events in statistically optimal way (see Embrechts et al. (1999)). However, it is often plausible to model large insurance claims as i.i.d. variables. But, claims amounts appear to have high correlations. The aim in this subsection is to introduce a new model called Generalized Extreme Value ACA Model (GEVACA),

which takes into account the temporal dependence of large losses. Indeed, a new specification is developed, based on ACA model assuming that standardized claims amounts follow a Generalized Extreme Value (GEV) distribution, with mean equal one.

The cumulative distribution function of a variable following the GEVACA model is:

$$G_{\theta}(y_t | \Omega_{t-1}) = \exp \left\{ - \left[1 + \frac{(\frac{y_t}{\psi_t} - \mu) [1 - \Gamma(1 - \xi)]}{(\mu - 1)} \right]^{-1/\xi} \right\}, \quad (2.10)$$

where μ is the location parameter and ξ is the shape parameter, and its probability density function is given by:

$$g_{\theta}(y_t | \Omega_{t-1}) = \frac{1 - \Gamma(1 - \xi)}{\xi(\mu - 1)\psi_t} \left[1 + \frac{(\frac{y_t}{\psi_t} - \mu)(1 - \Gamma(1 - \xi))}{(\mu - 1)} \right]^{-\frac{1}{\xi} - 1} \exp \left\{ - \left[1 + \frac{(\frac{y_t}{\psi_t} - \mu)[1 - \Gamma(1 - \xi)]}{(\mu - 1)} \right]^{-\frac{1}{\xi}} \right\} \quad (2.11)$$

See Appendix 2.8 for the proof of deriving these expressions. Consequently, the log-likelihood of the GEVACA model is then formalized as follows:

$$\begin{aligned} l(\theta; y_1, \dots, y_T) &= \sum_{t=1}^T \{ \log[1 - \Gamma(1 - \xi)] - \log[(\mu - 1)\psi_t\xi] \\ &\quad + \left(-\frac{1}{\xi} - 1 \right) \log \left(1 + \frac{(\frac{y_t}{\psi_t} - \mu)[1 - \Gamma(1 - \xi)]}{(\mu - 1)} \right) \\ &\quad - 1 - \frac{1}{(\mu - 1)} \left\{ \left(\frac{y_t}{\psi_t} - \mu \right) [1 - \Gamma(1 - \xi)] \right\}^{-\frac{1}{\xi}} \}. \end{aligned} \quad (2.12)$$

2.2.4 Implementing the mean squared errors in the case of time series

In order to check the estimates accuracy, the predictions precision is quantified via a cross validation procedure. It is a technique for assessing how accurately a predictive model will perform the data. Firstly, the claims amounts data are partitioned into two complementary samples: 70% (y_1, \dots, y_{t^*}) called the training sample, and 30% (y_{t^*+1}, \dots, y_T) called validation sample. Secondly, the parameters ($\hat{\omega}, \hat{\alpha}, \hat{\beta}$) of the ACA model are estimated on the training sample. Then, the conditional expected loss $\hat{\psi}_t$ is predicted for the period ($t^* + 1, \dots, T$). Finally, the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE) are computed on the validation sample given as follows:

$$MSE = \frac{1}{T - t^*} \sum_{t=t^*+1}^T (y_t - \hat{y}_t)^2, \quad (2.13)$$

$$RMSE = \sqrt{MSE}, \quad (2.14)$$

where

$$\hat{y}_t = \hat{\psi}_t = \hat{\omega} + \sum_{i=1}^p \hat{\alpha}_i y_{t-i} + \sum_{j=1}^q \hat{\beta}_j \hat{\psi}_{t-j}. \quad (2.15)$$

2.3 VaR ACA measure

One of the main goals of modelling the claim amount is to estimate the daily coverage amount for claims. This means to find a proper estimate of capital that allows insurers to be solvent in future years. Solvency 2 framework requires the use of Value at Risk (VaR) measure at a high level (99,5%). Our contribution is to develop a VaR ACA measure to forecast the coverage amount of claims.

The claims amounts are supposed to follow a parametric ACA distribution $G_\theta(y_t | \Omega_{t-1})$ depending on a parametric distribution for ϵ_t , denoted by $F_{\theta_\epsilon}(\epsilon)$. The Value at Risk forecast at the confidence level α and at time $t + 1$, denoted $VaR_{t+1}(\alpha)$, is defined by:

$$\Pr(y_{t+1} < VaR_{t+1}(\alpha) | \Omega_t) = \alpha. \quad (2.16)$$

Thus, $\Pr(\psi_{t+1}\epsilon_{t+1} < VaR_{t+1}(\alpha) | \Omega_t) = \alpha$ and $\Pr(\epsilon_{t+1} < \frac{VaR_{t+1}(\alpha)}{\psi_{t+1}} | \Omega_t) = \alpha$.

Consequently, it is straightforward to derive the VaR ACA:

$$VaR_{t+1}(\alpha) = \psi_{t+1} F_{\theta_\epsilon}^{-1}(\alpha). \quad (2.17)$$

To compute the estimated VaR ACA, denoted $\widehat{VaR}_{t+1}(\alpha)$, the unknown values ψ_{t+1} and θ_ϵ are replaced by their forecast $\hat{\psi}_{t+1}$ and $\hat{\theta}_\epsilon$:

$$\widehat{VaR}_{t+1}(\alpha) = \hat{\psi}_{t+1} F_{\hat{\theta}_\epsilon}^{-1}(\alpha), \quad (2.18)$$

From this expression, VaR ACA forecast can be deduced for various distributions for ϵ_t :

- VaR Exponential ACA (VaR EACA):

$$\widehat{VaR}_{t+1}(\alpha) = -\hat{\psi}_{t+1} \log(1 + \alpha). \quad (2.19)$$

- VaR Weibull ACA (VaR WACA):

$$\widehat{VaR}_{t+1}(\alpha) = -\hat{\psi}_{t+1} \frac{\log(1 - \alpha)^{1/\hat{k}}}{\Gamma(1 + 1/\hat{k})}, \quad (2.20)$$

where \hat{k} is the estimated shape parameter of Weibull distribution.

- VaR Gamma ACA (VaR GACA) :

$$\widehat{VaR}_{t+1}(\alpha) = \hat{\psi}_{t+1} G^{-1}(\alpha, \hat{k}), \quad (2.21)$$

where G^{-1} is the inverse of the Gamma function with shape parameter \hat{k} .

- VaR GEVACA :

One of the main objectives of this paper is to develop a methodology to obtain an expression for the Value at Risk of the GEVACA model for one day ahead predictive distribution of the claims (see Appendix 2.8).

$$\widehat{\text{VaR}}_{t+1}(\alpha) = \hat{\psi}_{t+1} \left\{ \frac{(\hat{\mu} - 1)}{1 - \Gamma(1 - \hat{\xi})} \left[\log(\alpha)^{-\hat{\xi}} - 1 \right] + \hat{\mu} \right\}, \quad (2.22)$$

where $\hat{\xi}$ is the estimated shape parameter and $\hat{\mu}$ is the estimated location parameter.

2.4 Empirical Analysis

2.4.1 Data description

Data were supplied by a French insurance company and consist of daily claims amounts of two lines of business Auto Damage and Auto Liability, over the period January 2002, to December 2007. We have 2061 and 2038 observations of claims amounts (in €), respectively of the Auto Damage and Auto Liability. The first guarantee covers all damaged caused to the car, even if the insured is responsible of the accident. The second is the guarantee of damage caused to others, belongs to the Civil Responsibility lines of business. The claims amounts are net of reinsurance, net of subrogation and the inflation is supposed to remain unchanged throughout the years. Focusing on the daily claims amounts of these two lines, the occurrences of claims are irregular. Indeed, claims amounts are observed daily but the interval is variable. As for instance, insurance company can pay today a total amount of claims equal to 1000€, after one, two or more days, it may pay a total amounts equal to 800€.

In the following, some summary statistics, the plots of the claims amounts corresponding to the two lines of business as well their autocorrelation functions are presented.

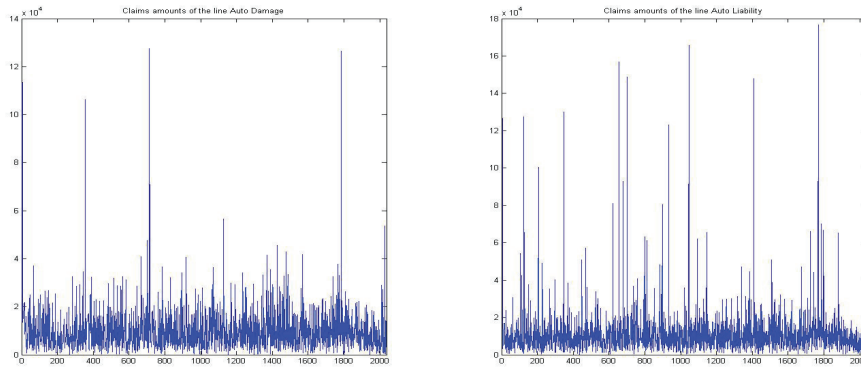


Figure 2.1: plots of the lines Auto Damage and Auto Liability

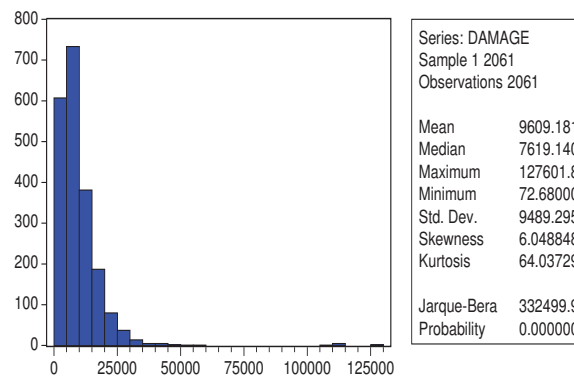


Figure 2.2: Histogram and Descriptive Statistics of the line Auto Damage

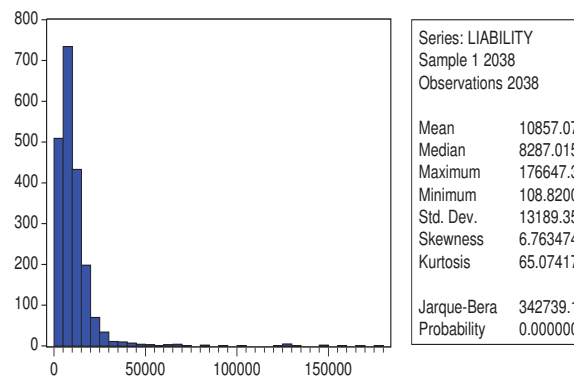


Figure 2.3: Histogram and Descriptive Statistics of the line Auto Liability

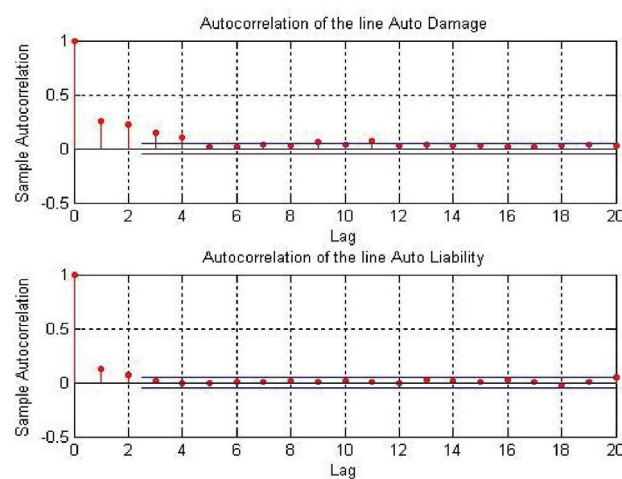


Figure 2.4: Autocorrelations of the lines Auto Damage and Auto Liability

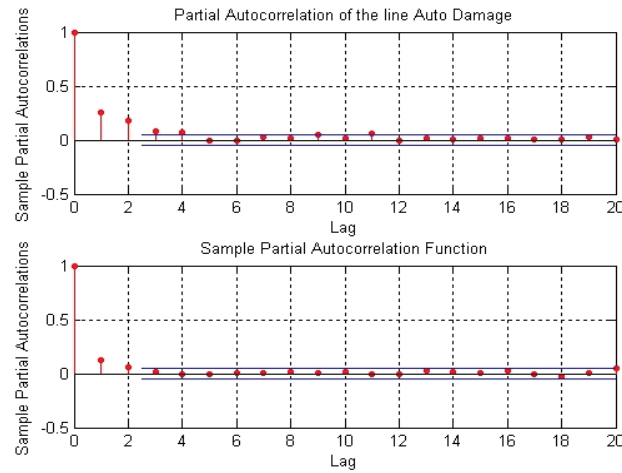


Figure 2.5: Partial Autocorrelations of the lines Auto Damage and Auto Liability

According to Figure 2.1, a phenomenon of claims amounts clustering can be observed: large (respectively small) amounts tend to be followed by large (respectively small) amounts. This stylized fact justifies the use of our autoregressive framework.

Moreover, claims amounts distributions of figure 2.3 reveals that both lines of business exhibit high positive skewness and excess kurtosis indicating that distributions are shifted to the left. The Jarque Bera test of normality rejects the normality of claims amounts distributions. Thus, our data sets are very heavy-tailed.

Furthermore, Figure 2.4 and 2.5 depict the correlograms of the Autocorrelations and the Partial autocorrelations of claims amounts. As can be seen, the correlations are not large in magnitude, but they clearly indicate serial dependence in the data. Also, the autocorrelations are positive, indicating that large or small claims amounts tend to occur together. This is consistent with the features usually exhibited by insurance company. In fact, this may be due to the arrival process of claims or the covered system adopted by the company. Indeed, due to weather events, claims may occur simultaneously. Also, insurance company may pay the large (respectively the small) losses in the same period.

These features of claims amounts challenge us to apply econometric techniques. For that, data stationary is checked using a bootstrap Augmented Dickey Fuller (ADF) test for a unit root, and dynamic process is estimated using the ACA models.

For the bootstrap Augmented Dickey Fuller test: a parametric version based on Gaussian distribution (which should not be appropriate in our case, as classical ADF test based on Dickey Fuller distribution), and a non parametric version based on the residuals distribution adjusted from mean, standard deviation and heteroscedasticity (which should fit accurately the final ADF test distribution). Two p-values are also calculated: classical unilateral p-value, and bilateral p-value (see de Peretti (2003)).

Number of bootstrap replications	Number of augmentations in the ADF regression	Residual independence	P value bootstrap (Unilateral P values, Bilateral P values)					
			Parametric bootstrap			non parametric bootstrap		
			AR(1)	AR(1)+c	AR(1)+c+t	AR(1)	AR(1)+c	AR(1)+c+t
999	0	Reject	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000
999	1	OK	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000
999	2	OK	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000

Table 2.1: Bootstrap Dickey Fuller test for the line Auto Damage

Number of bootstrap replications	Number of augmentations in the ADF regression	Residual independence	P value bootstrap (Unilateral P values, Bilateral P values)					
			Parametric bootstrap			non parametric bootstrap		
			AR(1)	AR(1)+c	AR(1)+c+t	AR(1)	AR(1)+c	AR(1)+c+t
999	0	Reject	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000
999	1	OK	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000
999	2	OK	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000	0.000, 0.000

c: constant
t: trend

Table 2.2: Bootstrap Dickey Fuller test for the line Auto Liability

Lines of Business	Statistic	P-value
Auto Damage	308.7	0.00
Auto Liability	42.2	6.910^{-9}

Table 2.3: Ljung Box Test

Results of table 2.1 and table 2.2 give evidence for stationarity of the two series. From table 2.4.1, it can be seen that claims amounts of both Auto damage and Auto Liability, exhibit a very large Ljung Box statistics. The null hypothesis of white noise is easily rejected for both lines of business based on the critical value 18.31.

To highlight the dynamic behaviour of these claims amounts, Autoregressive Conditional amount model is estimated in the following subsection.

2.4.2 Estimation Results with ACA model

In order to estimate claims amounts of the Auto Damage and Auto Liability with ACA model, various specification are considered at different lags : Exponential ACA (EACA), Weibull ACA (WACA), Gamma ACA (GACA) and Generalized extreme value ACA (GEVACA).

For each model, log-likelihood function is maximized to obtain parameters estimates. Only consistent models, which have significant parameters are presented. Parameters estimates with their Student test statistics (inside parenthesis) are reported in tables 2.4 and 2.5. We note that for the Ljung Box and Lagrange Multiplier (LM) test the statistics value and their p-value (inside parenthesis) are computed with 10 lags.

	EACA(1,1)	WACA(1,1)	GACA(1,1)	EACA(1,2)	WACA(1,2)	GACA(1,2)	GEVACA(1,1)
ω	0.042 [2.52] **	0.061 [6.16] ***	0.061 [4.3] ***	0.046 [4.61] ***	0.05 [4.8] ***	0.06 [26.6] ***	0.03 [16.18] ***
α_1	0.083 [6.78] ***	0.093 [10.15] ***	0.039 [7.2] ***	0.073 [4.7] ***	0.079 [6.33] ***	0.04 [17.7] ***	0, 03 [16.9] ***
β_1	0.91 [79.1] ***	0.9 [105.7] ***	0.96 [60.13] ***	0.635 [2.97] ***	0.678 [3.75] ***	0.75 [28.9] ***	0.97 [9.8] ***
β_2	-	-	-	0.29 [1.42] ***	0.24 [1.98] **	0.2 [13] ***	-
κ	-	1.3 [64.05] ***	1.7 [60.2] ***	-	1.316 [63.9] ***	1.7	-
ξ	-	-	-	-	-	-	0.24 [48.4] ***
μ	-	-	-	-	-	-	0, 596 [21.3] ***
Mean of ϵ_t	0.99	0.993	1.03	1.006	0.999	1.3	0.99
Variance	0.81	0.82	1.14	0.85	0.83	1.6	1.43
LJB Q(10)	21.17 (0.02)	24.5 (0.05)	10.32 (0.41)	25.45 (0.005)	21.7 (0.016)	39.4 (0.000)	65 (0.000)
LM (10)	22.7 (0.01)	26.13 (0.004)	10.02 (0.43)	15.33 (0.12)	17.12 (0.1)	11.3 (0.34)	5.6 (0.85)

[.] contains the corresponding Student test statistics,

(.) contains the corresponding p-values,

** : 5% significant statistics,

*** : 1% significant statistics.

Table 2.4: Estimation Results of the line Auto Damage

	EACA(1,1)	WACA(1,1)	GACA(1,1)	EACA(1,2)	WACA(1,2)	GACA(1,2)	GEVACA(1,1)
ω	0.054 [54.9] ***	0.052 [52.1] ***	0.055 [50.3] ***	0.059 [59.4] ***	0.052 [2.29] **	0.052 [37.9] ***	0.053 [4.2] ***
α_1	0.057 [7.5] ***	0.063 [9.27] ***	0.058 [5.13] ***	0.053 [5.11] ***	0.059 [6.41] ***	0.04 [44.8] ***	0.049 [2.8] ***
α_2	-	-	-	-	-	-	-
β_1	0.94 [131.18] ***	0.939 [149.07] ***	0.93 [96.7] ***	0.48 [2.3] **	0.45 [2.59] ***	0.45 [21.5] ***	0.95 [32.1] ***
β_2	-	-	-	0.46 [2.3] **	0.49 [2.86] ***	0.48 [67.5] ***	-
κ	-	1.16 [68.6] ***	1.6	-	1.18 [68.8] ***	1.7 [46.2] ***	-
ξ	-	-	-	-	-	-	0.3 [5.6] ***
μ	-	-	-	-	-	-	0.6 [9] ***
Mean of ϵ_t	0.999	0.99	1.007	1.008	0.998	1.55	1.03
Variance	1.52	1.47	1.64	1.57	1.51	3.56	1.67
LJB Q(10)	13.01 (0.22)	12.51 (0.25)	12.22 (0.27)	18.06 (0.04)	14.84 (0.13)	9.59 (0.47)	8.34 (0.6)
LM (10)	12.19 (0.27)	12.84 (0.98)	13.62(0.19)	7.22 (0.7)	7.98 (0.63)	9.15 (0.51)	9.4 (0.49)

[.] contains the corresponding Student test statistics,

(.) contains the corresponding p-values,

** : 5% significant statistics,

*** : 1% significant statistics.

Table 2.5: Estimation Results of the line Auto Liability

For the two lines of business, we remark that all coefficients of various models are significant. This observation is supported by the large t -statistics observed for all models on the parameters α_i and β_i , and $\sum(\alpha_i + \beta_i) < 1$, which allow the existence of conditional expected claims amounts.

The performance of the ACA models in detecting autocorrelations structure of loss payments, can be evaluated by examining the residuals (called the standardized claims amounts). Results reveal that residuals have a mean of approximately one.

Furthermore, it is clearly evident that the results of the Ljung Box test given in tables 2.4 and 2.5 show that the autocorrelations are significantly reduced for all ACA models.

From table 2.4 results of the line Auto Damage indicate that GACA(1,1) outperforms all the other models. Indeed, it is remarkable to see its high p-value for the Ljung Box test and regarding to LM test, the statistic exhibits no ARCH effect in residuals for this model.

In contrast, GEVACA(1,1) exhibits a p-value of Ljung Box test close to zero, indicating that this model fail to capture autocorrelations between residuals and is not adequate for the Auto Damage line of business. It is reasonable, because this line is related to the material damage of the car that does not present a large claims amounts.

For the Auto Liability, as shown in table 2.5, all models success to capture autocorrelations. There is no autocorrelations and no ARCH effect in the standardized claims amounts. Specifically, the GEVACA(1,1) exhibits a less significant Ljung Box and LM statistics than other models. Hence, this model could provide a more accurate description of claims behaviour than other models.

To select the appropriate model for the data, we are based on the log-likelihood function (LL), AIC, BIC criteria and RMSE. Results are summarized in table 2.6.

Model	Auto Damage				Auto Liability			
	LL	AIC	BIC	RMSE	LL	AIC	BIC	RMSE
EACA(1,1)	20942	-41878	-41861	6272	20962	-41932	-41949	6415
EACA(1,2)	20905	-41804	-41781	6256	20930	-41870	-41892	6417
WACA(1,1)	20810	-41613	-41591	6286	20917	-41842	-41865	6428
WACA(1,2)	20777	-41544	-41516	6261	20876,15	-41762	-41790	6422
GACA(1,1)	21138 *	-42268 *	-42245 *	6226 *	21061	-42114	-42092	6370
GACA(1,2)	20888	-41766	-41738	6318	18373	-36736	-36708	6775
GEVACA(1,1)	20727	-41442	-41408	6440	21110 *	-42208 *	-42174 *	5705 *

* indicates best model.

Table 2.6: Selection criteria for ACA models for the line Auto Damage and Auto Liability

According to table 2.6, GACA(1,1) model is the best one for modelling claims amounts for the line of business Auto Damage. Indeed, GACA(1,1) provides considerably higher log-likelihood value than other models. Furthermore, basing on information criteria, this model minimizes the AIC and BIC criteria. So, it outperforms all the other models and gives a better fit for the line Auto Damage.

For the line Auto liability, GEVACA(1,1) model maximizes the log-likelihood and minimizes the information criteria. Thus, this model is the most appropriate model for this line of business. This explained by the fact, that GEVACA model captures the temporal dependence between the large claims amounts, which are related to the corporal damage and all the consequences of accident on the health of a person.

In addition, we note that GACA(1,1) and GEVACA(1,1) present the smallest value of RMSE respectively for the Auto Damage and Auto Liability. Hence, they give more precision for forecasting claims amounts. Again, results favor these two models and confirm that our ACA approach might be appropriate.

2.4.3 Comparison with ARMA model

In the actuarial literature, the model used to detect the temporal dependence between claims amounts is particularly the Autoregressive Moving Average models (ARMA)(see Gerber (1982), Promislow (1991), Zhang (2005) and Cummins (1985)). To show the relevance of our ACA models in modelling claims amounts, we compare them to ARMA models. Results are reported in Table 2.7.

Auto Damage		Auto Liability	
ARMA(1,1)		ARMA(1,1)	
c	1563.13 (44.7) ***	c	2640.59 (38.2) ***
ϕ_1	0.83 (46.2) ***	ϕ_1	0.75 (19.65) ***
θ_1	-0.79 (-32.5) ***	θ_1	-0.74 (-17.6) ***
LJB Q(10)	11.56(0.32)	LJB Q(10)	3.67 (0.96)
LM Q(10)	8, 39(0.59)	LM Q(10)	2 (0.99)
RMSE	8021	RMSE	7128

Table 2.7: Estimation Results of the Auto Damage and the Auto Liability with ARMA(1,1)

From Table 2.7, it can be seen that ARMA(1,1) captures the autocorrelations between claims of the two lines of business. Indeed, Ljung Box statistics are < 18.3 , and there is no ARCH effect in the residuals (LM statistics < 18.3). Comparing Tables 2.7 and 2.6, the RMSE of the GACA(1,1) and GEVACA(1,1) are significantly lower than the RMSE of ARMA models, both for the line Auto Damage and Auto Liability. So, we can conclude that our GACA and GEVACA models are more appropriate than the ARMA models used in the actuarial literature.

Focusing on the residuals, the Quantile versus Quantile plot (QQplot) of residuals of ACA type and ARMA models versus respectively Gamma, GEV and the standard normal distribution are presented in the Appendix 4.7 .

As we can see from Figure 2.8 the QQplot of the standardized claims of GACA(1,1) of the Auto Damage against the assumed distribution Gamma with shape parameter 1.7 and scale equal to 1, is more close than the residuals of ARMA model against the standard normal distribution (same interpretation for the Auto Liability). It can be concluded that due to the normality assumption of the ARMA models, these models fail to describe accurately the dynamic behaviour of the claims, and our ACA models seem to be much more appropriate to forecasting the claims amounts.

In the next part, coverage amount is computed using the VaR measure, and a backtesting procedure is performed to show the validity of the $VaR_{99.5\%}$. Here, a conditional coverage test suggesting the null hypothesis H_0 : VaR model is correctly specified for coverage rate α at one day ahead, is used. For more details about this test, see Christoffersen (1998).

The statistic (LR_{cc}) and the p-value of the test for each ACA and ARMA models are presented in table 2.8. $VaR_{99.5\%}$ GACA(1,1) and $VaR_{99.5\%}$ GEVACA(1,1) respectively for Auto Damage and Auto Liability are illustrated in figure 2.6 and 2.7.

	EACA(1,1)	WACA(1,1)	GACA(1,1)	EACA(1,2)	WACA(1,2)	GACA(1,2)	GEVACA(1,1)	ARMA(1,1)
Auto Damage								
LR_{cc}	$8.38 \cdot 10^{-6}$	2.34	1.12	0.37	3.9	8.13	0.39	130.47
p-value	0.99	0.13	0.29	0.54	0.05	0.02	0.8	0
Auto Liability								
LR_{cc}	0.03	2.34	5.76	0.32	3.9	81.39	0.03	27.99
p-value	0.98	0.31	0.05	0.85	0.14	0	0.98	$8.33 \cdot 10^{-8}$

Table 2.8: Backtesting of $VaR_{99.5\%}$ of the lines Auto Damage and Auto Liability

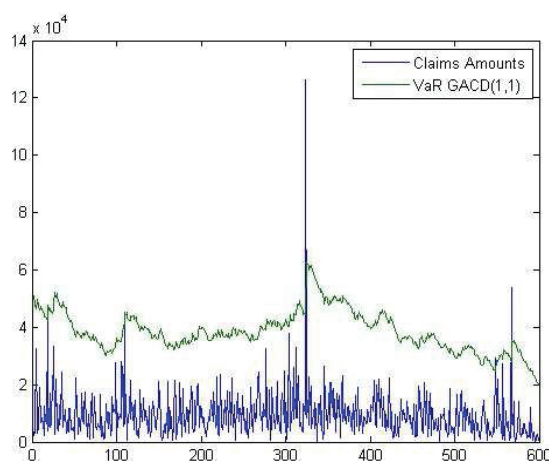


Figure 2.6: $VaR_{99.5\%}$ GACA(1,1) of the line Auto Damage

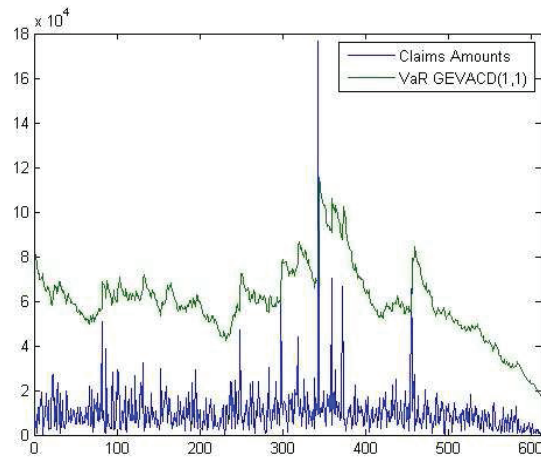


Figure 2.7: $VaR_{99.5\%}GEVACA(1,1)$ of the line Auto Liability

From table 2.8, for all $VaR_{99.5\%}$ ACD models except $VaR_{99.5\%}$ GACA(1,2) for the two lines of business, we accept the null hypothesis of the conditional coverage test. The LR_{cc} statistics are lower than the critical value = 5.99 at significance level 5%. The p-value are significantly higher than 5%, that reflect the validity of the $VaR_{99.5\%}$ ACA, therefore a good estimation of the risks.

From results of VaR ARMA for the two lines of business, the efficiency of the VaR ARMA forecasts is rejected. Indeed, the LR_{cc} statistics are higher than the critical value. So, VaR ARMA underestimates the risk, and gives a coverage amount lower than risks supported by the company.

From figure 2.6 and figure 2.7, we remark that $VaR_{99.5\%}$ GACA(1,1) and $VaR_{99.5\%}$ GEVACA(1,1) cover the claims of respectively, the lines Auto Damage and Auto Liability. Therefore, the evaluation of coverage amounts with these two models could ensure the solvency in the insurance company. Results are interesting, and actuaries may evaluate the amount of capital based on ACA models.

2.5 Conclusion

This paper proposes to analyse the dynamic behaviour of the claims amounts in the insurance company. Indeed, claims are viewed as inter-temporally correlated events. We treat dependence between the claims amounts as a stochastic process. For that purpose, we propose different ACA models and applied them to capture the autocorrelations for two lines of business, Auto Damage and Auto Liability. The Ljung Box test of 10 lags suggest that ACA models are more appropriate to take into account the temporal dependence between claims amounts than classical ARMA models presented in actuarial literature. Unfortunately, the ACA model may be too restrictive when claims amounts involve some extreme amounts.

To solve this problem, we introduced a new specification of a Generalized extreme value version of ACA model, which is denoted the GEVACA model. This specification looks very promising in modelling the line Auto liability, and allowed us to highlight the dynamic of extreme claims amounts.

Within this framework, our model can be viewed as a model explaining the variation of the loss payment process. The fluctuations in payment are explained by the past behaviour of the claims (autoregressive part of the model). The GACA(1,1) and GEVACA(1,1) model are respectively selected for the lines Auto Damage and Auto Liability because: they fit the data better than other ACA models.

Comparing our GACA and GEVACA models to ARMA models used in the literature, we remark that our selected models are more adequate than the ARMA models in fitting data.

Furthermore, as innovative, we derive a parametric ACA-VaR measure to evaluate the coverage amount of claims, and we observe that ACA-VaR measure provides an accurate estimation of risks.

Acknowledgement

The authors would like to thank Mr.Frederic Planchet professor at Institute of Financial and Insurance Sciences for useful comments to revise the present work.

2.6 Appendix

2.7 Model specifications

2.7.1 Exponential ACA model (EACA)

Engle and Russell (1998), introduced the EACD model as, $y_t = \psi_t \epsilon_t$ with $\epsilon_t \sim \exp(1)$. So $P(Y_t < y_t) = P(\psi_t \epsilon_t < y_t) = P(\epsilon_t < \frac{y_t}{\psi_t}) = 1 - \exp(-\frac{y_t}{\psi_t})$. Thus, the density function of the conditional claims amounts is:

$$f(y_t | \Omega_{t-1}) = \frac{1}{\psi_t} \exp(-\frac{y_t}{\psi_t}). \quad (2.23)$$

The log-likelihood of the model is given by:

$$L(y | \Omega_{t-1}, \theta) = - \sum_{t=1}^T [\log(\psi_t) + \frac{y_t}{\psi_t}], \quad (2.24)$$

where $y = (y_1, \dots, y_T)$.

2.7.2 Weibull ACA model (WACA)

Engle and Russel assume a model that the residuals follow a Weibull distribution. Following them, we define the density function of WACA as:

$$f(y_t | \Omega_{t-1}, \theta) = \left(\frac{y_t \Gamma(1 + \frac{1}{k})}{\psi_t}\right)^k \frac{k}{y_t} \exp\left(-\left(\frac{y_t \Gamma(1 + \frac{1}{k})}{\psi_t}\right)^k\right), \quad (2.25)$$

and the log-likelihood function is:

$$L(y | \Omega_{t-1}, \theta) = \sum_{t=1}^T \left[k \log\left(\frac{\Gamma(1 + \frac{1}{k}) y_t}{\psi_t}\right) + \log\left(\frac{k}{y_t}\right) - \left(\frac{y_t \Gamma(1 + \frac{1}{k})}{\psi_t}\right)^k \right], \quad (2.26)$$

where k is the shape parameter of the Weibull distribution.

2.7.3 Gamma ACA model (GACA)

The GACA model is a particular case of the Generalized Gamma Model introduced by Lunde (1999), so $P(Y_t < y_t) = P(\psi_t \epsilon_t < y_t) = P(\epsilon_t < \frac{y_t}{\psi_t}) = \frac{\Gamma(k, \frac{y_t}{\psi_t})}{\Gamma(k)}$. The density function and the log-likelihood are calculated as:

$$f(y_t | \Omega_{t-1}) = \frac{y_t^{k-1} k^k}{\psi_t^k \Gamma(k)} \exp\left(-\frac{y_t k}{\psi_t}\right), \quad (2.27)$$

and

$$L(y | \Omega_{t-1}, \theta) = \sum_{t=1}^T \left((k-1) \log(y_t) + k \log(k) - k \log(\psi_t) - \log(\Gamma(k)) - \frac{y_t k}{\psi_t} \right), \quad (2.28)$$

where k is the shape parameter of the Gamma distribution.

2.8 Proof

This appendix presents the proof of deriving the expression of the GEVACA distribution, GEVACA density and GEVACA-VaR model.

$$P(Y_t < y_t) = P(\psi_t \epsilon_t < y_t) = P(\epsilon_t < \frac{y_t}{\psi_t}) = \exp\left\{-\left[1 + \xi \left(\frac{y_t - \mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right\}$$

where

$\mu \in \mathfrak{R}$: location parameter,

$\sigma > 0$: scale parameter,

$\xi \in \mathfrak{R}$: shape parameter.

$$\text{We have } E(\epsilon_t) = 1 \Rightarrow \mu + \sigma \frac{\Gamma(1 - \xi) - 1}{\xi} = 1, \text{ so } \sigma = \frac{(\mu - 1)\xi}{1 - \Gamma(1 - \xi)}.$$

The distribution of the GEVACA model is given by:

$$F(y_t | \Omega_{t-1}) = \exp\left\{-\left[1 + \frac{(\frac{y_t}{\psi_t} - \mu)(1 - \Gamma(1 - \xi))}{(\mu - 1)}\right]^{\frac{-1}{\xi}}\right\}, \quad (2.29)$$

and the density function of GEVACA is:

$$\frac{\partial F(y_t | \Omega_{t-1})}{\partial y_t} = \frac{1 - \Gamma(1 - \xi)}{\xi(\mu - 1)\psi_t} \left[1 + \frac{(\frac{y_t}{\psi_t} - \mu)(1 - \Gamma(1 - \xi))}{(\mu - 1)}\right]^{\frac{-1}{\xi} - 1} \exp\left\{-\left[1 + \frac{(\frac{y_t}{\psi_t} - \mu)(1 - \Gamma(1 - \xi))}{(\mu - 1)}\right]^{\frac{-1}{\xi}}\right\}$$

The expression of the VaR GEVACA model is obtained as:

$$P(y_{t+1} < VaR_{t+1} | \Omega_t) = \alpha,$$

$$P(\hat{\psi}_{t+1} \epsilon_{t+1} < VaR_{t+1} | \Omega_t) = \alpha,$$

$$P(\epsilon_{t+1} < \frac{VaR_{t+1} | \Omega_t}{\hat{\psi}_{t+1}} | \Omega_t) = \alpha,$$

$$F\left(\frac{VaR_{t+1} | \Omega_t}{\hat{\psi}_{t+1}}\right) = \alpha,$$

where F is the GEVACA distribution.

$$\widehat{VaR}_{t+1|\Omega_t}(\alpha) = F^{-1}(\alpha)\hat{\psi}_{t+1}, \text{ with } F^{-1} \text{ is the inverse of the GEVACA distribution.}$$

We have:

$$\exp\left\{-\left[1 + \frac{(\frac{VaR_{t+1|\Omega_t}(\alpha)}{\hat{\psi}_{t+1}} - \hat{\mu})(1 - \Gamma(1 - \hat{\xi}))}{(\hat{\mu} - 1)}\right]^{\frac{-1}{\hat{\xi}}}\right\} = \alpha$$

$$\left[1 + \frac{\left(\frac{VaR_{t+1|\Omega_t}(\alpha) - \hat{\mu}}{\hat{\psi}_{t+1}}\right)(1 - \Gamma(1 - \hat{\xi}))^{-1}}{\hat{\mu} - 1}\right]^{\frac{-1}{\hat{\xi}}} = -\text{Log}(\alpha)$$

$$\widehat{VaR}_{t+1|\Omega_t}(\alpha) = \hat{\psi}_{t+1} \left[\frac{(\hat{\mu} - 1)}{1 - \Gamma(1 - \hat{\xi})} (\text{Log}(\alpha))^{-\hat{\xi}} - 1 \right] + \hat{\mu}$$

2.9 QQ plots

Figure 2.8: QQplot of standardized claims amounts of the line Auto Damage versus Gamma(1,1.7)

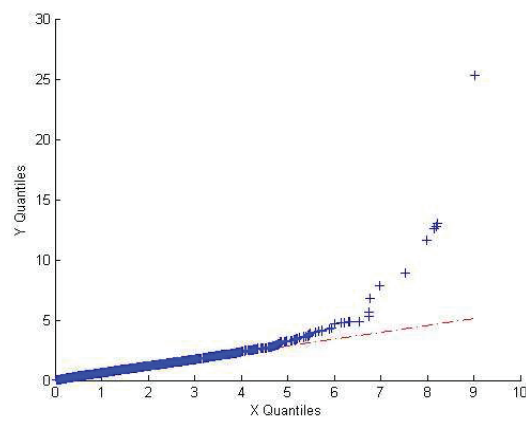


Figure 2.9: QQplot of standardized claims amounts of the line Auto Liability versus GEV(0.3,1,0.6)

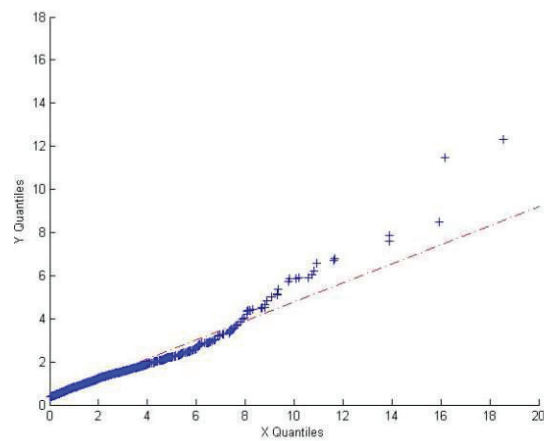


Figure 2.10: QQplot of Residuals from ARMA(1,1) of the Auto Damage

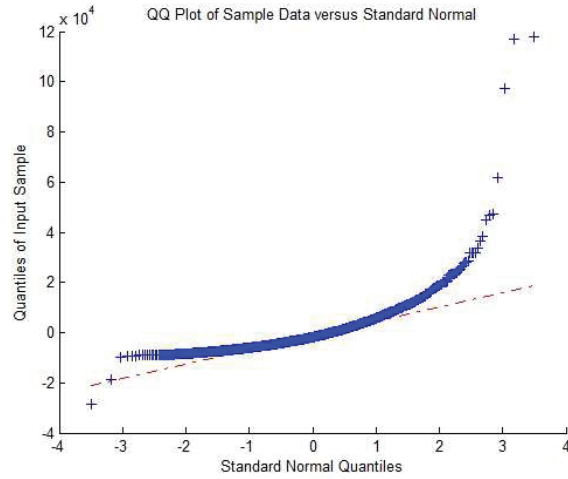
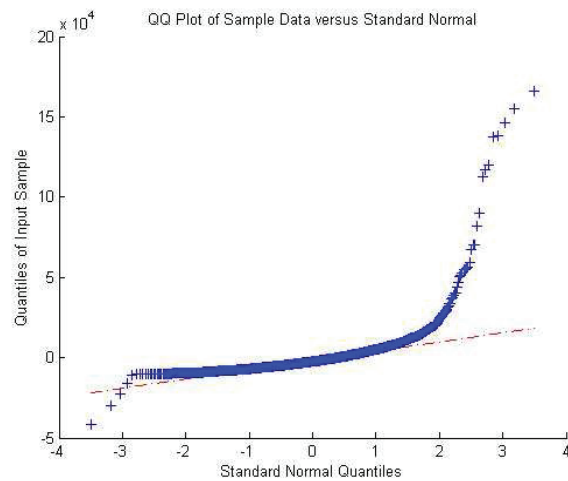


Figure 2.11: QQplot of Residuals from ARMA(1,1) of the Auto Liability



Bibliography

- Bauwens, L. and Giot, P. (2000). The logarithmic acd model: An application to the bid-ask quote process of three nyse stocks. *Annals of Economics and Statistics*, pages 117–149.
- Bolancé, C., Guillén, M., and Pinquet, J. (2003). Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on the random effects. *Insurance: Mathematics and Economics*, 33(2):273 – 282. Papers presented at the 6th {IME} Conference, Lisbon, 15-17 July 2002.
- Boucher, J.-P., Denuit, M., and Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, pages 135–162.
- Bühlmann, H. (1996). *Mathematical methods in risk theory*, volume 172 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin. Reprint of the 1970 original.
- Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, pages 841–862.
- Cooray, K. and Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scand. Actuar. J.*, (5):321–334.
- Cummins, J. (1985). Forecasting automobile insurance paid claim costs using econometric and ARIMA models. *International Journal of Forecasting North-Holland.*, pages 203–215.
- Cummins, J., Dionne, G., McDonald, J. B., and Pritchett, B. (1990). Applications of the GB2 family of distributions in modeling insurance loss processes. *Insurance: Mathematics and Economics*, 9(4):257 – 272.
- de Peretti, C. (2003). Bilateral bootstrap tests for long memory: An application to the silver market. *Computational Economics*, 22(2-3):187–212.
- El-Bassiouni, M. and El-Habashi, M. (1991). Forecasting compulsory motor insurance claims in kuwait. *Insurance: Mathematics and Economics*, 10(2):85 – 92.
- Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*.
- Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22.
- Engle, R. F. (2002). The econometrics of ultra-high-frequency data. *Journal of Applied Econometrics*, 17(5):425–446.
- Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration a new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162.

- Frees, E. W., Young, V. R., and Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24(3):229 – 247.
- Gerber, H. U. (1982). Ruin theory in the linear model. *Insurance Math. Econom.*, 1(3):177–184.
- Ghysels, E., Gouriéroux, C., and Jasiak, J. (2004). Stochastic volatility duration models. *Journal of Econometrics*, 119(2):413 – 433.
- Grammig, J. and Maurer, K.-O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal*, 3(1):16–38.
- Kremer, E. (1982). IBNR-claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal*.
- Lunde, A. (1999). A generalized gamma autoregressive conditional duration model. *discussion paper. Aalborg University*.
- Mikosch, T. (2006). *Non-life Insurance Mathematics*. Springer.
- Promislow, S. D. (1991). An axiomatic characterization of some measures of unfairness. *Journal of Economic Theory*, 53(2):345 – 368.
- Richard, J. V. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Mathematics and Economics*.
- Rolski, T., Schmidli, H., Schmidt, V., and Teugels, J. (1999). *Stochastic processes for insurance and finance*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.
- Zhang, L. (2005). Ruin probability in linear time series model. *Tsinghua Science and Technology*, 10(2):259 – 264.
- Zhang, M. Y., Russell, J. R., and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104(1):179 – 207.

Chapitre 3

Évaluation des réserves et du Capital de Solvabilité par le modèle de Sinistralité Conditionnel Généralisé (GACSM)

Résumé

Ce chapitre est composé d'un article intitulé "Generalized Autoregressive Conditional Sinistrality Model : A novel model for claims reserving in Non life insurance" coécrit avec Christian de Peretti et Lotfi Belkacem. (voir Araichi et al. (2015))

Le but de ce chapitre est de montrer l'existence d'une certaine forme de dépendance au sein d'un triangle de liquidation.

Nous proposons un nouveau modèle nommé *Generalized Autoregressive Conditional Sinistrality Model (GACSM)*. Ce modèle permet d'étendre le modèle linéaire généralisé (GLM), en incorporant la dépendance dynamique dans l'évaluation des réserves.

Ensuite, nous étudions l'impact de cette dépendance sur l'évaluation des réserves et du capital de solvabilité.

Finalement, les résultats obtenus du modèle GACSM sont comparés à ceux obtenus par la modélisation des méthodes classiques (Chain Ladder et GLM).

Introduction

Claims reserving is one of the main actuarial problems in non-life insurance. Insurers need to set up claims reserves to meet their obligations towards the policyholders. Several stochastic claims reserving models are provided in the literature, see England and Verrall (2002), Wüthrich and Merz (2008) for an overview. Among them, Generalized linear model (GLM) has become a very common approach for modelling actuarial data. This approach is based on the assumption that the claims amounts in different years are independent variables. However, in some situations, this assumption may be unrealistic. As for instance, a car accident may involve an important damages, and can hit several policies simultaneously. It has been pointed out that methods, which enable modelling the dependencies, are needed.

Recently, Antonio and Beirlant (2007) have suggested a semi parametric regression model, which is based on a Generalized Linear Mixed Model (GLMM), to handle possible dependence among the incremental claims in successive development years. However, this model is considered as too sophisticated in order to become popular in reserving practice, because it needs to incorporate external informations, such prior expert knowledge about the model parameters or market information. Also, it requires a more granular data (i.e., claim by claim data).

More recently, Hudecová and Pešta (2013) have proposed to apply the generalized estimating equations (GEE) for the estimation of the claims reserves. They have introduced a various correlation structures between the same accident year.

Another approach proposed by Pešta and Okhrin (2014) namely the conditional mean-variance model with a copula function that allows to model the dependence structure between claims. These approaches extend the classical GLM and are frequently used in panel (longitudinal) data.

In this paper, a novel model handling temporal dependence within one run-off triangle is proposed. For that purpose, a Generalized Autoregressive Conditional Sinistrality model (GACSM) framework is used, which takes into account the dependence between claims amounts and between development years. This model is a combination between the Autoregressive Conditional Amount Models (ACAM) proposed by the authors see Araichi et al. (2014) and the Generalized Linear models for reserves.

Moreover, a bootstrap procedure is suggested to achieve a full predictive distribution of reserves. Furthermore, a comparison between the results of the GACSM model and the classical approaches is provided.

This paper is organized as follows: Generalized Autoregressive conditional Sinistrality model is introduced in section 3.1. In section 3.2, a bootstrap technique is performed in order to predict the full distribution of reserves, and a mean squared error of the model is elaborated. Finally, in section 3.3 all the presented methods are applied on real data followed by a comparison between the proposed model and the classical Chain Ladder and GLM methods. Concluding remarks and some open questions are presented in Section 3.4.

3.1 Generalized Autoregressive Conditional Sinistrality Model

3.1.1 Claims reserving notation

In this section, the classical claims reserving notation is considered. Claims amounts relative to a line of business are classified in so called run-off triangle. Indeed, run-off triangle is used in non life insurance to forecast future claims amounts, see figure 3.1.

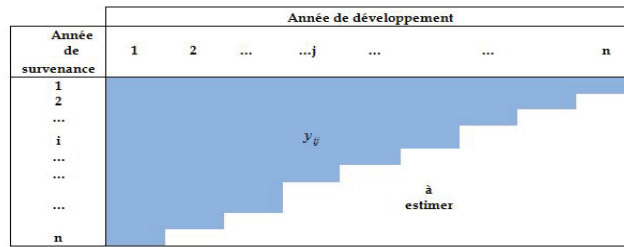


Figure 3.1: Run- off triangle for incremental claims amounts

Let us denote $y_{i,t}$ all the claims amounts relative to the accident year i , $\{y_{i,t}; i, t \in \nabla\}$ where $\nabla = \{i = 1, \dots, n, t = 1, \dots, T_i\}$, n is the most recent accident year, and $T_i = T + 1 - i$ is the last delay year. Hence, $y_{i,t}$ stand for all claims amounts in accident year i after t development periods.

The actuary’s goal is then to predict claims amounts in the lower unobserved future triangle, $\{y_{i,t}; i, t \in \Delta\}$, where $\Delta = \{i = 2, \dots, n; t = n - i + 2, \dots, T_i\}$. In the following, a Generalized Autoregressive Conditional Sinistrality model (GACSM), that allows to predict the lower triangle and the claims reserves as well is presented.

3.1.2 Autoregressive Conditional Amount model (ACAM)

In this subsection, a Generalized Autoregressive Conditional Sinistrality Model (GACSM) is presented, which is a combination of the Autoregressive Conditional Amount model (ACAM) and the GLM model of reserves.

The ACA model proposed by Araichi et al. (2014) is a model that allows to detect a temporal dependence structure between claims amounts, which their occurrences are spaced in time.

Let y_t represents the claims amount (loss) paid at the day t . The ACA model assumes that y_t follows a multiplicative structure:

$$y_t = \psi_t \epsilon_t, \tag{3.1}$$

$$\epsilon_t \sim \text{i.i.d non negative variable with } E(\epsilon_t) = 1. \tag{3.2}$$

The conditional expected loss is:

$$\psi_t = E(y_t | \Omega_{t-1}),$$

3.1. Generalized Autoregressive Conditional Sinistrality Model

where Ω_{t-1} is the set of the past observations until the time $t - 1$. An autoregressive structure for this conditional mean is considered:

$$\psi_t = \omega + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{j=1}^q \beta_j \psi_{t-j}, \quad (3.3)$$

where $\omega, \alpha_i, \beta_j > 0$ to insure that the amount is positive and $\sum(\alpha_i + \beta_j) < 1$ for stability purpose. Consequently, ψ_t depends on the p past observations and q past expected losses.

3.1.3 Generalized Autoregressive Conditional Sinistrality Model

In this subsection, a new model is presented. Contrary to the ACA model that consider a longitudinal data, our description of the Generalized Autoregressive Conditional Sinistrality Model (GACSM) is adapted to the context of claims reserving, when the data are classified in a run-off triangle.

Moreover, in order to have a more precision in calculating reserves, the accident quarter and development quarter are considered instead of accident year and development year. Indeed, the insurer will be able to know the amount of his liability after each quarter. The claims amounts $y_{i,t}$ where $i \in \{1, \dots, n\}$ refers to accident quarter, and $t \in \{1, \dots, T_i\}$ refers to development quarter are classified in run-off triangle which are ordered in time. It is therefore natural to suspect that these claims amounts may be dependent. On one hand we assume that the $y_{i,t}$ of a common accident quarter are correlated, they form a cluster. On the other hand, we assume that claims amounts of different accident quarters are independent. This assumption is similar to those of the Mack's Chain Ladder Mack (1993).

For these claims amounts $y_{i,t}$, a dynamic specification is considered:

$$y_{i,t} = \psi_{i,t} \epsilon_{i,t}, \quad (3.4)$$

$$\epsilon_{i,t} \sim \text{i.i.d non negative variable with } E(\epsilon_{i,t}) = 1. \quad (3.5)$$

where $\epsilon_{i,t}$ denote the standardized claims amounts (residuals), that are assumed to belong to the exponential family.

We denote the conditional expectation of $y_{i,t}$, $E(y_{i,t} | \Omega_{i,t-1}) = \psi_{i,t} = [\psi_{i,1}, \dots, \psi_{i,T_i}]$, where $\Omega_{i,t-1} = \{x_{i,1}, x_{i,2}, \dots, x_{i,t-1}, y_{i,1}, y_{i,2}, \dots, y_{i,t-1}\}$ are the previous informations until the development quarter $t - 1$.

Like the GLM model this conditional expectation of claims amounts is related to the predictor $\eta_{i,t}$ by a twice-differentiable one to one monotonic function which is called the link, such that $g(\psi_{i,t}) = \eta_{i,t}$. This link function could be: identity, logarithm, logit or reciprocal function. Alternatively, as pointed out by Merz and Wüthrich (2008) a log link is typically a natural choice in the insurance reserving context.

Unlike the standard GLM, where the linear predictor $\eta_{i,t} = X'_{i,t}\beta$, here we consider the problem of temporal dependence of development quarters, and we suppose a time dependent specification for $\eta_{i,t}$ defined by

$$\eta_{i,t} = X'_{i,t}\beta + \sum_{j=1}^p \phi_j \{g(y_{i,t-j}) - X'_{i,t-j}\beta\} + \sum_{j=1}^q \theta_j \{g(y_{i,t-j}) - \eta_{i,t-j}\} \quad (3.6)$$

where $X'_{i,t}$ is a $z \times p$ matrix of dummy covariates that arranges the impact of accident and development quarter on the claims amounts through model parameter $\beta \in \mathbb{R}^{p \times 1} = [\gamma, \alpha_2, \dots, \alpha_n, \delta_2, \dots, \delta_{T_i}]$, where α_i stands for the effect of accident quarter i , δ_t represents the effect of the development quarter t , and γ is the constant parameter of the model (taking $\alpha_1 = \delta_1 = 0$ to avoid over parametrization).

Therefore GACSM allows to model a time dependent claims amounts by taking into account the time dependent covariates. Indeed, relative to GLM, there is an additional component, that allows autoregressive and moving average terms to be included additively in the predictor. Here, autoregressive terms are the previous informations $\Omega_{i,t-1} = \{x_{i,1}, x_{i,2}, \dots, x_{i,t-1}, y_{i,1}, y_{i,2}, \dots, y_{i,t-1}\}$, and the moving average terms are the residuals of the predictor ($g(y_{i,t-1}) - \eta_{i,t-1}$). We note that GACSM model assumes that the data process is stationary and that the autocorrelation function decays exponentially.

Moreover in the GACSM framework it is necessary to specify the distribution of the standardized claims amounts $\epsilon_{i,t}$. Among parametric distributional families, the over dispersed Poisson and the Gamma distributions have been extensively studied in the claims reserving literature. The first one is used particularly in modelling count data (frequencies), the second has proved its success in modelling claims amounts. In our framework, the GACSM model estimation was done by maximum likelihood with Gamma distributed errors. We note that an over dispersed Poisson errors is not permitted with the GACSM model, because $\frac{\psi_{i,t}}{y_{i,t}} \notin \mathbb{N}$.

Hence we assume that the standardized claims amounts follow a Gamma distribution with $E(\epsilon_{i,t}) = 1$.

Given $\Omega_{i,t-1}$, the distribution and the density function of the Gamma GACSM are presented respectively as

$$\forall i = 1 \dots n, t = 2, \dots, T_i, T_i = T + 1 - i$$

$$F(y_{i,t}) = \frac{\Gamma(k, \frac{y_{i,t}}{\psi_{i,t}})}{\Gamma(k)} \quad (3.7)$$

$$f(y_{i,t} | \Omega_{i,t-1}) = \frac{y_{i,t}^{k-1} k^k}{\psi_{i,t}^k \Gamma(k)} \exp\left(-\frac{y_{i,t}}{\psi_{i,t}}\right) \quad (3.8)$$

where k is the shape parameter, $\psi_{i,t} = \exp(\eta_{i,t})$ and $\eta_{i,t}$ is given by

$$\eta_{i,t} = X'_{i,t}\beta + \sum_{j=1}^p \phi_j \{\log(y_{i,t-j}) - X'_{i,t-j}\beta\} + \sum_{j=1}^q \theta_j \{\log(y_{i,t-j}) - \eta_{i,t-j}\} \quad (3.9)$$

The log likelihood of the Gamma GACSM is then formalized as

$$L(y_{i,t} | \Omega_{i,t-1}) = \sum_{t=1}^T ((k-1)\log(y_{i,t}) + k\log(k) - k\log(\psi_{i,t}) - \log(\Gamma(k)) - \frac{y_{i,t}k}{\psi_{i,t}}) \quad (3.10)$$

The parameters of GACSM are estimated by maximizing the log likelihood function, the procedure yields a consistent and asymptotically normal estimates.

Here, in order to predict future claims amounts $\hat{\psi}_{i,t}$, we are interesting to specify the dynamic process of claims amounts at different lags that present the temporal dependence of claims.

The response variables $y_{i,t}$ of different origin quarters are independent, but dependent for different development quarters. This can be due to correlation patterns among repeated measurement of payments or to residual heterogeneity between claims amounts of the same origin quarter.

The dependence effect is included in the predictor; there is a part that takes account of the time dependent of claims amounts: $g(y_{i,t-j})$. Moreover, past value of the time dependent covariates $X'_{i,t-j}$ are also included in the predictor.

For this model, claims reserves for each origin quarter i and the total amount of reserves are respectively

$$R_i = \sum_{t \in \Delta_i} \hat{\psi}_{i,t} \quad (3.11)$$

$$R = \sum_{i=1}^n R_i \quad (3.12)$$

3.2 Bootstrap method and prediction errors

3.2.1 Bootstrap method

One of the main goal required by solvency 2 framework is to evaluate some quantities e.g., quantiles for the Value at Risk (VaR) at 99.5% level and prediction errors. This means that the whole predictive distribution of reserves is needed. For that, a non parametric bootstrap procedure for the suggested GACSM model in accordance with the method developed in Björkwall et al. (2009) is implemented.

Indeed, the bootstrap procedure proposed by Efron and Tibshirani (1993) is considered as the most powerful tools, and can be particularly interesting to assess the variability of the claims reserving prediction, and to construct the full distribution of reserves.

In addition the bootstrap is a particular method used to estimate the variability of a parameter. Applications of bootstrap in claims reserving can be found in England and Verrall (1999), England and Verrall (2002), Pinheiro et al. (2003), and Björkwall et al. (2009).

To implement this procedure we need to choose a model, to define an adequate residual and to use a bootstrap prediction procedure.

In the following, we present the steps of the bootstrap process suggested for Gamma GACSM:

First, we specify the lags (p,q) of the temporal dependence, we estimate the parameters of the Gamma GACSM(p,q) model of $y_{i,t}$ using maximum likelihood estimation, and we calculate the conditional mean $\hat{\psi}_{i,t}$.

Second, as this model produces iid residuals, they are computed as $\hat{\epsilon}_{i,t} = \frac{\hat{\psi}_{i,t}}{y_{i,t}}$.

Third, residuals are resampled B times in order to get B new triangles of residuals. The resulting bootstrapped residuals are denoted by $\nabla \epsilon_{i,t}^*$.

Then, B pseudo-triangles $\nabla y_{i,t}^*$ are generated by computing $y_{i,t}^* = \hat{\psi}_{i,t} \epsilon_{i,t}^*$, with $\hat{\psi}_{i,t} = \exp(\hat{\eta}_{i,t})$ and $\hat{\eta}_{i,t} = X'_{i,t} \hat{\beta} + \sum_{j=1}^p \hat{\phi}_j \{ \log(y_{i,t-j}) - X'_{i,t-j} \hat{\beta} \} + \sum_{j=1}^q \hat{\theta}_j \{ \log(y_{i,t-j}) - \hat{\eta}_{i,t-j} \}$

Finally, for these B pseudo triangles, GACSM(p,q) are re-estimated, and $\hat{\psi}_{i,t}^*$ are obtained by the same equation.

Estimators for the claims reserves of each quarter and the total amount of reserves are then derived for each triangle by respectively $\hat{R}_i^* = \sum_{t \in \Delta_i} \hat{\psi}_{i,t}$ and $\sum_{\Delta} \hat{R}_i^*$.

By repeating the last step for each pseudo-triangle, we obtain the full predictive distribution of reserve for each quarter and in total. These distributions will allow us to calculate the quantile and the uncertainty on reserves.

3.2.2 Mean squared error of prediction MSE

Prediction error is defined as the standard deviation of the distribution of possible reserve outcomes, and have a significant effect when the variability on reserve is large.

Our aim is to derive the mean squared error of prediction within the GACSM framework and bootstrap technique. Let us define the total distribution of reserves.

$$\hat{R}^* = \sum_{\Delta} \hat{R}_i^* = \sum_{\Delta} \sum_{t \in \Delta_i} \hat{\psi}_{i,t} = \sum_{\Delta} \sum_{t \in \Delta_i} g^{-1}(\hat{\eta}_{i,t}) \quad (3.13)$$

The bootstrap conditional MSE of prediction for each quarter and the bootstrap conditional MSE of total reserve are defined by

$$MSE(\hat{R}_i^*) = E(R_i^{**} - \hat{R}_i^*)^2 \quad (3.14)$$

$$MSE(\hat{R}^*) = E(R^{**} - \hat{R}^*)^2 \quad (3.15)$$

where R^{**} and \hat{R}^* are resampled reserves and resampled estimated reserves.

To get the resampled reserves for each quarter $\hat{R}_i^{**} = \sum_{\Delta_i} y_{i,t}^{**}$ and the resampled total reserve $R^{**} = \sum_{\Delta} \hat{R}_i^{**}$; the resampling is done once more from the empirical distribution function of the residuals to get B triangles of $\Delta \epsilon_{i,t}^{**}$, then solving $y_{i,t}^{**} = \hat{\psi}_{i,t} \epsilon_{i,t}^{**}$.

These prediction errors measure the deviation between the future reserves and the predicted ones. They are considered as conditional MSE, because reserves are evaluated by the GACSM model that take into account the temporal dependence structure.

3.3 Empirical results

Data were supplied by a French insurance company and consist of incremental claims amounts of two lines of business Auto Damage and Auto Liability, over the period January 2002, to December 2007. The first guarantee covers all damaged caused to the car, even if the insured is responsible of the accident. It is a short line of business, that once the claim is declared it will be paid after a short time. The second, is the guarantee of damage caused to others, belongs to the Civil Responsibility lines of business. It is called a long tail line of business, because there will be a time lag between the date of occurrence claim and the date of full refund. These claims amounts are classified quarterly into two run-off triangles. We have 300 observations for each lines of business spread over 24 accident quarters and 24 development quarters.

The claims amounts are net of reinsurance, net of subrogation and we suppose that inflation remains unchanged throughout the quarters. The convention is to examine historical claims amounts, to the extent reliable, as a source of knowledge for the dynamic process and parameter estimates.

To investigate the characteristics of these claims, summary statistics of claims amounts of the two lines of business are presented in table 3.1.

	Auto damage	Auto liability
max	517.129	497.472
min	0	0
mean	64.131	62.670
std	115.165	96.369
skewness	2.005	2.097
kurtosis	6.028	6.807
Jarque Bera test	315.7 (0.001)	401.02 (0.001)

(.) contains p-value of the Jarque Bera test

Table 3.1: Descriptive statistics of claims amounts (in thousand €)

For the lines Auto Damage and Auto Liability, the minimum and maximum claims amounts suggest that the distributions are widely spread. The standard deviation is larger than the mean, implying the skewness of the data. In fact, the two distributions of claims amounts exhibit a positive skewness and excess kurtosis indicating that are heavy tailed and right skewed.

According to Jarque Bera test, there is evidence that the two distributions of claims amounts are not normally distributed.

In order to detect the dynamic dependence of claims amounts, autocorrelation functions (ACF) of the two lines of business are presented. Then, a test of autocorrelations and a stationary test are applied.

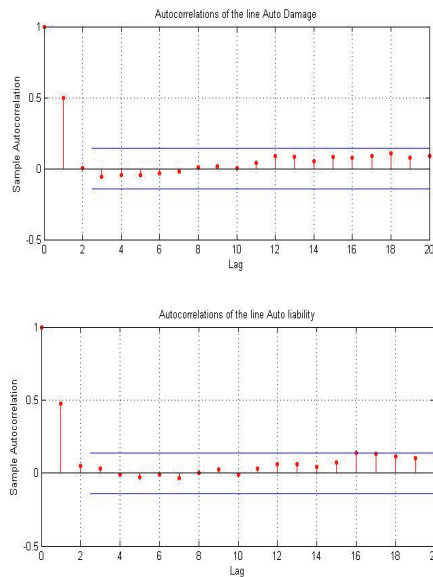


Figure 3.2: Autocorrelations of the line Auto Damage and Auto liability

Lines of Business	Statistic	P-value
Auto Damage	-8.58	0
Auto Liability	-8.002	0

Table 3.2: Augmented Dickey Fuller Test

Lines of Business	Statistic	P-value
Auto Damage	77.13	1.8210^{-12}
Auto Liability	69.82	0

Table 3.3: Ljung Box Test with 10 lags

According to the Augmented Dickey Fuller test (table 3.2), the two series of lines of business are stationary. Indeed, the $p\text{-value} = 0 < 5\%$, indicating no unit root in the data.

Furthermore, the ACF plot (figure 3.2) shows a significant lags, indicating that claims amounts exhibit temporal correlations. These correlations are positive, indicating that large or small claims amounts tend to occur together. This is consistent with the features usually exhibited by insurance company. Indeed, this may be due to the arrival process of claims or the covered system adopted by the company. In fact, due to weather events, claims may occur simultaneously.

Also, insurance company may pay the large (respectively the small) losses in the same period.

Moreover, from table 3.3 we remark that for the two lines of business, the Ljung Box statistics are large, so, claims amounts are highly correlated. Hence, the application of the proposed GACSM might be suitable.

Recall that we consider a Conditional Autoregressive Sinistrality Gamma model for parameters estimates. Indeed, a bootstrap procedure is conducted as in section 3.2.1. 1000 iterations of triangles are generated. The GACSM at different lags (1,1), (1,2), (2,1), (2,2) are fitted for each pseudo triangle. Here, only the GACSM(1,1) is presented because the estimated results are consistent with the constraints of this model (significant parameters and iid residuals). (All programs of estimates relative to GACSM functions are created by authors on MATLAB 7).

The average value of parameters with their relative standard errors are computed.

Moreover, a comparison of the estimated parameters is conducted with a GLM Gamma model, where a classical bootstrap procedure is also applied for this model. Results are listed in tables 3.4 and 3.5.

As can be seen from tables 3.4 and 3.5, temporal dependence parameters estimates from GACSM indicate dependence between claims of the same accident quarter. Indeed, we have $\sum \hat{\phi} + \hat{\theta} < 1$ that allows the existence of conditional expected claims amounts, and thus the GACSM might be appropriate.

Moreover, from tables 3.4 and 3.5, we remark that the average values and the standard errors of accident and development parameters from the GACSM Gamma and GLM Gamma are almost close for the two lines of business.

However, more detailed exploration shows that for the line Auto Liability, some efficiency could be gained by choosing the GACSM Gamma. Indeed, the minimal standard errors are obtained by using this model for 34 out of 47 of estimated parameters. However, for the line Auto Damage the GLM Gamma model provides a slightly better estimation of parameters than GACSM Gamma: minimal standard errors are observed for 34 parameters.

We can conclude that the two models are appropriate for fitting data and the difference in the standard errors are small.

Chapitre 3. Évaluation des réserves et du Capital de Solvabilité par le modèle (GACSM)

parameters	Damage				Liability			
	GACSM Gamma		GLM gamma		GACSM Gamma		GLM gamma	
	mean	std	mean	std	mean	std	mean	std
k	0.367***	0.027	2.697***	0.197	0.527***	0.059	1.919***	0.190
ϕ	0.039*	0.021			0.086*	0.047		
θ	-0.047*	0.026			-0.092*	0.049		
γ	12.309***	0.383	12.459***	0.298	11.657***	0.219	11.640***	0.258
α_2	-0.066*	0.036	-0.072*	0.040	0.300*	0.165	0.244*	0.136
α_3	-0.347*	0.210	-0.438*	0.265	0.239*	0.141	0.278*	0.167
α_4	-0.210*	0.125	-0.073*	0.045	0.260 *	0.154	0.198*	0.120
α_5	0.097*	0.054	0.057*	0.037	0.161*	0.084	0.152 *	0.090
α_6	-0.066*	0.037	-0.085*	0.044	0.154*	0.079	0.149*	0.087
α_7	-0.185*	0.099	-0.057*	0.034	0.188*	0.097	0.192*	0.115
α_8	-0.030*	0.017	-0.029*	0.018	0.208*	0.110	0.312 *	0.188
α_9	-0.032*	0.017	-0.035*	0.019	0.337*	0.200	0.360*	0.214
α_{10}	-0.159*	0.087	-0.025 *	0.015	0.214 *	0.127	0.234 *	0.135
α_{11}	-0.087*	0.05	-0.049*	0.025	0.261*	0.158	0.352*	0.153
α_{12}	-0.028*	0.016	-0.026*	0.015	0.195 *	0.110	0.246	0.148
α_{13}	0.229 *	0.135	0.101*	0.061	0.516*	0.271	0.484*	0.257
α_{14}	-0.070 *	0.041	-0.072*	0.038	0.232 *	0.139	0.364*	0.220
α_{15}	-0.151*	0.082	-0.116 *	0.064	0.216*	0.125	0.248 *	0.140
α_{16}	0.113 *	0.062	0.128*	0.078	0.344 *	0.182	0.353*	0.214
α_{17}	0.217*	0.125	0.132*	0.071	0.351*	0.200	0.360	0.210
α_{18}	-0.055*	0.031	-0.058*	0.035	0.304*	0.158	0.272*	0.165
α_{19}	-0.053 *	0.030	-0.055*	0.029	0.211 *	0.12	0.247	0.147
α_{20}	0.069*	0.04	0.033 *	0.014	0.455	0.272	0.418	0.290
α_{21}	0.208*	0.124	0.124*	0.071	0.413*	0.247	0.421*	0.253
α_{22}	-0.138 *	0.081	-0.124*	0.072	0.286*	0.167	0.391 *	0.235
α_{23}	-0.293*	0.165	-0.188*	0.116	-0.161 *	0.093	-0.153	0.09
α_{24}	-4.895*	2.951	-4.914*	2.815	-1.001 *	0.541	-1.296	0.777

Table 3.4: Average values and Standard errors of the estimated parameters of computed from 1000 iterations for GACSM Gamma and GLM Gamma

Wald test is applied to test whether estimated parameter is 0.

*:10% significant statistics.

** : 5% significant statistics,

***: 1% significant statistics.

3.3. Empirical results

parameters	Damage				Liability			
	GACSM Gamma		GLM gamma		GACSM Gamma		GLM gamma	
	mean	std	mean	std	mean	std	mean	std
δ_2	0.434*	0.247	0.178*	0.098	0.781***	0.271	0.766***	0.281
δ_3	-0.772*	0.463	-0.815**	0.393	-0.158	0.096	-0.187*	0.111
δ_4	-1.776***	0.473	-2.032***	0.403	-1.011***	0.282	-1.020***	0.288
δ_5	-2.122***	0.483	-2.310***	0.408	-0.964***	0.299	-1.095***	0.302
δ_6	-2.859***	0.489	-2.974***	0.440	-1.729***	0.294	-1.646***	0.294
δ_7	-2.852***	0.497	-3.049***	0.430	-1.565	0.307	-1.571	0.294
δ_8	-3.444***	0.497	-3.668***	0.422	-1.985***	0.306	-1.975***	0.324
δ_9	-2.985***	0.518	-2.897***	0.580	-2.788***	0.320	-2.737***	0.329
δ_{10}	-4.827***	0.519	-5.155***	0.456	-3.164***	0.317	-3.110***	0.318
δ_{11}	-5.717***	0.521	-5.964***	0.448	-4.040***	0.332	-4.021***	0.351
δ_{12}	-5.305***	0.551	-5.257***	0.526	-3.926***	0.319	-3.930***	0.348
δ_{13}	-2.827***	0.546	-2.823***	0.590	-3.614***	0.348	-3.586***	0.353
δ_{14}	-5.768***	0.554	-5.773***	0.675	-3.284***	0.373	-3.213***	0.361
δ_{15}	-6.138***	0.588	-6.101***	0.640	-3.983***	0.386	-3.444***	0.508
δ_{16}	-5.088***	0.593	-5.351***	0.512	-11.921***	0.395	-11.886***	0.393
δ_{17}	-5.547***	0.625	-6.172***	0.554	-3.485***	0.395	-3.398***	0.400
δ_{18}	-12.232***	0.672	-12.591***	0.579	-2.252***	0.423	-2.170***	0.425
δ_{19}	-12.459***	0.730	-12.628***	0.598	-2.320***	0.458	-2.203***	0.427
δ_{20}	-12.516***	0.763	-12.694***	0.722	-6.813***	0.489	-6.778***	0.473
δ_{21}	-8.043***	0.865	-8.342***	0.826	-6.831***	0.571	-6.698***	0.540
δ_{22}	-4.610***	1.161	-4.558***	1.264	-5.729***	0.780	-6.084***	0.681
δ_{23}	-12.888***	1.870	-13.329***	1.791	-12.140***	1.249	-12.092***	1.316
δ_{24}	-12.859***	3.247	-12.723***	3.314	-13.016***	3.110	-12.876***	2.960

Table 3.5: Average values and Standard errors of the estimated development quarter parameters, computed from 1000 iterations for GACSM Gamma and GLM Gamma

Wald test is applied to test whether estimated parameter is 0.

*:10% significant statistics.

** : 5% significant statistics,

***: 1% significant statistics.

Next we assess the adequacy of our models to the data, it is investigated via examination of residuals. For that, a diagnostic tests (Ljung Box (LJB) test and LM (Lagrange Multiplier test) are applied. Results are presented in table 3.6).

Further, to compare the GACSM Gamma and GLM Gamma, AIC (Akaike information criteria), BIC (Bayesian information criteria) and the loglikelihood values (LL) of the fitted models are computed. Results are reported in table 3.7.

	Auto Damage		Auto Liability	
	GACSM(1,1)	GLM Gamma	GACSM(1,1)	GLM Gamma
Mean	0.996	0.005	0.999	1.45 E-5
Variance	1.454	1.37	1.1	1.09
LJB test Q(10)	11.02 (0.356)	11.31 (0.333)	15.05 (0.13)	14.52 (0.15)
LM test Q(10)	2.079 (0.996)	1.66 (0.998)	1.15 (0.999)	0.95 (0.999)

Table 3.6: Mean, Variance and diagnostic tests of residuals (p-value inside parentheses) of Auto Damage and Auto Liability

	Auto Damage		Auto Liability	
	GACSM(1,1)	GLM Gamma	GACSM(1,1)	GLM Gamma
AIC	-5.35	-5.33	-6091	-6067
BIC	-5.16	-5.15	-5.9	-5.88
LL	2729.7	2718.6	3097.5	3083.6

Table 3.7: AIC, BIC and Log likelihood of the line Auto Damage and Auto Liability

It can be seen from table 3.6 that the GACSM(1,1) captures autocorrelations and overcomes all the constraints of this model. Indeed, results reveal that residuals have a mean of approximately one.

Further, considering Ljung Box test and LM test at 10 lags, we have a statistics value < 18.31 and p-value $> 5\%$ indicate respectively no autocorrelations and no ARCH effect in the residuals, both for the Auto Damage and Auto Liability. Therefore, the GACSM(1,1) model is able to fit the temporal dependence structure of claims.

For GLM Gamma, the two lines of business present a mean of residuals around zero. This is consistent with the constraints of this model.

In fact, GLM Gamma assumes that residuals follow a Gamma distribution with zero mean and variance equal to σ^2 . In sum, we can conclude that the two models are correctly specified.

From table 3.7, it can be seen that GACSM Gamma presents lower AIC and BIC criteria and higher loglikelihood value than GLM Gamma. So, the comparisons suggest that the GACSM Gamma gives a better fit than GLM Gamma. Therefore, GACSM

describes well the dynamic effect of development quarters as well as the dynamic effect of claims. It detects the temporal dependence of data through the conditional expected mean $\hat{\psi}_{i,t}$.

It seems that estimates obtained by this model might be used in the predictions of reserves.

In order to show the impact of this model on the estimated reserves, we predict reserves from the 1000 pseudo triangles generated with the bootstrap process (see section 3.2). Predicted reserves are presented in table 3.8, in figures 3.3 and 3.4. Prediction errors are exhibited in table 3.9, in figures 3.5 and 3.6.

For comparison purpose, table 3.8, table 3.9 and presented figures, also contain results from Chain Ladder model and GLM Gamma (for more details about these two models see Appendix 3.6). For these two models a two different bootstrap procedure are conducted in order to predict reserves and prediction errors. For more details about bootstrap for Chain Ladder model see Lowe (1994), Taylor (2000) and England and Verrall (2002), and see England and Verrall (1999), England and Verrall (2002), Pinheiro et al. (2003), and Björkwall et al. (2009) for bootstrap method based on GLM models.

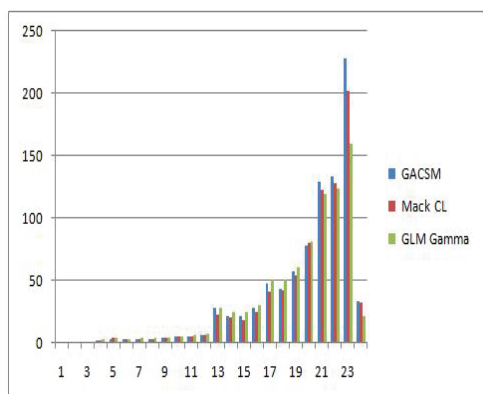


Figure 3.3: Reserves estimates of each quarter (in thousands €) for Auto Damage.

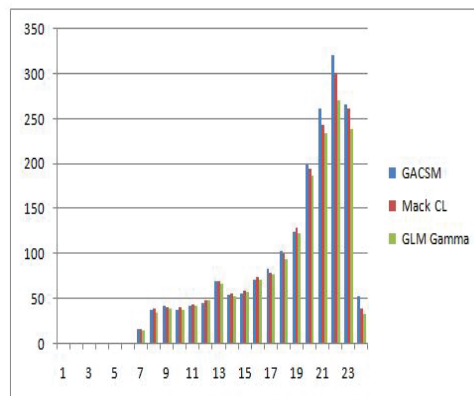


Figure 3.4: Reserves estimates of each quarter (in thousands €) for Auto Liability.

Accident quarter i	Auto Damage			Auto Liability		
	GCASM	Mack CL	GLM Gamma	GCASM	Mack CL	GLM Gamma
1	0	0	0	0	0	0
2	0.001	0.009	0.001	0.001	0.006	0.001
3	0.002	0.050	0.001	0.002	0.030	0.002
4	2.708	2.500	3.797	0.596	0.474	0.346
5	3.628	4.420	4.200	0.705	0.588	0.495
6	3.139	3.450	3.677	0.864	0.439	0.646
7	2.883	2.970	3.817	16.092	17.400	15.699
8	3.190	3.710	3.826	38.190	39.000	35.478
9	4.167	4.700	4.385	42.942	41.500	38.953
10	5.254	5.120	5.958	37.674	40.500	37.470
11	5.957	5.810	6.384	42.583	44.100	43.004
12	6.972	6.710	7.363	45.640	49.200	48.342
13	28.272	22.800	28.643	69.585	69.800	66.680
14	21.897	20.400	25.270	54.267	55.800	54.010
15	21.386	18.500	25.267	56.308	59.600	57.979
16	28.798	24.800	30.659	71.894	74.900	72.166
17	48.339	40.800	51.497	83.132	79.500	77.778
18	43.638	42.400	51.075	100.868	101.000	93.830
19	57.605	54.500	61.339	124.330	130.000	122.883
20	77.988	80.400	81.692	198.778	195.000	187.775
21	129.666	123.000	120.063	261.352	244.000	234.107
22	133.572	128.000	124.260	320.682	302.000	270.592
23	228.186	202.000	159.525	266.614	262.000	238.795
24	33.857	32.900	22.197	53.674	40.300	33.253
Total	891.104	829.949	824.896	1889.574	1847,137	1730.285
GACSM/Mack CL	7.369%			2.30%		
GACSM/GLM	8.026%			9.21%		

$$GACSM/CL = \frac{\text{Total reserve with GACSM} - \text{Total reserve with MackCL}}{\text{Total reserve with CL}} \times 100$$

$$GACSM/GLM = \frac{\text{Total reserve with GACSM} - \text{Total reserve with GLM}}{\text{Total reserve with GLM}} \times 100$$

Table 3.8: Reserves estimates (in thousands €) for Auto Damage and Auto Liability based on GACSM Gamma, Mack Chain Ladder (CL) and GLM Gamma.

Accident quarter i	Auto Damage			Auto Liability		
	GACSM	Mack	GLM Gamma	GACSM	Mack	GLM Gamma
1	0	0	0	0	0	0
2	85.83	395.77	190.098	164.63	101.786	122.33
3	71.35	208.20	125.05	106.80	5517.30	97.68
4	180.49	247.26	162.64	120.58	217.60	152.87
5	130.37	212.99	154.48	98.27	162.50	108.37
6	115.89	224.98	139.66	80.13	137.00	86.45
7	192.2	234.44	167.43	99.10	251.00	99.86
8	117.96	220.44	152.20	77.57	168.10	70.38
9	110.76	187.14	125.15	82.88	153.20	88.03
10	184.71	167.48	111.21	74.52	155.10	85.78
11	76.96	152.55	111.34	66.53	146.70	79.25
12	76.41	137.32	96.21	66.65	132.40	68.33
13	66.97	175.16	89.56	65.99	111.90	66.68
14	55.77	185.68	78.13	61.90	117.60	63.58
15	62.44	185.19	79.57	52.63	111.60	53.86
16	59.93	152.45	73.06	55.00	96.10	56.74
17	55.95	130.31	71.97	59.30	86.80	52.22
18	51.78	119.08	66.92	52.13	70.00	48.08
19	54.27	92.47	58.41	45.31	55.20	50.81
20	49.14	71.74	49.84	53.22	42.30	52.48
21	47.05	64.87	49.48	53.87	39.10	46.78
22	50.12	59.41	46.15	53.60	31.80	46.99
23	59.79	42.53	34.27	50.97	33.40	43.94
24	69.41	112.32	183.23	73.95	117.20	48.78
Total	24	29	28	21	28	24

Table 3.9: RMSE of prediction as % of reserve estimate for Auto Damage and Auto Liability from GACSM Gamma, Mack Chain Ladder (CL) and GLM Gamma model.

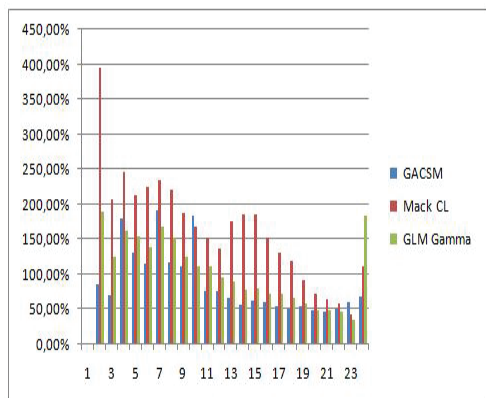


Figure 3.5: RMSE of prediction as % of reserve estimate of each quarter for Auto Damage.

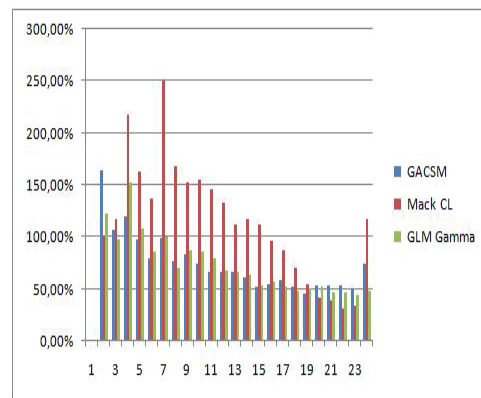


Figure 3.6: RMSE of prediction as % of reserve estimate of each quarter for Auto Liability.

At first sight, from table 3.8, from figures 3.3 and 3.4, it appears that for the two lines of business, the reserves estimates of GACSM are close to reserves of Mack Chain Ladder and GLM Gamma model, notably for the first accident quarters ($i < 11$). However, for the later accident quarters ($i > 11$), it will be seen that there are noticeable differences between the models.

Indeed, GACSM Gamma reserves estimates are significantly higher than Mack Chain Ladder and GLM Gamma, across accident quarters and in total.

Considering the line Auto Damage, a high difference is observed for the quarter 13 and 24. Indeed, for the quarter 13 reserves estimate with GACSM Gamma is 24% higher than reserves estimated with Mack Chain Ladder.

Moreover, for the quarter 24, reserves estimate with GACSM Gamma is 52.53% higher than reserves estimated with GLM Gamma model. The difference is much more significant when we look at the quarter 24 of the line Auto Liability, the GACSM reserve estimate is 33.2% and 61.41% higher than those estimated respectively by Chain Ladder and GLM Gamma.

These differences in the amount of reserves are mainly caused by specifying the really temporal dependence structure between development quarters. Indeed, predicting reserves are based on $\hat{\psi}_{i,t}$, and this conditional mean takes into account all the available information of the past (temporal dependence of the development quarters and claims amounts).

However, the difference is not very significant between the total GACSM reserves estimate and the total Chain Ladder and GLM Gamma reserve estimate.

For the line Auto Damage the total reserves is 7.36% and 8.026% higher than reserve estimated respectively under Chain Ladder and GLM Gamma.

For the Auto Liability total reserves is 2.30% and 9.21% higher than those estimated under these two classical models.

In addition, it is noticed from table 3.9, figures 3.5 and 3.6, that the prediction errors as percentage of reserve for the GACSM Gamma, are almost smaller than in the case of Chain Ladder and GLM Gamma. One reasons is the form of the MSE estimate, which is derived in a more efficient way compared to Chain Ladder and GLM Gamma model. Recall that the MSE derived with GACSM Gamma does not neglect the temporal dependence between claims and between development quarters.

Another remark involves the estimated RMSE of prediction for the two lines of business; relatively more weight is given to smaller accident quarter reserves, and less to larger accident quarter reserves giving greater weight to the tail of the reserve distribution, which is more variable. Thus the RMSE decrease as the accident quarters increase, giving more precision for the prediction of reserves, except for the last three accident quarters. As these last quarters present a large prediction of claims and are more uncertain, they exhibit a high prediction errors. This result is in line with that of England and Verrall (2002).

The main conclusion is that the prediction of reserves and their corresponding prediction errors are much more influenced by the temporal dependence in the tail of distribution of reserves.

For a complete comparison, we report an table 3.10 the evaluation of VaR at 99.5% level, which is the most frequent choice in practise, and the Solvency Capital Requirement (SCR)⁴ of the two lines of business.

Table 3.10: $VaR_{99.5\%}$ and SCR for Auto Damage and Auto Liability based on GACSM Gamma, Chain Ladder and GLM Gamma.

	Auto Damage			Auto Liability		
	GACSM	Mack CL	GLM Gamma	GACSM	Mack CL	GLM Gamma
$VaR_{99.5\%}$	1772.946	1430.062	1442.863	2991.343	2690.740	2503.021
SCR	881.8	600.352	617.967	1101.769	844.049	772.737
SCR GACSM/ SCR Mack CL	47%			30%		
SCR GACSM/SCR GLM	43%			43%		

⁴Solvency Capital Requirement of reserve risk for a line of business l , is given by

$$SCR_l = VaR_{99.5\%}^l - BE_l \quad (3.16)$$

where $VaR_{99.5\%}^l$ is the Value at Risk of the line business l at confidence level 99,5%, and BE_l is the best estimate of the total reserves of the line of business l , computed as the mean of the total reserves.

Investigating the effect of dependence between risks, it is noticed from table 3.10 that for the two lines of business, the GACSM Gamma provides remarkably higher $VaR_{99.5\%}$ than the Chain Ladder and GLM Gamma. But even more important, is the estimate of the SCR with GACSM, that are even higher than SCR of the two classical models. Indeed, the evaluation of SCR with GACSM model for the line Auto Damage is 47% and 43% higher than those evaluated by respectively Chain Ladder and GLM Gamma model. For the line Auto Liability, the difference is also significant; SCR with GACSM model is 30% and 43% higher than those evaluated by respectively Chain Ladder and GLM Gamma model. This is due to the tail of distribution. Indeed, the SCR is evaluated by the VaR at high level 99.5% that takes into account a major part of the tail of distribution, and as the temporal dependence structure affects principally the tail, it exhibits a high SCR amount.

We can conclude that the Chain Ladder and GLM Gamma model underestimate the risks of the insurance company and can cause a risk of error in evaluating the SCR.

In sum, our proposed model GACSM Gamma provides a considerable higher quantities (reserves, $VaR_{99.5\%}$, SCR) than the classical Chain Ladder and GLM Gamma model. In this configuration, we can say that taking into account the temporal dependence structure has a significant impact on the evaluation of reserve estimates for each accident quarters and in total as well. This can lead to evaluate the accurate SCR for lines of business.

Also, as already discussed GACSM Gamma provides lower prediction errors than GLM Gamma, so, better precision on the evaluation of reserves. Again, results favor GACSM Gamma and confirm that this approach might be appropriate.

3.4 Conclusion

This paper proposes a Conditional Autoregressive Sinistrality Model as a suitable model for claims reserving. Classical models assume that claims amounts of the run-off triangle are independent. This assumption is too restrictive since it was proved, that claims amounts exhibit different dependence structure. In this regard, GACSM Gamma presents a new approach that enables to model a specific dependence structure which is the temporal dependence between risks. The dependencies are modelled via a conditional expected mean.

The dynamic structure of the mean allows to take into account the temporal dependence structure of claims of each accident quarter.

We carried out an analysis of claims amounts of two lines of business Auto Damage and Auto Liability with the GACSM Gamma model. A comparison with the classical model is conducted in order to show the relevance of the obtained results.

For this model a bootstrap procedure are proposed as well as a conditional mean squared error of prediction. This procedure allows to construct the whole predictive distribution of reserves, and thus to evaluate the SCR.

An appropriate specification of the temporal dependence structure leads to predict higher reserves estimates and higher SCR than those estimated with Chain Ladder and GLM Gamma model. Moreover, the GACSM Gamma model improves the precision of prediction reserves (smaller prediction errors), and so it predicts an accurate reserves and SCR. This specification looks very promising and allowed us to highlight the temporal dependence of claims amounts.

Our study extends the classical GLM model and is consistent with prior works that model dependence structure between claims of the one run-off triangle Hudecová and Peřta (2013). Most notably, this is the first study to our knowledge to investigate temporal dependence structure in a run off triangle.

Despite of the promising results obtained from the mean squared errors with GACSM Gamma, but they remain high. Future studies should focus on elaborating a suitable expression of mean squared errors for temporal dependent claims amounts.

This study is a first step of modelling temporal dependence structure between claim reserving. Future work should focus on temporal dependence structure between several lines of business by incorporating a time varying copulas, which enable to fit dynamic dependence of several marginal distributions.

Moreover, our results are encouraging and should be validated with other distributions: Lognormal, Pareto, Generalized extreme value...

3.5 Appendix

3.6 Classical reserving models

3.6.1 Mack Chain Ladder model

Chain Ladder model is the traditional and the straightforward technique to estimate future claims amounts see Mack (1993) and Mack (1994) . This technique is applied to cumulative claims data. We assume that the data consist of a triangle of incremental claims, $\{y_{i,j} : i = 1, \dots, n; j = 1, \dots, n - i + 1\}$.

i : refers to the row, and indicates the accident quarter.

j : refers to the column and indicates the delay quarter (or development quarter).

We assume that these incremental claims are independent. Then, the cumulative claims data are obtained by

$$C_{i,j} = y_{i,j} + \sum_{k=1}^{j-1} y_{i,k} \quad (3.17)$$

The Chain Ladder methods consist of forecasting the future claims $C_{i,j}$, by estimating a development factors as

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}} \quad 1 \leq j \leq n - 1 \quad (3.18)$$

Then, the conditional expected loss, which is the best estimate of the future claims is

$$E(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = \hat{\lambda}_j C_{i,j} \quad (3.19)$$

For $1 \leq i \leq n$ and $1 \leq j \leq n$, we have $V(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = C_{i,j} \sigma_j^2$, with σ_j is the volatility of the development quarter defined as

$$\sigma_j^2 = \frac{1}{n-j-1} \sum_{i=1}^{n-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{\lambda}_j \right)^2 \quad (3.20)$$

The amount of the expected reserve $E(R_i = C_{i,n} - C_{i,n-i+1} | C_{i,1}, \dots, C_{i,j})$ is estimated by

$$R_i = \hat{C}_{i,n} - \hat{C}_{i,n-i+1} \quad (3.21)$$

The total amount of reserves is given by $R = \sum_{i=1}^n R_i$.

3.6.2 Generalized Linear Model (GLM)

In alternative model to the Stochastic Chain Ladder model was proposed by Nelder and Wedderburn (1972), the Generalized Linear Model (GLM). Contrary to Chain Ladder technique, this method is applied to incremental data $y_{i,j}$.

We assume that the response variable $y_{i,j}$ are independent and belong to the exponential family distributions defined as

$$f(y_{i,j}, \theta_{i,j}, \phi) = \exp\left\{\frac{y_{i,j}\theta_{i,j} - b(\theta_{i,j})}{a_{i,j}(\phi)} + c(y_{i,j}, \phi)\right\} \quad (3.22)$$

where $a_{i,j}(\phi)$, $b(\theta_{i,j})$ and $c(y_{i,j}, \phi)$ are functions specified in advance, $\theta_{i,j}$ is a parameter related to the mean and ϕ scale parameter related to the variance. for more details we refer the readers to McCullagh and Nelder (1989).

We denote the best estimate of claims amounts as $\mu_{i,j} = E(y_{i,j}) = b^{-1}(\theta_{i,j})$ and $v(y_{i,j}) = \phi v(\mu_{i,j}) = \phi b'(\theta_{i,j})$.

The mean $\mu_{i,j}$, is related to the covariates via the link function g , that is differentiable and monotonic, such that $g(\mu_{i,j}) = \eta_{i,j}$, this link function can be : identity, logarithm, logit and reciprocal. Alternatively, as pointed out by Merz and Wüthrich (2008), a log link is typically a natural choice in the insurance reserving context.

$\eta_{i,j}$ is the linear predictor and is defined by $\eta_{i,j} = X_{i,j}\beta$, where $X_{i,j}$ is a matrix of covariates and β is the vector of parameters.

The model parameters $\hat{\beta}$ are estimated using maximum likelihood, and future claims amounts is given by $E(y_{i,j}) = \hat{\mu}_{i,j} = g^{-1}(\hat{\eta}_{i,j})$. The amount of reserve for each accident quarter is calculated by $R_i = \sum_{j>n-i+1} \hat{\mu}_{i,j}$ and the best estimate of the total amount of reserves is given by $R = \sum_{i=1}^n R_i$.

Bibliography

- Antonio, K. and Beirlant, J. (2007). Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*.
- Araichi, S., de Peretti, C., and Belkacem, L. (2014). Modelling temporal dependence of claims in insurance using autoregressive conditional amount models. *ISFA working paper No 2013.9*, <http://docs.isfa.fr/labo/2013.9.pdf>.
- Björkwall, S., Hössjer, O., and Ohlsson, E. (2009). Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving. *Scandinavian Actuarial Journal*, (4):306–331.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- England, P. and Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics*.
- England, P. and Verrall, R. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*.
- Hudecová, Š. and Pešta, M. (2013). Modeling dependencies in claims reserving with {GEE}. *Insurance: Mathematics and Economics*, 53.
- Lowe, J. (1994). A practical guide to measuring reserve variability using: Bootstrapping, operational time and a distribution-free approach. Technical report, Institute of Actuaries and Faculty of Actuaries.
- Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserves estimates. *Astin Bulletin*.
- Mack, T. (1994). Which stochastic model is underlying the chain ladder method? *Insurance: Mathematics and Economics*.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC; 2nd edition.
- Merz, M. and Wüthrich, M. V. (2008). Prediction error of the multivariate chain ladder reserving method. *North American Actuarial Journal*.
- Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society*.
- Pešta, M. and Okhrin, O. (2014). Conditional least squares and copulae in claims reserving for a single line of business. *Insurance: Mathematics and Economics*.
- Pinheiro, P. J. R., e Silva, J. M. A., and Centeno, M. d. L. (2003). Bootstrap methodology in claim reserving. *The Journal of Risk and Insurance*.

- Taylor, G. (2000). *Loss reserving: an actuarial perspective*. Kluwer Academic Boston; London.
- Wüthrich, M. V. and Merz, M. (2008). *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons, Ltd, West Sussex, England.

Chapitre 4

Évaluation du Capital de Solvabilité : approche reposant sur les copules dynamiques

Résumé

Ce chapitre est composé d'un article en anglais, intitulé "*Time Varying Copula Model for claims reserving in Non life insurance*"(voir Araichi et al. (2015)).

Le but de ce chapitre est de modéliser conjointement une dépendance inter-sinistres et inter-branches.

Nous proposons d'abord le modèle des copules dynamiques comme alternatif du modèle des copules statiques, pour modéliser la dépendance entre deux branches d'assurance.

Nous choisissons le modèle GACSM pour estimer les marges.

Nous appliquons ces modèles sur deux branches d'assurance. Les résultats obtenus montrent qu'un effet de diversification des réserves peut se produire en choisissant le modèle de copule dynamique.

Introduction

Forecasting claims reserves and assessing the accurate solvency capital are a crucial issues for actuaries. Classical approaches assume that claims of different lines of business are independent. However, it is unlikely that claims to different policies are independent, as for instance; a car accident can induce claims on both auto and health policies, a fire may spread from one building to another resulting in claims from two or more policyholders, so one event may involve series of claims. Thus to assume that different policies of lines of business are independent can lead to over or under estimation of the aggregate loss. It has pointed out that methods, which enable modeling dependencies of different policies of lines of business are needed in order to improve the accuracy of the estimation of the total reserves and solvency capital.

Copulas functions have been proposed by several actuarial studies as a relevant tools to handle different dependence structure. They were introduced to actuarial mathematics by Frees and Valdez (1998). Since then, a large number of investigations have applied copulas functions to fully capture a wide range of dependence structures amongst different insured risks. Frees and Wang (2006) have used copulas for estimating the credibility of aggregate loss. In addition, Kaishev and Dimitrova (2006) have shown the importance of copulas in reinsurance. Antonin and Benjamin (2001) also used copulas to assess the amount of the reserve, they provided a model which combined at the same time the theory of copula and the theory of credibility in order to detect better the dependence between the lines of business. Furthermore, Belguise and Levi (2002), Faivre (2002) and Cadoux and Loizeau (2004) have shown that the model with copulas allows for an aggregation of risks and evaluates a capital higher than when assuming independence. In addition, Bargés et al. (2009) evaluated the capital allocation for the overall portfolio using the TVaR as a measure of risk and a FGM copula. Besides, Diers et al. (2012) have shown the flexibility of a Bernstein copula to model a several lines of business. Moreover Zhang and Dukic (2013) have introduced a Bayesian multivariate model based on the use of parametric copula to model dependencies between various lines of insurance claims.

More recently, a more relevant technique was introduced in actuarial sciences, that is the copula regression model, where distributions are specified conditional on a set of covariates. Some recent applications include the modeling of insurance claims (Frees and Valdez (2008), Frees et al. (2009)). The prediction of reserves (Zhao and Zhou (2010), Shi and Frees (2011)). The examination of asymmetric information (Shi and Valdez (2011), Shi and Valdez (2012)), and the analysis of insurance claims (Shi and Frees (2010), Shi and Valdez (2012), Krämer et al. (2013)), among others.

Despite the popularity of copulas in dependence modelling, their application in time series data is not considered. The majority of approaches proposed in actuarial science do not involve the temporal dependence problem, when a serial dependencies and a time varying dependence structure of claims can be captured.

Recently, a model proposed by Pešta and Okhrin (2014) for time series claims data, introduces a Generalized time series model for conditional mean and variance of one run-off triangle, and elaborates copula approach for dependence modelling within their proposed model. They proved, that their model improves the consistency of parameters estimates and the precision of reserve distribution as well. Therefore it is important to consider the temporal dependence structure of claims, and to model dependence structure with a more development copulas functions, which can follow the change in dependence over time.

Their paper deals with the temporal dependence structure of one run-off triangle, but time varying dependence structure of several run-off triangles is not considered.

In this paper, a new model handling time varying dependence structure between two run-off triangles is proposed. For that purpose, a conditional copula framework is used, which takes into account the dynamic behaviour of claims between two lines of business. Also, a time series Generalized linear model is developed to capture temporal dependence within each triangles. Furthermore, a simulation procedure is provided in order to estimate the aggregate reserves and the total solvency capital.

This paper is organized as follows: Section 4.1 presents the Autoregressive Conditional Sinistrality Model for marginals modelling. Section 4.2 provides the time varying copula model. Section 4.3 reports the empirical results, followed by concluding remarks in Section 4.4.

4.1 Marginal models

The primary goal of our study is to evaluate reserves and Solvency Capital taking into account the dynamic dependence structure between lines of business. For that, a suitable model for marginals must be identified. In this paper, a Generalized Autoregressive Conditional Sinistrality Model (GACSM) proposed by the authors see Araichi et al. (2015) is considered. This model assumes that claims amounts $y_{i,t}$ from the same accident quarter are correlated. A dynamic specification is considered

$$y_{i,t} = \psi_t \epsilon_{i,t}, \tag{4.1}$$

$$\epsilon_{i,t} \sim \text{i.i.d non negative variable with } E(\epsilon_{i,t}) = 1. \tag{4.2}$$

where $\epsilon_{i,t}$ denote the standardized claims amounts (residuals), that are assumed to belong to the exponential family.

We denote the conditional expectation of $y_{i,t}$, $E(y_{i,t} | \Omega_{i,t-1}) = \psi_{i,t} = [\psi_{i,1}, \dots, \psi_{i,T_i}]$, where $\Omega_{i,t-1} = \{x_{i,1}, x_{i,2}, \dots, x_{i,t-1}, y_{i,1}, y_{i,2}, \dots, y_{i,t-1}\}$ is the previous informations until the development quarter $t - 1$.

This conditional expectation of claims amounts is related to the predictor $\eta_{i,t}$ by the link function such that $g(\psi_{i,t}) = \eta_{i,t}$. As pointed out by Merz and Wüthrich (2008) a log link is typically a natural choice in the insurance reserving context.

Here, the problem of temporal dependence of development quarters is considered. For that, we suppose a time dependent specification for $\eta_{i,t}$ defined by

$$\eta_{i,t} = X_{i,t}^\top \beta + \sum_{j=1}^p \theta_j \{g(y_{i,t-j}) - X_{i,t-j}^\top \beta\} + \sum_{j=1}^q \phi_j \{g(y_{i,t-j}) - \eta_{i,t-j}\} \quad (4.3)$$

where $X_{i,t}^\top$ is a $z \times p$ matrix of dummy covariates that arranges the impact of accident and development quarter on the claims amounts through model parameter $\beta \in \mathbb{R}^{p \times 1} = [\gamma, \alpha_2, \dots, \alpha_n, \delta_2, \dots, \delta_{T_i}]$, where α_i stands for the effect of accident quarter i , and δ_t represents the effect of the development quarter t , and γ is the constant parameter of the model (taking $\alpha_1 = \delta_1 = 0$ to avoid over parametrization).

Standardized claims amounts are assumed to follow a Gamma distribution with $E(\epsilon_{i,t}) = 1$.

Given $\Omega_{i,t-1}$, the distribution and the density function of the Gamma GACSM are presented respectively as

$$\forall i = 1 \dots n, t = 2, \dots, T_i, T_i = T + 1 - i$$

$$F(y_{i,t}) = \frac{\Gamma(k, \frac{y_{i,t}}{\psi_{i,t}})}{\Gamma(k)} \quad (4.4)$$

$$f(y_{i,t} | \Omega_{i,t-1}) = \frac{y_{i,t}^{k-1} k^k}{\psi_{i,t}^k \Gamma(k)} \exp\left(-\frac{y_{i,t}}{\psi_{i,t}}\right) \quad (4.5)$$

where k is the shape parameter, $\psi_{i,t} = \exp(\eta_{i,t})$ and $\eta_{i,t}$ is given by

$$\eta_{i,t} = X_{i,t}^\top \beta + \sum_{j=1}^p \phi_j \{\log(y_{i,t-j}) - X_{i,t-j}^\top \beta\} + \sum_{j=1}^q \theta_j \{\log(y_{i,t-j}) - \eta_{i,t-j}\} \quad (4.6)$$

the log likelihood of the Gamma GACSM is then formalized as

$$L(y_{i,t} | \Omega_{i,t-1}) = \sum_{t=1}^T ((k-1) \log(y_{i,t}) + k \log(k) - k \log(\psi_{i,t}) - \log(\Gamma(k)) - \frac{y_{i,t} k}{\psi_{i,t}}) \quad (4.7)$$

The parameters of GACSM are estimated by maximizing the log likelihood function, the procedure yields a consistent and asymptotically normal estimates.

The GACSM model allows to predict future claims amounts and future reserves for each lines of business. When claims amounts are regularly time dependent distributed, one can ask for the aggregate total reserves of lines of business. In the next section, dynamic dependence structure is analysed between two lines of business in order to evaluate the aggregate reserves and Solvency Capital Requirement (SCR).

4.2 Time varying copula model

4.2.1 Motivation for analysing the time varying copula

It is often believed that dependence between risks has an impact on the level of aggregate losses of the company. For that, the solvency 2 framework requires adequate modelling of the dependences between different lines of business. Correlation coefficient used by solvency 2 is not sufficient to describe the dependence structure, since the distribution of claims amounts in insurance is non linear and non elliptical.

Actuarial research used copulas functions as a more informative measures of dependence between claims than linear correlation. These functions proved their success in allowing a more general dependence structure, than simple linear correlation, but it seems unrealistic to treat dependence as constant. It is questionable to assume that the structure of dependence between claims is the same over years. As for instance, a car accident or climates conditions (snowy, rainy) may create a positive dependence between the two lines of business Auto damage and Auto liability. They can cause both a material damage related to the car and a human damage related to the driver. These two lines of business may be weakly dependent as well as strongly dependent conditional on the intensity of the damage caused by the accident or the intensity of the climates conditions. Also, this dependence may vary through years. In order to capture the dynamic behaviour of dependence between these two lines of business, a time varying copulas models are considered, that allow the parameters of the copula to evolve across time, according to a dynamic process. To our knowledge is the first paper to investigate conditional copulas in insurance.

4.2.2 Conditional copulas models

To capture dependence between the two lines of business, the standardized claims amounts (residuals) obtained from GACSM estimation are considered to be modelled by dynamic copulas.

Let $\epsilon_{i,t}^1$ and $\epsilon_{i,t}^2$ denote the standardized claims amounts, classified in two run-off triangles of respectively, Auto Damage and Auto Liability, $\{\epsilon_{i,t}^1, \epsilon_{i,t}^2; i, t \in \nabla\}$, where $\nabla = \{i = 1, \dots, n, t = 1, \dots, T_i\}$.

i refers to accident quarter and t refers to development quarter with n is the most recent accident quarter, and $T_i = T + 1 - i$ is the last delay quarter. Let $\Omega_{i,t-1}$ denotes the information set available at time $\{i, t - 1\}$.

In order to capture the dynamic dependence structure existing between the cells of these two triangles, we assume a conditional cumulative distribution functions of $\epsilon_{i,t}^1$ and $\epsilon_{i,t}^2$ as $F(\epsilon_{i,t}^1 | \Omega_{i,t-1})$ and $G(\epsilon_{i,t}^2 | \Omega_{i,t-1})$, respectively.

Therefore, the conditional copula function, denoted as $C_{i,t}(u_{i,t}, v_{i,t} | \Omega_{i,t-1})$ is defined by the two cumulative distribution functions of random variables $u_{i,t}$ and $v_{i,t}$.

$u_{i,t} = F(\epsilon_{i,t}^1 | \Omega_{i,t-1})$ and $v_{i,t} = G(\epsilon_{i,t}^2 | \Omega_{i,t-1})$ are the probability integral transform of respectively $\epsilon_{i,t}^1$ and $\epsilon_{i,t}^2$. They are known to have the uniform distributions, regardless of the original distributions, F and G .

To define the dynamic copula function, a conditional version of the Sklar theorem is proposed by Patton (2001). It is given as

Theorem 1 *Let H be a conditional bivariate distribution function with continuous margins F and G , and let $\Omega_{i,t-1}$ be some conditioning set, then there exists a unique conditional copula $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that*

$$H(\epsilon_{i,t}^1, \epsilon_{i,t}^2 | \Omega_{i,t-1}) = C(u_{i,t}, v_{i,t} | \Omega_{i,t-1}) = C(F(\epsilon_{i,t}^2 | \Omega_{i,t-1}), G(\epsilon_{i,t}^1 | \Omega_{i,t-1}) | \Omega_{i,t-1}) \forall \epsilon_{i,t}^1, \epsilon_{i,t}^2 \in \overline{\mathbb{R}} \quad (4.8)$$

Conversely, if C is a conditional copula and F and G are the conditional distribution functions of $\epsilon_{i,t}^1$ and $\epsilon_{i,t}^2$ respectively, then the function H is a bivariate conditional distribution function with margins F and G .

The bivariate conditional density function of $\epsilon_{i,t}^1$ and $\epsilon_{i,t}^2$ can be constructed by the product of their copula density and their two marginal conditional densities, respectively denoted by f and g :

$$h(\epsilon_{i,t}^1, \epsilon_{i,t}^2 | \Omega_{i,t-1}) = c(u_{i,t}, v_{i,t} | \Omega_{i,t-1}) \times f(\epsilon_{i,t}^1 | \Omega_{i,t-1}) \times g(\epsilon_{i,t}^2 | \Omega_{i,t-1}) \quad (4.9)$$

where

$$c(u_{i,t}, v_{i,t} | \Omega_{i,t-1}) = \frac{\partial^2 C(u_{i,t}, v_{i,t} | \Omega_{i,t-1})}{\partial u_{i,t} \partial v_{i,t}},$$

$f(\epsilon_{i,t}^1 | \Omega_{i,t-1})$ is the conditional density function of $\epsilon_{i,t}^1$ and $g(\epsilon_{i,t}^2 | \Omega_{i,t-1})$ is the conditional density function of $\epsilon_{i,t}^2$.

In the insurance context, Frees and Valdez (1998) provide a number of parametric static copulas. In the present work, various time varying parametric copulas are considered: Specifically, time varying Normal copula, time varying Clayton copula, time varying symmetrised Joe-Clayton Copula and time varying Gumbel copula.

Time varying Normal copula

This copula is derived from the bivariate normal distribution and is defined by

$$C(u_{i,t}, v_{i,t}) = \int_{-\infty}^{\Phi^{-1}(u_{i,t})} \int_{-\infty}^{\Phi^{-1}(v_{i,t})} \frac{1}{2\pi\sqrt{(1-\rho_t^2)}} \exp\left\{-\frac{y_{i,t}^2 - 2\rho_t y_{i,t} z_{i,t} + z_{i,t}^2}{2(1-\rho_t^2)}\right\} dy_{i,t} dz_{i,t} \quad (4.10)$$

where ρ_t is the correlation coefficient that follows a time varying process

$$\rho_t = \Lambda(\omega + \beta\rho_{t-1} + \alpha | u_{i,t-1} - v_{i,t-1} |) \quad (4.11)$$

where $\Lambda(x)$ is defined as $(1 - e^{-x})(1 + e^{-x})$: is the modified logistic transformation to keep ρ_t in $(-1, 1)$ all times. This copula does not have a tail dependence, $\lambda_t^U = \lambda_t^L = 0$.

Time varying Clayton copula

This copula models positive dependence. It represents the risks which are more concentrated in the lower tail, so it correlates small losses. It is defined by

$$C(u_{i,t}, v_{i,t}) = (u_{i,t}^{-\theta_t} + v_{i,t}^{-\theta_t} - 1)^{-1/\theta_t} \quad \theta_t \in [-1, \infty) \setminus \{0\} \quad (4.12)$$

Time-varying dependence processes for the Clayton copula is described as

$$\theta_t = (\omega + \beta\theta_{t-1} + \alpha | u_{i,t-1} - v_{i,t-1} |) \quad (4.13)$$

This copula has lower tail dependence $\lambda_t^L = 2^{\frac{1}{\theta_t}}$ and $\lambda_t^U = 0$.

Time varying symmetrised Joe-Clayton copula

The symmetrized Joe-Clayton copula introduced by Patton (2006) is a flexible two-parameters copula, that is parametrized in terms of λ^U and λ^L . A time-varying version of this copula, as considered in Patton (2006), allows for changing degrees of asymmetry, as well as a time-varying overall level of dependence. It is defined as

$$C(u_{i,t}, v_{i,t}) = 0.5(C_{JC}(u_{i,t}, v_{i,t}) + (C_{JC}(1 - u_{i,t}, 1 - v_{i,t}) + u_{i,t} + v_{i,t} - 1)) \quad (4.14)$$

where C_{JC} is the Joe-Clayton copula, also called BB7, given by

$$C(u_{i,t}, v_{i,t}) = 1 - \left(1 - \left\{ [1 - (1 - u_{i,t})^{a_t}]^{-b_t} \right\} + [1 - (1 - v_{i,t})^{a_t}]^{-b_t} - 1 \right)^{\frac{-1}{b_t}} \quad (4.15)$$

with

$$a_t = \frac{1}{\text{Log}(2 - \lambda_t^U)} \quad \text{and} \quad b_t = \frac{-1}{\text{Log}(\lambda_t^L)}, \quad \lambda_t^U, \lambda_t^L \in (0, 1)$$

The time dynamics equations for the parameters λ_t^U and λ_t^L are

$$\lambda_t^U = \Lambda(\omega_U + \beta_U \lambda_{t-1}^U + \alpha | u_{i,t-1} - v_{i,t-1} |) \quad (4.16)$$

$$\lambda_t^L = \beta(\omega_L \beta_L \lambda_{t-1}^L + \alpha | u_{i,t-1} - v_{i,t-1} |) \quad (4.17)$$

where $\Lambda(x) = (1 - e^{-x})^{-1}$.

Time varying Gumbel copula

This copula models a positive dependence and represents the risks which are more concentrated in the upper tail. It is defined as

$$C(u_{i,t}, v_{i,t}) = \exp\{-[(-\ln u_{i,t})^{\theta_t} + (-\ln v_{i,t})^{\theta_t}]^{1/\theta_t}\} \quad \theta_t \geq 1 \quad (4.18)$$

The dynamic process for the dependence parameter is given

$$\theta_t = (\omega + \beta\theta_{t-1} + \alpha | u_{i,t-1} - v_{i,t-1} |) \quad (4.19)$$

where $\Lambda(x) = (1 - e^{-x})^{-1}$.

Time varying Gumbel copula has an upper tail dependence $\lambda_t^U = 2 - 2^{1/\theta_t}$ and $\lambda_t^L = 0$.

For these copulas functions, dynamic process allows dependence between claims of the cells of triangles. Past information is taken from claims of the same accident quarter, that means between development quarters, and we assume that claims from different accident quarters are independent.

4.2.3 Copula estimation

For parameter estimation, we are based on Canonical Maximum Likelihood (CML) method. This method uses empirical probability integral transform in order to obtain the uniform marginals needed to estimate the copula parameter. Copula parameter can then be estimated by maximizing the log likelihood function of the copula density using the uniform variables given by

$$L_c = \sum_{t=1}^T \text{Log}(c(\hat{F}(\epsilon_{i,t}^1), \hat{G}(\epsilon_{i,t}^2))) = \sum_{t=1}^T \text{Log}(c(\hat{u}_{i,t}, \hat{v}_{i,t})) \quad (4.20)$$

4.2.4 Simulation method

In order to evaluate the Solvency Capital, we need to estimate the whole predictive distribution of reserves. Here, a simulation procedure that allows to predict reserves taking into account the dynamic dependence between the two lines of business is presented:

Step 1: The lags (p,q) of the temporal dependence are specified, then the parameters of the Gamma GACSM(p,q) model of $y_{i,t}$ are estimated, using maximum likelihood estimation for each lines of business.

Step 2: Using parameters estimates from step 1, future claims amounts of the lower unobserved triangle are predicted, we obtain $\hat{\psi}_{i,t}^1$ and $\hat{\psi}_{i,t}^2$, respectively for Auto Damage and Auto liability.

where:

$$\hat{\psi}_{i,t}^k = \exp(\hat{\eta}_{i,t}), \forall k = 1, 2 \text{ and}$$

$$\hat{\eta}_{i,t} = X'_{i,t} \hat{\beta} + \sum_{j=1}^p \hat{\phi}_j \{\log(y_{i,t-j}) - X'_{i,t-j} \hat{\beta}\} + \sum_{j=1}^q \hat{\theta}_j \{\log(y_{i,t-j}) - \hat{\eta}_{i,t-j}\}.$$

Step 3: We simulate the joint vectors $(u_{i,t}, v_{i,t})$ from the subjacent ⁵ copula of the lower unobserved triangle, $\{u_{i,t}, v_{i,t}; i, t \in \Delta\}$, where $\Delta = \{i = 2, \dots, n; t = n - i + 2, \dots, T_i\}$.

⁵The subjacent copula is the best copula for the data: we estimate parameters of various copulas. Based on the information criteria and a test of copula, we are able to select the best copula.

Step 4: Residuals vectors $(\hat{\epsilon}_{i,t}^1, \hat{\epsilon}_{i,t}^2)$ can be obtained as $\hat{\epsilon}_{i,t}^1 = \hat{F}^{-1}(u_{i,t})$ and $\hat{\epsilon}_{i,t}^2 = \hat{G}^{-1}(v_{i,t})$, where F and G are the estimated standardized claims amounts distributions.

Step 5: Future claims amounts for Auto Damage and Auto Liability are computed as, $\hat{y}_{i,t}^1 = \hat{\psi}_{i,t}^1 \hat{\epsilon}_{i,t}^1$ and $\hat{y}_{i,t}^2 = \hat{\psi}_{i,t}^2 \hat{\epsilon}_{i,t}^2$.

The aggregate future claims reserves for each origin quarter i and the aggregate total reserves are respectively

$$\hat{R}_i = \sum_{t \in \Delta_i} (\hat{y}_{i,t}^1 + \hat{y}_{i,t}^2) \quad (4.21)$$

$$\hat{R} = \sum_{i=1}^n \hat{R}_i \quad (4.22)$$

Step 6: Repeat step 3, step 4 and step 5 B times to obtain the aggregate total reserves distribution.

4.3 Empirical results

4.3.1 Marginal modelling

Data were supplied by a French insurance company and consist of incremental claims amounts of two lines of business Auto Damage and Auto Liability, over the period January 2002, to December 2007. The motivation for considering these two lines of business is that they present a large number of claims in the insurance company, and they are highly dependent.

The first guarantee covers all damaged caused to the car, even if the insured is responsible of the accident. It is a short line of business, that once the claim is declared it will be paid after a short time. The second, is the guarantee of damage caused to others, belongs to the Civil Responsibility lines of business. it is called a long tail line of business, because there will be a time lag between the date of occurrence claim and the date of full refund. These claims amounts are classified quarterly into two run-off triangles. We have 300 observations for each lines of business spread over 24 accident quarters and 24 development quarters.

The claims amounts are net of reinsurance, net of subrogation and we suppose that inflation remains unchanged throughout the quarters.

To investigate the characteristics of these claims, summary statistics of claims amounts of the two lines of business are presented in table 4.1.

	Auto damage	Auto liability
max	517.129	497.472
min	0	0
mean	64.131	62.670
std	115.165	96.369
skewness	2.005	2.097
kurtosis	6.028	6.807
Jarque Bera test	315.7 (0.001)	401.02 (0.001)

(.) contains p-value of the Jarque Bera test

Table 4.1: Descriptive statistics of claims amounts (in thousand €)

For the lines Auto Damage and Auto Liability, the minimum and maximum claims amounts suggest that the distributions are widely spread. The standard deviation is larger than the mean, implying the skewness of the data. In fact, the two distributions of claims amounts exhibit a positive skewness and excess kurtosis indicating that are heavy tailed and right skewed. According to Jarque Bera test, there is evidence that the two distributions of claims amounts are not normally distributed.

The ACF plot (figure 4.5) presented in the Appendix 4.6 shows a significant lags, indicating that claims amounts exhibit temporal correlations. These correlations are positive, indicating that large or small claims amounts tend to occur together.

For these two lines of business, the GACSM model is fitted. Results are reported in tables 3.4 and 3.5 in Chapter 4.

From tables 3.4 and 3.5 in Chapter 4, we remark that parameters estimates are significantly different from zero. Moreover, from table 4.5 in the Appendix 4.6, there is no autocorrelation in the residuals as shown by the statistics of the Ljung Box test.

Also, figure 4.6 and 4.7 in the Appendix 4.7 show that the QQplot of residuals of the Auto Damage and Auto Liability versus respectively the Gamma (0.367,2.75) and Gamma (0.527,1.9) distributions are very close. So the two lines of business are well specified.

These two lines of business present a Pearson's correlation coefficient equal to 0.93, they are highly dependent. Therefore, a car accident can affect the line Auto Damage as well as the line Auto Liability. But, one can ask if this linear correlation is constant over time. For that, in the next section the two marginals of residuals obtained from the estimation with GACSM are estimated with a dynamic copula model, which can capture a non linear as well as a time varying dependence structure.

4.3.2 Modelling copulas

For the lines of business Auto Damage and Auto Liability, the standardized claims amounts (residuals) are transformed into uniform marginals using the empirical probability integral transform. The log likelihood function of each copula density is maximized to obtain parameters estimates. (All estimates of copulas functions were made using the copula toolbox provided by Andrew Patton (University of Oxford-Department of economics)) and some functions created by authors on MATLAB 7.

Results are reported in table 4.2. Figure 4.1 illustrates the time varying parameters estimates corresponding to each copula function.

	tvar Normale	tvar Clayton	tvar SJC		tvar Gumbel
ω	0.035 (31.36)***	0.382 (62.4)***	0.932 (3.96)**	-2.820 (12.96.6)***	2.172 (146.1)***
β	0.093 (10.24)***	-1.723 (5.76)**	-24.915 (3.92)**	-0.870 (73.96)***	-1.264 (140.9)***
α	1.809 (29.7)***	0,968 (5.29)**	-5.565 (10.3)***	8.218 (15.3)***	-1.510 (17.64)***
LL	7.581	11.726	13.469		15.498
AIC	-15.142	-17.452	-26.898		-30.976
BIC	-15.105	-6.341	-26.824		-30.939

(.) contains the corresponding Wald test statistics,

** : 5% significant statistics,

*** : 1% significant statistics.

Table 4.2: Parameters estimates of time varying copulas (tvar). Log likelihood function (LL), AIC and BIC criteria.

As we can see from table 4.2, parameters estimates of the time varying copulas are statistically different from zero. This indicates that the dependence between the two lines of business changes in time. This dynamic behavior of dependence is captured by the specific properties of each copula function (equations (4.11), (4.13), (4.16), (4.17) and (4.19)).

Indeed, from figure 4.1 Normal copula panel (a) displays a dynamic correlation coefficient that shows a significant dynamic dependence between claims amounts. The value of the constant parameter is 0.194 and time evolution takes place around this value and becomes almost constant over the recent period.

Clayton copula in Panel (b), which measures the dependence in the lower tail, captures a constant coefficient of 0.289. The time evolution of the parameter fluctuates almost below the constant and equals even zero during certain periods. This result shows that small amounts of the two lines of business are weakly dependent. A small car damage does not lead necessarily to a claims in the line Auto Liability.

This behavior is also captured by the SJC copula, which measures dependence both in the lower and upper tail. The coefficient of lower tail dependence is 0.0156. The evolution in time of this parameter remains around this value during the period of study, except for recent period when it raises and achieves 0.972. This result is informative, showing certain standard behavior for the lower dynamic dependence of small claims amounts. The coefficient of upper tail dependence for SJC copula is shown in Panel (d). Constant coefficient of upper tail is equal to 0.1403 higher than the lower tail. This result is confirmed by the evolution in time of this parameter. The upper tail exhibits a high time varying dependence between claims amounts that achieves to 0.7.

Gumbel copula captures also a highly time varying dependence that evolves from 1 to 2.28, indicating that large claims amounts are strongly dependent. This means that a car accident can induce a large claims amounts into the line Auto Damage as well as into the line Auto Liability, as it causes a material damage, and it can frequently cause physical injuries related to the civil responsibility.

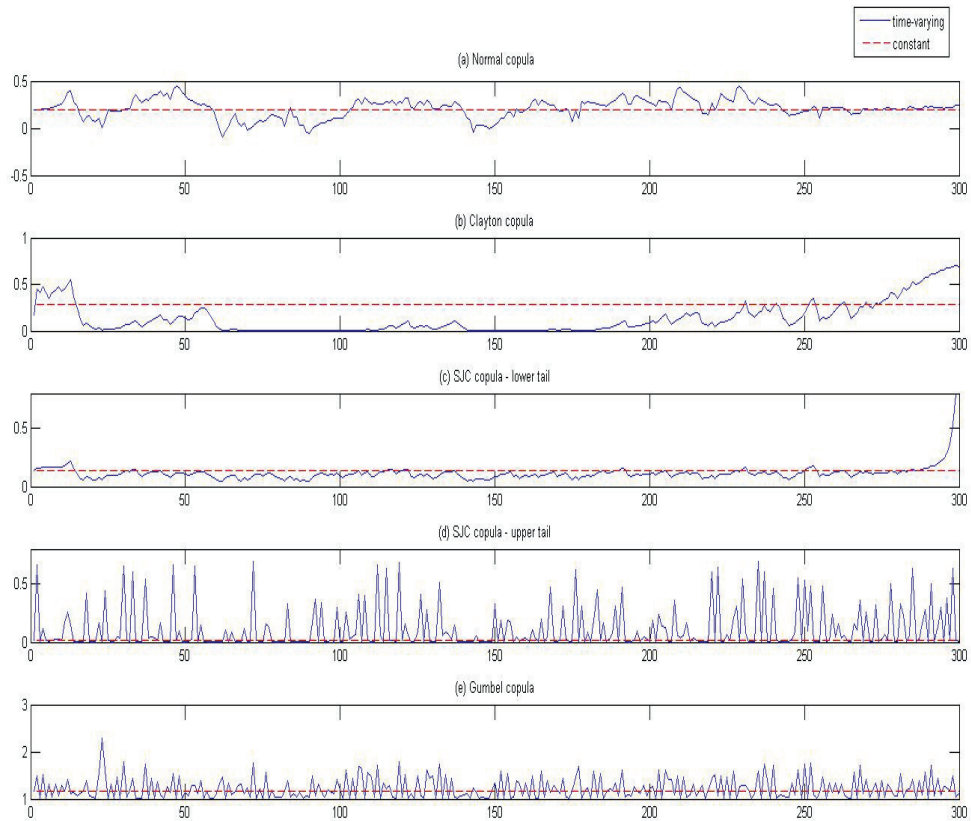


Figure 4.1: Time varying parameters estimates corresponding to Normal, Clayton, SJC and Gumbel copulas

To choose the best copula model to our data, we are based on the loglikelihood function (LL), AIC and BIC criteria. From table 4.2 we can conclude that the time varying Gumbel copula model presents a higher LL and a lower AIC and BIC values. Hence, this model gives a better fit for data than the other models and this copula provides a more accurate description of claims amounts behavior.

To validate this choice, a Vuong test ⁶is performed for each pair of copula. This test is based on the likelihood ratio and is proposed by Vuong (1989) to non nested models (see Vuong (1989) for more details on the Vuong test). This test is appropriate as our copula models are non nested. Table 4.3 displays the results of the Vuong test.

		Model B			
		tvar Gumbel	tvar SJC	tvar clayton	tvar Normale
Model A	tvar Gumbel		(1.41)	tvar Gumbel (4.28)	tvar Gumbel(5.9)
	tvar SJC			tvar SJC (4.93)	tvar SJC (4.47)
	tvar Clayton				tvar Normale (-1.4)

Table 4.3: Pairwise Vuong tests. We display the copula family that is selected on a $\alpha = 0.05$ level. In parentheses, we display the value of the Vuong test statistic.

For each pair, we present the copula family that is selected on a $\alpha = 0.05$ level. In parentheses, we display the value of the Vuong test statistics. Note that a statistic value > 2 indicates that we select model A, and a value < -2 indicates that we select model B. Otherwise, no decision among the two copula families is possible. As shown in table 4.3, the tvar Gumbel copula model significantly outperforms the tvar Clayton and the tvar Normal copulas models. But, no decision among the tvar Gumbel and the tvar SJC copulas could be made. So, there is no substantial difference between these two copulas models. Based on results reported in table 4.2 (LL , AIC , BIC), tvar Gumbel copula is preferred to tvar SJC copula model. Therefore, in the following section, we continue our analysis with the tvar Gumbel copula model.

4.3.3 Impact of time varying dependence on total loss

To investigate the impact of time varying dependence on the evaluation of reserves and solvency capital, a simulation study is conducted. 1000 replications of the lower unobserved triangle are drawn from the tvar Gumbel copula to the two lines of business as in section 4.2.4). The simulated claims are summed up to an aggregated total reserves distribution as in equation 4.22. Then, some quantities are computed on these aggregate total reserves: Mean, Value at Risk ($VaR_{99,5\%}$), Conditional Value at Risk ($CVaR_{99,5\%}$) and Solvency Capital Requirement (SCR)⁷.

⁶ $H_0: \frac{LR(\hat{\beta}, \hat{\gamma})}{n^{1/2}\hat{\omega}}$ where
 $LR(\hat{\beta}, \hat{\gamma}) = Ln f(y|x; \hat{\beta}) - Ln g(y|x; \hat{\gamma})$
 $\hat{\omega}^2 = \sum_{i=1}^n \left[Ln \frac{f(y|x; \hat{\beta})}{g(y|x; \hat{\gamma})} \right]^2 - \left[\frac{1}{n} \sum_{i=1}^n Ln \frac{f(y|x; \hat{\beta})}{g(y|x; \hat{\gamma})} \right]^2$

⁷Solvency Capital Requirement of reserve risk for a line of business l , is given by

$$SCR_l = VaR_{99,5\%}^l - BE_l \tag{4.23}$$

where $VaR_{99,5\%}^l$ is the Value at Risk of the line business l at confidence level 99,5%, and BE_l is the best estimate of the total reserves of the line of business l , computed as the mean of the total reserves.

For comparison purposes, we present below the results of modelling with the static Symmetrized Joe Clayton copula model ⁸, Generalized Autoregressive Conditional sinistrality model and the classical Chain Ladder model (CL). Indeed, data are fitted, and aggregate total reserves distributions are predicted under each model.

We specify that model (1) is the time varying Gumbel copula model with GACSM marginals.

Model (2) considers that the two marginals are modelled and predicted with a GACSM model, and the joint distribution of reserves is predicted from a SJC static copula, where the dependence structure is constant over time. (simulation procedure for this copula is the same as in 4.2.4).

Model (3) is the GACSM without copula. The two marginals are predicted independently using the GACSM model and the joint distribution is the sum of the two marginals distributions. Indeed, temporal dependence structure is identified among claims amounts of each line of business, and we assume that claims amounts between the two lines of business are independent.

Model (4) is the Chain Ladder model (CL) usually considered in the literature as the benchmark model. Here, a classical bootstrap version of Chain Ladder is conducted to obtain the total reserves distributions (see England and Verrall (2002)).

The Results of the proposed approach time varying copula model and the copula static model, the Generalized Conditional Sinistrality model and the traditional Chain Ladder model are compared numerically in table 4.4 and graphically in figures 4.2, 4.3 and 4.4.

	tvar Gumbel	static SJC	GACSM	CL
Mean	2569.867	2449.716	2780.678	2660.9
$VaR_{99.5\%}$	3903.742	3575.782	4478,167	3726.1
$CVaR_{99.5\%}$	4185.886	3746.795	4603.496	3801.4
SCR	1333.876	1126,066	1697.489	1107.2

Table 4.4: Mean, $VaR_{99.5\%}$, $CVaR_{99.5\%}$ and SCR (in thousand €) of the aggregate total reserves

⁸We note that we estimated four static copulas corresponding to the Normal, Clayton, Symmetrized Joe Clayton and the Gumbel copulas (more details about static copulas are presented in 4.8). According to LL, AIC and BIC criteria the best copula was the Symmetrized Joe Clayton (see table 4.6 in Appendix 4.8).

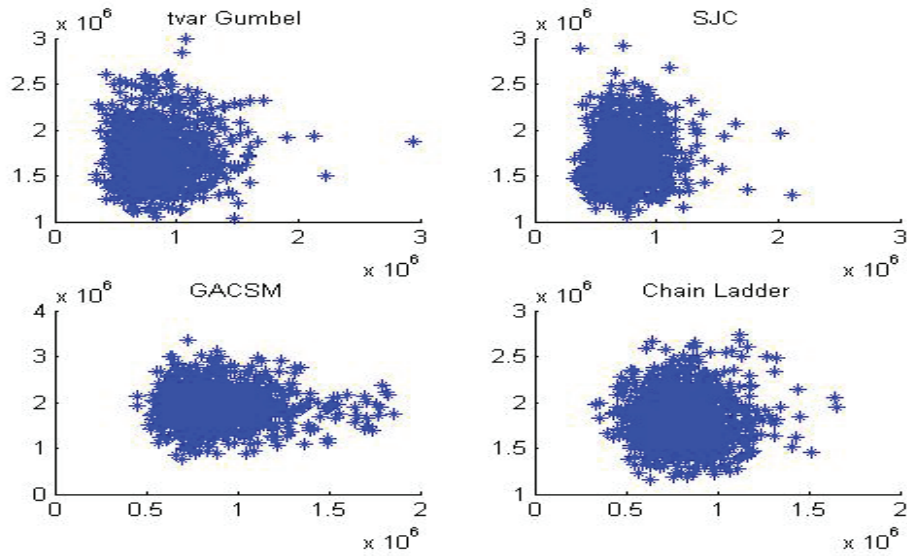


Figure 4.2: Scatter plots of the aggregate total reserves

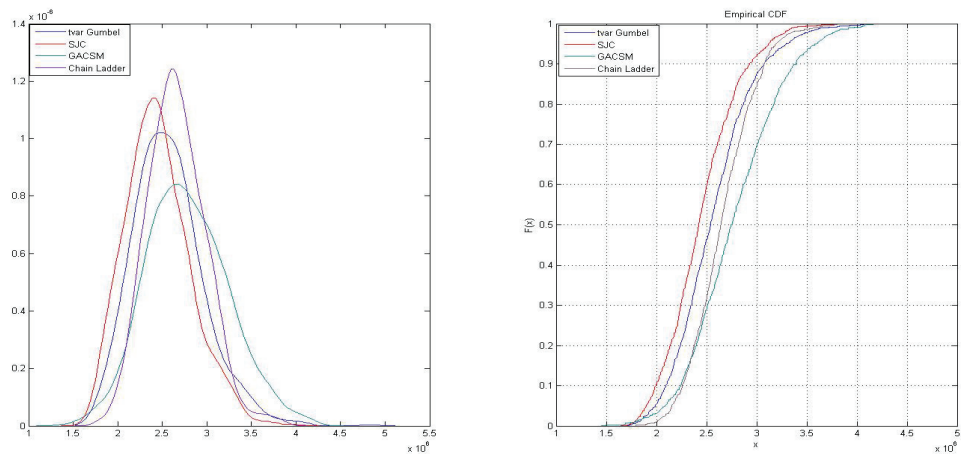


Figure 4.3: Total aggregate reserves distributions from different models

It is clearly visible from figure 4.2 that the three scatter plots of models (1), (2) and (3), considerably differ in range and shape from model (4). Total reserves under these models are more stretched than total reserves under model (4), when each line of business has been modelled independently of the other line.

Indeed, in the tvar Gumbel case the scatter plot shows a dependence in the upper right corner, while in the SJC case there is a reasonable dependence both in the lower and upper corner. In contrast to these models, Chain Ladder model presents a random pattern of the pairs of total reserves between the two business lines.

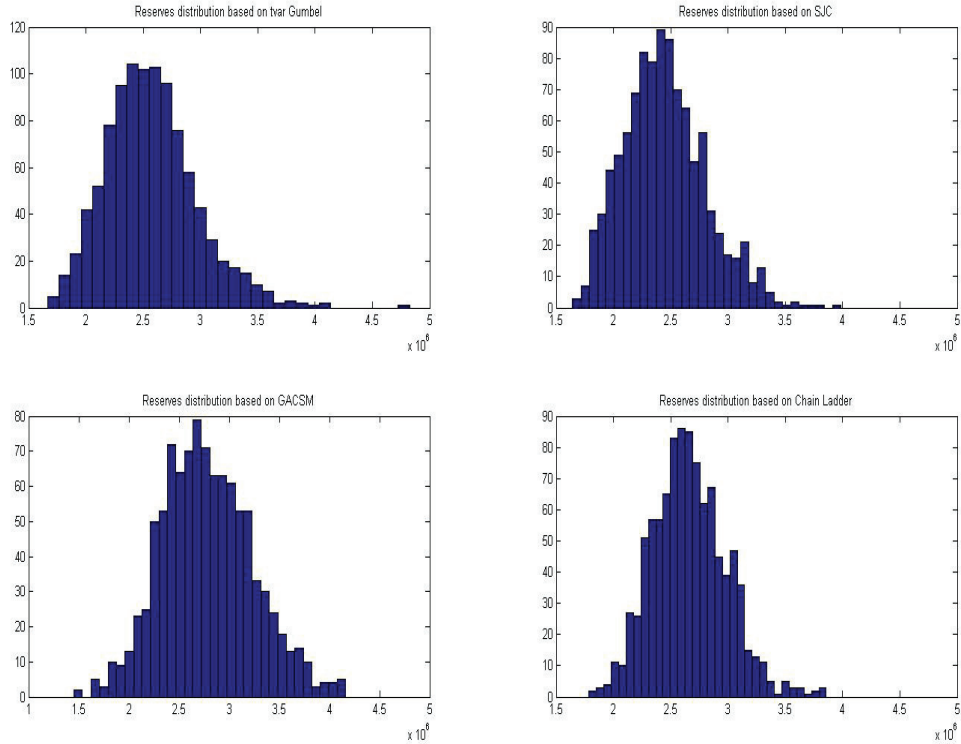


Figure 4.4: Total aggregate reserves distributions

Further features are captured by the cumulative and density distributions of figure 4.3. As we can see, predictive distributions of the aggregate reserves from Chain Ladder model is more tighter than other models. This is explained by the fact that this method ignores the dependence structure both among claims amounts and between the two lines. When modelling with GACSM, this model provides a fatter aggregate reserves distribution than copula models. Hence, considering the association between the two lines leads to estimate an aggregate total reserves lower than when we treat the two lines of business as unrelated.

In addition, figure 4.4 displays a right long tail distributions of the aggregate reserves under tvar Gumbel and SJC copula models. This is explained by the fact that these two copulas models take into account the dependence structure into the upper tail of the reserves distribution.

To show numerically the effect of dependence structure, some quantities are computed on the aggregate total reserves distributions for each model. Results are summarized in table 4.4.

At first sight, The results shed some light on the practical effects of dependence modelling. Specifically, tvar Gumbel copula provides a total aggregate amount of reserves significantly lower than the GACSM model and the Chain Ladder model.

However, it provides a 4.9% higher amount of reserves than the SJC model. Hence, a diversification effect could be gained on the amount of reserves when considering copula models. But, when we analyse the effect of the tvar Gumbel on the measures of risk, this copula produces considerably a higher $VaR_{99.5\%}$ and $CVaR_{99.5\%}$ and therefore a significantly higher SCR than the SJC copula and the Chain Ladder model. This explained by the fact that this copula takes into account the time dependence structure of risks, which are more concentrated in the upper tail, and therefore, it generates in additional amount of capital.

Moreover, SCR under tvar Gumbel copula is 21.42% lower than SCR under GACSM. Hence a diversification effect could be gained when we consider a time varying dependence structure between the two lines of business. However, SCR under this copula is 18.5% and 20.47% higher than SCR evaluated respectively under SJC copula model and Chain Ladder model. Hence, time varying dependence structure generate in additional amount of capital, contrary to the constant dependence structure that can induce an underestimation of risk. This effect is even more pronounced for the Chain Ladder model which ignores dependence among claims amounts as well between lines of business. So, this model could underestimate the risk.

4.4 Conclusion

Prior actuarial works have suggested a constant dependence structure to accommodate relation between lines of business. In this paper a time varying dependence structure was proposed to model dependencies between claims amounts of two lines of business. Such framework was demonstrated to be suitable for stochastic claims reserving of non life insurance.

The proposed model takes the advantage on traditional methods for being dynamic and it captures the evolution in time of the dependence structure. Indeed, contrary to classical reserving techniques, a temporal dependence structures are captured both between claims amounts of each line of business and between claims of the two lines of business. In this regard, GACSM enables us to specify a temporal dependence between development quarters of a line of business, and the time varying Gumbel copula model allows to capture the dynamic behavior of dependence between the two cells of triangles.

For these two lines of business a simulation study was performed to show the implication of such dependencies on predictive reserves and some measures such that VaR, CVaR and SCR. The time varying Gumbel copula with the GACSM provide a significant lower total reserves and SCR than the GACSM without copula.

Therefore, we can conclude that this dynamic behaviour of the structure of dependence provides a diversification effect between the two lines, and a potential benefit for in insurance company may be gained by reducing the cost of capital. In this regard, insurer might consider expanding the Auto Damage or shrinking the Auto Liability to take best advantage of the diversification effect.

4.5 Appendix

4.6 Autocorrelations of the lines Auto Damage and Auto liability

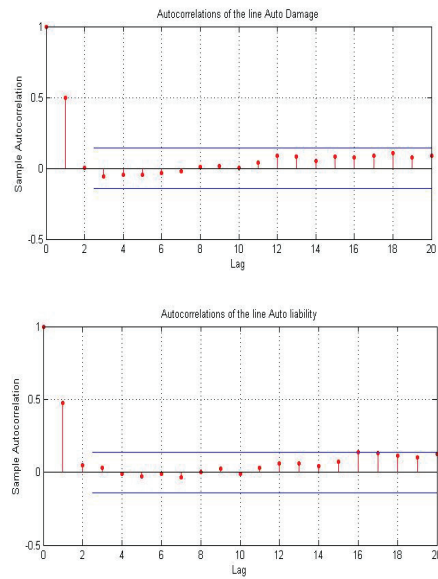


Figure 4.5: Autocorrelations of the line Auto Damage and Auto liability

Table 4.5: Mean, Variance and diagnostic tests of residuals (p-value inside parentheses) of Auto Damage and Auto Liability

	Auto Damage	Auto Liability
	GACSM(1,1)	GACSM(1,1)
Mean	0.996	0.999
Variance	1.454	1.1
LJB test Q(10)	11.02 (0.356)	15.05 (0.13)
LM test Q(10)	2.079 (0.996)	1.15 (0.999)

4.7 QQ plots

Figure 4.6: QQplot Auto damage residuals versus Gamma (0.367,2.75)

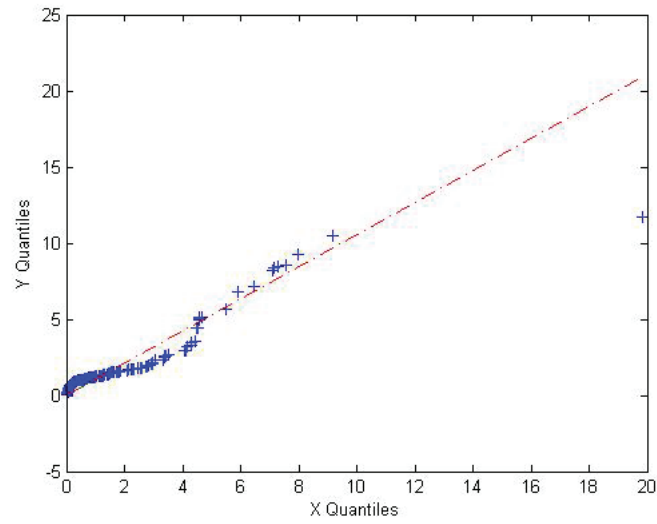
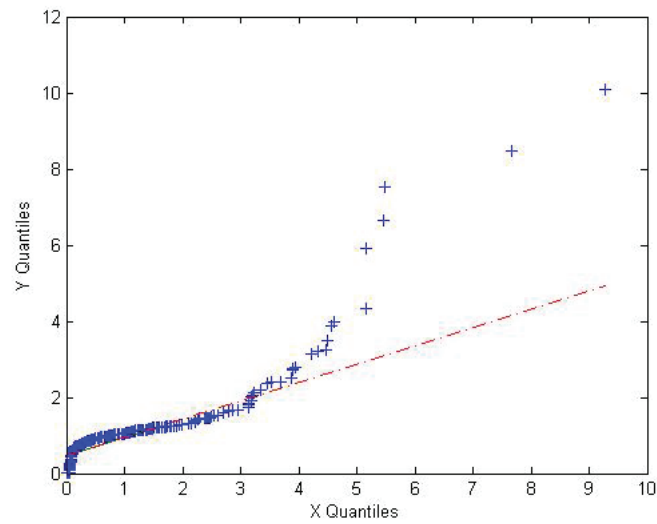


Figure 4.7: QQplot Auto liability residuals versus Gamma (0.527,1.9)



4.8 Copulas

In this section, we present statistic copulas for dependent claims amounts relative to the two lines of business. Copulas functions have been introduced in the insurance context by Frees and Valdez (1998). A copula is based on an assumption that both marginal distributions are known. For more details about copula functions, we refer readers to Nelsen (2006).

Let $F_X(x)$ and $F_Y(y)$ denote the marginal distribution functions of the variables X corresponding to the claims amounts of the line Auto Damage and Y corresponding to the claims amounts of the line Auto Liability.

The joint distribution function $F_{X,Y}(x, y)$ is then obtained as

$$F_{X,Y}(x, y) = C[F_X(x), F_Y(y)] \quad (4.24)$$

where $C(u, v)$ is the copula, a cumulative distribution function for a bivariate distribution with support on the unit square and uniform marginals.

In this paper, we assume that the marginal distributions are continuous with density functions $f_X(x)$ and $f_Y(y)$. Then, the joint density function is

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)C_{12}[f_X(x), f_Y(y)] \quad (4.25)$$

where

$$C_{12}(u, v) = \frac{\partial \partial C(u, v)}{\partial u \partial v}$$

The conditional distribution function of $Y | X = x$ is

$$F_{Y|X}(y | x) = C_1[F_X(x), F_Y(y)] \quad (4.26)$$

where

$$C_1(u, v) = \frac{\partial C(u, v)}{\partial u}$$

In the insurance context, Frees and Valdez (1998) provide a number of copulas. In each case, the parameter α is a constant parameter that measures the degree of association. In the present work, we used several static copulas in order to select the appropriate one, that describe the structure of dependence between the two lines of business.

We investigate Normal copula, Clayton copula, Symmetrized Joe Clayton copula and Gumbel copula.

Normal copula

This copula is derived from the bivariate normal distribution and is defined by

$$C(u_{i,t}, v_{i,t}) = \int_{-\infty}^{\Phi^{-1}(u_{i,t})} \int_{-\infty}^{\Phi^{-1}(v_{i,t})} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{y_{i,t}^2 - 2\rho y_{i,t}z_{i,t} + z_{i,t}^2}{2(1-\rho^2)}\right\} dy_{i,t} dz_{i,t} \quad (4.27)$$

where the correlation coefficient $\rho \in (-1, 1)$.

Clayton copula

This copula models positive dependence, it represents the risks which are more concentrated in the lower tail, so it correlates small losses. It is defined as

$$C(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha} \quad \alpha \in [-1, \infty) \setminus \{0\} \quad (4.28)$$

The Clayton copula has lower tail dependence $\lambda^L = 2^{-1/\alpha}$ and $\lambda^U = 0$.

Symmetrised Joe-Clayton copula

The symmetrized Joe-Clayton copula introduced by Patton (2006) is a flexible two-parameter copula that is parametrized in terms of λ^U and λ^L . A time-varying version of this copula, as considered in Patton (2006), allows for changing degrees of asymmetry, as well as a time-varying overall level of dependence. It is defined as

$$C(u_{i,t}, v_{i,t}) = 0.5(C_{JC}(u_{i,t}, v_{i,t}) + (C_{JC}(1 - u_{i,t}, 1 - v_{i,t}) + u_{i,t} + v_{i,t} - 1)) \quad (4.29)$$

where C_{JC} is the Joe-Clayton copula, also called BB7, given by

$$C(u_{i,t}, v_{i,t}) = 1 - \left(1 - \left\{ [1 - (1 - u_{i,t})^a]^{-b} \right\} + [1 - (1 - v_{i,t})^a]^{-b} - 1 \right)^{\frac{-1}{b}} \quad (4.30)$$

where

$$a = \frac{1}{\text{Log}(2 - \lambda^U)} \quad \text{and} \quad b = \frac{-1}{\text{Log}(\lambda^L)}, \quad \lambda^U, \lambda^L \in (0, 1)$$

Gumbel copula

This copula models a positive dependence and represents the risks which are more concentrated in the upper tail. It is defined as

$$C(u, v) = \exp\{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{1/\alpha}\} \quad \alpha \geq 1 \quad (4.31)$$

where α is the parameter dependence.

The Gumbel copula has upper tail dependence $\lambda^U = 2 - 2^{1/\alpha}$, but no lower tail dependence.

By using these copulas, we are able to detect dependence in the tails.

In order to estimate copulas model, we are based on the Canonical maximum likelihood method. This method consists in transforming the data of claims amount (x_1^t, \dots, x_N^t) into uniform variates $(\hat{u}_1^t, \dots, \hat{u}_N^t)$ using the empirical distribution functions, and then estimate the parameter in the following way

$$\hat{\alpha} = \underset{\alpha}{\text{argmax}} \sum_{t=1}^T \text{ln}c(\hat{u}_1^t, \dots, \hat{u}_N^t; \alpha) \quad (4.32)$$

The parameters of copulas are estimated by maximizing the log likelihood function of each copula, the procedure yields a consistent and asymptotically normal estimates.

	Normale	Clayton	SJC	Gumbel
κ	0.194 (2.04)**	0.289 (2.27)**	0.016 (3.12)***	0.140 (5.7)***
LL	5.777	6.719	6.955	6.7
AIC	-11.546	-13.431	-13.897	-13.394
BIC	-11.534	-13.419	-13.872	-13.381

(.) contains the corresponding Student test statistics,

** : 5% significant statistics,

*** : 1% significant statistics.

Table 4.6: Parameters estimates of static copulas. Log likelihood function (LL), AIC and BIC criteria.

Bibliography

- Antonin, G. and Benjamin, S. (2001). Effets de la dépendance entre différentes branches sur le calcul des provisions. *Astin Bulletin*.
- Araichi, S., de Peretti, C., and Belkacem, L. (2015). Generalized autoregressive conditional sinistrality model : a novel model for claims reserving in non life insurance. *ISFA working paper No 2015.4*.
- Bargés, M., Cossette, H., and Marceau, E. (2009). Tvar-based capital allocation with copulas. *Insurance: Mathematics and Economics*.
- Belguise, O. and Levi, C. (2002). Tempêtes : Étude des dépendances entre les branches automobile et incendie à l'aide de la théorie des copulas. *Bulletin Français d'actuariat*.
- Cadoux, D. and Loizeau, J.-M. (2004). Copules et dépendances : application pratique à la détermination du besoin en fonds propres d'un assureur non vie. *Astin Bulletin*.
- Diers, D., Eling, M., and Marek, S. D. (2012). Dependence modeling in non-life insurance using the bernstein copula. *Insurance: Mathematics and Economics*.
- England, P. and Verrall, R. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*.
- Faivre, F. (2002). Copula: A new vision for economic capital and application to a four line of business company. *Astin Bulletin*.
- Frees, E. W., Shi, P., and Valdez, E. A. (2009). Actuarial Applications of a Hierarchical Insurance Claims Model. *Astin Bulletin*, 39:165–197.
- Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*.
- Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457–1469.

- Frees, E. W. and Wang, P. (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics*.
- Kaishev, V. K. and Dimitrova, D. S. (2006). Excess of loss reinsurance under joint survival optimality. *Insurance: Mathematics and Economics*.
- Krämer, N., Brechmann, E. C., Silvestrini, D., and Czado, C. (2013). Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics*, 53(3):829 – 839.
- Merz, M. and Wüthrich, M. V. (2008). Prediction error of the multivariate chain ladder reserving method. *North American Actuarial Journal*.
- Nelsen, R. B. (2006). *An Introduction to Copulas (Springer Series in Statistics)*. Springer, New York, second édition.
- Patton, A. (2001). Modelling time-varying exchange rate dependence using the conditional copula. Technical report, Department of Economics, UC San Diego.
- Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.
- Pešta, M. and Okhrin, O. (2014). Conditional least squares and copulae in claims reserving for a single line of business. *Insurance: Mathematics and Economics*, 56(0):28 – 37.
- Shi, P. and Frees, E. W. (2010). Long-tail longitudinal modeling of insurance company expenses. *Insurance: Mathematics and Economics*, 47(3):303 – 314.
- Shi, P. and Frees, E. W. (2011). Dependent loss reserving using copulas. *Astin Bulletin - Journal of the International Actuarial Association*.
- Shi, P. and Valdez, E. A. (2011). A copula approach to test asymmetric information with applications to predictive modeling. *Insurance: Mathematics and Economics*, 49(2):226 – 239.
- Shi, P. and Valdez, E. A. (2012). Longitudinal modeling of insurance claim counts using jitters. *Scandinavian Actuarial Journal*, 2012(2):159–179.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):pp. 307–333.
- Zhang, Y. and Dukic, V. (2013). Predicting multivariate insurance loss payments under the bayesian copula framework. *Journal of Risk and Insurance*.
- Zhao, X. and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*.

Conclusion et perspectives

Dans ce travail de thèse, nous avons apporté des contributions supplémentaires aux thèmes de la modélisation et de l'estimation des risques en dimension plus grande que un. Tout au long de cette thèse, nous avons porté une attention particulière aux problématiques liées à la structure de dépendance entre les risques.

Notre objectif dans cette thèse était de contribuer à la littérature existante sur la modélisation des risques en assurance non vie. Afin d'atteindre cet objectif, nous avons modélisé différentes structures de dépendance entre les risques en assurance, dans un cadre dynamique, en mettant en évidence les effets de cette dépendance sur l'évaluation des réserves et du capital de solvabilité.

Le chapitre 2 tente de répondre à la question que se pose un actuaire qui dispose des sinistres corrélés au sein d'une même branche d'assurance. En effet, nous avons proposé un modèle dynamique appelé "*Autoregressive Conditional Amount model (ACA)*", qui permet d'analyser la dépendance temporelle entre les sinistres. Nous avons montré l'importance de considérer l'autocorrélation, aussi bien entre les montants faibles qu'entre les montants élevés au niveau des queues de distribution. C'est d'ailleurs, le fait de considérer la dépendance dynamique entre les montants des sinistres qui fait l'originalité de notre travail par rapport aux études précédentes ayant analysées la dépendance au sein d'une même branche d'assurance.

Nous avons proposé également une mesure de risque adaptée à cette problématique de dépendance temporelle, par la mise en place d'une VaR dynamique appelée VaR ACA adéquate aux caractéristiques des risques souscrits.

Le chapitre 3 étudie l'effet de considérer la dépendance dynamique au sein d'une même branche d'assurance sur l'évaluation des réserves et du capital de solvabilité.

Nous avons proposé un modèle dynamique d'évaluation des réserves "*Generalised Autoregressive Sinistrality Model (GACSM)*" et qui permet de prendre en compte la dépendance temporelle au sein d'un triangle de liquidation. Une comparaison des résultats empiriques est effectuée avec les modèles classiques Chain Ladder et GLM, montre que le modèle proposé permet d'évaluer des montants de réserves et du capital de solvabilité plus élevés que les modèles classiques. Par conséquent, les modèles de la littérature peuvent sous estimer les risques souscrits.

De plus, notre modèle fournit un risque d'erreur global sur l'évaluation des réserves faible, ce qui émane de la pertinence du modèle proposé.

Le chapitre 4 étudie une autre forme de dépendance, qui est la dépendance entre deux branches d'assurance. En effet, l'agrégation des risques permet d'avoir une vision globale sur le risque. Particulièrement, on s'intéresse à évaluer le montant agrégé des sinistres, en considérant une forme de dépendance dynamique qui change au fil du temps.

Nous avons pu répondre à cette problématique par la mise en place d'un modèle qui permet de prendre en compte conjointement la dépendance au sein de la même branche ainsi qu'entre les deux branches. Nous avons prouvé que ces modèles permettent de suivre l'évolution de la dépendance à travers le temps.

Les résultats empiriques ont montré que le modèle proposé permet à un actuare de gagner un effet de diversification sur le montant des réserves agrégées.

Les résultats obtenus dans notre travail suggèrent d'introduire la dépendance dynamique dans l'analyse des risques en assurance et l'évaluation des réserves et du capital. Ainsi le projet solvabilité 2 qui incite les sociétés d'assurance à améliorer leurs modèles internes peut trouver un fondement dans notre analyse des risques.

Des extensions peuvent être envisagées pour approfondir l'analyse des risques. En effet, nous pouvons remarquer dans le chapitre 3 que les erreurs quadratiques moyennes sont calculées par l'expression $E(\hat{y}_{i,t} - y_{i,t})$. Cependant, Hudecová and Pešta (2013) ont montré que cette expression est biaisée, et qu'elle donne une erreur moyenne biaisée si les données sont dépendantes. Une première piste d'extension intéressante serait de procéder à développer une expression adéquate qui permet de tenir compte de la dépendance temporelle entre les risques.

Une seconde piste d'extension, qui pourrait étendre notre travail, est de généraliser notre étude au cas multivarié. En effet, il serait intéressant de considérer plusieurs branches d'assurance et voir leurs effet de diversification.

En ce qui concerne le capital de solvabilité, une perspective de recherche serait envisageable. Il serait intéressant de développer une méthode d'allocation optimale du SCR en tenant compte d'une certaine forme de dépendance temporelle entre les branches d'assurance.

Bibliographie

- Anatolyev, S. (2009). Dynamic modeling under linear-exponential loss. *Economic Modelling*, 26(1) :82 – 89.
- Antonin, G. and Benjamin, S. (2001). Effets de la dépendance entre différentes branches sur le calcul des provisions. *Astin Bulletin*.
- Antonio, K. and Beirlant, J. (2007). Actuarial statistics with generalized linear mixed models. *Insurance : Mathematics and Economics*.
- Araichi, S. and Belkacem, L. (2014). Solvency capital for insurance company : Modelling dependence using copula. *Journal of Applied Economic Sciences*, IX Issue 4(30) Winter 2014.
- Araichi, S., de Peretti, C., and Belkacem, L. (2013). Modelling temporal dependence of claims in insurance using autoregressive conditional amount models. *ISFA working paper No 2013.9*, <http://docs.isfa.fr/labo/2013.9.pdf>.
- Araichi, S., de Peretti, C., and Belkacem, L. (2015). Generalized autoregressive conditional sinistrality model : a novel model for claims reserving in non life insurance. *ISFA working paper No 2015.4*. <http://docs.isfa.fr/labo/2015.4.pdf>.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Math. Finance*, 9(3) :203–228.
- Bargés, M., Cossette, H., and Marceau, E. (2009). Tvar-based capital allocation with copulas. *Insurance : Mathematics and Economics*.
- Bauwens, L. and Giot, P. (2000). The logarithmic acd model : An application to the bid-ask quote process of three nyse stocks. *Annals of Economics and Statistics*, pages 117–149.
- Belguise, O. and Levi, C. (2002). Tempêtes : Etude des dépendances entre les branches automobile et incendie à l’aide de la théorie des copulas. *Bulletin Français d’Actuariat*.
- Björkwall, S., Hössjer, O., and Ohlsson, E. (2009). Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving. *Scandinavian Actuarial Journal*, (4) :306–331.

- Bolancé, C., Guillén, M., and Pinquet, J. (2003). Time-varying credibility for frequency risk models : estimation and tests for autoregressive specifications on the random effects. *Insurance : Mathematics and Economics*, 33(2) :273 – 282. Papers presented at the 6th {IME} Conference, Lisbon, 15-17 July 2002.
- Boucher, J.-P., Denuit, M., and Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, pages 135–162.
- Bühlmann, H. (1996). *Mathematical methods in risk theory*, volume 172 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin. Reprint of the 1970 original.
- Cadoux, D. and Loizeau, J. M. (2004). Copules et dépendances : application pratique à la détermination du besoin en fonds propres d’un assureur non vie. *Astin Bulletin*.
- Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, pages 841–862.
- Commission, E. (2003). Conception d’un futur système de contrôle prudentiel applicable dans l’union européenne recommandation des services de la commission. markt/2509/03. Technical report, Institutions Financières Assurance.
- Commission, E. (2010). Qis5 technical specifications. Technical report, Financial Institution Insurance and Pensions.
- Cooray, K. and Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scand. Actuar. J.*, (5) :321–334.
- Cummins, J. (1985). Forecasting automobile insurance paid claim costs using econometric and ARIMA models. *International Journal of Forecasting North-Holland.*, pages 203–215.
- Cummins, J., Dionne, G., McDonald, J. B., and Pritchett, B. (1990). Applications of the GB2 family of distributions in modeling insurance loss processes. *Insurance : Mathematics and Economics*, 9(4) :257 – 272.
- De Peretti, C. (2003). Bilateral bootstrap tests for long memory : An application to the silver market. *Computational Economics*, 22(2-3) :187–212.
- Denuit, M. and Charpentier, A. (2004). *Mathématiques de l’assurance non-vie. Tome 1 : Principes fondamentaux de théorie du risque*. Economica.
- Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial Theory for Dependent Risks*. Wiley.
- Diers, D., Eling, M., and Marek, S. D. (2012). Dependence modeling in non-life insurance using the bernstein copula. *Insurance : Mathematics and Economics*.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

-
- El-Bassiouni, M. and El-Habashi, M. (1991). Forecasting compulsory motor insurance claims in kuwait. *Insurance : Mathematics and Economics*, 10(2) :85 – 92.
- Embrechts, P., McNeil, A., and Straumann, D. (1999a). Correlation : Pitfalls and alternatives. *A short, non-technical article, RISK Magazine*.
- Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999b). Extreme value theory as a risk management tool. *North American Actuarial Journal*.
- England, P. and Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance : Mathematics and Economics*.
- England, P. and Verrall, R. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*.
- Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1) :1–22.
- Engle, R. F. (2002). The econometrics of ultra-high-frequency data. *Journal of Applied Econometrics*, 17(5) :425–446.
- Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration a new model for irregularly spaced transaction data. *Econometrica*, 66(5) :1127–1162.
- Faivre, F. (2002). Copula : A new vision for economic capital and application to a four line of business company. *Astin Bulletin*.
- Frees, E. W., Shi, P., and Valdez, E. A. (2009). Actuarial Applications of a Hierarchical Insurance Claims Model. *Astin Bulletin*, 39 :165–197.
- Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*.
- Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484) :1457–1469.
- Frees, E. W. and Wang, P. (2006). Copula credibility for aggregate loss models. *Insurance : Mathematics and Economics*.
- Frees, E. W., Young, V. R., and Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance : Mathematics and Economics*, 24(3) :229 – 247.
- Gerber, H. U. (1982). Ruin theory in the linear model. *Insurance Math. Econom.*, 1(3) :177–184.
- Ghysels, E., Gouriéroux, C., and Jasiak, J. (2004). Stochastic volatility duration models. *Journal of Econometrics*, 119(2) :413 – 433.
- Grammig, J. and Maurer, K.-O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal*, 3(1) :16–38.

- Hudecová, Š. and Pešta, M. (2013). Modeling dependencies in claims reserving with {GEE}. *Insurance : Mathematics and Economics*, 53.
- Jaworski, P., Durante, F., Härdle, W., and Rychlik, T. (2010). *Copula Theory and Its Applications*. Lecture Notes in Statistics-Proceedings. Springer.
- Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73. Chapman & Hall.
- Kaishev, V. K. and Dimitrova, D. S. (2006). Excess of loss reinsurance under joint survival optimality. *Insurance : Mathematics and Economics*.
- Krämer, N., Brechmann, E. C., Silvestrini, D., and Czado, C. (2013). Total loss estimation using copula-based regression models. *Insurance : Mathematics and Economics*, 53(3) :829 – 839.
- Kremer, E. (1982). IBNR-claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal*.
- Lowe, J. (1994). A practical guide to measuring reserve variability using : Bootstrapping, operational time and a distribution-free approach. Technical report, Institute of Actuaries and Faculty of Actuaries.
- Lunde, A. (1999). A generalized gamma autoregressive conditional duration model. *discussion paper. Aalborg University*.
- Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserves estimates. *Astin Bulletin*.
- Mack, T. (1994). Which stochastic model is underlying the chain ladder method? *Insurance : Mathematics and Economics*.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC ; 2nd edition.
- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management : concepts, techniques and tools*. Princeton series in finance. Princeton University Press, Princeton (N.J.).
- Merz, M. and Wüthrich, M. V. (2008). Prediction error of the multivariate chain ladder reserving method. *North American Actuarial Journal*.
- Mikosch, T. (2006). *Non-life Insurance Mathematics*. Springer.
- Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society*.
- Nelsen, R. B. (2006). *An Introduction to Copulas (Springer Series in Statistics)*. Springer, New York, second édition.

-
- Patton, A. (2001). Modelling time-varying exchange rate dependence using the conditional copula. Technical report, Department of Economics, UC San Diego.
- Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2) :527–556.
- Pešta, M. and Okhrin, O. (2014). Conditional least squares and copulae in claims reserving for a single line of business. *Insurance : Mathematics and Economics*, 56(0) :28 – 37.
- Pinheiro, P. J. R., e Silva, J. M. A., and Centeno, M. d. L. (2003). Bootstrap methodology in claim reserving. *The Journal of Risk and Insurance*.
- Promislow, S. D. (1991). An axiomatic characterization of some measures of unfairness. *Journal of Economic Theory*, 53(2) :345 – 368.
- Richard, J. V. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Mathematics and Economics*.
- Rolski, T., Schmidli, H., Schmidt, V., and Teugels, J. (1999). *Stochastic processes for insurance and finance*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.
- Shi, P. and Frees, E. W. (2010). Long-tail longitudinal modeling of insurance company expenses. *Insurance : Mathematics and Economics*, 47(3) :303 – 314.
- Shi, P. and Frees, E. W. (2011). Dependent loss reserving using copulas. *Astin Bulletin - Journal of the International Actuarial Association*.
- Shi, P. and Valdez, E. A. (2011). A copula approach to test asymmetric information with applications to predictive modeling. *Insurance : Mathematics and Economics*, 49(2) :226 – 239.
- Shi, P. and Valdez, E. A. (2012). Longitudinal modeling of insurance claim counts using jitters. *Scandinavian Actuarial Journal*, 2012(2) :159–179.
- Taylor, G. (2000). *Loss reserving : an actuarial perspective*. Kluwer Academic Boston ; London.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2) :pp. 307–333.
- Wüthrich, M. V. and Merz, M. (2008). *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons, Ltd, West Sussex, England.
- Zhang, L. (2005). Ruin probability in linear time series model. *Tsinghua Science and Technology*, 10(2) :259 – 264.
- Zhang, M. Y., Russell, J. R., and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104(1) :179 – 207.

Bibliographie

Zhang, Y. and Dukic, V. (2013). Predicting multivariate insurance loss payments under the bayesian copula framework. *Journal of Risk and Insurance*.

Zhao, X. and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance : Mathematics and Economics*.