Etude des marchés d'assurance non-vie à l'aide d'équilibre de Nash et de modèle de risques avec dépendance

Christophe Dutang

I.S.F.A. **École Doctorale Sciences Économiques et de Gestion**

Etude des marchés d'assurance non-vie à l'aide d'équilibres de Nash et de modèles de risques avec dépendance

THÈSE

Numéro d'ordre 70-2012

présentée et soutenue publiquement le 31 Mai 2012

pour l'obtention du

Doctorat de l'Université Claude Bernard Lyon I

(mathématiques appliquées)

par

Christophe Dutang

Composition du jury

Président :	Jean-Noël Bacro, Professeur à l'Université Montpellier 2
Rapporteurs :	Etienne Marceau, Professeur à l'Université Laval de Québec Patrice Bertail, Professeur à l'Université Paris X
Examinateurs :	Stéphane Loisel, Professeur à l'Université Lyon 1 (co-directeur de thèse) Véronique Maume-Deschamps, Professeur à l'Université Lyon 1 (directrice de thèse) Christian-Yann Robert, Professeur à l'Université Lyon 1

Laboratoire Science Actuarielle Financière — EA 2429

Remerciements

Je voudrais commencer par remercier l'initiateur de ce projet de thèse, Stéphane Loisel, sans qui rien n'aurait été possible. Un jour de décembre 2007, Stéphane m'a fait une proposition de thèse CIFRE, me permettant de garder un pied en entreprise et l'autre en université, que je ne pouvais tout simplement pas refuser. Je le remercie tout particulièrement pour les travaux que nous avons effectués ensemble et pour nos discussions sans fin sur l'actuariat non-vie. Je tiens aussi à remercier mon autre directrice de thèse, Véronique Maume-Deschamps, toujours disponible pour répondre à mes questionnements. J'espère que ma collaboration avec mes directeurs de thèse ne s'arrêtera pas avec ma soutenance.

Je suis très honoré qu'Etienne Marceau et Patrice Bertail aient accepté d'être les rapporteurs de ma thèse. Merci à eux pour le temps consacré à parcourir ce manuscrit. Je tiens à remercier également Christian-Yann Robert et Jean-Noël Bacro d'avoir accepté de faire partie de mon jury.

Un grand merci à Hansjoerg Albrecher pour m'avoir aidé quand j'en avais besoin. Il ne m'a pas abandonné sur le projet de théorie des jeux et cycles de marché, malgré mon long apprentissage du sujet. De plus, je remercie Claude Lefèvre pour avoir lancé l'idée du sujet de théorie de la ruine. Nous avons eu d'agréables moments de discussion ensemble, tant sur ce projet que sur d'autres sujets.

Je tiens aussi à remercier mes chefs successifs au GRM d'AXA : Emmanuel Pierron et ses citations mythiques, Valérie Gilles et son sang-froid à toute épreuve, et Gilles Hornecker et son sens de l'humour. Ils m'ont chacun aidé, à leur manière, à la réalisation de cette thèse. J'en profite aussi pour remercier les membres du GRM non-vie PAP Valeur que j'ai pu côtoyer : les deux Mathieu, Jean, Francis, Nicolas, Antoine, Maria, Jérôme, Amélie, Yannick, Cyril, Pierre, Sara, Valérie et Jean-François. J'espère que la bonne ambiance de ce côté du plateau perdurera après mon départ. De plus, j'ai beaucoup apprécié les liens qui se sont créés avec les autres thésards CIFRE. Merci à Nabil, Théodora et Madeleine-Sophie pour nos longues discussions et nos repas trimestriels dont l'organisation nécessitait des doodles ! Un grand merci à toute l'équipe du GRM non-vie qui m'a toujours bien accueilli.

Je n'oublierai pas d'adresser un petit clin d'oeil aux autres thésards, à commencer par ceux de ma génération Xavier, Yahia, Elena, Manel, Abdou, Alain, Anisa, Soffana, à ceux qui m'ont précédé, Romain, Mathieu, Areski, Florent, et à ceux qui me suivent, les deux Julien, Andrès, Erwan, Jean-Charles et à ceux que j'aurais oubliés. Je remercie tout particulièrement l'équipe de recherche du laboratoire SAF au plot 4, ainsi que les membres du personnel administratif du plot 1, toujours prêts à faciliter mes recherches et mes démarches.

On dit souvent qu'une thèse est un long périple. Pour ma part, cela a pris la forme d'un long voyage de 154 671 kilomètres au prix de 9,94 centimes le kilomètre... Un grand merci donc à la SNCF pour maintenir l'illusion que Paris est proche de toutes les grandes villes de France. Par monts et par vaux, j'ai pu travailler sur ma thèse dans ses trains, sans subir perte de temps et fatigue sur les routes.

Un grand merci au créateur de TEX, Donald Knuth, pour avoir inventé ce fabuleux logiciel dans les années 70, qui fut plus tard repris par Leslie Lamport pour devenir LATEX. Je n'ose pas compter les heures de souffrance évitées grâce à LATEX comparé aux logiciels de type WYSIWYG. Sur le même thème, je remercie également Martin Maechler d'avoir encouragé Robert Gentleman et Ross Ihaka de rendre libre les sources de R en 1995. J'ai énormément utilisé ce logiciel gratuit et libre durant toute ma thèse.

De plus, je tiens à remercier tous mes amis, à commencer par ceux de Grenoble, Jérémy, Quentin, Benoit, Fabrice, Vincent, Olivier, Horté, Julie. Les moments de détente en leur compagnie ont été très bénéfiques pour ne jamais perdre le moral ! Je remercie aussi les $\rho\oplus$ de Paris, Onur, Fabrice, Jean-François, Tugdual, Romain, Marie, Julien, Samuel, Guillaume, avec qui j'ai passé des dimanches soirs tout simplement mémorables.

Merci à ma belle-famille pour avoir établi un pont aérien entre l'Ariège et Paris 13$^{\text{ème}}$ afin de m'envoyer à flot quasi-continu des spécialités vernaculaires ! Cela m'a permis de rester en forme durant toute cette thèse. Un grand merci aussi à Alain et Christine pour leur soutien et leur relecture.

J'ai une pensée particulière pour mon frère, qui m'a accueilli de nombreuses fois dans son appartement les veilles de séminaire Lyon-Lausanne. Je suis aussi pour toujours redevable à mes parents, qui m'ont inculqué les valeurs morales pour que je réussisse dans la vie. Merci à eux pour leurs relectures durant le dernier mois.

Le mot de la fin est adressé à ma tendre Isabelle, à qui je dédie cette thèse. Elle a toujours été là à mes côtés pour me soutenir et m'encourager durant ces trois longues années. Je ne pourrai jamais assez la remercier.

À Isabelle

Résumé

Etude des marchés d'assurance non-vie à l'aide d'équilibre de Nash et de modèle de risques avec dépendance

L'actuariat non-vie étudie les différents aspects quantitatifs de l'activité d'assurance. Cette thèse vise à expliquer sous différentes perspectives les interactions entre les différents agents économiques, l'assuré, l'assureur et le marché, sur un marché d'assurance. Le chapitre 1 souligne à quel point la prise en compte de la prime marché est importante dans la décision de l'assuré de renouveler ou non son contrat d'assurance avec son assureur actuel. La nécessité d'un modèle de marché est établie.

Le chapitre 2 répond à cette problématique en utilisant la théorie des jeux non-coopératifs pour modéliser la compétition. Dans la littérature actuelle, les modèles de compétition se réduisent toujours à une optimisation simpliste du volume de prime basée sur une vision d'un assureur contre le marché. Partant d'un modèle de marché à une période, un jeu d'assureurs est formulé, où l'existence et l'unicité de l'équilibre de Nash sont vérifiées. Les propriétés des primes d'équilibre sont étudiées pour mieux comprendre les facteurs clés d'une position dominante d'un assureur par rapport aux autres. Ensuite, l'intégration du jeu sur une période dans un cadre dynamique se fait par la répétition du jeu sur plusieurs périodes. Une approche par Monte-Carlo est utilisée pour évaluer la probabilité pour un assureur d'être ruiné, de rester leader, de disparaître du jeu par manque d'assurés en portefeuille. Ce chapitre vise à mieux comprendre la présence de cycles en assurance non-vie.

Le chapitre 3 présente en profondeur le calcul effectif d'équilibre de Nash pour n joueurs sous contraintes, appelé équilibre de Nash généralisé. Il propose un panorama des méthodes d'optimisation pour la résolution des n sous-problèmes d'optimisation. Cette résolution se fait à l'aide d'une équation semi-lisse basée sur la reformulation de Karush-Kuhn-Tucker du problème d'équilibre de Nash généralisé. Ces équations nécessitent l'utilisation du Jacobien généralisé pour les fonctions localement lipschitziennes intervenant dans le problème d'optimisation. Une étude de convergence et une comparaison des méthodes d'optimisation sont réalisées.

Enfin, le chapitre 4 aborde le calcul de la probabilité de ruine, un autre thème fondamental de l'assurance non-vie. Dans ce chapitre, un modèle de risque avec dépendance entre les montants ou les temps d'attente de sinistre est étudié. De nouvelles formules asymptotiques de la probabilité de ruine en temps infini sont obtenues dans un cadre large de modèle de risques avec dépendance entre sinistres. De plus, on obtient des formules explicites de la probabilité de ruine en temps discret. Dans ce modèle discret, l'analyse structure de dépendance permet de quantifier l'écart maximal sur les fonctions de répartition jointe des montants entre la version continue et la version discrète.

Mots-clés: Comportement client, cycles de marché, théorie de la ruine, actuariat non-vie, théorie des jeux, calcul d'équilibre de Nash généralisé, montants de sinistres dépendants

Abstract

Studying non-life insurance markets with Nash equilibria and dependent risk models

In non-life actuarial mathematics, different quantitative aspects of insurance activity are studied. This thesis aims at explaining interactions among economic agents, namely the insured, the insurer and the market, under different perspectives. Chapter 1 emphasizes how essential the market premium is in the customer decision to lapse or to renew with the same insurer. The relevance of a market model is established.

In chapter 2, we address this issue by using noncooperative game theory to model competition. In the current literature, most competition models are reduced to an optimisation of premium volume based on the simplistic picture of an insurer against the market. Starting with a one-period model, a game of insurers is formulated, where the existence and uniqueness of a Nash equilibrium are verified. The properties of premium equilibria are examined to better understand the key factors of leadership positions over other insurers. Then, the derivation of a dynamic framework from the one-period game is done by repeating of the one-shot game over several periods. A Monte-Carlo approach is used to assess the probability of being insolvent, staying a leader, or disappearing of the insurance game. This gives further insights on the presence of non-life insurance market cycles.

A survey of computational methods of a Nash equilibrium under constraints is conducted in Chapter 3. Such generalized Nash equilibrium of n players is carried out by solving a semismooth equation based on a Karush-Kuhn-Tucker reformulation of the generalized Nash equilibrium problem. Solving semismooth equations requires using the generalized Jacobian for locally Lipschitzian function. Convergence study and method comparison are carried out.

Finally, in Chapter 4, we focus on ruin probability computation, another fundemantal point of non-life insurance. In this chapter, a risk model with dependence among claim severity or claim waiting times is studied. Asymptotics of infinite-time ruin probabilities are obtained in a wide class of risk models with dependence among claims. Furthermore, we obtain new explicit formulas for ruin probability in discrete-time. In this discrete-time framework, dependence structure analysis allows us to quantify the maximal distance between joint distribution functions of claim severity between the continuous-time and the discrete-time versions.

Keywords: Customer behavior, market cycles, ruin theory, non-life insurance, game theory, generalized Nash equilibrium computation, dependent claim severity models

Table des matières

Théorie des jeux

Théorie de la ruine

Conclusion

Introduction générale

Introduction

Tout ce qui augmente la liberté
augmente la responsabilité.
Victor Hugo (1802-1885)

L'actuariat non-vie étudie les différents aspects mathématiques de l'activité d'assurance. L'objet de cette thèse est d'expliquer sous différentes perspectives les interactions entre les différents agents économiques, l'assuré, l'assureur et le marché, sur un marché d'assurance, tant au niveau de la modélisation des primes que des sinistres. Cette introduction vise à présenter et à remettre dans leur contexte quatre articles en cours de soumission, qui constituent les chapitres de cette thèse. Ceux-ci couvrent trois grands sujets de l'activité d'assurance : la modélisation des résiliations en prenant en compte le marché, la modélisation des primes dans un environnement de compétition et l'évolution de la richesse d'une compagnie d'assurance.

En souscrivant une police d'assurance, un individu souhaite se prémunir contre les conséquences d'évènements extérieurs (incendies, accidents, etc...) envers un de ses biens (voiture, logement, etc...) ou sa personne (responsabilité civile). En contrepartie de cette assurance, l'assuré paie une prime d'assurance en début de période. L'assureur quant à lui peut être amené à fournir une prestation si un certain type de sinistre survient pendant la période considérée. A ces deux agents économiques s'ajoutent une troisième composante, impersonnelle, le marché. Dans le cadre de cette thèse, nous excluons la réassurance de notre étude. Dans ce schéma à trois agents, l'assureur fait donc face en premier lieu au risque d'avoir peu ou pas d'assurés, dans le cas où ses prix sont excessifs ou simplement très supérieurs à ceux des autres assureurs. Le marché agit à la fois sur l'assuré en pouvant l'inciter à résilier son contrat d'assurance pour se couvrir chez un autre assureur, et sur l'assureur en le contraignant dans une certaine mesure à rendre ses primes d'assurance à des niveaux acceptables. Le risque de prime regroupe donc deux composantes : les résiliations et la compétition. Les modèles de résiliation reposent sur les modèles statistiques de régression dont le plus connu est le modèle

linéaire généralisé. Cette thèse propose une revue de tels modèles et étudie leur pertinence dans le schéma à trois agents.

La compétition sur les marchés d'assurance peut être modélisée de deux façons : une approche agrégée modélisant le marché dans sa globalité, une vision plus fine visant à modéliser chacun des assureurs composant le marché. La première approche consiste à étudier l'évolution de variables macroéconomiques telles que la prime moyenne marché, le ratio sinistre sur prime du marché. La seconde approche repose sur l'utilisation de la théorie des jeux pour modéliser les interactions entre assureurs. En théorie des jeux, le concept de solution est soustrait au concept d'équilibre. Un des apports de cette thèse est de proposer un modèle de théorie des jeux pour comprendre les interactions entre assureurs et assurés sur un marché d'assurance. En plus de l'application d'un jeu pour le marché d'assurance, nous proposons un panorama des méthodes de calcul d'équilibre de la théorie des jeux.

Par ailleurs, les assureurs sont confrontés au risque propre d'assurance une fois des contrats souscrits, en plus du risque de prime. En effet, l'assureur est tenu de respecter ses engagements envers les assurés. Pour ce faire, il détient un capital initial auquel se rajoutent les primes et se retranchent les montants des sinistres au cours du temps. La théorie de la ruine s'intéresse à l'évolution du niveau de richesse d'une compagnie d'assurances à l'aide de processus stochastiques. Parmi les mesures généralement considérées, nous nous intéressons à la probabilité de ruine en temps infini, qui dépend du capital initial de l'assureur. Un dernier apport de la thèse est de proposer de nouvelles formules asymptotiques de probabilités de ruine, dans un cadre de dépendance entre les sinistres, lorsque le capital initial est élevé. De plus, nous obtenons des formules explicites pour la probabilité de ruine en temps discret.

La suite de cette introduction va développer les points évoqués ci-dessus, en commençant par décrire les modèles de résiliation et de cycles de marché, puis en présentant les modèles de théories des jeux usuels et enfin les modèles de théorie de la ruine.

Modèles de résiliation et cycles de marché

En assurance non-vie, l'individu se prémunit contre les conséquences financières d'un risque envers un de ses biens ou sa personne en achetant une police d'assurance. En contrepartie de cette assurance, l'assuré paie une prime d'assurance en début de période de couverture, par exemple un an en assurance automobile. En France et dans de nombreux pays d'Europe continentale, les contrats d'assurance sont renouvelés par tacite reconduction, c'est à dire si l'assuré ne manifeste pas son intention de mettre un terme à son contrat d'assurance, celui-ci est automatiquement renouvelé pour la même durée de couverture.

D'une part, la compétition sur les marchés d'assurance empêche de demander des prix très élevés pour une couverture d'assurance. En effet, la prime d'assurance est calculée en fonction des caractéristiques propres de l'assuré, mais dépend tout autant des conditions de marché. Il n'y a rien de plus naturel que l'assuré soit tenté de résilier son contrat d'assurance s'il peut trouver moins cher chez un autre assureur. A couverture équivalente, il semble donc raisonnable de supposer que les individus vont chercher à s'assurer à moindre prix. Notons un premier biais de perception de la part de l'assuré, les couvertures proposées par les assureurs ne sont pas forcément équivalentes, franchise et limite pouvant être différentes, sans parler des couvertures additionnelles comme l'assistance, la perte de valeur, la garantie du contenu, etc... Constatons donc que le prix de l'assurance n'explique pas entièrement le comportement d'un client.

D'autre part, les assureurs, les mutuelles et autres structures d'assurances cherchent dans une certaine mesure à maximiser leur volume de primes, défini comme la somme des primes émises sur l'ensemble des contrats pendant une année. Au premier ordre, le volume de prime peut être approché par le produit du nombre de contrats et de la prime moyenne. Il paraît logique que ces deux grandeurs évoluent en sens inverse : une prime moyenne élevée entraînera une diminution du nombre de contrats souscrits, et vice versa.

Une autre caractéristique de l'assurance non-vie est que certaines assurances sont obligatoires, par exemple, la garantie responsabilité civile en assurance automobile, qui permet l'indemnisation des dommages causés par un automobiliste aux tiers. Ainsi, certains produits d'assurance non-vie sont plus concurrentiels que d'autres, du fait de leur caractère obligatoire. La résiliation d'un contrat d'assurance résulte donc de plusieurs décisions dépendant de différents agents : l'assuré, son assureur actuel et les concurrents.

A première vue, la survie d'un assureur est menacée par deux phénomènes : une sinistralité accrue qui pourrait le rendre insolvable ou même le ruiner, et une baisse d'attractivité auprès des individus cherchant à s'assurer. Etant nécessaire à toutes études prospectives sur le volume de prime, le taux de résiliation est donc une variable clé pour les assureurs. Le taux de résiliation peut se définir soit en nombre de contrats, soit en volume de prime, ou encore en nombre de risques. Cette thèse ciblant l'assurance de particuliers, nous choisirons comme taux de résilitation, le nombre de contrats résiliés (sur une période) divisé par le nombre total de contrats en début de période.

Ce constat étant établi, il apparait tout à fait pertinent de chercher à modéliser la résiliation des assurés et les cycles de marché. Les deux modélisations sont généralement réalisées indépendamment l'une de l'autre, comme nous allons le présenter dans cette partie de l'introduction. Le chapitre 2 de cette thèse vise à modéliser conjointement le comportement des assurés et des assureurs à l'aide de la théorie des jeux.

Causalités de la résiliation et comportement des clients

Dans cette thèse, nous nous concentrons sur la résiliation des contrats d'assurance par l'assuré au moment du renouvellement du contrat. Cela exclut donc les résiliations avant terme, par exemple, la disparition du risque suite à la revente du véhicule en assurance auto. Nous écartons aussi les résiliations du contrat par l'assureur.

Tout d'abord, les études purement descriptives, voir Bland *et al.* (1997); Kelsey *et al.* (1998), révèlent le premier constat : pour une génération donnée de polices, le taux de résiliation décroît au cours du temps. Autrement dit, plus l'assuré reste longtemps en portefeuille, plus sa probabilité de résiliation diminue.

Une autre variable clé de la résiliation est sans grande surprise la prime proposée. Ainsi la politique de prix de l'assureur, c'est à dire la combinaison d'un tarif technique, d'un éventuel rabais commercial et de négotiation, va faire évoluer la prime d'assurance. Par conséquent, la résiliation est très fortement impactée par la politique tarifaire de l'assureur.

On peut légitemement supposer qu'un assuré résilie s'il a trouvé mieux ailleurs, soit en terme de prix soit en terme de couverture. La comparaison avec la concurrence est indissociable du processus de résiliations. Par conséquent, malgré internet, le géomarketing et l'information visible sur la concurrence ne sont pas à négliger : par exemple, un assureur spécialisé pour les zones rurales est moins inquiété par un assureur généraliste des zones urbaines qu'un assureur spécialisé ciblant aussi cette portion du marché. Ainsi si les garanties sont adaptées à sa clientèle, l'assureur spécialisé n'a pas besoin d'être le moins cher pour garder ses clients en

portefeuille.

A cela se rajoutent l'inertie des habitudes, la subjectivité de la couverture d'assurance et l'image de marque de l'assureur. Lorsque les assurés comparent les primes d'assurance, la comparaison est souvent biaisée du fait de la segmentation et de la valeur du risque assuré. Cependant, l'arrivée des comparateurs de prix sur internet ajoute un peu de transparence sur les prix proposés par les assureurs.

A garantie et profil de risque comparables, la résiliation dépend majoritairement de l'élasticité de l'assuré au prix. Cette sensibilité au prix dépend du caractère psychologique des prix en premier lieu. Sur cet aspect, la loi psychophysique de Webner-Fechner précise que la sensation varie comme le logarithme de l'excitation. Sa transcription en termes d'élasticité prix nous laisse penser qu'une hausse successive des prix entraîne moins de résiliations qu'une hausse brutale et unique des prix. Vice versa, une baisse successive de prix devrait favoriser le renouvellement des contrats par rapport à une baisse unique des prix.

Modèle logistique

Au delà de ces considérations économiques et de marketing, il convient de vérifier a posteriori les causalités envisagées pour la résiliation et l'élasticité prix des clients. La résiliation s'exprime à l'aide d'une variable aléatoire de Bernoulli Y_i valant 1 si l'assuré i résilie son contrat, et 0 s'il le renouvelle. Notons $\boldsymbol{X_i}$ le vecteur des variables explicatives de l'assuré i. Un des modèles les plus simples est le modèle logistique qui repose sur l'équation suivante :

$$P(Y_i = 1) = \frac{1}{1 + e^{-\boldsymbol{\beta}^T \boldsymbol{X_i}}}, \tag{1}$$

où $\boldsymbol{\beta}$ est le vecteur de paramètres et M^T est la transposée de la matrice M. Il est facile de vérifier que le membre de droite, la fonction logistique, reste toujours compris dans l'intervalle [0,1].

La loi de probabilité étant spécifiée, nous pouvons donc estimer le paramètre $\boldsymbol{\beta}$ par maximum de vraisemblance. Le modèle logistique fait partie de la grande classe des modèles linéaires généralisés introduits par Nelder et Wedderburn (1972), pour lesquels de nombreuses propriétés ont été démontrées. Le modèle logistique est donc un choix de modèles très naturel. Le chapitre 1 analyse les modèles de régression pour expliquer la résiliation en assurance non-vie, notamment les modèles linéaires généralisés et une de leurs extensions les modèles additifs généralisés. Ce premier chapitre souligne tout particulièrement les variables explicatives indispensables pour obtenir des résultats cohérents en terme de prédiction des taux de résiliation et met en exergue le role clé de la prime moyenne marché. Le chapitre conclut sur la pertinence de modéliser le marché.

Modèles de choix

La modélisation du marché peut se faire de deux façons différentes : (i) soit nous modélisons le marché comme un seul compétiteur, (ii) soit nous modélisons l'ensemble des assureurs constituant le marché. Dans le premier cas, le modèle logistique suffit (i.e variable de décision $Y_i \in \{0, 1\}$) puisque si l'assuré résilie, c'est pour aller vers le marché. Il suffit donc de modéliser le marché à l'aide de séries chronologiques (cf. la sous-section suivante), les résiliations à l'aide d'un modèle logistique et d'utiliser la théorie du contrôle optimal pour déterminer une prime répondant à certains critères. Dans le second cas, chaque assureur présent sur le marché va être

modélisé, ainsi la variable de décision des clients n'aura plus deux modalités. Ainsi, il nous faut un modèle de choix (entre chaque assureur) pour modéliser une variable $Y_i \in \{0, \ldots, c-1\}$.

L'étude des modèles de choix est un thème bien connu en économétrie. Le livre de Manski et McFadden (1981) constitue un ouvrage de référence dans ce domaine. McFadden (1981) * présente en profondeur les modèles de choix dans le chapitre 5 de ce livre. C'est une extension probabiliste du modèle de l'*homo economicus* de l'économie classique.

Les modèles de choix reposent sur deux composantes : (i) un système de probabilité de choix et (ii) un cadre de maximisation d'utilité aléatoire. Soit P la probabilité de choix pour un individu. P est une fonction de l'ensemble $I \times \mathcal{B} \times S$ dans l'intervalle [0,1], où I désigne l'ensemble des alternatives, $\mathcal{B} \subset I$ l'ensemble des choix possibles offerts à l'individu et s une caractéristique mesurée de l'individu.

$P(i|\ B, s)$ désigne la probabilité de choisir l'alternative i parmi la sélection B pour un individu de caractéristique s. De plus, nous rajoutons une fonction d'attribut observé $\xi : I \mapsto Z$, telle que $\xi(i)$ représente les attributs observés. Le système de probabilité de choix est donc le vecteur $(I, Z, \xi, \mathcal{B}, S, P)$. Sur ce système, deux hypothèses sont faites : (i) la sommation $\forall B \in \mathcal{B}, P(B|B, s) = 1$, (ii) la caractérisation totale $\forall B = \{i_1, \ldots, i_n\}, \forall \tilde{B} = \{\tilde{i}_1, \ldots, \tilde{i}_n\}$, $\xi(i_k) = \xi(\tilde{i}_k)$ entraîne $P(i_k|B, s) = P(\tilde{i}_k|B, s)$.

Outre le système de probabilité de choix, McFadden (1981) se place dans un cadre de maximisation d'utilité aléatoire. En effet, comme l'utilité d'un individu n'est pas quelque chose de très facilement mesurable, il est cohérent de considérer son caractère aléatoire d'un point de vue de l'observateur.

La deuxième composante des modèles de choix, une hypothèse de maximisation d'utilité aléatoire, est définie par un vecteur (I, Z, ξ, S, μ) où (I, Z, ξ, S) vient du système de probabilité et μ est la mesure de probabilité sur l'espace des fonctions d'utilités définies sur I, dépendant $s \in S$. $\mu(., s)$ représente donc la loi de probabilité des "goûts" pour la population de caractéristique s. Ainsi, la probabilité $P(i_k|B, s)$ de choisir l'alternative i_k s'écrit

$$\mu(\{U \in \mathbb{R}^I | \forall j = 1, \ldots, n, U(i_k) \geq U(i_j)\}, s).$$

Des hypothèses additionnelles complètent la composante (I, Z, ξ, S, μ) pour que l'équation précédente soit toujours définie.

Deux modèles paramétriques sont très utilisés dans ce cadre d'étude : les modèles de Luce (1959) et de Thurstone (1927). Luce (1959) considère la forme paramétrique suivante pour la probabilité de choisir l'alternative i parmi B

$$P(i|\boldsymbol{z}_B, \boldsymbol{\beta}) = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{z}_i}}{\sum_{j \in B} e^{\boldsymbol{\beta}^T \boldsymbol{z}_j}},$$

où $\boldsymbol{z}_B = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m)$ correspond au vecteur d'attributs observés pour les alternatives dans B et $\boldsymbol{\beta}$ un vecteur de paramètres. Cette forme paramétrique présuppose l'indépendance des alternatives non pertinentes, c'est à dire, pour tout $i \in A \subset B$, $P(i|\boldsymbol{z}_B, \boldsymbol{\beta}) = P(i|\boldsymbol{z}_A, \boldsymbol{\beta}) P(A|\boldsymbol{z}_B, \boldsymbol{\beta})$. Très souvent, une catégorie i_0 de référence est considérée pour laquelle $\boldsymbol{z}_{i_0} = 0$ entrainant l'apparition de 1 dans la fraction ci-dessus. Le modèle logistique est un cas particulier de ce modèle avec deux alternatives.

∗. Daniel L. McFadden a reçu le prix Nobel d'économie pour ces travaux sur les choix discrets le 10 décembre 2000.

Thurstone (1927) considère la forme suivante

$$P(i|\boldsymbol{z}_B, \boldsymbol{\beta}) = \Phi_{0,\Omega}(-\boldsymbol{z}_{B-i}\boldsymbol{\beta}),$$

où $\boldsymbol{z}_{B-i} = (\boldsymbol{z}_1 - \boldsymbol{z}_i, \ldots, \boldsymbol{z}_{i-1} - \boldsymbol{z}_i, \boldsymbol{z}_{i+1} - \boldsymbol{z}_i, \ldots, \boldsymbol{z}_m - \boldsymbol{z}_i)$ et $\Phi_{0,\Omega}$ est la fonction de répartition de la loi normale multivariée de moyenne 0 et de matrice de variance-covariance $\Omega = \boldsymbol{z}_{B-i}AA^T\boldsymbol{z}_{B-i}^T$. Ici, la catégorie de référence est i. Ces deux paramétrisations sont généralement appelées modèles logit multinomial et probit multinomial pour leur lien avec les modèles linéaires généralisés multivariés.

Par la suite, le chapitre 2, centré sur un modèle de compétition entre assureurs, modélise le choix des assurés par une paramétrisation logit multinomiale. L'ensemble des alternatives est l'ensemble des I assureurs $B = \{1, \ldots, I\}$, tandis que les caractéristiques de l'individu se limiteront au numéro de son assureur actuel j et au vecteur des prix proposés par chaque assureur $\boldsymbol{x} = (x_1, \ldots, x_I)$. Ainsi, $P(i|(\boldsymbol{x}, j), \boldsymbol{\beta})$ représentera la probabilité de choisir l'assureur i sachant que l'assuré est chez l'assureur j, une gamme de prix $\boldsymbol{x}$ et un vecteur de paramètre $\boldsymbol{\beta}$. Nous renvoyons au chapitre 2 pour plus de détails.

Dynamique de prix et cycles de marché

La dynamique des prix a un grand impact sur le comportement des assurés, aussi bien en valeur absolue qu'en valeur relative par rapport au marché. Les modèles de cycles de marché sont donc très utilisés par les praticiens pour estimer le prix du marché de l'année prochaine. Puis, chaque assureur va essayer de se positionner par rapport au prix du marché attendu. En plus du prix de marché attendu, le ratio sinistre sur prime espéré doit être en adéquation avec les réserves courantes en capital. Il est temps maintenant d'expliquer les raisons de la présence de cycles sur les marché d'assurance non-vie. Ensuite, nous présenterons les modèles de séries temporelles les plus utilisés, avant de conclure sur les problèmes de cette approche.

Causes du cycle

L'étude des cycles de marché est un thème depuis longtemps débattu en économie d'assurance. Les articles fondamentaux remontent à Venezian (1985) et Cummins et Outreville (1987). Dans cette introduction, nous nous basons sur la revue bibliographique très exhaustive de Feldblum (2001).

Cet article traite de diverses théories classiques expliquant la causalité des cycles de marché en assurance non-vie. L'auteur crédibilise et décrédibilise une à une les principales théories. Au final, il conclut que la présence de cycles est dû l'effet conjugué de quatre causes : (i) la tarification actuarielle, (ii) la philosophie de la souscription, (iii) les fluctuations des taux d'intérêts et (iv) la stratégie compétitive. Nous précisons ici chacune des causes avant de présenter les modèles de séries temporelles utilisés pour montrer la pertinence d'une cause plutôt qu'une autre.

La tarification actuarielle peut engendrer des cycles résultant de l'effet combiné de l'incertitude sur les sinistres et de la contre-cyclicité des coûts réels. L'incertitude est inhérente à toute activité d'assurance et est due à l'inversion de cycle de production. Donc sans même évoquer les sinistres les plus volatiles commes les catastrophes naturelles, l'assureur ne connaîtra jamais avec certitude le montant de la charge sinistre à venir au moment où il tarifie ses contrats. La contre-cyclicité provient du décalage entre la survenance des sinistres et leur intégration dans les nouveaux tarifs. Ce retard d'information dû à des contraintes légales et techniques est généralement de 2 ou 3 ans.

L'antithèse de cette vision est que les cycles de marché ne sont généralement pas contracycliques aux conditions macroéconomique (prospérité vs. récession). Les actuaires sont capables d'apprendre de leurs erreurs passées et d'intégrer prospectivement les tendances actuelles dans leur tarif. Les cycles de souscription ne seraient pas dus à la tarification actuarielle mais à la résistance des souscripteurs à appliquer de nouveaux tarifs via les rabais commerciaux qu'ils peuvent accorder.

Une psychologie de masse des souscripteurs crée des effets d'emballement. En phase profitable, les assureurs sont optimistes et se font la guerre des prix en quête de parts de marché. A l'opposé, lorsque les résultats de souscription se dégradent, les assureurs sont contraints au bout d'un certain temps de remonter les prix de manière à retrouver un niveau de profitabilité acceptable. Le jeu consiste à détecter le plus tôt possible les changements de cycles pour en tirer profit, car malgré une psychologie de masse, les souscripteurs n'agissent pas de manière concertée.

Et, c'est là le problème de cette thèse : si la compétition est aussi intense que l'on sous-entend, l'offre et la demande ne devraient-elles pas converger vers un équilibre ? La coordination entre assureurs étant interdite, les assureurs sont incapables de déterminer le prix et la quantité d'assurance d'une demande globale en perpétuel déséquilibre. Les consensus sur les changements de phase restent toujours un mystère et pourtant les résultats de souscription d'un assureur à l'autre sont très corrélés.

Une caractéristique fondamentale du marché de l'assurance est le temps assez long s'écoulant entre l'encaissement des primes et le paiement effectif des sinistres (tout particulièrement sur les garanties de responsabilité civile). La fluctuation des intérêts n'est qu'une couche supplémentaire aux contraintes que subissent les assureurs. En période de taux d'intérêts forts, les souscripteurs peuvent se permettre de souscrire au coût attendu des sinistres et espérer les profits sur l'actualisation des sinistres. D'un autre côté, en période de taux bas, les assureurs sont forcés de remonter les prix.

Or, les cycles ne perdent rarement de leur intensité lorsque les taux d'intérêts fluctuent peu. De plus, les taux augmentent à la fois les primes et les coûts des sinistres, donc les bénéfices de souscription devraient être faiblement impactés par ses changements de taux. L'expertise financière des assureurs rend l'hypothèse de fluctuations des taux d'intérêts peu crédible, car elle suppose des méthodes de tarification naïves, des souscripteurs simplistes, des régulateurs rigides et des gestionnaires de fond décevants. En environnement compétitif, les joueurs irrationnels disparaissent.

Enfin, la théorie économique sur la compétition garantit que les firmes vendent au coût marginal. Cependant de nombreux effets contrecarrent cette belle théorie, les assureurs ne connaissent ni l'offre ni la demande en assurance et encore moins le coût d'un produit d'assurance (au moment de sa vente). A cela, se rajoutent les barrières à l'entrée non pas financières mais opérationnelles et techniques : coût de distribution, définition d'un standard de souscription, segmentation, etc... Comme le souligne Feldblum (2001), il est facile d'entrer sur un marché d'assurance, mais il est nettemment plus difficile d'y entrer avec succès.

Comme déjà énoncé, la faible différentiation des produits et la relative inélasticité des clients rend difficile l'arrivée de nouveaux entrants. Les caractéristiques intrinsèques de l'assurance rendent les stratégies compétitives à la fois indispensables et stabilisantes. De plus, sur les marchés matures, seules les fusions-acquisitions permettent de gagner de grosses parts de marché sans pratiquer une brutale chute des prix. Par conséquent, le nombre de risque par assureur reste en général stable en régime de croisière.

Modèles de séries temporelles

Nous présentons dans cette sous-section les modèles de séries temporelles les plus classiques utilisées dans la littérature des cycles de marché pour affirmer ou infirmer une conjecture. L'objet des séries temporelles est l'étude des processus stochastiques en temps discret $(X_t)_{t\in\mathbb{N}}$. Le modèle de base est le modèle autorégressif d'ordre p. Pour un processus faiblement stationnaire (X_t) (c'est à dire espérance constante et fonction d'autocovariance dépendant seulement de l'écart de temps), $(X_t)_t$ est dit autorégressif d'ordre p s'il existe des coefficient $a_1, \dots, a_p$ et un processus bruit blanc $(\mathcal{E}_t)$ tels que

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \mathcal{E}_t.$$

Pour le cas particulier $p = 1$, on retrouve une marche aléatoire, tandis que pour $p = 2$, (X_t) peut être périodique si $a_2 < 0$ et $a_1^2 + 4a_2 < 0$ avec une période

$$P = 2\pi \arccos\left(\frac{a_1}{2\sqrt{-a_2}}\right).$$

Malgré sa simplicité et le caractère très fort des hypothèses, les modèles autoregressifs ont été appliqués dans beaucoup d'articles de cycles de marché, par exemple Cummins et Outreville (1987). Cummins et Outreville (1987) cherchent à montrer l'hypothèse de tarification actuarielle avec X_t le résultat de souscription de l'année t. Les périodes de cycles oscillent entre 6 et 10 ans suivant les pays.

Une hypothèse centrale des modèles autorégressifs est la stationnarité du processus. Elle peut par exemple être testée par le test de racine unitaire de Dickey-Fuller (voir, par exemple, Gourieroux et Monfort (1997)). En pratique, l'hypothèse de stationnarité est rarement vérifiée. Pour compenser cela, Fields et Venezian (1989) et Gron (1994a) vont introduire des variables explicatives pour obtenir une régression temporelle du type

$$X_t = a_1 X_{t-1} + b_0 Y_t + c_0 Z_t + \mathcal{E}_t,$$

où Y_t, Z_t sont des indicateurs temporels indépendants et $a_0, b_0, c_0 \neq 0$. Une modélisation plus intéressante est proposée par Haley (1993) en utilisant les modèles de cointégration.

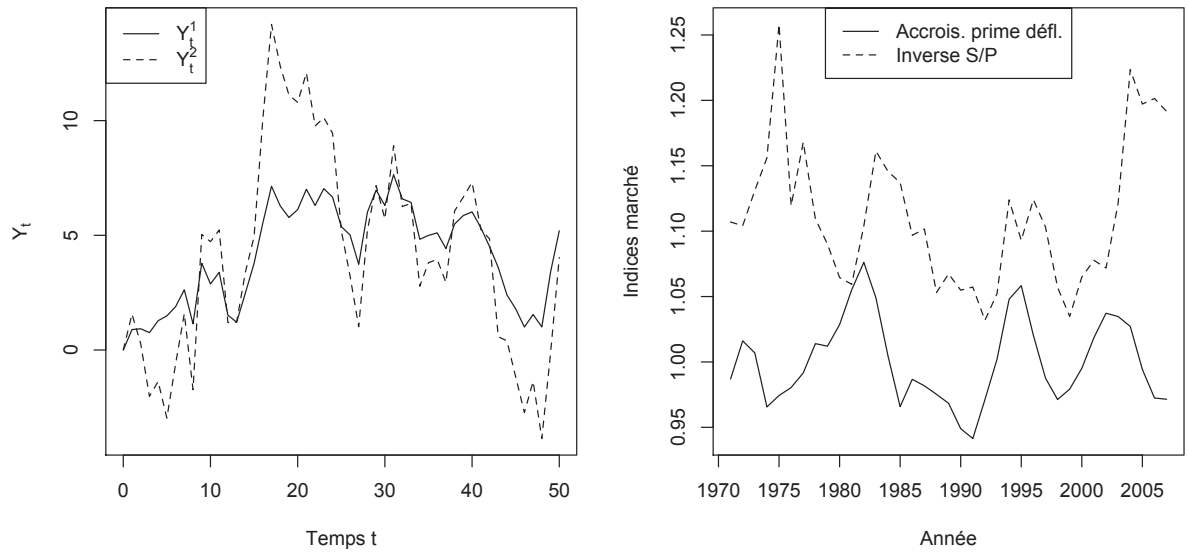
Les modèles cointégrés ont été proposés par Granger et Engle (1987) *, voir aussi Committee (2003). Deux processus $(X_t)_t$ et $(Y_t)_t$ sont cointégrés s'il existe une combinaison linéaire de ces deux processus qui est stationnaire, c'est à dire $(\alpha X_t + \beta Y_t)_t$ est stationnaire pour un couple (α, β) non nul. Cette notion de cointégration est cruciale car elle permet de modéliser des tendances long-terme entre les deux séries $(X_t)_t$ et $(Y_t)_t$.

Prenons l'exemple de Hamilton (1994) avec

$$X_t = X_{t-1} + \mathcal{E}_t \quad \text{et} \quad Y_t = 2X_t + \widetilde{\mathcal{E}}_t.$$

où $\mathcal{E}_t, \widetilde{\mathcal{E}}_t$ sont deux bruits blancs. Il est facile de voir que $(Y_t - 2X_t)$ est un processus stationnaire. Néanmoins, les trajectoires des processus $(X_t)_t$ et $(Y_t)_t$ sont fortement corrélées, comme l'illustre la figure 1a. Sur la figure 1, on a aussi tracé l'évolution de la prime moyenne marché et l'inverse du ratio sinistre sur prime (S/P) marché pour le marché auto français. Le graphique 1b montre à quel point ces deux grandeurs sont liées et valide le fait que la prime marché est inversement proportionnelle au ratio S/P.

∗. Clive W.J. Granger et Robert F. Engle ont reçu le prix Nobel d'économie pour leurs travaux sur les séries temporelles macroéconomiques et financières le 10 décembre 2003.



(a) Exemple de modèles cointégrés (b) Marché auto français

FIGURE 1 – Séries temporelles

Si on désigne l'opérateur différence par Δ, le système d'équations ci-dessus peut se réécrire

$$\Delta X_t = \mathcal{E}_t \quad \text{et} \quad \Delta Y_t = 2X_t - Y_t + 2\mathcal{E}_t + \widetilde{\mathcal{E}}_t.$$

L'exemple précédent est en fait un cas particulier du théorème de réprésentation de Granger. Pour deux séries X_t, Y_t cointégrées d'ordre 1, il existe α_1, α_2 non tous nul et $\beta \neq 0$, telle que soit $\Delta X_t - \alpha_1(Y_{t-1} - \beta X_{t-1})$ soit $\Delta Y_t - \alpha_2(Y_{t-1} - \beta X_{t-1})$ sont stationnaires.

Plus généralement, les modèles autorégressifs vectoriels avec cointégration d'ordre p admettent la représentation suivante :

$$\Delta \mathbf{X}_t = AB^T \mathbf{X}_{t-1} + \sum_{j=1}^{p} \Gamma_j \Delta \mathbf{X}_{t-j} + \mathcal{E}_t,$$

où $\mathbf{X}_t$ est un processus de dimension n, AB^T une matrice $n \times n$ produit de matrices $n \times r$, Γ_j des matrices $n \times n$ et $\mathcal{E}_t$ un bruit blanc corrélé de dimension n. Ces modèles sont largement utilisés en économétrie. Dans le cadre des modèles de cycles de marché, ils sont aussi abondamment utilisés (voir, par exemple, Haley (1993); Grace et Hotchkiss (1995); Doherty et Garven (1995); Blondeau (2001)). Typiquement, nous modélisons la prime moyenne marché en fonction des ratios sinistres sur primes, de l'inflation et d'autres indicateurs macroéconomiques comme les taux d'intérêts court et long termes.

Problèmes de cette approche

Pour prendre des actions pertinentes, l'assureur doit savoir se positionner par rapport au marché. L'utilisation de modèles cointégrés apparaît nécessaire, car les modèles de séries chronologiques auto-régressifs sont rarement adaptés pour modéliser la prime marché ou le ratio marché sinistre sur prime. Une fois après avoir calibré un modèle de séries chronologiques

et un modèle de résiliation, l'assureur peut donc juger en théorie de sa politique tarifaire pour l'année suivante.

En pratique, il ne suffit pas de modéliser la prime moyenne marché pour faire un tarif. L'assureur doit être capable de décliner les changements de tarif par segment. C'est à dire, il doit pouvoir modéliser les plus grands assureurs de la place (par exemple les 5 premiers) individuellement et pas seulement comme un seul acteur. Et c'est là que l'approche par série temporelle est problématique : il faut avoir un historique assez important par assureur pour la variable considérée. En pratique, c'est rarement le cas voire impossible.

L'objectif du chapitre 2 est de proposer une réponse à ce problème. Nous modélisons conjointement le comportement des assurés et la compétition entre assureurs. Le comportement des assurés sera modélisé par une paramétrisation multinomiale logit, tandis que la compétition sera modélisée par la théorie des jeux non-coopérative, que nous présentons dans la prochaine section.

Théorie des jeux et modèles de compétition

La théorie des jeux est l'étude des interactions entre plusieurs agents (hommes, entreprises, animaux, etc...) et regroupe l'ensemble des outils mathématiques nécessaires à la compréhension du phénomène de prise de décision pour un problème donné. Le principe fondamental sous-jacent à la théorie des jeux est que les joueurs tiennent compte, d'une manière ou d'une autre, des comportements des autres joueurs dans leur prise de décision, à l'opposé d'une vision individualiste de la théorie du contrôle optimal.

La théorie des jeux prend ses racines dans les études économiques d'oligopoles réalisées par Cournot (1838); Edgeworth (1881) et Bertrand (1883). Elle a été popularisée et est devenue une discipline à part entière grâce au livre de von Neumann et Morgenstern (1944), qui pose les bases des jeux à somme nulle à plusieurs joueurs, non coopératifs et coopératifs. Quelques années plus tard, Nash (1950a,b, 1951, 1953) * a transformé la théorie des jeux en proposant un nouveau concept d'équilibre et étudié l'existence de tels équilibres. Depuis, la théorie des jeux n'a cessé de croître dans de multiples directions.

Le champ d'application de la théorie des jeux ne se restreint pas à l'économie. Elle s'applique notamment à la biologie, l'ingénierie, les transports, les réseaux, etc... La présentation, qui suit, se base sur les ouvrages de référence suivants : Fudenberg et Tirole (1991), Basar et Olsder (1999), Osborne et Rubinstein (2006).

Un jeu est une description formelle d'une interaction entre plusieurs joueurs. Il est constitué d'un ensemble de joueurs $E = \{1, \dots, I\}$, d'une fonction objective ou d'une fonction coût pour chacun des joueurs $O_i : X \mapsto \mathbb{R}$, et d'un ensemble d'actions possibles par joueur $X_i \subset \mathbb{R}^{n_i}$ pour $i \in E$, où $X = X_1 \times \cdots \times X_I$. Notons que l'ensemble des actions X_i du joueur i n'est pas nécessairement fini ni nécessairement discret. Un profil d'action x regroupe un ensemble d'actions x_i des I joueurs. Un concept de solution va spécifier un critère selon lequel un profil d'action $\boldsymbol{x}$ est plus préférable que $\boldsymbol{y}$ pour un joueur.

Il existe de multiples classes de jeux permettant de préciser le type d'intéractions étudiées : les actions des joueurs sont elles simultanées ou séquentielles (description normale ou extensive des jeux) ; cherche-t-on à maximiser un bien-être global (coopération) ou les joueurs sont-ils non coopératifs ; l'information entre les joueurs est-elle parfaite (chaque joueur connait les objectifs de ses compétiteurs) ou seule une partie de l'information est révélée aux concurrents ;

∗. John F. Nash a reçu le prix Nobel d'économie pour ces travaux en théorie des jeux le 8 décembre 1994.

joue-t-on sur une ou plusieurs périodes ; la fonction objective dépend-elle d'un phénomène aléatoire ? Nous ne présenterons dans cette introduction que les jeux simultanés non-coopératifs déterministes à information parfaite.

Jeux statiques

Nous supposons que l'information est parfaite, c'est à dire chaque joueur i connait les fonctions objectives/coûts O_j des autres joueurs $j \neq i$. Les joueurs choisissent leur action simultanément : personne ne peut tirer profit en jouant après les autres. De plus, chaque joueur cherche son propre bien-être et ne peut coopérer, c'est à dire on exclut les jeux coopératifs.

Jeux finis

Considérons le cas où l'ensemble des actions possibles est fini, c'est à dire les ensembles X_i sont discrets. L'ensemble des actions possibles est donc fini et contient $\text{Card}(X_1) \times \cdots \times \text{Card}(X_I)$ éléments. Pour toutes ces possibilités, on peut calculer la valeur des fonctions objectives pour chacun des joueurs. Ainsi, la fonction objective de chaque joueur peut être décrite dans un tableau multidimensionnel. Par simplification, on se restreint aux jeux à deux joueurs, $I = 2$.

Commençons par un exemple, le dilemne du prisonnier. Deux suspects (complices d'un délit) sont retenus dans des cellules séparées, dans lesquelles ils ne peuvent pas communiquer. Les enquêteurs leurs proposent de passer aux aveux pour réduire leur éventuelle peine de prison. Si un et seulement un des deux prisonniers dénonce l'autre, il est remis en liberté alors que le second écopera de la peine maximale (par exemple 10 ans). Si les deux se dénoncent entre eux, ils seront condamnés à une peine plus légère (par exemple 5 ans). Enfin si les deux refusent de se dénoncer, la peine sera minimale (par exemple 6 mois), faute d'éléments au dossier.

Les coûts des joueurs sont représentés par la double matrices suivantes, où le joueur 1 joue sur les lignes et le joueur 2 sur les colonnes.

J1 | J2	se tait	dénonce
se tait	(-1/2, -1/2)	(-10, 0)
dénonce	(0, -10)	(-5, -5)

S'ils coopéraient, les deux joueurs écoperaient seulement de 6 mois de prison. Mais comme ils ne peuvent coopérer, chacun va chercher à minimiser sa peine potentielle, c'est à dire le joueur 1 cherche le minimum des maximum des lignes, tandis que le joueur 2 cherche le minimum des maximum des colonnes. Par conséquent, chaque joueur va choisir de dénoncer l'autre joueur.

Les jeux finis à deux joueurs sont appelés les jeux bimatrices. En effet, les coûts des deux joueurs sont représentés par deux matrices A et B de taille $\text{Card}(X_1) \times \text{Card}(X_2)$. Dans cette configuration a_{ij} et b_{ij} représentent le coût des joueurs lorsque pour un profil d'action (i, j) où i (respectivement j) désigne le $i^{\text{ème}}$ ($j^{\text{ème}}$) élément de l'ensemble fini X_1 (X_2). Nous introduisons maintenant l'équilibre de Nash.

Définition. *Une paire de stratégies $(i^\star, j^\star)$ constitue un équilibre de Nash au jeu bimatrice A, B si les inégalités suivantes sont vérifiées*

$$a_{i^\star j^\star} \leq a_{ij^\star} \quad et \quad b_{i^\star j^\star} \leq b_{i^\star j} \tag{2}$$

pour tout $i = 1, \ldots, Card(X_1)$ et $j = 1, \ldots, Card(X_2)$.

Si A représente les gains plutôt que les coûts, il suffit d'inverser les inégalités. Un équilibre de Nash s'interprète comme un point où aucun des joueurs n'a d'intérêt à changer d'actions tant que son opposant ne change pas.

Une question légitime qui peut se poser maintenant est l'existence d'équilibre de Nash pour toutes matrices A, B. Malheureusement, il existe des matrices pour lesquelles aucun équilibre n'existe. Par exemple, pour

$$A = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix}.$$

La raison de l'inexistence d'équilibre de Nash est la discontinuité des actions $i = 1, \ldots, \mathrm{Card}(X_1)$ et $j = 1, \ldots, \mathrm{Card}(X_2)$.

Un cas particulier des jeux finis à deux joueurs est le cas où les objectifs des deux joueurs sont antagonistes, c'est à dire $B = -A$. Ces sont les jeux à somme nulle. L'équation (2) de l'équilibre de Nash se réduit à la double inégalité suivante

$$a_{i^\star j} \leq a_{i^\star j^\star} \leq a_{ij^\star}.$$

L'équilibre de Nash $(i^\star, j^\star)$ est appelé point col. Définissons le minimax et le maximin par $\underline{V}(A) = \max_j \ \min_i \ a_{ij}$ et $\overline{V}(A) = \min_i \ \max_j \ a_{ij}$. Si $\overline{V}(A) = \underline{V}(A)$ alors il existe un équilibre de Nash. Pour les jeux à somme non nulle, on a vu que ce n'était pas aussi simple.

L'astuce proposée par Nash lui-même pour garantir l'existence est de considérer des stratégies mixtes, où les joueurs choisissent leur action en fonction d'un événement aléatoire. Par exemple, dans l'exemple précédent, le joueur 1 peut choisir 2 fois sur 3 de jouer la première action et 1 fois sur 3 la deuxième. Une telle stratégie est notée $(2/3, 1/3)$. Les stratégies mixtes du joueur i sont par définition une loi de probabilité parmi les actions de l'ensemble X_i. Pour ne pas confondre, les stratégies non mixtes sont appelées stratégies pures et sont des cas particuliers des stratégies mixtes où la loi de probabilité est dégénérée, par exemple (1,0) dans l'exemple précédent.

Définition. *Une paire de stratégies $(\boldsymbol{x}^\star, \boldsymbol{y}^\star)$ constitue un équilibre de Nash au jeu bimatrice (A, B) en stratégie mixte si pour tous vecteurs de probabilité x, y, on a*

$$\boldsymbol{x}^{\star T} A \boldsymbol{y}^\star \leq \boldsymbol{x}^T A \boldsymbol{y}^\star \quad \textit{et} \quad \boldsymbol{x}^{\star T} B \boldsymbol{y}^\star \leq \boldsymbol{x}^{\star T} B \boldsymbol{y}.$$

On peut maintenant énoncer un théorème d'existence.

Théorème. *Tous les jeux bimatrices admettent un équilibre de Nash en stratégie mixte.*

La démonstration est basée sur le théorème de point de fixe de Brouwer, qui suit.

Théorème (Brouwer (1912))**.** *Soient B^n la boule unité d'un espace euclidien de dimension n et $T : B^n \mapsto B^n$ une application. Si T est continue, alors T admet au moins un point fixe.*

L'ensemble B^n peut être remplacé par n'importe quel ensemble compact, convexe, non vide. Dans notre cas, on considèrera le simplexe de dimension 2 ou supérieure. Il est assez facile de comprendre pourquoi l'équilibre de Nash en stratégies pures n'existe pas forcément : l'ensemble $X_1 \times X_2$ n'est pas convexe.

Le calcul d'équilibre en stratégie mixte est assez complexe. Néanmoins, on peut le reformuler au problème d'optimisation bilinéaire

$$\min_{\boldsymbol{x}, \boldsymbol{y}, p, q} \ \boldsymbol{x}^T A \boldsymbol{y} + \boldsymbol{x}^T B \boldsymbol{y} + p + q,$$

sous contrainte

$$A\boldsymbol{y} \geq -p\mathbf{1}, B^T\boldsymbol{x} \geq -q\mathbf{1}, \boldsymbol{x} \geq 0, \boldsymbol{y} \geq 0, \boldsymbol{x}^T\mathbf{1} = 1, \boldsymbol{y}^T\mathbf{1} = 1,$$

$p, q \in \mathbb{R}$ sont des variables auxiliaires telles que si $(\boldsymbol{x}^\star, \boldsymbol{y}^\star, p^\star, q^\star)$ sont solutions du problème précédent alors $p^\star = \boldsymbol{x}^{\star T} A\boldsymbol{y}^\star$ et $q^\star = \boldsymbol{x}^{\star T} B\boldsymbol{y}^\star$, voir la section 3.6 de Basar et Olsder (1999). Une autre approche basée sur l'itération des stratégies à l'aide de pivots est l'algorithme de Lemke-Howson.

Les jeux finis à deux joueurs peuvent être généralisés à des jeux à I joueurs. Les matrices se transforment en tableaux à I dimensions et les deux inégalités définissant un équilibre deviennent I inégalités. Nous renvoyons le lecteur intéressé vers les ouvrages de référence précédemment listés.

Jeux continus

Traitons maintenant le cas des jeux continus où les ensembles de stratégies X_i sont continus et non plus discrets. On omet volontairement le cas des jeux où l'espace X_i est dénombrable et infini, où $\mathbb{N}$, car il ne représente d'utilité dans le cadre de cette thèse. On peut par exemple penser à un intervalle de prix, un intervalle de quantités, etc... Les ensembles X_i sont généralement supposés compact, convexe et non vide. L'équilibre de Nash se définit de la manière suivante.

Définition. *Pour un jeu à deux joueurs où O_1, O_2 désignent le coût des joueurs, un couple de stratégie $(x_1^\star, x_2^\star) \in X_1 \times X_2$ est un équilibre de Nash si les inégalités suivantes sont respectées*

$$O_1(x_1^\star, x_2^\star) \leq O_1(x_1, x_2^\star) \quad et \quad O_1(x_1^\star, x_2^\star) \leq O_1(x_1^\star, x_2), \tag{3}$$

pour tout $(x_1, x_2) \in X_1 \times X_2$.

Lorsqu'on travaille avec des fonctions de gains O_1, O_2 plutôt que des fonctions de coûts, il suffit de renverser les inégalités. Si le jeu est à somme nulle, c'est à dire $O_2 = -O_1$, alors un équilibre de Nash (équation (3)) est un point col

$$O_1(x_1^\star, x_2) \leq O_1(x_1^\star, x_2^\star) \leq O_1(x_1, x_2^\star).$$

Pour un jeu à I joueurs, on introduit les notations suivantes. Soit $i \in E$ un joueur : x_i désigne l'action du joueur i, tandis que $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_I)$ les actions des autres joueurs. L'équilibre de Nash se définit comme suit.

Définition. *Pour un jeu à I joueurs où $O_i, i \in E$ désignent le coût du joueur i, un vecteur de stratégie $(x_1^\star, \ldots, x_I^\star) \in X$ est un équilibre de Nash si pour tout $i \in E$, on a*

$$O_i(x_i^\star, x_{-i}^\star) \leq O_i(x_i, x_{-i}^\star), \quad pour\ tout \quad x_i \in X_i. \tag{4}$$

Pour mieux comprendre les théorèmes d'existence qui suivent, il faut comprendre que l'équation (4) est en fait un problème d'optimisation. Un équilibre de Nash $x^\star$ vérifie les I sous-problèmes d'optimisation suivant

$$x_i^\star \in \underset{x_i \in X_i}{\arg\min}\ O_i(x_i, x_{-i}^\star).$$

Le problème d'optimisation ci-dessus admet (au moins) une solution si la fonction $x_i \mapsto O_i(x_i, x^\star_{-i})$ est quasiconvexe. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est quasiconvexe si pour tout $x, y \in \mathbb{R}$, et pour tout $\lambda \in]0,1[$, on a $f(\lambda x + (1-\lambda)y) \leq \max(f(x), f(y))$. Géométriquement parlant, une fonction univariée quasiconvexe est unimodale, par exemple monotone ou décroissante et croissante.

On énonce maintenant un premier théorème d'existence.

Théorème (Nikaido et Isoda (1955))**.** *Soit un jeu à I joueurs où les espaces de stratégie X_i sont non-vides, convexes et compacts. Supposons que les fonctions de coût $O_i : X \mapsto \mathbb{R}$ sont continus. Si les fonctions $x_i \mapsto O_i(x_i, x_{-i})$ sont quasiconvexes, alors il existe un équilibre de Nash (en stratégie pure).*

Si on travaille avec des fonctions de gain, alors la quasiconvexité devient la quasiconcavité, définie par $f(\lambda x + (1-\lambda)y) \geq \min(f(x), f(y))$, pour tout $\lambda \in]0,1[$.

Le concept de quasiconvexité est plus faible que celui de convexité. En fait, il existe plusieurs variantes allant de la quasiconvexité à la stricte convexité. Nous rappellons ci-dessous certains concepts, renvoyons le lecteur vers Diewert *et al.* (1981) détaillant les neuf sortes de quasiconvexité. Soit $f : \mathbb{R}^n \mapsto \mathbb{R}$ une fonction. On dit que

- f est quasiconvexe : $\forall x, y \in \mathbb{R}, \forall \lambda \in]0,1[$, on a $f(\lambda x + (1-\lambda)y) \leq \max(f(x), f(y))$.
- f est convexe : $\forall x, y \in \mathbb{R}, \forall \lambda \in]0,1[$, on a $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$.
- f est strictement convexe : $\forall x, y \in \mathbb{R}, \forall \lambda \in]0,1[$, on a $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$.

Un concept manquant est la pseudoconvexité, mais qui requiert une fonction au moins différentiable directionnellement. Pour une fonction C^1, on a

- f est quasiconvexe : $\forall x, y \in \mathbb{R}$, on a $f(x) \geq f(y) \Rightarrow \nabla f(x)^T (y - x) \leq 0$.
- f est pseudoconvexe : $\forall x, y \in \mathbb{R}$, on a $f(x) > f(y) \Rightarrow \nabla f(x)^T (y - x) < 0$.
- f est convexe : $\forall x, y \in \mathbb{R}$, on a $f(y) - f(x) \leq \nabla f(x)^T (y - x)$.
- f est strictement convexe : $\forall x, y \in \mathbb{R}$, on a $f(y) - f(x) < \nabla f(x)^T (y - x)$.

Pour une fonction C^2, on a

- f est quasiconvexe : $\forall x \in \mathbb{R}, \forall d \in \mathbb{R}^n, d^T \nabla f(x) = 0 \Rightarrow d^T \nabla^2 f(x) d \geq 0$.
- f est pseudoconvexe : $\forall x \in \mathbb{R}, \forall d \in \mathbb{R}^n, d^T \nabla f(x) = 0 \Rightarrow d^T \nabla^2 f(x) d > 0$.
- f est convexe : $\forall x \in \mathbb{R}, \forall d \in \mathbb{R}^n, d^T \nabla^2 f(x) d \geq 0$, c'est à dire $\nabla^2 f$ semidéfinie positive.
- f est strictement convexe : $\forall x \in \mathbb{R}, \forall d \in \mathbb{R}^n, d^T \nabla^2 f(x) d > 0$, c'est à dire $\nabla^2 f$ définie positive.

Toutes les définitions sont incrémentales, ainsi la stricte convexité implique la convexité, impliquant la pseudoconvexité, impliquant la quasiconvexité. Par conséquent, on constate que le théorème d'existence d'équilibre de Nash de Nikaido et Isoda (1955) requiert une des conditions les plus faibles de convexité sur la fonction $x_i \mapsto O_i(x_i, x_{-i})$.

Pour avoir l'unicité de l'équilibre de Nash, il faut cependant requérir beaucoup plus que la quasiconvexité. Le théorème 2 de Rosen (1965) donne un résultat d'unicité dans un cadre légèrement plus général que l'équation (4). Il considère le problème suivant

$$\min_{x_i \in X_i} O_i(x_i, x^\star_{-i}) \quad \text{tel que} \quad g^i(x_i) \leq 0, \tag{5}$$

où $g_i : x_i \mapsto g^i(x_i)$ est la fonction contrainte du joueur i supposée continue. L'ensemble de stratégies possibles se réduit donc à l'ensemble $\widetilde{X}_i = \{x_i \in X_i, g^i(x_i) \leq 0\}$.

Théorème (Rosen (1965))**.** *Soit un jeu continu à I joueurs avec des espaces X_i non-vides, convexes, compacts, des fonctions contraintes g^i convexes et des fonctions coûts O_i telles que $x_i \mapsto O_i(\boldsymbol{x})$ sont convexes.*

S'il existe $r > 0$ telle que la fonction $g_O : \mathbb{R}^n \times \mathbb{R}^I \mapsto \mathbb{R}^n$ définie par

$$g_O(\boldsymbol{x}, r) = \begin{pmatrix} r_1 \nabla_{x_1} O_1(\boldsymbol{x}) \\ \vdots \\ r_I \nabla_{x_I} O_I(\boldsymbol{x}) \end{pmatrix},$$

vérifie la propriété suivante

$$(\boldsymbol{x} - \boldsymbol{y})^T g_O(\boldsymbol{y}, r) + (\boldsymbol{y} - \boldsymbol{x})^T g_O(\boldsymbol{x}, r) > 0, \tag{6}$$

alors l'équilibre de Nash vérifiant l'équation (5) est unique.

Une condition suffisante pour l'équation (6) est que la matrice par bloc suivante soit définie positive

$$G_O(\boldsymbol{x}, r) = \begin{pmatrix} r_1 \nabla_{x_1} \nabla_{x_1} O_1(\boldsymbol{x}) & \dots & r_1 \nabla_{x_I} \nabla_{x_1} O_1(\boldsymbol{x}) \\ \vdots & & \vdots \\ r_I \nabla_{x_1} \nabla_{x_I} O_I(\boldsymbol{x}) & \dots & r_I \nabla_{x_I} \nabla_{x_I} O_I(\boldsymbol{x}) \end{pmatrix}.$$

L'équation (6) se réécrit

$$\sum_{i=1}^{I} r_i (x_i - y_i)^T \nabla_{x_i} O_i(\boldsymbol{y}) + \sum_{i=1}^{I} r_i (y_i - x_i)^T \nabla_{x_i} O_i(\boldsymbol{x}) > 0.$$

Rappelons que pour une fonction f strictement convexe, on a $\nabla f(x)^T (y - x) > f(y) - f(x)$, ce qui est équivalent à $\nabla f(y)^T (x - y) > f(x) - f(y)$. Ainsi, une condition suffisante (mais pas nécessaire) pour que l'équation (6) soit vérifée est la stricte-convexité des fonctions $x_i \mapsto O_i(x_i, x_{-i})$.

Une autre condition garantissant cette fois-ci le caractère définie positive de la matrice $G_O(\boldsymbol{x}, r)$ est la dominance diagonale , c'est à dire $\forall i \in E, \forall m = 1, \dots, n_i$,

$$\left| \frac{\partial^2 O_i(\boldsymbol{x})}{\partial x_{im}^2} \right| > \sum_{j=1}^{I} \sum_{k=1}^{n_j} \left| \frac{\partial^2 O_i(\boldsymbol{x})}{\partial x_{jk} \, \partial x_{im}} \right|.$$

Ainsi, guarantir l'unicité de l'équilibre de Nash requiert la convexité des fonctions $x_i \mapsto O_i(\boldsymbol{x})$ et que l'équation (6) soit vérifiée, tandis que l'existence nécessite seulement la quasiconvexité de ces fonctions.

Les méthodes de calcul d'équilibre de Nash sont complexes, puisqu'il ne suffit pas de réaliser I optimisations basées sur l'équation (5). Les équations sont toutes liées puisque les fonctions objectives O_i dépendent des actions des autres joueurs. Nous présentons ici le cas plus simple des jeux à deux joueurs et renvoyons au chapitre 3 pour le cas général.

Notons R_1 (resp. R_2) la fonction de meilleure réponse du joueur 1 (joueur 2) pour les actions du joueur 2 (joueur 1) : $R_1(x_2) = \{x_1 \in X_1, \forall y_1 \in X_1, O_1(x_1, x_2) \leq O_1(y_1, x_2)\}$ (resp. $R_2(x_1) = \{x_2 \in X_2, \forall y_2 \in X_2, O_2(x_2, x_1) \leq O_2(y_2, x_1)\}$). Dans les définitions précédentes, on suppose implicitement que la réaction d'un joueur par rapport à l'autre soit unique, par exemple lorsque les fonctions objectives sont convexes. Un équilibre de Nash est un point

d'intersection des courbes $(x_1, R_2(x_1))$ et $(R_1(x_2), x_2)$ pour $x_1 \in X_1$ et $x_2 \in X_2$. C'est à dire un point fixe de l'équation $x_1 = R_1(R_2(x_1))$.

De la même manière que pour les jeux finis, on peut définir des stratégies mixtes pour les jeux continus. Les actions sont des fonctions de répartition μ_i et les fonctions objectives deviennent

$$\overline{O}_i(\mu_1, \ldots, \mu_I) = \int_{X_1} \ldots \int_{X_I} O_i(x_1, \ldots, x_I) d\mu_1(x_1) \ldots d\mu_I(x_I).$$

On peut étendre la définition d'un équilibre de Nash aux stratégies mixtes.

Définition. *Pour un jeu à I joueurs où $O_i, i \in E$ désignent le coût du joueur i, un vecteur de probabilité $(\mu_1^\star, \ldots, \mu_I^\star) \in X$ est un équilibre de Nash en stratégie mixte si pour tout $i \in E$, et pour toute fonction de répartion μ_i sur X_i, on a*

$$\overline{O}_i(\mu_i^\star, \mu_{-i}^\star) \leq \overline{O}_i(\mu_i, \mu_{-i}^\star). \tag{7}$$

Nous donnons ci-dessous le théorème de Glicksberg (1950).

Théorème. *Soit un jeu à I joueurs où les espaces de stratégie X_i sont compacts. Si les fonctions de coût $O_i : X \mapsto \mathbb{R}$ sont continues, alors il existe un équilibre de Nash en stratégie mixte.*

Jeux généralisés

Les jeux généralisés proposent une extension des équilibres de Nash suggérée dans l'équation (5). Pour un jeu à I joueurs, nous introduisons une fonction contrainte rendant les actions possibles d'un joueur dépendantes non seulement de son action mais aussi des actions des autres joueurs. Soit $g^i : X \mapsto \mathbb{R}^{m_i}$ la fonction contrainte d'un joueur telle que les actions possibles du joueur i appartiennent à l'ensemble

$$\{x_i \in X_i, g^i(x_i, x_{-i}) \leq 0\}.$$

L'équilibre de Nash généralisé se définit comme suit.

Définition. *Pour un jeu à I joueurs où $O_i, i \in E$ désignent le coût du joueur i, un vecteur de stratégie $(x_1^\star, \ldots, x_I^\star) \in X$ est un équilibre de Nash généralisé si pour tout $i \in E$, on a*

$$O_i(x_i^\star, x_{-i}^\star) \leq O_i(x_i, x_{-i}^\star), \quad \textit{pour tout} \quad x_i \in X_i, g^i(x_i, x_{-i}) \leq 0. \tag{8}$$

La différence entre les équations (5) et (8) est le fait que la fonction contrainte dépend des actions de tous les joueurs et pas seulement de l'action x_i du joueur i. Un équilibre de Nash généralisé $\boldsymbol{x}^\star$ vérifie donc I sous-problèmes d'optimisation

$$\min_{x_i \in X_i} O_i(x_i, x_{-i}^\star) \ \text{ such that } \ g^i(x_i, x_{-i}^\star) \leq 0, \tag{9}$$

pour tout $i \in E$. Pour donner des théorèmes d'existence d'équilibres généralisés, nous introduisons les correspondances.

Une correspondance $F : X \mapsto 2^Y$ est une application telle que $\forall x \in X$, $F(x)$ est un sous-ensemble de Y. Les correspondances sont parfois notées $F : X \mapsto \mathcal{P}(Y)$ ou encore $F : X \rightrightarrows Y$. Pour les correspondances, le domaine de F se définit par $dom(F) = \{x \in X, F(x) \neq \emptyset\}$, la portée par $rg(F) = \bigcup_x F(x)$, le graphe de F par $Gr(F) = \{(x, y) \in X \times Y, y \in F(x)\}$. Deux

exemples typiques de correspondances sont $F : x \mapsto [-|x|, |x|]$; l'inverse d'une fonction f, $F : x \mapsto f^{-1}(x)$.

Maintenant, nous définissons deux types de continuités pour les correspondances : semicontinuité inférieure et supérieure, abrégées l.s.c. et u.s.c.. Dans la littérature, deux définitions s'opposent : la semicontinuité au sens de Berge (voir (Berge, 1963, page 109)) et la semicontinuité au sens de Hausdorff (voir (Aubin et Frankowska, 1990, page 38-39)).

Cependant, ces définitions sont équivalentes si la correspondance F est à valeur compacte. Dans ce cas, les semicontinuités u.s.c/l.s.c se caractérisent sur le graphe de F. Nous rapportons ici ces définitions, voir Hogan (1973).

Définition. *F est semicontinue supérieurement (u.s.c.) en x, si $\forall (x_n) \in X^{\mathbb{N}}, x_n \to x, \forall y_n \in T(x_n)$, et $\forall y \in Y$,*

$$y_n \to y \Rightarrow y \in T(x).$$

F est semicontinue inférieurement (l.s.c.) en x, si $\forall (x_n)_n \in X^{\mathbb{N}}, x_n \to x, \forall y \in T(x), \exists (y_k) \in Y^{\mathbb{N}}$ et $\forall k \in \mathbb{N}$,

$$y_k \in T(x_k) \quad et \quad y_k \to y.$$

F est semicontinue supérieurement (resp. inférieurement) sur X, si F est semicontinue supérieurement (inférieurement) en tout point de X.

Introduisons maintenant la correspondance de contraintes liés aux équilibres de Nash généralisés $C_i : X_{-i} \mapsto 2^{X_i}$ représentant les contraintes du joueur i par

$$C_i(x_{-i}) = \{x_i \in X_i, g^i(x_i, x_{-i}) \leq 0\}.$$

Nous pouvons maintenant énoncer le théorème d'existence.

Théorème (Ichiishi (1983))**.** *Soit un jeu à I joueurs caractérisé par des espaces de stratégies $X_i \subset \mathbb{R}^{n_i}$, des correspondances de contrainte C_i et des fonctions objectives $O_i : \mathbb{R}^{n_i} \mapsto \mathbb{R}$. Si pour tout joueur i, on a*

- *X_i est non vide, convexe et compact,*
- *C_i est u.s.c. et l.s.c. sur X_{-i},*
- *$\forall x_{-i} \in X_{-i}$, $C_i(x_{-i})$ est non vide, fermé et convexe,*
- *O_i est continue sur le graphe $Gr(C_i)$,*
- *$\forall x \in X$, $x_i \mapsto O_i(x_i, x_{-i})$ est quasiconcave sur $C_i(x_{-i})$,*

Alors il existe un équilibre de Nash généralisé.

La démonstration repose sur le théorème de point fixe pour les correspondances * de Kakutani (Kakutani (1941)) et sur le théorème du maximum de Berge. Nous renvoyons le lecteur vers Ichiishi (1983), Aubin (1998) ou Ok (2005) pour une démonstration de ce théorème.

Nous analysons maintenant les conséquences de la semicontinuité l.s.c. et u.s.c. de la correspondance C_i sur les fonctions contraintes. Une propriété de Rockafellar et Wets (1997) permet d'obtenir facilement la semicontinuité u.s.c. lorsque les fonctions g^i sont continues.

Proposition (Rockafellar et Wets (1997))**.** *Soit $C_i : X_{-i} \mapsto 2^{X_i}$ la correspondance de contraintes définie précédemment. Si les ensembles X_i sont fermés et que toutes les composantes g^i_j sont continues sur $X_i \times X_{-i}$, alors C_i est une u.s.c. sur X_{-i}.*

∗. L'existence d'équilibre de Nash sans contrainte repose sur le théorème de Brouwer.

Cependant, il est plus ardu de montrer la semicontinuité inférieure d'une correspondance. Rockafellar et Wets (1997) suppose l'existence d'un point à l'intérieur du domaine de contraintes, c'est à dire $\exists(\bar{x}_i, \bar{x}_{-i}) \in X_i \times X_{-i}, g^i(\bar{x}_i, \bar{x}_{-i}) > 0$. Mais en utilisant le théorème 13 de Hogan (1973), nous avons une condition plus faible.

Proposition (Hogan (1973))**.** *Soit $C_i : X_{-i} \mapsto 2^{X_i}$ la correspondance de contrainte du jeu généralisé. Soit $\widetilde{C}_i$ définie par $\widetilde{C}_i(x_{-i}) = \{x_i \in X_i, g^i(x_i, x_{-i}) > 0\}$. Si les composantes de g^i sont semicontinues (c'est à dire fermeture de l'épigraphe) et si $C_i(\bar{x}_{-i}) \subset cl(\widetilde{C}_i(\bar{x}_{-i}))$, alors C_i est l.s.c.*

Par conséquent, si les fonctions contraintes g^i sont continues alors C_i est bien semicontinue inférieurement et supérieurement. Néanmoins, nous devons aussi garantir que C_i renvoie des ensembles convexes, fermés et non-vides. Si les fonctions $x_i \mapsto g^i(x_i, x_{-i})$ sont quasiconvexes, alors la convexité est garantie. En effet, la quasiconvexité d'une fonction f est équivalente à ce que tous les ensembles $U_f(r) = \{x \in X, f(x) \geq r\}$ soient convexes pour tout r, voir Diewert *et al.* (1981). La continuité de g^i va garantir la fermeture des ensembles. Mais, il est difficile de trouver des conditions garantissant que les ensembles $C_i(\bar{x}_{-i})$ soient non-vides, autres que de garantir l'existence d'un point $(\bar{x}_i, \bar{x}_{-i}) \in X_i \times X_{-i}, g^i(\bar{x}_i, \bar{x}_{-i}) > 0$ pour tout x_{-i}.

L'unicité de l'équilibre de Nash généralisé est un sujet nettement plus complexe que pour les équilibre de Nash standard. Pour appréhender ce problème, nous devons introduire les conditions d'optimisation du premier ordre des I sous-problèmes.

En supposant que les fonctions objectives O_i et contraintes g^i soient continûment différentiable, les conditions nécessaires de Karush-Kuhn-Tucker (KKT) pour le sous-problème d'équation (9) sont données ci-dessous. Si $x^\star$ résout le problème (9) pour tout $i \in E$ et que pour chaque joueur, une qualification des contraintes est satisfaite, alors pour tout $i \in E$, il existe un multiplicateur de Lagrange $\lambda^{i\star} \in \mathbb{R}^{m_i}$ tel que

$$\begin{aligned} \nabla_{x_i}\theta_i(\boldsymbol{x}^\star) + \sum_{1\leq j\leq m_i} \lambda_j^{i\star}\nabla_{x_i} g_j^i(\boldsymbol{x}^\star) = 0 \qquad & (\in \mathbb{R}^{n_i}). \\ 0 \leq \lambda^{i\star},\ -g^i(\boldsymbol{x}^\star) \geq 0,\ g^i(\boldsymbol{x}^\star)^T\lambda^{i\star} = 0 \qquad & (\in \mathbb{R}^{m_i}). \end{aligned} \tag{10}$$

Pour que les conditions KKT soient aussi suffisantes, il faut requérir des conditions supplémentaires. Celles-ci sont données dans le théorème 4.6 de Facchinei et Kanzow (2009).

Théorème. *Soit un problème d'équilibre de Nash généralisé vérifiant l'équation (8) et telles que les fonctions objective et contrainte soient continûment différentiable.*

(i) Si $\boldsymbol{x}^\star$ est un équilibre de Nash généralisé et que tous les sous-problèmes (9) satisfassent une qualification de contrainte, alors il existe $\lambda^\star \in \mathbb{R}^m$ tel que $\boldsymbol{x}^\star, \lambda^\star$ résolvent les I systèmes (10).

(ii) Si $\boldsymbol{x}^\star, \lambda^\star$ résolvent les I systèmes (10), que les fonctions $x_i \mapsto O_i(x)$ sont pseudoconvexes et que les ensembles $C_i(x_{-i})$ sont fermés et convexes, alors $x^\star$ résout un équilibre de Nash généralisé.

Jusqu'ici nous n'avons pas explicité les contraintes de qualification, nous le faisons ci-dessous. Les contraintes de qualification ont pour but d'assurer que la version linéarisée de l'ensemble contraint est une bonne approximation locale de l'ensemble original (nonlinéaire) contraint. L'ensemble des contraintes actives au point $\boldsymbol{x}$ se définit par $A_i(\boldsymbol{x}) = \{j = 1, \ldots, m_i, g_j^i(\boldsymbol{x}) = 0\}$. Deux contraintes de qualification sont très utilisées, nous les énonçons ci-dessous.

La qualification de contrainte (CQ) d'indépendance linéaire (LICQ) est satisfaite lorsque l'ensemble des gradients des contraintes actives, $\{\nabla g_j^i(\boldsymbol{x}), i \in A_i(\boldsymbol{x})\}$, est linéairement indépendant. La qualification de contrainte de Slater (SCQ) est satisfaite lorsque toutes les contraintes actives sont strictement actives, c'est à dire $\lambda_j^{i\star} > 0$ et $g_j^i(\boldsymbol{x}) = 0$ pour tout $j \in A_i(\boldsymbol{x})$.

En pratique, des critères simples permettent de vérifier de telles conditions : (i) les contraintes sont toutes linéaires, (ii) les contraintes sont convexes et (iii) les contraintes ont un gradient non nul lorsqu'elles sont actives. Nous renvoyons le lecteur vers le chapitre 12 de Nocedal et Wright (2006) et Arrow et Enthoven (1961).

Maintenant, nous avons les éléments pour présenter un résultat d'unicité pour une sous-classe d'équilibres de Nash. Rosen (1965) s'intéresse aux jeux conjointement convexes où les fonctions contraintes g^i sont communes à tous les joueurs, c'est à dire $g^1 = \cdots = g^I = g^0 : \mathbb{R} \mapsto \mathbb{R}^{m_0}$. Ainsi, les ensembles de stratégies sont tels que pour tout $i \in E$,

$$C_i(x_{-i}) = \{x_i \in X_i, g^0(x_1, \ldots, x_i, \ldots, x_I) \leq 0\}.$$

L'ensemble global des actions possibles se simplifie

$$K = \{\boldsymbol{x} \in X, \forall i \in E, x_i \in C_i(x_{-i})\} = \{\boldsymbol{x} \in X, g^0(\boldsymbol{x}) \leq 0\}.$$

De plus, Rosen (1965) suppose que la fonction g^0 est convexe pour garantir la convexité de cet ensemble K. Les I systèmes (10) pour ce cas particulier se simplifient légèrement en remplaçant g^i par g^0 et $\lambda^{i\star} \in \mathbb{R}^{m_0}$. Rosen (1965) définit un équilibre de Nash normalisé pour les jeux conjointement convexes lorsque $x^\star$ vérifie les I systèmes (10) tels qu'il existe $\lambda^\star \in \mathbb{R}^{m_0}$ et $r_i > 0$,

$$\lambda^{i\star} = \lambda^{0\star}/r_i. \tag{11}$$

En d'autres termes, les multiplicateurs de Lagrange $\lambda^{i\star}$ de chaque joueur i sont reliés par un seul multiplicateur de Lagrange $\lambda^{0\star}$ commun à tous les joueurs et le paramètre $r \in]0, +\infty[^I$. r s'interprète comme un paramètre d'échelle sur les fonctions objectives O_i.

Théorème (Rosen (1965))**.** *Soit un jeu conjointement convexe à I joueurs, où la fonction contrainte g^0 est convexe. Si les fonctions objectives O_i sont convexes alors pour tout $r \in]0, +\infty[^I$, il existe un équilibre de Nash généralisé vérifiant (11).*

Si de plus, pour $r = \bar{r} > 0$ donné, l'inéqualité suivante est vérifiée

$$(\boldsymbol{x} - \boldsymbol{y})^T g_O(\boldsymbol{y}, \bar{r}) + (\boldsymbol{y} - \boldsymbol{x})^T g_O(\boldsymbol{x}, \bar{r}) > 0, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,$$

pour g_O définie par

$$g_O(\boldsymbol{x}, r) = \begin{pmatrix} r_1 \nabla_{x_1} O_1(\boldsymbol{x}) \\ \vdots \\ r_I \nabla_{x_I} O_I(\boldsymbol{x}) \end{pmatrix},$$

alors l'équilibre de Nash généralisé vérifiant (11) est unique pour $r = \bar{r}$.

Ce théorème de Rosen (1965) garantissant l'unicité d'équilibre de Nash généralisé est très similaire au théorème équivalent pour les équilibres de Nash simples, mais à une différence importante, l'équilibre de Nash généralisé dépend de la valeur du coefficient r. Cette classe d'équilibre de Nash vérifiant l'équation vérifiant (11) est appelée l'ensemble des équilibres normalisés. Depuis l'introduction des équilibres normalisés, un consensus sur le choix du paramètre r semble être formé (voir, par exemple, Harker (1991); Facchinei et Kanzow (2009);

von Heusinger *et al.* (2010), pour lesquels les équilibres normalisés sont calculés pour $r = 1$). Les équilibres normalisés ont une interprétation particulière via les inégalités variationnelles, voir Facchinei *et al.* (2007).

Dans le chapitre 2, nous utiliserons des équilibres de Nash simples et généralisés. Nous verrons à quel point les équilibres de Nash généralisés sont plus difficiles à manier du fait qu'ils ne sont pas uniques.

Jeux dynamiques

Dans cette sous-section, nous portons une brève attention aux jeux dynamiques, bien que dans le chapitre 2 nous utilisions un jeu statique. Dans la sous-section précédente, nous avons présenté des jeux statiques, mais dans beaucoup de cas, cela ne reflète pas la réalité. Les agents prennent une suite d'action au cours du temps plutôt qu'une seule. Il existe quatre grandes classes de jeux dynamiques : les jeux répétés, les jeux dynamiques à variable d'état (en temps discret ou en temps continu) et les jeux de la théorie de l'évolution.

Nous nous concentrons sur les jeux répétés et les jeux à équation d'état en temps discret. Les premiers trouvent leurs applications dans les jeux à réputation, par exemple, Alesina (1987) ou dans la définition de politique publique, par exemple Sleet (2001). Les seconds ont été utilisés pour modéliser l'allocation de ressource en eau Ganji *et al.* (2007); Krawczyk et Tidball (2005), d'émission carbone Haurie et Viguier (2002), ou de ressource en énergie Bompard *et al.* (2008); Genc et Sen (2008).

Jeux répétés

Les jeux répétés s'intéressent aux interactions à long terme entre des joueurs au cours de la répétition d'un jeu ordinaire de période en période. Les conditions du jeu (nombre de joueurs, ensemble de stratégies, fonctions objectives) sont constantes au cours du temps. Notons I le nombre de joueurs, $X_i, i \in E$ les ensembles de stratégies supposés finis et O_i les fonctions objectives des joueurs, c'est à dire $O_i(x_1, \dots, x_I)$ représente le gain du joueur i pour $\boldsymbol{x} \in X$. Contrairement au jeu statique, les objectifs des joueurs dans les jeux répétés sont majoritairement présentés en terme de gain plutôt qu'en terme de coût.

Définition. *Un jeu répété basé sur le jeu ordinaire caractérisé par $(I, (X_i)_i, (O_i)_i)$ est une forme extensive d'un jeu avec information parfaite et actions simultanées, telle que les actions se définissent en profil de stratégies au cours du temps $\sigma_i = (x_{i,1}, \dots, x_{i,t}, \dots) \in X^\infty$ et que le joueur i peut comparer la suite de gains $(O_i(x_{1,t}, \dots, x_{I,t}))_t$ pour deux profils différents $\sigma_i, \tilde{\sigma}_i$.*

Une stratégie pour le joueur i est donc une règle de décision permettant de choisir une suite d'actions $\sigma_i = (x_{i,1}, \dots, x_{i,t}, \dots)$ dépendant de l'histoire passée du jeu au temps t. Nous pouvons en imaginer trois grands types : les stratégies à boucle ouverte (ou *open-loop*) dans lesquelles la suite d'actions ne tient pas compte de l'histoire du jeu, les stratégies feedback ou markoviennes où les actions en t ne dépendent que des actions passées en $t-1$ et enfin les stratégies à boucle fermée (ou *closed-loop*) dans lesquelles les joueurs utilisent toute l'histoire passée à n'importe quelle période.

Pour comparer deux stratégies $\sigma_i, \tilde{\sigma}_i$, nous utilisons la somme actualisée des gains

$$G_i(\sigma_1, \dots, \sigma_I) = \sum_{t=0}^{T} \delta^t O_i(x_{1,t}, \dots, x_{I,t}),$$

où δ est facteur d'actualisation. Cette somme permet de caractériser différentes situations suivant la valeur du facteur d'actualisation $\delta < 1$ vs. $\delta = 1$ et le nombre de période $T < \infty$ ou $T = \infty$. Un équilibre de Nash pour le jeu répété est un ensemble de profils $(\sigma_1^\star, \ldots, \sigma_I^\star)$ tel que pour tout joueur i

$$G_i(\sigma_1^\star, \ldots, \sigma_i^\star, \ldots, \sigma_I^\star) \geq G_i(\sigma_1^\star, \ldots, \sigma_i, \ldots, \sigma_I^\star),$$

pour tout profil σ_i. Les jeux répétés possèdent néanmoins des difficultés qui leur sont propres : les profils de stratégie σ_i appartiennent à un espace de dimension infinie, les équilibres de Nash du jeu ordinaire (le constituant) ne sont pas forcément des équilibres pour les jeux répétés.

La litérature académique s'intéresse à caractériser l'ensemble des coûts totaux possibles G_i. On définit pour ce faire l'ensemble des gains possibles par l'enveloppe convexe de l'ensemble des gains possibles

$$\text{co}\left\{(O_1, \ldots, O_I) \in \mathbb{R}^I, \forall i \in E, \forall x_i \in X_i, O_i = O_i(x_1, \ldots, x_I)\ \right\},$$

et l'ensemble des gains individuellement rationnels

$$R = \left\{ g_i \in \mathbb{R}, g_i \geq \min_{m_{-i} \in M(X_i)} \max_{x_i \in X_i} O_i(x_i, m_{-i}) \right\},$$

où $M(X_i)$ représente l'ensemble des stratégies mixtes sur l'ensemble fini X_i et m_i une stratégie mixte, c'est à dire un vecteur de probabilité. Maintenant, nous pouvons présenter les "folk" théorèmes.

Théorème (Folk théorème)**.** *Pour un jeu répété infiniment et sans actualisation, c'est à dire $T = \infty$ et $\delta = 1$, l'ensemble des gains d'équilibre est l'ensemble des gains possibles et individuellement rationnels.*

Des versions du "folk" théorème existent dans le cas d'un jeu actualisé représentant des joueurs plus ou moins impatients et/ou d'un jeu répété un nombre fini de fois, voir Osborne et Rubinstein (2006); Tomala et Gossner (2009).

Jeux à temps discret

Enfin, nous présentons les jeux dynamiques en temps discret avec équation d'état en se basant sur le chapitre 5 de Basar et Olsder (1999). Pour définir de tels jeux, nous introduisons les notations suivantes : un nombre de joueurs I, un nombre d'étapes T, un espace d'état $X \subset \mathbb{R}^d$, des espaces d'actions $U_t^i \subset \mathbb{R}^{m_i}$. Dans cette sous-section, les actions des joueurs ne sont plus notées x_i mais $u_i^t \in U_i^t$ pour la période t.

Définition. *Un jeu dynamique en temps discret est caractérisé par une équation d'état initialisée par $x_1 \in X$*

$$\boldsymbol{x}_{t+1} = f_t(\boldsymbol{x}_t, u_t^1, \ldots, u_t^I),$$

pour une fonction $f_t : X \times U_t^1 \times \cdots \times U_t^I \mapsto X$, des fonctions coûts $L^i : S_1 \times \cdots \times S_T \mapsto \mathbb{R}$, où $S_t = X \times U_t^1 \times \ldots U_t^I$, une structure d'information $\eta_i^t \subset \{x_1^1, \ldots, x_t^I, u_1^1, \ldots, u_{t-1}^I\}$, et un ensemble de fonctions $\gamma_t^i : X \times S_1 \times \cdots \times S_{t-1} \mapsto U_t^i$.

Des exemples de structure d'information sont similaires à ceux définis pour les jeux répétés : boucle ouverte (ou *open-loop*) $\eta_t^i = \{x_1\}$, feedback ou markovien $\eta_t^i = \{x_1, x_t\}$ et boucle fermée (ou *closed-loop*) $\eta_t^i = \{x_1, \dots, x_t\}$. Une stratégie pour le joueur i est donc un ensemble de fonctions $(\gamma_t^i)_t$ spécifiant l'action à jouer $\gamma_t^i(\eta_i^t)$ en t pour une information η_i^t.

Pour simplifier, la fonction de coût L^i a généralement une forme additive

$$L^i((u_t^1, \dots, u_t^N)_t) = \sum_{t=1}^{T} g_t^i(x_{t+1}, u_t^1, \dots, u_t^N, x_t).$$

Un équilibre de Nash dans un tel jeu est un ensemble de fonctions $\gamma^\star$ tel que pour tout $\eta_i^t \in X^{It} \times U_1^1 \times \cdots \times U_t^I$, et pour toute fonction $\gamma_t^i : X \times S_1 \times \cdots \times S_{t-1} \mapsto U_t^i$,

$$L^i\left(\left(\gamma_t^{1\star}\left(\eta_1^t\right), \dots, \gamma_t^{i\star}\left(\eta_i^t\right), \dots, \gamma_t^{N\star}\left(\eta_N^t\right)\right)_t\right) \leq L^i\left(\left(\gamma_t^{1\star}\left(\eta_1^t\right), \dots, \gamma_t^{i}\left(\eta_i^t\right), \dots, \gamma_t^{N\star}\left(\eta_N^t\right)\right)_t\right).$$

Nous parlons d'équilibre de Nash *open-loop*, *feedback* ou *closed-loop* suivant la structure d'information choisie. Dans le cas d'équilibre de Nash *open-loop*, le jeu se réduit à un jeu statique puisque la variable d'état x_t n'a pas d'incidences sur les actions choisies. La stratégie optimale est obtenue à l'aide de la théorie du contrôle optimal et de la programmation dynamique, voir théorème 6.1 de Basar et Olsder (1999). Dans le cas des stratégies feedback et *closed-loop*, des équations rétrogrades du même type donnent des conditions d'optimalités, voir théorèmes 6.5 et 6.6 de Basar et Olsder (1999).

Modèle de compétition en assurance non-vie

Nous présentons dans cette sous-section brièvement le jeu répété du chapitre 2. Considérons un marché d'assurance non-vie composé de I assureurs. Chaque assureur propose des couvertures d'assurance à une population de $n \gg I$ clients. Connaissant la sinistralité passée, au temps t, le jeu consiste à fixer un prix de police. Notons $x_{j,t}$ le prix proposé par l'assureur j au temps t et $n_{j,t}$ le nombre de clients en portefeuille pour la période t. La séquence de jeu pour la période t est la suivante

1. Les assureurs maximisent leur fonction objective

$$\sup_{x_{j,t}} O_{j,t}(x_{j,t}, x_{-j,t}) \quad \text{tel que} \quad g_{j,t}(x_{j,t}) \geq 0,$$

   où $g_{j,t}(x_{j,t}) \geq 0$ représente la contrainte de solvabilité, fonction du capital $K_{j,t-1}$..

2. Une fois la prime d'équilibre calculée $x_t^\star$, les assurés choisissent de résilier ou de renouveler leur contrat selon une loi multinomiale logit de vecteur de probabilité $p_{l\to j}(x_t^\star)$. Une réalisation $n_{j,t}$ de la taille de portefeuille est obtenue.

3. Ensuite, les sinistres pour chaque assuré sont tirés aléatoirement selon un modèle fréquence – sévérité.

4. Enfin, on détermine le résultat de souscription et en déduit le nouveau capital disponible $K_{j,t}$.

Le chapitre 2 analyse les propriétés statiques et dynamiques de ce jeu répété. Nous renvoyons au prochain chapitre de l'introduction pour plus de détails sur les résultats obtenus.

Théorie de la ruine

La théorie du risque s'intéresse à tous les aspects d'un portefeuille d'assurance non-vie, tarification, provisionnement, gestion du risque, etc. . . , voir, par exemple,Bowers *et al.* (1997), Marceau (2012). La théorie de la ruine se concentre sur la solvabilité à moyen et long terme d'un assureur. Nous présentons ci-dessous les grands résultats de la théorie de ruine sans preuve et renvoyons le lecteur vers les ouvrages de référence : Grandell (1991),Rolski *et al.* (1999),Asmussen (2000),Asmussen et Albrecher (2010). Nous suivrons plus particulièrement la présentation d'Asmussen et Albrecher (2010).

L'étude du niveau de richesse d'une compagnie d'assurance, introduite par Lundberg (1903), est une problématique centrale de la théorie de la ruine. Au début du XX$^{\text{ème}}$ siècle, l'école Suédoise pose les fondamentaux de cette théorie, sous l'impulsion de Filip Lunderg puis d'Harald Cramér. Cramér (1930) propose le modèle collectif (plus tard appelé le modèle de Cramér-Lundberg) dans lequel la richesse de l'assureur $(U_t)_t$ au temps t est modélisée par le processus stochastique suivant

$$U_t = u + ct - \sum_{i=1}^{N_t} X_i, \tag{12}$$

où $u > 0$ est le capital initial, $c > 0$ le taux de prime par unité de temps, $(N_t)_{t\geq 0}$ représentant le nombre de sinistres au temps t et X_i le montant du $i^{\text{ème}}$ sinistre. Notons $S_t = \sum_{i=1}^{N_t} X_i$ la perte agrégée au temps t.

Dans le modèle de Cramér-Lundberg, les hypothèses suivantes sont faites : (i) les montants des sinistres $(X_i)_i$ sont indépendants et identiquement distribués, (ii) les montants sont indépendants de N_t, (iii) $(N_t)_{t\geq 0}$ un processus de Poisson (d'intensité λ). Notons $(T_i)_i$ les temps d'attente entre deux sinistres. Pour un processus de Poisson, les temps $(T_i)_i$ sont de lois exponentielles $\mathcal{E}(\lambda)$.

Définition. *La probabilité de ruine (en temps infini) est définie comme étant le premier instant où le processus de richesse $(U_t)_t$ est strictement négatif*

$$\psi(u) = P(\exists t \geq 0, U_t < 0). \tag{13}$$

De manière similaire, on définit la probabilité de ruine en temps fini par

$$\psi(u, T) = P(\exists t \in [0, T], U_t < 0), \tag{14}$$

où $T > 0$ est l'horizon de gestion.

Un exemple de trajectoire du processus $(U_t)_t$ est donné en figure 2, où les temps d'inter-occurrence sont de loi exponentielle $\mathcal{E}(3)$, les sinistres de loi exponentielle $\mathcal{E}(2)$ et le taux de prime $c = 2$. Le capital initial u est le point de départ du processus, la pente est donnée par le taux de prime c, représentant l'acquisition des primes au cours du temps. Ensuite chaque sinistre X_i produit un saut vers le bas. Sur cet exemple, la ruine intervient au bout du 6$^{\text{ème}}$ sinistre.

Le modèle de Cramér-Lundberg a rapidement été généralisé en considérant des processus de renouvellement pour $(N_t)_{t\geq 0}$ par Andersen (1957), plus tard appelé modèle de Sparre Andersen. Ainsi, les temps d'inter-occurrence ne sont plus nécessairement de loi exponentielle mais simplement indépendants et identiquement distribués. Le lien avec la théorie des files d'attente est encore plus clair que dans le modèle de Cramér-Lundberg.

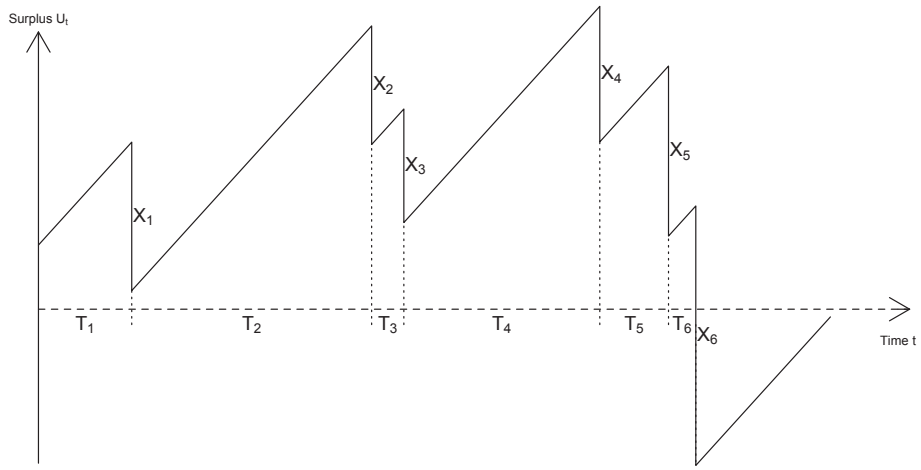


FIGURE 2 – Une trajectoire du processus $(U_t)_t$

Processus de risque à accroissements indépendants et stationnaires

La théorie de la ruine cherche à déterminer le taux de prime c et le niveau de capital u répondant à un critère de ruine. Pour éviter la ruine certaine, le taux de prime c doit déjà vérifier la contrainte dite de profit net. Celle-ci est déduite de la proposition suivante, voir la proposition IV.1.2 d'Asmussen et Albrecher (2010).

Proposition (Dérive et oscillation)**.** *Dans le modèle de Cramér-Lundberg avec des sinistres indépendants et identiquement distribués d'espérance $E(X)$, notons $\rho = c - \lambda E(X)$ le gain espéré par unité de temps. On a presque sûrement*

$$\lim_{t\to+\infty} \frac{U_t - u}{t} = \rho.$$

Si $\rho > 0$, alors presque sûrement

$$\lim_{t\to+\infty} U_t = +\infty,$$

c'est à dire la ruine n'est pas certaine, $\psi(u) < 1$. Dans le cas contraire, si $\rho < 0$, alors presque sûrement

$$\lim_{t\to+\infty} U_t = -\infty,$$

c'est à dire la ruine est certaine $\psi(u) = 1$. Dans le cas particulier, où $\rho = 0$, la ruine est aussi certaine, puisque le processus $(U_t)_t$ vérifie

$$\limsup_{t\to+\infty} U_t = +\infty, \quad et \quad \liminf_{t\to+\infty} U_t = -\infty.$$

La condition de profit net $\rho > 0$ peut se réécrire $c = (1 + \eta)\lambda E(X)$ avec un chargement $\eta > 0$. Nous pouvons maintenant énoncer la première formule fermée pour la probabilité de ruine dans le modèle de Cramér-Lundberg, voir le corollaire IV.3.2 d'Asmussen et Albrecher (2010).

Théorème. *Dans le modèle de Cramér-Lundberg avec des sinistres de loi exponentielle $\mathcal{E}(1/\mu)$ (de moyenne μ),*

$$\psi(u) = \frac{\lambda\mu}{c} e^{-u(1/\mu - \lambda/c)},$$

si la condition de profit net est vérifiée $\rho = c - \lambda\mu > 0$.

Cette formule de probabilité de ruine a la caractéristique de décroitre exponentiellement en fonction du capital initial u. Cette propriété est vérifiée pour une large classe de modèles de sinistres. Plus précisément, la décroissance exponentielle est encore valide si la loi des sinistres $(X_i)_i$ possède une fonction génératice des moments $M_X(t) = E(e^{tX})$ pour $t > 0$. Nous regroupons ici les théorèmes IV.5.2 et IV.5.3 d'Asmussen et Albrecher (2010).

Théorème (Borne et approximation de Cramér-Lundberg)**.** *Soit γ la solution strictement positive de l'équation (en r) de Lundberg*

$$M_X(r)\frac{\lambda}{\lambda + rc} = 1.$$

On a alors pour tout $u \geq 0$

$$\psi(u) \leq e^{-\gamma u} \quad et \quad \psi(u) \underset{u\to+\infty}{\sim} Ce^{-\gamma u},$$

où la constante C est donnée par $C = (c - \lambda\mu)/(\lambda M_X'(\gamma) - c)$, $e^{-\gamma u}$ est appelée borne de Lundberg, $Ce^{-\gamma u}$ approximation de Lundberg et γ coefficient d'ajustement.

La décroissance exponentielle est aussi constatée dans le modèle de Sparre Andersen où le processus du nombre de sinistres $(N_t)_{t\geq 0}$ (de l'équation (12)) est un processus de renouvellement. Les temps d'inter-occurrence des sinistres $(T_i)_i$ sont indépendants et identiquement distribués selon une variable générique T. Nous rapportons ci-dessous une extension du théorème de Cramér-Lundberg pour le modèle de Sparre Andersen, voir, par exemple, théorème 6.5.4 de Rolski *et al.* (1999).

Théorème. *Dans le modèle de Sparre Andersen, notons Y les incréments de la perte agrégée $Y = X - cT$. x_0 est défini comme le supremum de l'ensemble $\{x, F_Y(x) < 1\}$. Pour $u \geq 0$, nous disposons de l'encadrement suivant*

$$b_- e^{-\gamma u} \leq \psi(u) \leq b_+ e^{-\gamma u},$$

où γ est solution de l'équation $M_X(r)M_T(-rc) = 1$, les constantes b_-, b_+ ont pour expression

$$b_- = \inf_{x\in[0,x_0[} \frac{e^{\gamma x}\bar{F}_Y(x)}{\int_x^{+\infty} e^{\gamma y} d\bar{F}_Y(y)} \quad et \quad b_+ = \sup_{x\in[0,x_0[} \frac{e^{\gamma x}\bar{F}_Y(x)}{\int_x^{+\infty} e^{\gamma y} d\bar{F}_Y(y)}.$$

En plus de ces asymptotiques, d'autres formules explicites de probabilité de ruine sont disponibles pour d'autre lois de sinistres, notamment les mélanges de lois exponentielles, les lois Erlang (c'est à dire loi gamma avec un paramètre de forme entier), les mélanges de lois Erlang. Ces lois font partie de la grande classe des lois phase-type introduite par Neuts (1975), et popularisée dans la théorie des files d'attente par notamment Neuts (1981).

Soit $m \in \mathbb{N}^\star$ un entier positif. Considérons un processus de Markov $(M_t)_t$ en temps continu et à valeurs dans l'ensemble fini $\{0, 1, \ldots, m\}$, où 0 est un état absorbant. Une loi phase-type

est la loi du temps d'absorption du processus $(M_t)_t$ dans l'état 0 partant d'un état initial de l'ensemble $\{1, \ldots, m\}$.

Ces lois sont paramétrées par une matrice de sous-intensité J, une dimension m et un vecteur de probabilité initial $\pi \in [0,1]^m$. La matrice d'intensité Λ du processus sous-jacent $(M_t)_t$ est donnée par la matrice par bloc

$$\Lambda = (\lambda_{ij})_{ij} = \begin{pmatrix} 0 & 0 \\ j_0 & J \end{pmatrix},$$

où j_0 est le vecteur des intensités de sortie $j_0 = -J\mathbf{1}_m$ et $\mathbf{1}_m$ le vecteur rempli de 1 de $\mathbb{R}^m$. Cela signifie que les probabilités de changement d'état du processus $(M_t)_t$ sont donnés par $P(M_{t+h} = j / M_t = i) = \lambda_{ij}h + o(h)$ si $i \neq j$ et $1 + \lambda_{ii}h + o(h)$ si $i = j$ avec des probabilités initiales $P(M_0 = i) = \pi_i$.

Pour de telles lois phase-type $PH(\pi, J, m)$, les fonctions de répartition et de densité sont données par

$$F(x) = 1 - \pi e^{Jx}\mathbf{1}_m, \quad \text{et} \quad f(x) = \pi e^{Jx} j_0,$$

où e^{Jx} correspond à l'exponentielle de matrice définie la série $\sum_{n=0}^{+\infty} \frac{T^n x^n}{n!}$, voir Moler et Van Loan (2003) pour une revue récente de son calcul. La loi exponentielle $\mathcal{E}(\lambda)$ est obtenue par la paramétrisation $PH(1, \lambda, 1)$, le mélange de n lois exponentielles est obtenue par $PH(m, \pi, J)$ où $m = n$, $\pi = (p_1, \ldots, p_n)$, et une matrice de sous-intensité diagonale $J = -\text{diag}[(\lambda_1, \ldots, \lambda_n)]$.

Les lois phase-type $PH(\pi, J, m)$ font partie des lois à queue de distribution légère, au même titre que la loi exponentielle, la loi gamma, au vue de la décroissance exponentielle de sa queue de distribution. Ainsi, la fonction génératrice des moments et le moment d'ordre n possèdent des formules explicites. Dans ce contexte, Asmussen et Rolski (1991) proposent des formules explicite de probabilité de ruine lorsque les montants des sinistres $(X_i)_i$ et les temps d'inter-occurrence $(T_i)_i$ sont de lois phase type.

Théorème (Asmussen et Rolski (1991))**.** *Dans le modèle de Cramér-Lundberg, lorsque les montants des sinistres sont phase-type $PH(\pi, J, m)$, la probabilité de ruine est donnée par*

$$\psi(u) = \pi_+ e^{Qu}\mathbf{1}_m,$$

où la matrice s'écrit $Q = J + j_0\pi_+$, le vecteur $\pi_+ = -\lambda/c\pi J^{-1}$ et $j_0 = -J\mathbf{1}_m$ est le vecteur des taux de sortie. En d'autre termes, la probabilité de ruine admet une représentation phase-type $PH(\pi_+, Q, m)$.

Dans le modèle de Sparre Andersen, lorsque les temps d'inter-occurrence ont une fonction de répartition F_T, la probabilité de ruine admet toujours une représentation phase-type $PH(\pi_+, Q, m)$ mais π_+ est la solution de l'équation de point fixe

$$\pi_+ = \pi M_T(J + j_0\pi_+),$$

où M_T correspond la fonction génératrice des moments (avec un argument matriciel).

Une loi de probabilité admet une représentation phase-type s'il existe une fonction génératrice des moments rationnelle, voir, par exemple, Hipp (2005). Nécessairement, une loi phase-type a une queue de distribution légère. Ainsi, le théorème montre pour la grande classe des lois phase-type* de montant de sinistre à queue distribution légère, que la probabilité de

∗. Comme l'ensemble des lois phase-type est dense dans l'ensemble des lois de probabilité à support positif, il est possible en théorie approcher à n'importe quel lois à support positif pour un degré de précision donné.

ruine décroît exponentiellement vite. Une question légitime est donc de savoir si ce principe est- toujours respecté pour des lois à queue de distribution plus épaisse.

Jusqu'ici les lois de sinistre X étaient telles que la fonction génératrice des moments $M_X(t) = \int_0^{+\infty} e^{tx} d\bar{F}_X(x)$ existait pour certains $t > 0$. Cette classe est appelée la classe des lois à queue de distribution légère. De nombreuses lois n'appartiennent pas à cette classe, c'est à dire $M_X(t)$ est infini pour $t > 0$, par exemple, la loi lognormale, la loi de Weibull ou encore la loi de Pareto.

Comme la classe des lois pour lesquelles il n'existe pas de lois de fonction de génératrice de moments est vaste et peu explicite, la classe de lois sous-exponentielles a été introduite. Une fonction de répartition F_X appartient à la famille sous-exponentielle si pour deux variables aléatoires X_1, X_2 indépendantes et identiquement distribuées de fonction de répartition F_X, elles vérifient

$$\frac{P(X_1 + X_2 > x)}{P(X_1 > x)} \underset{x \to +\infty}{\longrightarrow} 2.$$

Pour mieux comprendre cette définition, il est de bon rappeler que pour toute variable aléatoire positive X, on a $P(\max(X_1, X_2) > x) \sim 2\bar{F}_X(x)$ lorsque $x \to +\infty$. Par conséquent, la classe des lois sous-exponentielle est telle que $P(X_1 + X_2 > x) \sim P(\max(X_1, X_2) > x)$ pour des grandes valeurs de x. La propriété se généralise pour une somme de n variables indépendantes.

En théorie de la ruine, l'application des lois sous-exponentielle a été faite par Teugels et Veraverbeke (1973) ou encore Embrechts et Veraverbeke (1982). Pour cette classe de montant de sinistre, la probabilité de ruine décroît comme l'inverse d'un polynome. Nous rapportons ci-dessous la version proposée dans Asmussen et Albrecher (2010).

Théorème (Embrechts et Veraverbeke (1982))**.** *Dans le modèle de Sparre Andersen, où les espérances des montants $(X_i)_i$ et des temps d'attente des sinistres $(T_i)_i$ sont finis et tels que $E(X) < cE(T)$. Notons $F_{X,0}(x) = \int_0^x \overline{F}_X(y)dy/E(X)$. Si F_X et $F_{X,0}$ appartiennent à la classe sous-exponentielle, on a alors*

$$\psi(u) \underset{u \to +\infty}{\sim} \frac{1}{cE(T) - E(X)} \int_u^{+\infty} \bar{F}_X(y)dy.$$

Ce théorème donne lieu aux cas particuliers suivant. Considérons des montants de sinistre Pareto $\mathcal{P}a(k, \alpha)$, c'est à dire $P(X > x) = (k/x)^\alpha$ avec $\alpha > 1$. On a alors

$$\psi(u) \underset{u \to +\infty}{\sim} \frac{k}{cE(T)(\alpha - 1) - \alpha k} \left(\frac{k}{u}\right)^{\alpha - 1}.$$

De manière similaire, pour les lois à variations régulières dont les queues de distribution vérifient $P(X > x) \sim L(x)/x^\alpha$ pour des grandes valeurs de x et L une fonction à variation lente, telle que $L(xt)/L(x) \to 1$ pour $t > 0$ et $x \to +\infty$, nous obtenons

$$\psi(u) \underset{u \to +\infty}{\sim} \frac{1}{cE(T) - E(X)} \times \frac{L(u)}{(\alpha - 1)u^{\alpha - 1}}.$$

Lorsque X suit une loi de Weibull avec $P(X > x) = \exp(-x^\beta)$ (resp. une loi lognormale $P(X > x) = 1 - \Phi((\log x - \mu)/\sigma)$), alors on a $\psi(u) \sim u^{1-\beta} e^{-u^\beta}$ (resp. $\psi(u) \sim u e^{-\log^2(u)}/\log^2(u)$). Toutes ces formules, sauf celle pour le cas Weibull, présentent une décroissance en puissance de u du type C/u^α, qui contraste nettement avec une décroissance exponentielle pour les lois à queue légère $Ce^{-\gamma u}$.

Jusqu'à maintenant, le seul indicateur de risque introduit est la probabilité de ruine. De nombreuses autres quantités sont intéressantes à étudier. Notons $\tau = \inf(t > 0, u+Ct-S_t < 0)$ le (premier) temps de ruine pour le processus de richesse $(U_t)_t$ de valeur initiale u. La valeur du niveau de richesse juste avant la ruine et le déficit au moment de la ruine, respectivement U_{τ^-} et $|U_\tau|$, sont des exemples de mesures de ruine, voir Dickson (1992), Lin et Willmot (2000).

Gerber et Shiu (1998) proposent un cadre unique d'étude pour ces mesures de ruine par la fonction de pénalité suivante (plus tard appelée fonction de Gerber-Shiu)

$$m_\delta(u) = E\left[e^{-\delta\tau} w(U_{\tau^-}, |U_\tau|)\mathbb{1}_{(\tau<+\infty)} | U_0 = u\right],$$

où δ est le taux d'intérêt et w une fonction de pénalité positive. Pour $w(x, y) = 1$ et $\delta = 0$, la fonction de Gerber-Shiu $m_\delta(u)$ se réduit à la probabilité de ruine en temps infini. Notons que δ peut jouer le role d'une variable pour la transformée de Laplace du terme $w(U_{\tau^-}, |U_\tau|)$ conditionnelement à un temps de ruine fini.

Dans le modèle de Cramér-Lundberg, la fonction de Gerber-Shiu vérifie l'équation intégrale suivante

$$m_\delta = m_\delta * g + h,$$

avec $*$ le produit de convolution et g, h données par

$$g(x) = \frac{\lambda}{c}\int_0^{+\infty} e^{-\gamma z} f_X(x+z)dz \quad \text{et} \quad h(x) = \frac{\lambda}{c}\int_x^{+\infty} e^{-\gamma(z-x)} \int_z^{+\infty} w(z, x-z) f_X(x)dxdz,$$

où γ est la solution positive de l'équation de Lundberg $\delta + \lambda - c\xi = \lambda\widehat{f_X}(\xi)$ et $\widehat{f_X}$ est la transformée de Laplace de la densité f_X. En prenant la transformée de Laplace de l'équation précédente, l'équation se résout facilement par $\widehat{m}_\delta(\xi) = \widehat{h}(\xi)/(1-\widehat{g}(\xi))$. Avec cette formulation, il est possible d'obtenir des asymptotiques de la fonction de pénalité pour des grandes valeurs de u. Gerber et Shiu (1998) ont introduit cette fonction de pénalité dans le modèle de Cramér-Lundberg, mais ensuite l'approache a été généralisée dans d'autres modèles, voir, par exemple, Gerber et Shiu (2005) et Song *et al.* (2010).

Avant de présenter des modèles de risque avec dépendance, nous listons d'autres extensions possibles du modèle de Sparre Andersen : intégration de la réassurance, voir, par exemple, Centeno (2002a),Centeno (2002b),Barges *et al.* (2012), modélisation de plusieurs de lignes d'affaire, par exemple, Collamore (1996),Cai et Li (2005),Biard *et al.* (2010), ou encore la modélisation des crises de corrélation, Biard *et al.* (2008).

Processus de risque avec dépendance

Jusqu'à présent, nous avons travaillé avec trois hypothèses d'indépendance : (i) une indépendance entre les montants des sinistres $(X_1, X_2, \dots)$, (ii) une indépendance entre les temps d'attente $(T_1, T_2, \dots)$ et (iii) une indépendance entre le montant X_i et le temps d'attente T_i du $i^{\text{ème}}$ sinistre. Cela résulte du fait que le processus d'arrivée des sinistres $(N_t)_{t\geq 0}$ est supposé à acroissements indépendants et stationnaires, puisque c'est un processus de renouvellement.

Une première extension consiste à considérer un processus de Poisson non homogène $(N_t)_{t\geq 0}$ de paramètre $(\lambda_t)_t$ pour le processus de risque de l'équation (12). Lu et Garrido (2005) ont montré l'intérêt de tels processus dans la modélisation des ouragans aux Etats-Unis, où deux fonctions déterministes pour λ_t sont testées : une fonction simplement périodique et une doublement périodique.

Albrecher et Asmussen (2006) utilisent les processus de Poisson non-homogènes où le paramètre d'intensité $(\lambda_t)_t$ est aussi un processus stochastique. Plus précisement, il suppose que le processus d'arrivée des sinistres $(N_t)_{t\geq 0}$ a pour paramètre d'intensité le processus stochastique suivant

$$\lambda_t = \lambda + \sum_{n\in\mathbb{N}} h(t - U_n, Y_n) + \nu_t, \tag{15}$$

où $\lambda > 0$ une constante, $(U_n)_n$ sont les temps d'occurrence d'un processus de Poisson de paramètre ρ, $(Y_n)_n$ une suite de variables aléatoires positives indépendantes et identiquement distribuées, $h(.,.)$ une fonction positive et $(\nu_t)_t$ est un processus stochastique représentant les pertubations du passé.

Albrecher et Asmussen (2006) obtiennent différents résultats pour la probabilité de ruine en temps fini et infini avec des sinistres à queue de distribution lourde et légère. Nous donnons ici que deux de leurs résultats (théorèmes 4.2 et 5.2) et renvoyons le lecteur vers leur article pour plus de détails.

Théorème (Albrecher et Asmussen (2006))**.** *Considérons le processus de risque de l'équation (12) où le processus d'arrivée $(N_t)_{t\geq 0}$ a pour paramètre d'intensité le processus de l'équation (15). Nous supposons que la condition de profit net est vérifiée par $c > E(X)\mu$, où $\mu = \lambda + \rho E(H(\infty, Y))$ et $H(t, y) = \int_0^t h(s, y)ds$.*

Supposons que les montants des sinistres $(X_i)_i$ possèdent une fonction génératrice des moments $M_X(\alpha)$ pour $\alpha > 0$ proche de 0, que le processus $(\nu_t)_t$ vérifie

$$\frac{\log E\left(\exp((M_X(\alpha) - 1)\int_0^t \nu_s ds)\right)}{t} \to 0,$$

lorsque t tend vers $+\infty$, et que $E(\exp(\alpha H(\infty, Y)))$ existe pour $\alpha > 0$. On a alors

$$\psi(u) \underset{u\to+\infty}{\sim} e^{-\gamma u},$$

où γ est la solution positive d'une certaine équation $\kappa(\alpha) = 0$ avec κ une fonction de $M_X(\alpha)$.

Si les montants des sinistres $(X_i)_i$ appartiennent à une classe sous-exponentielle et que pour $\alpha > 0$, $E(\exp(\alpha H(\infty, Y)))$ et $E(\exp(\alpha \int_0^t \nu_s ds))$ existent, alors on a

$$\psi(u) \underset{u\to+\infty}{\sim} \frac{\mu}{(c-\mu)E(X)} \int_u^{+\infty} \bar{F}_X(x)dx.$$

Malgré l'augmentation de la variabilité sur le processus de risque et bien que la probabilité de ruine ait augmentée, la forme des asymptotiques demeurent inchangée par rapport au modèle de Sparre Andersen.

Dans le même esprit, Asmussen (1989), Asmussen et Rolski (1991), Asmussen *et al.* (1994) considèrent un processus d'arrivée des sinistres controlé par un processus Markovien $(J_t)_t$ à valeurs finies. Conditionnellement à $J_t = i$, le taux de prime est c_i, la loi des sinistres est $F_{i,X}$, et le taux d'arrivées de sinistre λ_i. Cela correspond aussi un processus de Poisson doublement stochastique de processus d'intensité $(\lambda_{J_t})_t$.

Notons $\psi_i(u)$ la probabilité de ruine sachant $J_0 = i$. Asmussen (1989),Asmussen et Rolski (1991) fournissent des formules exactes (à l'aide des lois phase-type) et des asymptotiques dans le cas de lois de sinistres à queue de distribution légère, tandis que Asmussen *et al.* (1994)

traitent le cas des lois à queue lourde. Les asymptotiques dans le cas des lois à queue légère sont du type

$$\psi_i(u) \underset{u\to+\infty}{\sim} C_i e^{-\gamma u},$$

tandis que pour les lois de la classe sous-exponentielle avec $F_{i,X} = F_X$

$$\psi_i(u) \underset{u\to+\infty}{\sim} a_i \int_u^{+\infty} \bar{F}_X(x)dx.$$

Nous renvoyons le lecteur vers les articles pour les expressions des constantes a_i et C_i.

Les extensions précédentes au modèle de Sparre-Andersen se sont portées sur la modification du processus de sinistres $(N_t)_{t\geq 0}$. Nous présentons maintenant les extensions où on suppose explicitement une dépendance entre le montant X_i et le temps d'attente T_i du $i^{\text{ème}}$ sinistre. Albrecher et Boxma (2004) est un premier exemple où la densité du temps d'attente du $i+1^{\text{ème}}$ sinistre dépend du montant du $i^{\text{ème}}$ sinistre. Ils supposent que

$$f_{T_{i+1}}(x) = P(X_i > \tau_i)\lambda_1 e^{-\lambda_1 x} + P(X_i \leq \tau_i)\lambda_2 e^{-\lambda_2 x},$$

où τ_i est une variable aléatoire représentant un seuil de gravité modifiant le temps d'attente du sinistre. En d'autres termes, T_i est un mélange de loi exponentielle $\mathcal{E}(\lambda_1)$, $\mathcal{E}(\lambda_2)$ dont la probabilité de mélange est $P(X_i > \tau_i)$. Les variables de seuil (τ_i) forment une suite de variables aléatoires indépendantes et identiquement distribuées. Le processus de risque considéré est maintenant un processus Markovien puisque le temps d'attente du $i+1^{\text{ème}}$ sinistre dépend (uniquement) du montant du $i^{\text{ème}}$ sinistre, c'est à dire les incréments $cT_i - X_i$ ne sont plus stationnaires ou indépendents.

La condition de profit net est $E(X) < c(P(X > \tau)/\lambda_1 + P(X \leq \tau)/\lambda_2)$, où X, τ sont les variables génériques. Dans le cas d'une loi de sinistre à queue de distribution légère et si le premier temps d'attente est de loi $\mathcal{E}(\lambda_j)$, nous pouvons obtenir une expression explicite de la transformée de Laplace de la probabilité de survie conditionnelle $1-\psi_j(u)$ sous la forme d'un ratio de fonctions. Une inversion numérique de la transformée de Laplace est possible si la transformée de Laplace est rationnelle. Le fait que la transformée de Laplace ait une unique solution à partie réelle positive garantit une décroissance exponentielle de la probabilité de ruine du type $e^{-\sigma u}$.

Dans la même idée, Boudreault *et al.* (2006) considèrent une structure de dépendance dans laquelle les incréments $cT_i - X_i$ sont toujours indépendants et stationnaires. Mais les variables (X_i, T_i) ne sont plus indépendantes : ils supposent que

$$f_{X_i}^{T_i}(x) = e^{-\beta T_i} f_1(x) + (1 - e^{-\beta T_i}) f_2(x),$$

où f_1, f_2 sont deux densités.

Toujours en travaillant avec des lois de sinistre à queue légère, Boudreault *et al.* (2006) expriment explicitement la transformée de Laplace en termes de ratios. Ils obtiennent une expression explicite pour la fonction de Gerber-Shiu sous forme de combinaisons exponentielles $e^{R_i u}$ où R_i correspondent aux racines du dénominateurs de la transformée de Laplace.

Dépendance à l'aide des copules

Albrecher et Teugels (2006) s'intéressent aussi à une modélisation directe du couple (X_i, T_i) et à l'impact sur la probabilité de ruine en temps fini et infini. La dépendance est modélisée à

l'aide d'une copule bivariée $(u_1, u_2) \mapsto C(u_1, u_2)$. Pour une dimension d fixée, une copule est une fonction multivariée $C : [0,1]^d \mapsto [0,1]$ vérifiant certaines propriétés de manière à ce que C puisse être interprétée comme une fonction de répartition d'un vecteur aléatoire $(U_1, \ldots, U_d)$ à marginale uniforme.

Définition. *Soit C une fonction multivariée de $[0,1]^d \mapsto [0,1]$, où $d \geq 2$ est une dimension fixée. C est une copule si la fonction vérifie les propriétés suivantes :*

1. $\forall i \in \{1, \ldots, d\}, \forall \boldsymbol{u} \in [0,1]^d, C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0$,
2. $\forall i \in \{1, \ldots, d\}, \forall \boldsymbol{u} \in [0,1]^d, C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$,
3. $\forall i \in \{1, \ldots, d\}, \forall \boldsymbol{u} \in [0,1]^d, \forall (a_i \leq b_i)_i, \Delta^1_{a_1,b_1} \ldots \Delta^d_{a_d,b_d} C(\boldsymbol{u}) \geq 0$, *où* $\Delta^i_{a_i,b_i}$ *est la différence d'ordre i, c'est à dire*

$$\Delta^i_{a_i,b_i} C(\boldsymbol{u}) = C(u_1, \ldots, u_{i-1}, b_i, u_{i+1}, \ldots, u_d) - C(u_1, \ldots, u_{i-1}, a_i, u_{i+1}, \ldots, u_d).$$

La propriété 3 est appelée croissance d'ordre d.

L'interprétation probabiliste d'une copule est la suivante

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d),$$

où $(U_i)_i$ sont des variables aléatoires uniformes $\mathcal{U}(0,1)$. Pour toute copule C, on a l'inégalité suivante

$$W(\boldsymbol{u}) = \max(u_1 + \cdots + u_d - (d-1), 0) \leq C(u_1, \ldots, u_d) \leq \min(u_1, \ldots, u_d) = M(\boldsymbol{u}), \quad (16)$$

où les bornes sont appelées bornes de Fréchet W, M. Les inégalités (16) donnent des bornes inférieure et supérieure quelque soit la structure de dépendance considérée. Notons que M est toujours une copule quelque soit la dimension d, tandis que W ne l'est qu'en dimension $d = 2$. Une dernière copule particulière, qui a toute son importance, est la copule d'indépendance $\Pi(\boldsymbol{u}) = u_1 \times \cdots \times u_d$. Cette copule sera une copule limite de la plupart des copules paramétriques.

Un théorème fondamental liant un vecteur aléatoire avec des fonctions de répartition marginales données à une copule est le théorème de Sklar (1959).

Théorème (Sklar (1959))**.** *Soit $F : \mathbb{R}^d \mapsto [0,1]$ une fonction de répartition multivariée avec des marginales $F_1, \ldots, F_d$, alors il existe une copule C telle que pour tout $(x_1, \ldots, x_d) \in \mathbb{R}^d$, $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$. C est unique sur l'ensemble $S_1 \times \cdots \times S_d$ où S_i est le support de la $i^{ème}$ marginale.*

Notons que si les variables aléatoires marginales sont des variables continues, alors la copule est unique sur $\mathbb{R}^d$. Sinon elle n'est unique qu'en certains points. En supposant que C est différentiable sur $[0,1]^d$, la densité de la loi jointe d'un vecteur aléatoire avec des marginales de densité f_i est donnée par

$$c(x_1, \ldots, x_d) = \frac{\partial^d C(F_1(x_1), \ldots, F_d(x_d))}{\partial u_1 \ldots \partial u_d} f_1(x_1) \ldots f_d(x_d).$$

Nous nous arrêtons ici pour la présentation des copules et renvoyons le lecteur vers les ouvrages de référence : Marshall (1996),Joe (1997),Nelsen (2006).

Retournons à notre problème de ruine. Albrecher et Teugels (2006) montrent le théorème suivant à l'aide de la marche aléatoire sous-jacente au processus de risques.

Théorème (Albrecher et Teugels (2006))**.** *Pour le processus de risque donné en équation (12), avec des montants X_i et des temps d'attente de sinistre T_i identiquement distribués et des queues de distribution légères. Notons γ le coefficient d'ajustement solution positive de l'équation $E(\exp r(X - cT)) = 1$, existant si la condition de profit net $E(X) < cE(T)$ est vérifiée. On a alors*

$$\psi(u) \underset{u\to+\infty}{\sim} Ce^{-\gamma u},$$

où $C = e^{-B}/\gamma E(S_N \exp(RS_N))$, $N = \inf(n > 0, S_n > 0)$, S_n la somme cumulée des incréments et B une constante positive.

Albrecher et Teugels (2006) étudient ensuite l'équation de Lundberg pour différentes copules : copules de Spearman, copule EFGM ou les copules Archimédiennes. Nous détaillons ces dernières, qui vont être utilisées dans cette introduction et au chapitre 4.

Les copules archimédiennes sont caractérisées par un générateur $\phi : \mathbb{R}_+ \mapsto [0, 1]$, qui est une fonction infinement différentiable et complètement monotone, c'est à dire, pour tout $k \in \mathbb{N}, (-1)^k \phi^{(k)}(t) \geq 0$. Une copule archimédienne est caractérisée de la manière suivante

$$C(u_1, \dots, u_d) = \phi^{-1}\left(\phi(u_1) + \cdots + \phi(u_d)\right).$$

Les exemples les plus classiques sont la copule de Clayton $\phi(t) = (t^{-\alpha} - 1)/\alpha$, la copule de Gumbel $\phi(t) = (-\log t)^\alpha$ ou encore la copule de Frank $\phi(t) = \log(e^{-\alpha} - 1) - \log(e^{-\alpha t} - 1)$, voir le chapitre 4 de Nelsen (2006).

Dépendance par mélange

Enfin, nous présentons un dernier type de dépendance introduit par Albrecher *et al.* (2011). Cela consiste à introduire une dépendance dans la suite des sinistres, soit sur les montants $(X_1, X_2, \dots)$ ou les temps d'attente de sinistres $(T_1, T_2, \dots)$ à l'aide d'une variable latente. Soit Θ une variable aléatoire positive représentant une certaine hétérogénéité sur le portefeuille d'assurance. Tout d'abord, nous supposons que les montants de sinistre X_i sont indépendants et identiquement distribués conditionnellement à $\Theta = \theta$ de loi exponentielle $\mathcal{E}(\theta)$. De plus, les temps d'attente sont eux-aussi de loi exponentielle λ et indépendents des montants de sinistre.

Conditionnellement à $\Theta = \theta$, c'est le modèle de Cramér-Lundberg. Ainsi, on a

$$\psi(u, \theta) = \min\left(\frac{\lambda}{\theta c} e^{-u(\theta - \frac{\lambda}{c})}, 1\right),$$

où le minimum utilisé ci-dessus est équivalent à la condition de profit net $\theta > \lambda/c$. En intégrant par rapport à la variable θ, on obtient

$$\psi(u) = F_\Theta(\theta_0) + I(u, \theta_0),$$

où $\theta_0 = \lambda/c$ et

$$I(u, \theta_0) = \int_{\theta_0}^{+\infty} \frac{\theta_0}{\theta} e^{-u(\theta - \theta_0)} dF_\Theta(\theta).$$

Notons dès à présent que la probabilité de ruine est strictement positive quelque soit le niveau de capital initial u.

Dans un tel modèle, une copule archimédienne se cache entre les montants des sinistres. En effet, la probabilité de survie est donnée par

$$P(X_1 > x_1, \ldots, X_n > x_n | \Theta = \theta) = \prod_{i=1}^{n} e^{-\theta x_i}.$$

Par conséquent,

$$P(X_1 > \bar{F}_X^{-1}(u_1), \ldots, X_n > \bar{F}_X^{-1}(u_n)) = \int_0^{+\infty} e^{-\theta \sum_{i=1}^{n} \bar{F}_X^{-1}(u_i)} dF_\Theta(\theta) = L_\Theta \left(\sum_{i=1}^{n} \bar{F}_X^{-1}(u_i) \right),$$

où L_Θ est la transformée de Laplace de la variable aléatoire Θ et $\bar{F}_X$ la fonction de survie de X. Il est facile d'identifier une structure archimédienne avec un générateur $\phi(t) = L_\Theta^{-1}(t)$. Néanmoins, la dépendance a lieu sur la copule de survie, c'est à dire sur la fonction $\overline{C}(\boldsymbol{u})$ définie par

$$P(U_1 > u_1, \ldots, U_n > u_n) = \overline{C}(u_1, \ldots, u_n).$$

Albrecher *et al.* (2011) donnent trois résultats explicites de probabilités de ruine pour trois lois. Par exemple, lorsque Θ est une loi gamma $\mathcal{G}a(\alpha, \lambda)$, alors on a

$$\psi(u) = \frac{\gamma(\alpha - 1, \theta_0 \lambda)}{\Gamma(\alpha)} + \frac{\lambda^\alpha \theta_0}{\Gamma(\alpha)} e^{\theta_0 u} \times \frac{\Gamma(\alpha - 1, \theta_0(\lambda + u))}{(\lambda + u)^{\alpha - 1}},$$

où $\Gamma(.,.)$ est la fonction gamma incomplète supérieure. En utilisant un développement asymptotique de cette fonction, voir par exemple Olver *et al.* (2010), on obtient

$$\psi(u) = \frac{\gamma(\alpha - 1, \theta_0 \lambda)}{\Gamma(\alpha)} + \frac{\lambda^\alpha \theta_0}{\Gamma(\alpha)} e^{\lambda \theta_0} \left(\frac{1}{\lambda + u} + o\left(\frac{1}{u}\right) \right).$$

Cette formule laisse entrevoir une nouvelle forme d'asymptotique du type $A + B/u + o(1/u)$. En effet, le chapitre 4 va montrer que cette forme asymptotique est valable quelque soit la loi pour Θ. Ce chapitre traitera en profondeur les asymptotiques de la probabilité de ruine $\psi(u)$ lié à ce modèle de dépendance.

Le modèle de risque en temps discret

Nous considérons une version discrète du processus de risque présenté jusqu'à présent. Le modèle de ruine en temps discret, introduit par Gerber (1988), suppose que les primes, les sinistres et les arrivées sont à valeurs discrètes. Par changement de l'échelle de temps, nous supposons généralement que les primes sont normées à 1. Le processus de risque est le suivant

$$U_t = u + t - \sum_{i=1}^{t} X_i,$$

où les montants des sinistres sont à valeurs entières et la condition de profit net $E(X) < 1$. Le modèle le plus simple considère des montants de sinistres indépendants et identiquement distribués.

Gerber (1988) considère la ruine comme le premier instant t où le processus $(U_t)_t$ atteint 0. C'est à dire

$$\psi_G(u) = P(\inf(t \in \mathbb{N}^+, U_t \leq 0) < +\infty | U_0 = u).$$

A l'opposé, Shiu (1989) considère la ruine comme le premier instant t où le processus $(U_t)_t$ devient strictement négatif

$$\psi_S(u) = P(\inf(t \in \mathbb{N}^+, U_t < 0) < +\infty | U_0 = u).$$

Géométriquement parlant, ψ_G regarde le premier temps d'atteinte de l'axe des abcissses, tandis que ψ_S considère le premier temps d'atteinte de la droite horizontale $y = -1$. Nous pouvons facilement passer d'une définition à l'autre en changeant le capital initial, en utilisant la relation $\psi_G(u) = \psi_S(u-1)$.

Pour le différencier de sa version continue, le processus de risque discret est généralement affichée en deux composantes : les primes $u+t$ et la perte agrégée $\sum_{i=1}^{t} X_i$. Sur la figure 3, les primes sont traçées en pointillées, les sinistres en trait plein et les sinistres non-nuls sont notés en lettre. Sur cette trajectoire, selon la définition de Shiu, la ruine intervient en $t = 14$, tandis que selon la définition de Gerber, elle a lieu en $t = 13$.

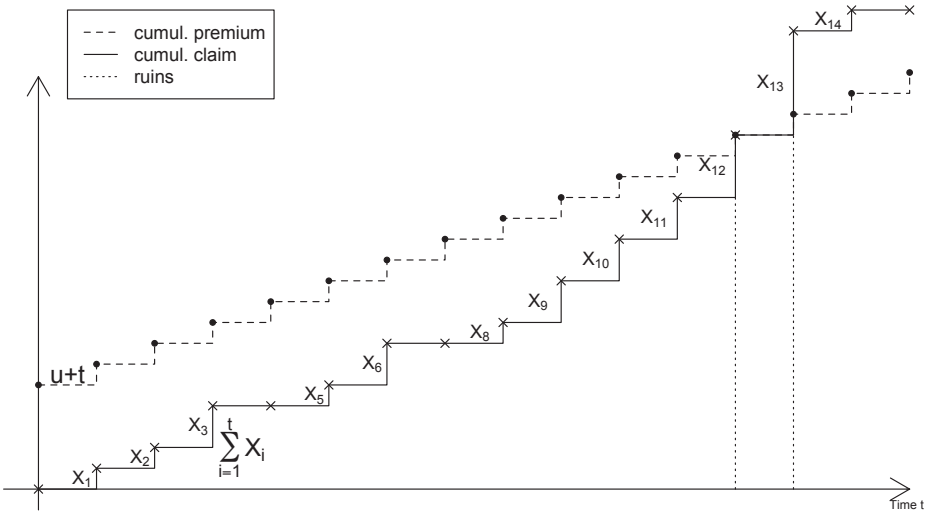


FIGURE 3 – Une trajectoire du processus $(U_t)_t$

Dans le cas de sinistre géométrique, $P(X = k) = p(1-p)^k$ pour $k \in \mathbb{N}^*$, la probabilité de ruine peut s'obtenir facilement. Nous présentons ici une version améliorée de Sundt et dos Reis (2007) utilisant la loi géométrique zéro et un modifiés. La loi géométrique zéro modifiée a pour fonction de masse de probabilité

$$P(X = k) = q\delta_{0,k} + (1-q)(1-\delta_{0,k})\rho(1-\rho)^{k-1},$$

où $\delta_{i,j}$ est le produit de Kronecker. En prenant $q = \rho$, nous retrouvons la loi géométrique simple. Quant à la version zéro et un modifée de la loi géométrique, les probabilités élémentaires sont

$$P(X = 0) = q = 1 - p \quad \text{et} \quad P(X = k/X > 0) = \rho\delta_{k,1} + (1-\rho)(1-\alpha)\alpha^{k-2}(1-\delta_{k,1}).$$

$*$. Certains auteurs présentent une version zéro tronquée $p(1-p)^{k-1}$ pour $k \geq 1$.

En prenant $\alpha = 1 - \rho$, on retrouve la loi zéro modifiée. Les moments sont donnés par

$$E(X) = p\left(1 + \frac{1-\rho}{1-\alpha}\right) \quad \text{et} \quad Var(X) = q\rho + (1-q)(1-\rho)\frac{3-\alpha}{(1-\alpha)^2} - p^2\left(1 + \frac{1-\rho}{1-\alpha}\right)^2.$$

Pour cette dernière loi, le paramètre α contrôle la queue de distribution, tandis que les paramètres q, ρ fixent la probabilité en 0 et 1 respectivement. Sur la figure 4, nous avons tracé un exemple pour chacune des lois.

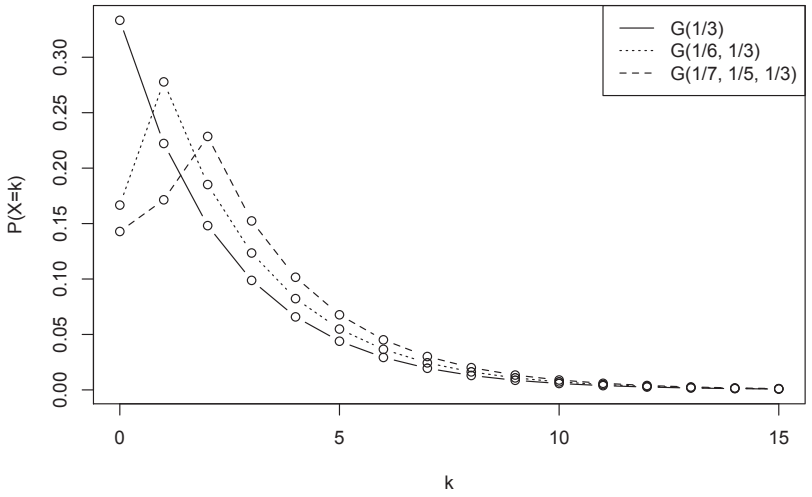


FIGURE 4 – Loi géométrique simple et modifiée

Le raisonnement de Sundt et dos Reis (2007) consiste à imposer une forme pour la probabilité de ruine et ensuite de déduire la loi de sinistre correspondante. Ils supposent $\psi_S(u) = kw^{-u}$ et trouvent que les sinistres doivent être de loi géométrique zéro et un modifés. Ils utilisent l'équation de récurrence suivante obtenue à partir de la définition de la probabilité de ruine en conditionnant par rapport au premier sinistre X_1

$$\psi_S(u) = P(X_1 > u+1) + \sum_{x=0}^{u+1} P(X_1 = x)\psi_S(u+1-x),$$

qui peut être réécrite comme

$$\psi_S(u) = p\bar{F}(u+1) + q\psi_S(u+1) + p\sum_{x=1}^{u+1} f(x)\psi_S(u+1-x),$$

où $q = P(X_1 = 0)$, $p = P(X_1 > 0)$, $F(u+1) = P(X_1 \leq u+1 | X_1 > 0)$ et $f(x) = P(X_1 = x|X_1 > 0)$. Cela fonctionne correctement car en soustrayant astucieusement l'équation de récurrence en u et $u+1$, la somme disparait. Si nous choisissions une probabilité de ruine du type $\alpha/(u+\beta)$, il est beaucoup plus difficile de retrouver la loi des sinistres. De cette manière, ils obtiennent le résultat suivant.

Théorème (Sundt et dos Reis (2007))**.** *Dans le modèle discret avec des sinistres géométriques $\mathcal{G}e(q, \rho, 1-\alpha)$, la probabilité de ruine a pour expression*

$$\psi_S(u) = \min\left(\frac{(1-q)(1-\rho)}{q(1-\alpha)}\left(\frac{1-q}{q}(1-\rho) + \alpha\right)^u, 1\right).$$

Le minimum garantit la condition de profit net $(1-q)(1-\rho)/q + \alpha < 1$.

Plusieurs extensions ont été proposées au modèle en temps discret, voir, par exemple, Cossette et Marceau (2000),Cossette *et al.* (2004),Marceau (2009), où une dépendance entre les montants de sinistre est introduite. Nous renvoyons le lecteur vers Li *et al.* (2009) pour une revue complète des modèles en temps discret.

Dans le chapitre 4, nous considérons un modèle à mélange basé sur ce modèle en temps discret. Les montants des sinistres sont supposés indépendants et identiquement distribués de loi géométrique zéro-modifiée $\mathcal{G}e(q, e^{-\theta})$ conditionnellement à $\Theta = \theta$. Dans un tel modèle, le chapitre 4 donnera des formules explicites de la probabilité de ruine $\psi_S(u)$ pour certaines lois Θ et des asymptotiques valable pour toute loi de Θ. Nous renvoyons au prochain chapitre pour plus de détails.

Principaux résultats

Cette thèse se décompose en quatre chapitres indépendants dont le fil conducteur est la modélisation du marché d'assurance non-vie. Chaque chapitre correspond à un article et étudie une composante du marché de l'assurance.

Comportement d'un client

Le chapitre 1 est constitué de l'article Dutang (2012b), dans lequel nous cherchons à modéliser la résiliation des contrats d'assurance par l'assuré. Nous présentons une application des modèles linéaires généralisés, ainsi que des modèles additifs généralisés, à ce problème. Ces derniers utilisent des termes non-linéaires dans le prédicteur et peuvent permettre une plus grand souplesse. Néanmoins, le but de ce chapitre est de mettre en garde contre une utilisation abusive et simpliste de ces modèles de régression pour prédire les taux de résiliation. Etant donné leur simplicité d'application, on peut en effet être tenté de les utiliser brutalement sans faire attention à la pertinence des taux prédits. Ce chapitre montre à quel point ces estimations peuvent être erronées si la régression n'utilise pas le pourcentage de rabais (accordé par le vendeur d'assurance) et surtout une estimation du prix marché par contrat. D'autre part, le chapitre propose une méthode simple pour tester l'éventuelle présence d'asymmétrie d'information reposant sur des hypothèses de forme fonctionnelle.

Compétition et cycles en assurance non-vie

L'article Dutang *et al.* (2012a) forme le chapitre 2 et a pour but de combler les déficiences de l'approche traditionnelle à trois agents : assuré, assureur et marché, où les cycles de marché sont modélisés d'un côté (voir par exemple, Haley (1993)) et d'un autre côté, un positionnement de l'assureur est choisi (voir Taylor (1986) et ses extensions). Nous commençons par l'introduction d'un jeu simple non-coopératif à plusieurs joueurs pour modéliser un marché d'assurance sur une période. Dans ce jeu, les assureurs maximisent leur fonction objective $O_j(x_j, x_{-j})$ sous contrainte de solvabilité $g_j(x_j) \geq 0$. Les assurés choisissent de résilier ou de renouveler leur contrat selon une loi multinomiale logit de vecteur de probabilité dépendant du vecteur de prix x. L'existence et l'unicité de l'équilibre de Nash est établie, ainsi que sa sensibilité aux les paramètres initiaux, voir les propositions 2.2.1 et 2.2.2, respectivement. Un jeu plus complexe est ensuite proposé en modélisant plus finement la fonction objective $O_j(x_j, x_{-j})$ et la fonction contrainte des joueurs $g_j(x_j, x_{-j})$. Bien que l'existence d'équilibre de Nash généralisé soit toujours garantie, l'unicité est perdue, voir la proposition 2.3.1. Ainsi, cette version améliorée peut se révéler moins utile en pratique. Une sensibilité aux paramètres est obtenue dans la proposition 2.3.2.

De plus, une version dynamique du jeu est proposée en répétant le jeu simple sur plusieurs périodes tout en mettant à jour les paramètres des joueurs et en tenant de compte de la sinistralité observée sur la période t, c'est à dire nous étudions dans ce jeu $O_{j,t}(x_{j,t}, x_{-j,t})$ et $g_{j,t}(x_{j,t})$. En temps infini, la proposition 2.4.1 démontre que le jeu se termine avec au plus un gagnant. La proposition 2.4.2 donne un ordre stochastique sur le résultat de souscription par police pour un assureur, permettant de mieux comprendre ce qui favorise la position d'un leader. Enfin, par une approche Monte-Carlo, nous estimons la probabilité pour un joueur d'être ruiné ou de se retrouver leader après un nombre fini de périodes sur un grand nombre de simulation. Une cyclicité de la prime marché d'environ dix périodes est observée sur la plupart des simulations. L'utilisation de la théorie des jeux non-coopératifs pour modéliser des problématiques marché est relativement nouvelle : Taksar et Zeng (2011) utilise un jeu continu à deux joueurs à somme nulle, tandis que Demgne (2010) se sert de modèles standards de jeux économiques. Par conséquent, ce chapitre apporte une nouvelle preuve de l'utilité de la théorie des jeux non-coopératifs dans la modélisation des marchés de l'assurance.

Calcul d'équilibres de Nash généralisés

Le chapitre 3 se compose de l'article Dutang (2012a). Ce chapitre montre que le calcul effectif d'équilibre de Nash généralisé n'est pas limité aux seuls jeux à deux joueurs, comme c'est généralement proposé dans les ouvrages de théorie des jeux. Nous nous intéressons aux jeux généralisés les plus génériques et excluons les jeux généralisés conjointement convexes de notre étude. D'une part, ce chapitre a pour but de faire un panorama des méthodes d'optimisation les plus avancées pour calculer un équilibre de Nash généralisé pour un jeu à plusieurs joueurs. Ces méthodes se basent sur une reformulation semi-lisse des équations de Karush-Kuhn-Tucker du problème d'équilibre de Nash. Elles nécessitent l'utilisation du jacobien généralisé, une extension du jacobien classique aux fonctions semi-lisses. D'autre part, nous passons en revue les principaux théorèmes de convergence pour les méthodes de résolution d'équation semi-lisse et étudions leur application dans le contexte des équilibres de Nash généralisé. Une comparaison numérique de ces méthodes (notamment les méthodes de Newton et Broyden généralisées) est réalisée sur un jeu test possédant plusieurs équilibres. Le panorama proposé dans ce chapitre est à comparer à Facchinei et Kanzow (2009) étudiant les jeux généralisés (généraux et conjointement convexes).

Asymptotiques de la probabilité de ruine

Enfin, le chapitre 4 basé sur l'article Dutang *et al.* (2012b) étudie une classe de modèles de risque avec dépendance introduite par Albrecher *et al.* (2011). En temps continu, le modèle de risque considéré est basé sur une approche mélange où les montants de sinistre X_i sont conditionnellement indépendants et identiquement distribués de loi exponentielle $\mathcal{E}(\theta)$ par rapport la valeur d'une variable latente Θ. Ceci est équivalent à supposer que les montants de sinistres ont une copule de survie archimédienne. Au sein de ce modèle, nous démontrons l'existence d'une nouvelle forme d'asymptotique $A+B/u$ pour la probabilité de ruine en temps infini $\psi(u)$.

Ce nouveau type d'asymptotique en $A + B/u$ est démontré dans le théorème 4.3.1, dont nous citons ci-dessous l'item 2 : si Θ suit une loi continue de densité f_Θ telle que f_Θ est presque partout différentiable et Lebesgue intégrable, alors la probabilité de ruine vérifie

$$\psi(u) = F_\Theta(\theta_0) + \frac{f_\Theta(\theta_0)}{u} + o\left(\frac{1}{u}\right),$$

où $\theta_0 = \lambda/c$ et pour un capital initial $u > 0$. Remarquons qu'un résultat similaire peut s'obtenir lorsqu'on ajoute l'hétérogénéité sur les temps d'attente inter-sinistres T_i plutôt que les montants de sinistres X_i. Ce type d'asymptotique est nouveau par rapport à la littérature actuelle, voir Asmussen et Albrecher (2010).

Dans un second temps, le chapitre 4 analyse une version en temps discret du modèle de ruine. Ainsi, nous considérons des montants de sinistres de loi géométrique zéro-modifée $\mathcal{G}e(q, e^{-\theta})$ conditionnellement à l'événement $\Theta = \theta$. De nouveau, nous pouvons montrer qu'une formule asymptotique pour la probabilité de ruine en $A+B/u$ prévaut, voir le théorème 4.3.4. Nous donnons ci-dessous l'item 2 de ce dernier : si Θ suit une loi continue de densité f_Θ telle que f_Θ est presque partout différentiable avec une dérivée f'_Θ bornée, alors la probabilité de ruine vérifie

$$\psi(u) = \bar{F}_\Theta(\theta_0) + \frac{1}{u+2} \times \frac{q f_\Theta(\theta_0)}{1-q} + o\left(\frac{1}{u+2}\right),$$

où $\theta_0 = -\log(1-q)$ et pour un capital initial $u \geq 0$. Le chapitre se termine par une analyse de la dépendance induite par l'approche mélange sur les montants de sinistres dans le modèle en temps discret. Le cas discret pose des problèmes d'identifiabilité de la copule, qui sont abordés dans la section 4.4. La proposition 4.4.6 quantifie la distance maximale en termes de fonctions de répartition jointe des sinistres entre la version continue et la version discrète. Des applications numériques sont proposées. Pour les modèles discrets, ce type d'approche par mélange est là nouveau et permet d'obtenir de nouvelles formules fermées pour la probabilité de ruine.

Modèles de régression

Chapitre 1

Sur la nécessité d'un modèle de marché

—

The customer, the insurer and the market

If you're playing within your capability, what's the point ?
If you're not pushing your own technique to its own
limits with the risk that it might just crumble
at any moment, then you're not really doing your job.
Nigel Kennedy

Ce chapitre se base sur l'article Dutang (2012) soumis au Bulletin Français d'Actuariat.

1.1 Introduction

In price elasticity studies, one analyzes how customers react to price changes. In this paper, we focus on its effect on the renewal of non-life insurance contracts. The methodologies developed can also be applied to new business. Every year insurers face the recurring question of adjusting premiums. Where is the trade-off between increasing premium to favour higher projected profit margins and decreasing premiums to obtain a greater market share? We must strike a compromise between these contradictory objectives. The price elasticity is therefore a factor to contend with in actuarial and marketing departments of every insurance company.

In order to target new market shares or to retain customers in the portfolio, it is essential to assess the impact of pricing on the whole portfolio. To avoid a portfolio-based approach, we must take into account the individual policy features. Moreover, the methodology to estimate the price elasticity of an insurance portofolio must be sufficiently refined enough to identify customer segments. Consequently the aim of this paper is to determine the price sensitivity of non life insurance portfolios with respect to individual policy characteristics constituting the portfolio.

We define the price elasticity as the customer's sensitivity to price changes relative to their current price. In mathematical terms, the price elasticity is defined as the normed derivative $e_r(p) = \frac{dr(p)}{dp} \times \frac{p}{r(p)}$, where $r(p)$ denotes lapse rate as a function of the price p. However, in this paper, we focus on the additional lapse rate $\Delta_p(dp) = r(p + dp) - r(p)$ rather $e_r(p)$ since the results are more robust and easier to interpret. In the following, we abusively refer to $\Delta_p(dp)$ as the price elasticity of demand.

Price elasticity is not a new topic in actuarial literature. Two ASTIN * workshops (see Bland et al. (1997); Kelsey et al. (1998)) were held in the 90's to analyze customer retention and price/demand elasticity topics. Shapiro and Jain (2003) also devote two chapters of their book to price elasticity: Guillen et al. (2003) use logistic regressions, whereas Yeo and Smith (2003) consider neural networks.

In the context of life insurance, the topic is more complex as the lapse can occur at any time, whereas for non-life policies, most lapses occur at renewal dates. There are some trigger effects due to contractual constraints: penalties are enforced when lapses occur at the beginning of the policy duration, while after that period, penalties no longer applied. Another influencial feature is the profit benefit option of some life insurance policies allowing insurers to distribute part of benefits to customers in a given year. This benefit option stimulates customers to shop around for policies with higher profit benefits.

In terms of models, Kagraoka (2005); Atkins and Gallop (2007) use counting process to model surrenders of life insurance, while Kim (2005) uses a logistic regression to predict the lapse. Milhaud et al. (2011) point out relevant customer segments when using Classification And Regression Trees models (CART) and logistic regression. Finally, Loisel and Milhaud (2011) study the copycat behavior of insureds during correlation crises.

In non-life insurance, generalized linear models have been the main tool to analyze price-sensitivity, e.g., Dreyer (2000); Sergent (2004); Rabehi (2007). However, generalized linear model outputs might underestimate the true price sensitivity. This could lead to irrelevant conclusions, and therefore gross premium optimization based on such results may lead to biased and sometimes irrelevant pricing decisions, see, e.g., (Hamel, 2007, Part 5), (Bella and Barone, 2004, Sect. 3).

∗. ASTIN stands for Actuarial STudies In Non-Life insurance.

What makes the present paper different from previous research on the topic is the fact that we tackle the issue of price elasticity from various points of view. Our contribution is to focus on price elasticity of different markets, to check the impact of distribution channels, to investigate the use of market proxies and to test for evidence of adverse selection. We have furthermore given ourselves the dual objective of comparing regression models as well as identifying the key variables needed.

In this paper, we only exploit private motor datasets, but the methodologies can be applied to other personal non-life insurance lines of business. After a brief introduction of generalized linear models in Section 1.2, Section 1.3 presents a naive application. Based on the dubious empirical results of Section 1.3, the Section 1.4 tries to correct the price-sensitivity predictions by including new variables. Section 1.5 looks for empirical evidence of asymmetry of information on our datasets. Section 1.6 discusses the use of other regression models, and Section 1.7 concludes. Unless otherwise specified, all numerical applications are carried out with the R statistical software, R Core Team (2012).

1.2 GLMs, a brief introduction

The Generalized Linear Models (GLM*) were introduced by Nelder and Wedderburn (1972) to deal with discrete and/or bounded response variables. A response variable on the whole space of real numbers $\mathbb{R}$ is too retrictive, while with GLMs the response variable space can be restricted to a discrete and/or bounded sets. They became widely popular with the book of McCullagh and Nelder, cf. McCullagh and Nelder (1989).

GLMs are well known and well understood tools in statistics and especially in actuarial science. The pricing and the customer segmentation could not have been as efficient in non-life insurance as it is today, without an intensive use of GLMs by actuaries. There are even books dedicated to this topic, see, e.g., Ohlsson and Johansson (2010). Hence, GLMs seem to be the very first choice of models we can use to model price elasticity. This section is divided into three parts: (i) theoretical description of GLMs, (ii) a clear focus on binary models and (iii) explanations on estimation and variable selection within the GLM framework.

1.2.1 Theoretical presentation

In this section, we only consider fixed-effect models, i.e. statistical models where explanatory variables have deterministic values, unlike random-effect or mixed models. GLMs are an extension of classic linear models, so that linear models form a suitable starting point for discussion. Therefore, the first subsection shortly describes linear models. Then, we introduce GLMs in the second subsection.

Starting from the linear model

Let $X \in M_{np}(\mathbb{R})$ be the matrix where each row contains the value of the explanatory variables for a given individual and $Y \in \mathbb{R}^k$ the vector of responses. The linear model assumes the following relationship between X and Y:

$$Y = X\Theta + \mathcal{E},$$

∗. Note that in this document, the term GLM will never be used for general linear model.

where Θ denotes the (unknown) parameter vector and $\mathcal{E}$ the (random) noise vector. The linear model assumptions are: (i) white noise: $E(\mathcal{E}_i) = 0$, (ii) homoskedasticity: $Var(\mathcal{E}_i) = \sigma^2$, (iii) normality: $\mathcal{E}_i \sim \mathcal{N}(0, \sigma^2)$, (iv) independence: $\mathcal{E}_i$ is independent of $\mathcal{E}_j$ for $i \neq j$, (v) parameter identification: $rank(X) = p < n$. Then, the Gauss-Markov theorem gives us the following results: (i) the least square estimator $\hat{\Theta}$ of Θ is $\hat{\Theta} = (X^T X)^{-1} X^T Y$ and $\hat{\sigma}^2 = ||Y - X\Theta||^2/(n-p)$ for σ^2, (ii) $\hat{\Theta}$ is a Gaussian vector independent of the random variable $\hat{\sigma}^2 \sim \chi^2_{n-p}$, (iii) $\hat{\Theta}$ is an unbiased estimator with minimum variance of Θ, such that $Var(\hat{\Theta}) = \sigma^2 (X^T X)^{-1}$ and $\hat{\sigma}^2$ is an unbiased estimator of σ^2.

Let us note that first four assumptions can be expressed into one single assumption $\mathcal{E} \sim \mathcal{N}(0, \sigma^2 I_n)$. But splitting the normality assumption will help us to identify the strong differences between linear models and GLMs. The term $X\Theta$ is generally referred to the linear predictor of Y.

Linear models include a wide range of statistical models, e.g. the simple linear regression $y_i = a + bx_i + \epsilon_i$ is obtained with a 2-column matrix X having 1 in first column and $(x_i)_i$ in second column. Many properties can be derived for linear models, notably hypothesis tests, confidence intervals for parameter estimates as well as estimator convergence, see, e.g., Chapter 6 of Venables and Ripley (2002).

We now focus on the limitations of linear models resulting from the above assumptions. The following problems have been identified. When X contains near-colinear variables, the computation of the estimator $\widehat{\Theta}$ will be numerically unstable. This would lead to an increase in the variance estimator [*]. Working with a constrained linear model is not an appropriate answer. In pratice, a solution is to test models with omitting one explanatory variable after another to check for near colinearity. Another stronger limitation lies in the fact that the response variance is assumed to be the same (σ^2) for all individuals. One way to deal with this issue is to transform the response variable by the nonlinear Box-Cox transformation. However, this response transformation can still be unsatifactory in certain cases. Finally, the strongest limitation is the assumed support of the response variable. By the normal assumption, Y must lies in the whole set $\mathbb{R}$, which excludes count variable (e.g. Poisson distribution) or positive variable (e.g. exponential distribution). To address this problem, we have to use a more general model than linear models.

In this paper, Y represents the lapse indicator of customers, i.e. Y follows a Bernoulli variable with 1 indicating a lapse. For Bernoulli variables, there are two main pitfalls. Since the value of $E(Y)$ is contained within the interval $[0, 1]$, it seems natural the expected values $\hat{Y}$ should also lie in $[0, 1]$. However, predicted values $\widehat{\theta} X$ may fall out of this range for sufficiently large or small values of X. Furthermore, the normality hypothesis of the residuals is clearly not met: $Y - E(Y)$ will only take two different values, $-E(Y)$ and $1 - E(Y)$. Therefore, the modelling of $E(Y)$ as a function of X needs to be changed as well as the error distribution. This motivates to use an extended model that can deal with discrete-valued variables.

Toward generalized linear models

A Generalized Linear Model is characterized by three components:

1. a random component: Y_i follows a specific distribution of the exponential family $\mathcal{F}_{exp}(\theta_i, \phi_i, a, b, c)$ [†],

∗. This would be one way to detect such isssue.

†. See Appendix 1.8.1.

2. a systematic component: the covariate vector X_i provides a linear predictor * $\eta_i = X_i^T \beta$,
3. a link function: $g : \mathbb{R} \mapsto S$ which is monotone, differentiable and invertible, such that $E(Y_i) = g^{-1}(\eta_i)$,

for all individuals $i \in \{1, \ldots, n\}$, where θ_i is the shape parameter, ϕ_i the dispersion parameter, a, b, c three functions and S a set of possible values of the expectation $E(Y_i)$. Let us note that we get back to linear models with a Gaussian distribution and an identity link function ($g(x) = x$). However, there are many other distributions and link functions. We say a link function to be canonical if $\theta_i = \eta_i$.

Distribution	Canonical link	Mean	Purpose
Normal $\mathcal{N}(\mu, \sigma^2)$	identity: $\eta_i = \mu_i$	$\mu = X\beta$	standard linear regression
Bernoulli $\mathcal{B}(\mu)$	logit: $\eta_i = \log(\frac{\mu}{1-\mu})$	$\mu = \frac{1}{1+e^{-X\beta}}$	rate modelling
Poisson $\mathcal{P}(\mu)$	log: $\eta_i = \log(\mu_i)$	$\mu = e^{X\beta}$	claim frequency
Gamma $\mathcal{G}(\alpha, \beta)$	inverse: $\eta_i = \frac{1}{\mu_i}$	$\mu = (X\beta)^{-1}$	claim severity
Inverse Normal $\mathcal{I}(\mu, \lambda)$	squared inverse: $\eta_i = -\frac{1}{\mu_i^2}$	$\mu = (X\beta)^{-2}$	claim severity

Table 1.1: Family and link functions

There are many applications of GLM in actuarial science. Table 1.1 below lists the most common distributions with their canonical link and standard applications. Apart from the identity link function, the log link function is the most commonly used link function in actuarial applications. In fact, with this link function, the explanatory variables have multiplicative effects on the observed variable and the observed variable stays positive, since $E(Y) = \prod_i e^{\beta_i x_i}$. For example, the effect of being a young driver and owning an expensive car on average loss could be the product of the two separate effects: the effect of being a young driver and the effect of owning an expensive car. The logarithm link function is a key element in most actuarial pricing models and is used for modelling the frequency and the severity of claims. It makes possible to have a standard premium and multiplicative individual factors to adjust the premium.

1.2.2 Binary regression

Since the insurer choice is a Bernoulli variable, we give further details on binary regression in this subsection.

Base model assumption

In binary regression, the response variable is either 1 or 0 for success and failure, respectively. We cannot parametrize two outcomes with more than one parameter. So, a Bernoulli distribution $\mathcal{B}(\pi_i)$ is assumed, i.e. $P(Y_i = 1) = \pi_i = 1 - P(Y_i = 0)$, with π_i the parameter. The mass probability function can be expressed as

$$f_{Y_i}(y) = \pi_i^y (1 - \pi_i)^{1-y},$$

∗. For GLMs, the name 'linear predictor' is kept, despite η_i is not a linear predictor of Y_i.

which emphasizes the exponential family characteristic. Let us recall the first two moments are $E(Y_i) = \pi_i$ and $Var(Y_i) = \pi_i(1-\pi_i) = V(\pi_i)$. Assuming Y_i is a Bernoulli distribution $\mathcal{B}(\pi_i)$ implies that π_i is both the parameter and the mean value of Y_i. So, the link function for a Bernoulli model is expressed as follows

$$\pi_i = g^{-1}(x_i^T\beta).$$

Let us note that if some individuals have identical covariates, then we can group the data and consider Y_i follows a binomial distribution $\mathcal{B}(n_i, \pi_i)$. However, this is only possible if all covariates are categorical. As indicated in McCullagh and Nelder (1989), the link function and the response variable can be reformulated in term of a latent variable approach $\pi_i = P(Y_i = 1) = P(x_i^T\beta - \epsilon_i > 0)$. If ϵ_i follows a normal distribution (resp. a logistic distribution), we have $\pi_i = \Phi(x_i^T\beta)$ ($\pi_i = F_{\text{logistic}}(x_i^T\beta)$).

Now, the log-likelihood is derived as

$$\ln(\mathcal{L}(\pi_1, \ldots, \pi_n, y_1, \ldots, y_n)) = \sum_{i=1}^{n} \left[y_i \ln(\pi_i) + (1-y_i)\ln(1-\pi_i) \right],$$

plus an omitted term not involving π_i. Further details can be found in Appendix 1.8.1.

Link functions

Generally, the following three functions are considered as link functions for the binary variable

1. logit link: $g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right)$ with g^{-1} being the standard logistic distribution function,
2. probit link: $g(\pi) = \Phi^{-1}(\pi)$ with g^{-1} being the standard normal distribution function,
3. complementary log-log link: $g(\pi) = \ln(-\ln(1-\pi))$ with g^{-1} being the standard Gumbel II distribution function *.

On Figure 1.4 in Appendix 1.8.1, we plot these three link functions and their inverse functions. All these three functions are the inverses of a distribution function, so other link functions can be obtained using inverses of other distribution function. Let us note that the first two links are symmetrical, while the last one is not.

In addition to being the canonical link function for which the fitting procedure is simplified, cf. Appendix 1.8.1, the logit link is generally preferred because of its simple interpretation as the logarithm of the odds ratio. Indeed, assume there is one explanatory variable X, the logit link model is $p/(1-p) = e^{\mu+\alpha X}$. If $\hat{\alpha} = 2$, increasing X by 1 will lead to increase the odds by $e^2 \approx 7.389$.

1.2.3 Variable selection and model adequacy

As fitting a GLM is quick in most standard software, then a relevant question is to check for its validity on the dataset used.

*. A Gumbel of second kind is the distribution of $-X$ when X follows a Gumbel distribution of first kind.

Model adequacy

The deviance, which is one way to measure the model adequacy with the data and generalizes the R^2 measure of linear models, is defined by

$$D(y,\hat{\pi}) = 2(\ln(\mathcal{L}(y_1,\dots,y_n,y_1,\dots,y_n)) - \ln(\mathcal{L}(\hat{\pi}_1,\dots,\hat{\pi}_n,y_1,\dots,y_n))),$$

where $\hat{\pi}$ is the estimate of the beta vector. The "best" model is the one having the lowest deviance. However, if all responses are binary data, the first term can be infinite. So in practice, we consider the deviance simply as

$$D(y,\hat{\pi}) = -2\ln(\mathcal{L}(\hat{\pi}_1,\dots,\hat{\pi}_n,y_1,\dots,y_n)).$$

Furthermore, the deviance is used as a relative measure to compare two models. In most softwares, in particular in R, the GLM fitting function provides two deviances: the null deviance and the deviance. The null deviance is the deviance for the model with only an intercept or if not offset only, i.e. when $p = 1$ and X is only an intercept full of 1 *. The (second) deviance is the deviance for the model $D(y,\hat{\pi})$ with the p explanatory variables. Note that if there are as many parameters as there are observations, then the deviance will be the best possible, but the model does not explain anything.

Another criterion introduced by Akaike in the 70's is the Akaike Information Criterion (AIC), which is also an adequacy measure of statistical models. Unlike the deviance, AIC aims to penalized overfitted models, i.e. models with too many parameters (compared to the length of the dataset). AIC is defined by

$$\mathrm{AIC}(y,\hat{\pi}) = 2k - \ln(\mathcal{L}(\hat{\pi}_1,\dots,\hat{\pi}_n,y_1,\dots,y_n)),$$

where k the number of parameters, i.e. the length of β. This criterion is a trade-off between further improvement in terms of log-likelihood with additional variables and the additional model cost of including new variables. To compare two models with different parameter numbers, we look for the one having the lowest AIC.

In a linear model, the analysis of residuals (which are assumed to be identical and independent Gaussian variables) may reveal that the model is unappropriate. Typically we can plot the fitted values against the fitted residuals. For GLMs, the analysis of residuals is much more complex, because we loose the normality assumption. Furthermore, for binary data, i.e. not binomial data, the plot of residuals exhibits straight lines, which are hard to interpret, see Appendix 1.8.2. We believe that the residual analysis is not appropriate for binary regressions.

Variable selection

From the normal asymptotic distribution of the maximum likelihood estimator, we can derive confidence intervals as well as hypothesis tests for coefficents. Therefore, a p-value is available for each coefficient of the regression, which help us to keep only the most significant variable. However, as removing one variable impacts the significance of other variables, it can be hard to find the optimal set of explanatory variables.

There are two approaches: either a forward selection, i.e. starting from the null model, we add the most significant variable at each step, or a backward elimination, i.e. starting from the full model, we remove the least significant variable at each step.

∗. It means all the heterogeneity of data comes from the random component.

Another way to select significant explanatory variables is to use the analysis of deviance. It consists in looking at the difference of deviance $\ln \mathcal{L}$ between two models, i.e. ratios of likelihood. Using an asymptotic distribution, either chi-square or Fisher-Snedecor distributions, a p-value can be used to remove or to keep an explanatory variable. Based on this fact, statistical softwares generally provide a function for the backward and the forward selection using an automatic deviance analysis

In conclusion, GLM is a well-known statistical method in actuarial science. This fact motivates its use to model lapse rate. Since it is a classic among statistical models, fitting method and variable selection use state-of-art algorithms providing robust estimator. So, there is absolutely no problem in applying GLMs for a daily use. In the following section, we apply GLMs to explain the customer price-sensitivity.

1.3 Simplistic applications and biased business conclusions

This section is intended to present quite naive GLM applications and to show how they can lead to inconclusive or even biased findings. First, we use a dataset with poor and limited data, and then a larger dataset with more comprehensive data. Finally, we summarize the issues encountered. It may seem obvious, but to study customer price-sensitivity, insurers need to collect the premium proposed to customers when renewing policy, especially for those who lapse.

For confidential reasons, the country names are not revealed, but we study two continental European insurance markets. In this part of the world, the insurance penetration rate is considered high, e.g., 8.6% in France, 7% in Germany, 7.6% in Italy, according to Cummins and Venard (2007). Thus, the insurance markets studied are mature as well as competition level is intense. Furthermore, data outputs presented in this paper have been perturbed, but original conclusions have been preserved.

1.3.1 An example of poor data

In this subsection, we work with a (representative) subset of a 1-year lapse history database in 2003. Each line of the dataset represents a policy for a given vehicle. The dataset suffers a major problem because only few variables are available.

Descriptive analysis

To better understand interactions between lapses, the premium and other explanatory variables, we start with a short descriptive analysis. As a general comment, all variables in the dataset are dependent to the lapse variable according to a Chi-square test. At our disposal, we have the last year premium and the proposed premium. Computing the premium ratio, we observe that most of the portfolio experienced a price decrease, probably due to the ageing and the market conditions. We expect to slightly underestimate the true price sensitivity of clients, since customers attention will be released.

Turning to customer variables, we focus on gender and driver age variables, reported in Table 1.2. As the age of the customer increases, the lapse rate decreases. So, the most sensitive clients seem to be the youngest clients. The gender * does not have any particular impact of

∗. In a near future, insurers will no longer have the right to discreminate premium against the gender of the policyholder according to the directive 2004/113/CE from the European comission.

the lapse. However the GLM analysis may reveal some links between the gender and lapses.

	(30,47.5]	(47.5,62.5]	(62.5,77.5]	(77.5,92.5]	FEMALE	MALE
Lapse rate (%)	20	17	14	14.6	18	19
Prop. of total (%)	38	42	17	3	20	80

Table 1.2: Driver age and Gender

We also have a categoric variable containing a lapse type with three possible values: lapse by insured, lapse by company and payment default. We observe a total lapse rate of 18%, of which 11% is a payment default, 6% a lapse by the customer, only 1% a lapse by the company. The lapse by company has to be removed, because those lapses generally result from the pruning strategy of insurers. However, default of payment must be taken with care since it might represent a hidden insured decision. It may result from a too high premium that the customer can't afford. Thus, we choose to keep those policies in our study. Note that the lapse motive cannot be used in the regression because its value is not known in advance, i.e. the lapse motive is endogeneous.

The last variables to explore are policy age and vehicle age. According to Table 1.3, some first conclusions can be derived. As the policy age increases, the remaining customers are more and more loyal, i.e. lapse rates decrease. Unlike the policy age, when the vehicle age increase, the lapse rate increases. One explanation may be that the customer may shop around for a new insurer when changing the vehicle.

	(1, 5]	(5,9]	(9,13]	(13,17]	(1,8]	(8,14]	(14,20]	(20,26]
Lapse rate (%)	21	17	18	16.9	17.6	19.4	21	39
Prop. of total (%)	38	33	22	7	36	37	22	4

Table 1.3: Policy age and vehicle age

GLM analysis

For the GLM analysis of this dataset, we use a backward selection. The explanatory variables are driver age, gender, policy age, vehicle age, the last year premium and the price ratio, i.e. ratio of the premium proposed and the premium paid last year. In order to have better fit and predictive power, all explanatory variables are crossed with the price ratio: crossing variable x_j with price ratio p consists in creating a dummy variable $x_{ji} \times p_i$ for all observations $1 \leq i \leq n$.

Note that variable x_j might be categorical, i.e. valued in $\{0, \dots, d\}$, which allows to zoom in on some particular features of individuals. The linear predictor for observation i is thus given by

$$\eta_i = \mu \times 1 + (x_{1i}, \dots, x_{ki})^T \beta_{-p} + (z_{1i}, \dots, z_{ki})^T \beta_{+p} \times p_i,$$

where μ is the intercept, β_{-p} (resp. β_{+p}) the coefficient for price-noncross variables (resp. price-cross), x_i price-noncross variables, z_i price-cross variables and p_i the price ratio.

Yet not reported here, we test two models: (i) a GLM with original (continuous) variable and (ii) a GLM with categorized variables. We expect the second model with categorized data to be better. Using continuous variables limits the number of parameters: 1 parameter for a

continuous variable and $d-1$ parameters for a categorical variable with d categories. Cutting the driver age, for example, into three values $]18, 35]$, $]35, 60]$ and $]60, 99]$ enables to test for the significance of the different age classes.

The numerical application reveals that a GLM with categorical data is better in terms of deviance and AIC. Hence, we only report this model in Appendix 1.8.2, first column is the coefficient estimates $\hat{\mu}$, $\hat{\beta}_{-p}$ and $\hat{\beta}_{+p}$. The GLM with continuous variables also has business inconsistent fitted coefficients, e.g. the coefficient for the price ratio was negative. This also argues in favor of the GLM with categorized variables. We also analyze (but do not report) different link functions to compare with the (default) logit link function. But the fit gives similar estimate for the coefficients $\hat{\mu}$, $\hat{\beta}_{-p}$ and $\hat{\beta}_{+p}$, as well as similar predictions.

To test our model, we want to make lapse rate predictions and to compare against observed lapse rates. From a GLM fit, we get the fitted probabilities $\hat{\pi}_i$ for $1 \leq i \leq n$. Plotting those probabilities against the observed price ratios does not help to understand the link between a premium increase/decrease and the predicted lapse rate. Recall that we are interested in deriving a portfolio elasticity based on individual policy features, we choose to use an average lapse probability function defined as

$$\hat{\pi}_n(p) = \frac{1}{n} \sum_{i=1}^{n} g^{-1} \left(\hat{\mu} + x_i(p)^T \hat{\beta}_{-p} + z_i(p)^T \hat{\beta}_{+p} \times p \right), \tag{1.1}$$

where $(\hat{\mu}, \hat{\beta}_{-p}, \hat{\beta}_{+p})$ are the fitted parameters, x_i price-noncross explanatory variables, z_i price-cross explanatory variables* and g the logit link function, i.e. $g^{-1}(x) = 1/(1+e^{-x})$. Note that this function applies a price ratio constant to all policies. For example, $\hat{\pi}_n(1)$ the average lapse rate, called central lapse rate, if the premium remains constant compared to last year for all our customers.

Computing this sum for different values of price ratio is quite heavy. We could have use a prediction for a new obsveration $(\tilde{x}, \tilde{y}, \tilde{p})$,

$$g^{-1} \left(\hat{\mu} + \tilde{x}^T \hat{\beta}_{-p} + \tilde{y}^T \hat{\beta}_{+p} \times \tilde{p} \right),$$

where the covariate $(\tilde{x}, \tilde{y}, \tilde{p})$ corresponds to the average individual. But in our datasets, the ideal average individual is not the best representative of the average behavior. Equation (1.1) has the advantage to really take into account portfolio specificities, as well as the summation can be done over a subset of the overall data.

In Table 1.4, we put the predicted lapse rates, i.e. $\hat{\pi}_n(1)$. We also present a measure of price sensitivity, the delta lapse rate defined as

$$\Delta_{1-}(\delta) = \hat{\pi}_n(1-\delta) - \hat{\pi}_n(1) \quad \text{and} \quad \Delta_{1+}(\delta) = \hat{\pi}_n(1+\delta) - \hat{\pi}_n(1), \tag{1.2}$$

where δ represents a premium change, for example 5%.

As mentioned in the introduction, this measure has many advantages compared to the price elasticity† ($e_r(p) = \frac{dr(p)}{dp} \times \frac{p}{r(p)}$): it is easier to compute, more robust‡, easier to interpret.

∗. Both x_i and y_i may depend on the price ratio, e.g. if x_i represents the difference between the proposed premium and a technical premium.

†. It is the customer's sensitivity to price changes relative to their current price. A price elasticity of e means that an increase by 1% of p increase the lapse rate by e%.

‡. Price elasticity interpretation is based on a serie expansion around the point of computation. So, price elasticity is not adapted for large δ.

	$\Delta_{1-}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$
All	-0.745	14.714	0.772
Old drivers	-0.324	9.44	0.333
Young pol., working male	-0.585	15.208	0.601
Young drivers	-1.166	19.784	1.211

Table 1.4: Central lapse rates (%) and deltas (pts)

In Table 1.4, we report the predicted lapse rates and deltas for the whole dataset (first line) as well as for three subsets: old drivers, young policies and working male, young drivers. This first result exhibits the wide range of behaviors among a portfolio: young vs. old drivers. However, delta values seem unrealistic: a 5% premium increase will increase the lapse rate only by 0.772 pts. Based only on such predictions, one will certainly not hesitate to increase premium.

As this small dataset only provides the driver age, GLM outputs lead to inconclusive or dubious results. The old versus young segmentation alone cannot in itself substantiate the lapse reasons. We conclude that the number of explanatory variables are too few to get reliable findings with GLMs, and probably with any statistical models.

1.3.2 A larger database

In this subsection, we study another dataset from a different country in continental Europe in 2004. As for the other dataset, a record is a policy purchased by an individual, so an individual may have different records for the different covers he bought.

Descriptive analysis

This dataset is very rich and contains much more variables than the previous set. The full list is available in Appendix 1.8.2. In Table 1.5, we present some explanatory variables. The dataset contains policies sold through different distribution channels, namely tied-agents, brokers and direct platforms, cf. first line of Table 1.5. Obviously, the way we sell insurance products plays a major role in the customer decision to renew or to lapse. The coverage types (Full Comprehensive, Partial Comprehensive and Third-Part Liability) have a lesser influence on the lapse according to the first table.

The dataset also contains some information on claim history, e.g. the bonus/malus or the claim number. In Table 1.5, we present a dummy variable for the bonus evolution (compared to last year). From this table, we observe that a non-stable bonus seems to increase the customer propency to lapse. This could be explained by the fact that decreasing or increasing bonus implies the biggest premium difference compared to last year premium, raising the customer attention.

At this stage, the claim number does not seem to influence the lapse. The policy age has the same impact as in the previous dataset (cf. Table 1.3). The older is the policy the lower the customer lapses. However, the opposite effect is observed for the vehicle age compared to previous dataset.

Coverage	FC	PC	TPL			Channel	Agent	Broker	Direct	
prop. size	36.16	37.61	26.23			prop. size	65.1	20.1	6.1	
lapse rate	14.26	12.64	12.79			lapse rate	7.4	10.3	12.1	
Claim nb.	0	1	2	3	(3 - 13]	Bonus evol.	down	stable	up	
prop. size	70.59	25.29	3.60	0.44	0.09	prop. size	33.32	62.92	3.76	
lapse rate	13.75	13.37	16.03	12.82	35.16	lapse rate	16.69	11.53	12.02	
Policy age	(0,1]	(1,2]	(2,7]	(7,34]		Vehicle age	(0,6]	(6,10]	(10,13]	(13,18]
prop. size	24.97	16.79	34.38	23.86		prop. size	26.06	31.01	21.85	21.08
lapse rate	17.43	15.27	11.26	8.78		lapse rate	15.50	13.56	12.72	10.67

Table 1.5: Impact on lapse rates (%)

GLM analysis

Now, we go to the GLM analysis. We apply a backward selection to select statistically significant variables. The regression summary is put in Appendix 1.8.2. The sign of coefficient β_{+p} are positive for the two categories of last year premium level *, thus this is business consistent. The most significant variables † are the region code, the distribution channel and the dummy variable indicating the relative difference between the technical premium and the proposed premium and the dummy variable checking whether the policyholder is also the car driver.

In terms of prediction, the results presented in Table 1.6 are similar to the previous subsection. As for the "poor" dataset, we use the average lapse function $\hat{\pi}_n(p)$ and delta lapse rate $\Delta_{1+}(\delta)$ defined in Equations (1.1) and (1.2), respectively. The overall central lapse rate is low compared to the previous set but the customers on that market seems more price sensitive, with bigger deltas for a 5% decrease or increase. Taken into account the distribution channel, the differences are huge: around 8.7% vs. 11.6% for agent and direct, respectively. Despite observing higher deltas, we think these estimates still underestimate the true price sensitivity.

	$\Delta_{1-}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$
All	-0.833	8.966	1.187
Channel agent	-0.759	7.732	0.75
Channel broker	-1.255	9.422	1.299
Channel direct	-1.18	11.597	1.268
Coverage Full Comp.	-0.622	7.723	0.97
Coverage Part. Comp.	-0.714	9.244	1.063
Coverage TPL	-0.899	10.179	1.178

Table 1.6: Central lapse rates (%) and deltas (pts)

Looking at the bottom part, the impact of cover type on central lapse rates is considerably lower. Central rates are between 8% and 10%, regardless of the product purchased. Delta

∗. See `lastpremgroup2(0,500]` and `lastpremgroup2(500, 5e+3]`.
†. See `diff2tech`, `region2`, `channel`, `diffdriverPH7`.

lapse rates Δ_{1+} are again surprisingly low around 1 pt. In Appendix 1.8.2, we also compare the observed lapse rate by channel and coverage type against the fitted lapse rate, see Table 1.17. The results are unsatisfactory.

1.3.3 Issues

The price-sensitivity assessment appears to be difficult. Getting outputs is easy but having reliable estimates is harder. We are not confident on the lapse prediction as well as the additional lapse rates Δ_{1+}. A first answer is shown in Table 1.18 of Appendix 1.8.2, where we present the predicted results when the dataset is split according to the distribution channel or the coverage type. This split provides more realistic lapse rates, each fit better catches the specificity of the distribution channel. Thus we choose to fit nine regressions in the following in order to catch the full characteristics of the distribution channel and the coverage type.

However, this section reveals major issues of a rapid and simplistic application of GLMs. We miss something as it does not really make sense that a 5% premium increase on the whole portfolio leads to a lapse rate increase less than 1pt. In such situation, the insurer has all reasons to increase premium by 5% and to get a higher gross written premium. The market competition level drives the level of customer price-sensitivity that we can estimate. Therefore, high cautions is needed when using GLMs predictions with few variables.

1.4 Incorporating new variables in the regression

This section focuses on identifying new key variables needed in the GLM regression in order to get more reliable results. Attentive readers have probably noticed that some variables have been forgotten in this first analysis. As we will see, they have a major impact on the GLM outputs. Furthermore, taking into account previous conclusions on the large dataset of Subsection 1.3.2, all results presented in this section are obtained by nine different regressions, one for each channel and each coverage type.

1.4.1 Rebate levels

Firstly, we add to all regressions the rebate level variable, specifying the amount of rebate granted by the agent, the broker or the client manager to the customer. As reported in Table 1.7, the number of customers having rebates is considerably high. The broker channel grants a rebate to a majority of customers. Then comes the tied-agent channel and finally the direct channel.

	Full Comp.	Part. Comp.	TPL
Agent	56.62	36.84	22.26
Broker	62.25	52.5	36.24
Direct	23.05	22.89	10.37

Table 1.7: Proportion of granted rebates (%)

It seems logical that the direct channel does not grant rebates since the premium is generally lower through the direct channel than with other distribution channels. The influence

of the coverage type is also substantial: it is harder to get a rebate for a third-part liability (TPL) product than a full comprehensive coverage product.

In order to catch the most meaningful features of the rebate on the lapse decision, the rebate variable has been categorized. Despite the dataset is subdivided into 9 parts, this variable is always statistically significant. For example in the TPL broker subgroup, the estimated coefficients $\hat{\beta}$ for the rebate variable are $\hat{\beta}_{10-20} = -0.368879$, $\hat{\beta}_{25+} = -0.789049$. In that case, the variable has three categories (0, 10-20 and 25+), thus two coefficients for two categories plus the baseline integrated in the intercept. The negative sign means that the rebate level has a negative impact on the lapse, i.e. a rebate of 15 decreases the linear predictor (hence the predicted lapse rate). This is perfectly natural.

Furthermore, when predicting lapse rate with the average lapse function $\hat{\pi}_n$, we force the rebate level to zero. That is to say, in the equation

$$\hat{\pi}_n(p) = \frac{1}{n}\sum_{i=1}^{n} g^{-1}\left(\hat{\mu} + x_i(p)^T\hat{\beta}_{-p} + z_i(p)^T\hat{\beta}_{+p} \times p\right),$$

the explanatory variables $x_i(p), z_i(p)$ are updated depending on the price ratio p. The rebate variable appearing in the vector $(x_i(p), z_i(p))$ is set to zero when predictions are carried out. So that a 5% increase really means such premium increase, and not 5% minus the rebate that the customer got last year.

	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$
Agent	7.278	0.482	8.486	0.896	8.549	0.918
Broker	10.987	2.888	9.754	2.776	10.972	3.437
Direct	12.922	1.154	11.303	1.263	11.893	1.490
	Full Comp.		Part. Comp.		TPL	

Table 1.8: Central lapse rates (%) and deltas (pts)

Table 1.8 presents GLM predictions for the nine subgroups. We can observe the major differences compared to the situation where the rebate level was not taken into account, cf. Table 1.6. Notably for the broker channel, the delta lapse rates are high and represent the broker's work for the customer to find the cheapest premium. The central lapse rates also slightly increase in most cases compared to the previous fit. This subsection shows how important the rebate variable is when studying customer price-sensitivity.

1.4.2 Market proxy

In this subsection, we add another variable to regressions, a market premium proxy by policy. The proxy is computed as the tenth lowest premium among competitor premiums of a standard third-part liabibility * coverage product. Such computation is carried out on a market premium database which is filled by all insurers of the market. However, we don't have the choice of the market proxy. It would have been a good study to see the influence of the market proxy choice, e.g., the fifth, the first lowest or the mean premium, in the GLM fit.

Unfortunately, the market proxy information is only available on two subsets of the database, namely TPL agent and TPL direct subsets. As for the technical premium, we

∗. There is no deductible with this product.

choose to insert that variable in the regression via the relative difference compared to the proposed premium. We consider

$$m_i = \frac{\text{market}_i - \text{premium}_i}{\text{premium}_i},$$

where market_i and premium_i denote the market premium and the proposed premium for the ith policy, respectively. In Table 1.9, we give a basic cross-table of lapse and relative market premium variables. Among the lapsed policies, 65% of them have a higher premium than the market proxy, whereas for renewed policies it drops to 57%.

m	(-0.75,-0.5]	(-0.5,-0.25]	(-0.25,0]	(0,0.25]	(0.25,0.5]	(0.5,0.75]	(0.75,1]
Renew	0.69	18.484	33.248	28.254	9.735	0.571	0.066
Lapse	0.079	1.326	4.158	2.637	0.327	0.032	0.006

Table 1.9: Percentage of policies (%)

However, we cannot conclude that lapses result from a higher premium compared to the market, just based on this table. In fact, the market proxy is just a proxy for the third-part liability coverage, computed as the tenth lowest premium. Moreover, the market proxy is a theoretical premium based on the risk characteristics. If a client goes to another company, it may have a lower or a higher premium depending if he get a *rebate* or choose an *add-on* cover. As the indemnification procedure also varies between two insurers, the market proxy should be seen as a probe of the market level rather than a true premium.

Now, that we have described the new explanatory variable, we turn our attention to the GLM regression. The residual deviance and Akaike Information Criterion (AIC) have slightly decreased with the addition of the market proxy (8866 to 8728 and 8873 to 8735, respectively). Regression summary for the GLM with market variable is available on request to the author.

The most instructive results are the average lapse prediction. Comparing the Table 1.10 with Table 1.8 reveals that the addition of the market proxy has a major impact of the delta lapse rate $\Delta_{1+}(5\%)$, cf. bolded figures. For the TPL agent subset, it goes from 0.918 to 1.652 pts, while for the TPL direct subset, from 1.490 to 2.738. Central lapse rates before and after the market proxy inclusion are consistent.

	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$
Agent	7.278	0.482	8.486	0.896	**8.548**	**1.652**
Broker	10.987	2.888	9.754	2.776	10.972	3.437
Direct	12.922	1.154	11.303	1.263	**11.958**	**2.738**
	Full Comp.		Part. Comp.		TPL	

Table 1.10: Central lapse rates (%) and deltas (pts)

The predicted results are plotted on Figure 1.1, where the x-axis represents central lapse rates ($\hat{\pi}_n(1)$), the y-axis delta lapse rates for a 5% premium increase ($\Delta_{1+}(5\%)$). The bubble radius are determined by the proportion of the subet in the whole dataset. The text order in the legends is the decreasing order of bubble radius.

On Figure 1.1, we clearly observe the difference between those two channels both in terms of central lapse rates and delta lapse rates. These two differences can be explained again by

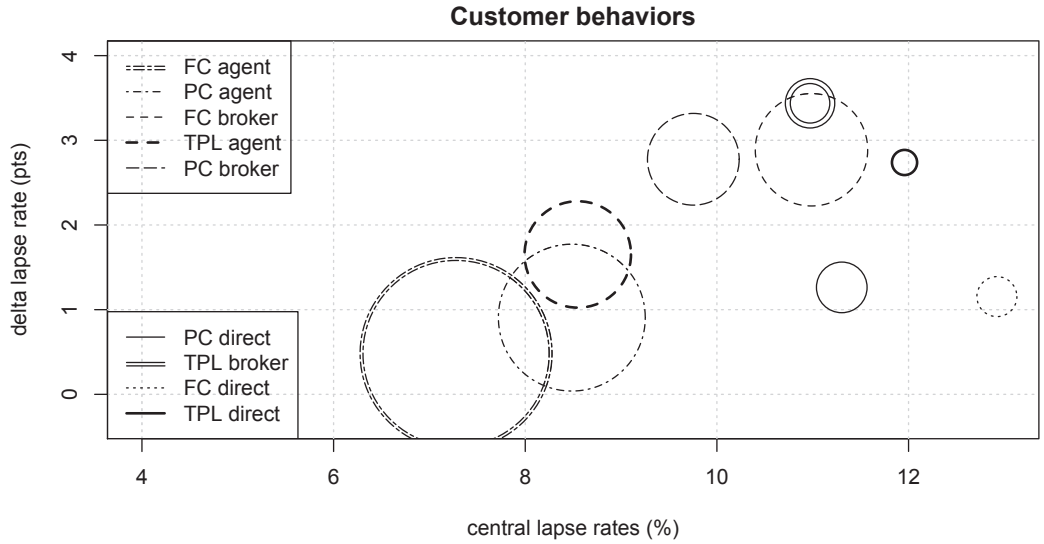


Figure 1.1: Comparison of distribution channels and cover types

the fact the brokers are paid to find the cheapest premium. The direct channel shows higher central lapse rates $\hat{\pi}_n(1)$, but the *estimated* delta lapse rates are lower than those for Broker channel. Direct channel are designed for customers shopping around on internet, so it seems logical that their propency to lapse should be higher. We would have expected the same to hold for delta lapse rates $\Delta_{1+}(5\%)$. The estimated delta rate of the direct channel might still be underestimated. In addition to the absence of market proxies in the TPL direct database, the direct channel is also small in size. Hence, higher uncertainty on those estimates might explain low delta lapse rates for FC/PC direct subsets.

1.4.3 Backtesting

In this subsection, we present backtesting results for the fitted GLMs. We start by looking only at an aggregate level: channel and coverage. The results are given in Table 1.11, reporting observed and fitted lapse rates. The observed lapse rate r_j for the jth group is computed as the average lapse rate variable over the jth group, whereas fitted lapse rate is the average of the fitted probabilities $\hat{\pi}_i$ over the jth group given the observed explanatory variables for each individual $\frac{1}{n_j}\sum_{i=1}^{n_j}\hat{\pi}_i$.

The fitted results are good, since for each subgroup, the deviation is below one percentage point. Compared to the previous backfit table, the improvements with rebate level, market proxy and datasplit are amazing. The two subgroups for which we use market proxy, the results are even better (deviation $<$ 0.1pt), see TPL agent and direct. However, we must recognize that observed price ratio are relatively small: for 85% of the portfolio, the difference is below 5%. Hence, the model appropriately catches the lapse phenomenon when the variation in premium remains *reasonable*.

To further assess the predictive power of our GLM fits, we focus on the TPL coverage product. We consider three subpopulations representing three different behaviors: (i) old drivers with at least two contracts in the household, (ii) working class with a decreasing

	Observed	Fitted	Observed	Fitted	Observed	Fitted
Full Comp.	7.361	7.124	10.167	10.468	12.958	12.881
Part. Comp.	8.123	8.084	9.971	10.09	11.258	11.193
TPL	8.579	8.569	10.867	11.042	11.153	11.171
	Agent		Broker		Direct	

Table 1.11: Central lapse rates (%) and deltas (pts)

bonus-malus and an old vehicle, (iii) young drivers. We expect the population 3 to be the most price-sensitive.

	Prop.	Obs.	Fit.	Std.Err.	Prop.	Obs.	Fit.	Std.Err.	Prop.	Obs.	Fit.	Std.Err.
Pop. 1	13	4.98	5.16	0.22	5	7.99	8.24	0.49	3	6.98	8.187	0.65
Pop. 2	13	8.45	8.65	0.32	16	11.59	12.36	0.50	17	12.44	13.02	0.61
Pop. 3	10	10.01	9.91	0.42	14	13.25	12.45	0.62	13	14.91	14.184	0.74
	Agent				Broker				Direct			

Table 1.12: Lapse rates and proportions (%)

In Table 1.12, we report the backfit results for the three selected populations separating each distribution channel. Each block presents the proportion of population i in the total subset, the observed lapse rate for population i, the mean of fitted lapse rates and standard deviations. As expected the difference between the three populations is high whatever the channel. Population 1 can be tagged as a sluggish behavior, Population 2 a kind of medium behavior, while Population 3 represents highly sensitive customers.

1.4.4 Market scenarios

Having a market variable in the database allows us to perform market scenarios. In this subsection, we briefly present this topic particularly interesting for business line managers. We perform two basic scenarios: a 5% increase of market premium and a 5% decrease of market premium.

	Insurer			Insurer		
	-5%	0%	+5%	-5%	0%	+5%
Market -5%	8.007	8.763	10.481	12.538	14.143	17.589
Market 0%	7.801	8.548	10.152	9.604	11.958	14.696
Market +5%	7.645	8.359	9.916	8.638	10.943	13.589
	Agent			Direct		

Table 1.13: Market scenarios (%)

The results are summarized in Table 1.13. It is surprising to see how the tied-agent customers react very slowly when premium fluctuates. In particular when market decrease of 5% and the proposed premium increases by 5%, then the lapse rate goes only from 8.548% to 10.481%. While for the direct channel, the lapse rate rockets from 11.958% to 17.589%. Actually for any difference in premium, the lapse rate fluctuates largely for the direct channel

The two previous sections demonstrate that GLMs are easy to implement, but care on the variable selection and appropriate data are needed to ensure reliable outputs. In this section, we show how incorporating new key variables in the GLM regression substantially improves the lapse rate predictions in the different premium scenarios. The rebate level partially reveals the agent or the broker actions on the customer decisions, while the use of market proxies illustrates how decisive the competition level is when studying customer price-sensitivity.

In conclusion, the GLM methodology, when used on appropriate data, fulfills the initial objective to derive average lapse rate prediction taking into account individual features. Furthermore, using the predicted lapse rate values of GLMs, it has been easy to identify customer segments, which react differently to premium changes. The back-fit of the GLMs on the identified populations is correct. At a customer segment level, GLMs provide a fair estimate of lapse rate and price sensitivity for reasonable premium changes. But at a policy level, we think lapse predictions should be treated carefully.

1.5 Testing asymmetry of information

Asymmetry of information occurs when two agents (say a buyer and a seller of insurance policies) do not have access to the same amount of information. In such situations, one of the two agents might take advantage of his additional information in the deal. Typically, two problems can result from this asymmetry of information : adverse selection and moral hazard. In insurance context, moral hazard can be observed when individuals behave in risker ways, when they are insured. Insurers cannot control the policyholder's actions to prevent risk.

Adverse selection depicts a different situation where the buyer of insurance coverage has a better understanding and knowledge of the risk he will transfer to the insurer than the insurer himself. Generally, the buyer will choose a deductible in his favor based on its own risk assessment. Hence, high-risk individuals will have the tendency to choose lower deductibles. Adverse selection is caused by hidden information, whereas moral hazard is caused by hidden actions.

Joseph Stiglitz was awarded the Nobel price in economics in 2001 for his pioneer work in asymmetric information modelling. In insurance context, Rothschild and Stiglitz (1976) models the insurance market where individuals choose a "menu" (a couple of premium and deductible) from the insurer offer set. Within this model, they show that high-risk individuals choose contracts with more comprehensive coverage, whereas low-risk individuals will choose higher deductibles.

1.5.1 Testing adverse selection

The topic is of interest when modelling customer behaviors, since a premium increase in hard market cycle phase, i.e. an increasing premium trend, may lead to a higher loss ratio. Indeed if we brutally increase the price for all the policies by 10%, most of high-risk individuals will renew their contracts (in this extreme case), while the low-risk will just run away. Therefore the claim cost will increase per unit of sold insurance cover.

In this paper, we follow the framework of Dionne et al. (2001), which uses GLMs to test for the evidence of adverse selection [*]. Let X be an exogenenous variable vector, Y an

∗. Similar works on this topic also consider the GLMs, see Chiappori and Salanié (2000) and Dardanoni and Donni (2008).

endogeneous variable and Z a decision variable. The absence of adverse selection is equivalent to the prediction of Z based on the joint distribution of X and Y coincides with prediction with X alone. This indirect characterization leads to

$$l(Z|X,Y) = l(Z|X), \tag{1.3}$$

where $l(.|.,.)$ denotes the conditional probability density function.

One way to test for the conditionnal independence of Z with respect to Y is to regress the variable Z on X and Y and see whether the coefficient for Y is significant. The regression model is $l(Z|X,Y) = l(Z; aX + bY)$. However, to avoid spurious conclusions, Dionne et al. (2001) recommend to use the following econometric model

$$l(Z|X,Y) = l\left(Z|aX + bY + c\widehat{E}(Y|X)\right), \tag{1.4}$$

where $\widehat{E}(Y|X)$, the conditionnal expectation of Y given the variable X, will be estimated by a regression model initially. The introduction of the estimated expectation $\widehat{E}(Y|X)$ allows to take into account nonlinear effects between X and Y, but not nonlinear effects between Z and X, Y *.

Summarizing the testing procedure, we have first a regression Y on X to get $\widehat{E}(Y|X)$. Secondly, we regress the decision variable Z on X, Y, and $\widehat{E}(Y|X)$. If the coefficient for Y is significant in the second regression, then risk adverse selection is detected. The relevant choice for Z is the insured deductible choice, with X rating factors and Y the observed number of claims. $\widehat{E}(Y|X)$ will be estimated with a Poisson or more sophisticated models, see below.

1.5.2 A deductible model

The deductible choice takes values in the discrete set $\{d_0, d_1, \ldots, d_K\}$. The more general model is a multinomial model $\mathcal{M}(1, p_0, \ldots, p_K)$, where each probability parameter p_j depends on covariates through a link function. If we assume that variables Z_i are independent and identically distributed random variables from a multinomial distribution $\mathcal{M}(1, p_0, \ldots, p_K)$ and we use a logit link function, then the multinomial regression is defined by

$$P(Z_i = d_j) = \frac{e^{x_i^T \beta_j}}{1 + \sum_{l=1}^{K} e^{x_i^T \beta_l}},$$

for $j = 1, \ldots, K$ where 0 is the baseline category and x_i covariate for ith individual, see, e.g., McFadden (1981), Faraway (2006) for a comprehensive study of discrete choice modelling.

When reponses ($d_0 < d_1 < \cdots < d_K$) are ordered (as it is for deductibles), one can also use ordered logistic models for which

$$P(Z_i = d_j) = \frac{e^{\theta_j - x_i^T \beta}}{1 + e^{\theta_j - x_i^T \beta}} - \frac{e^{\theta_{j-1} - x_i^T \beta}}{1 + e^{\theta_{j-1} - x_i^T \beta}}.$$

Note that the number of parameters substantially decreases since the linear predictor for multinomial logit regression, we have $\eta_{ij} = x_i^T \beta_j$, whereas for the ordered logit, $\eta_{ij} = \theta_j - x_i^T \beta$.

∗. See Su and White (2003) for a recent procedure of conditional independence testing.

The parameters θ, called thresholds, have a special interpretation since they link the response variable Z with a latent variable U by the equation $Z = d_k \Leftrightarrow \theta_{k-1} < U \leq \theta_k$. Hence, the trick to go from a Bernoulli model to a polytomous model is to have different ordered intercept coefficients θ_k's for the different categorical values.

As in Dionne et al. (2001), our choice goes to the ordered logit model for its simplicity. So, Z is modelled by the following equation

$$P(Z_i \leq j/X_i, Y_i) = g^{-1}\left(\theta_j + X_i^T\beta + Y_i\gamma + \widehat{E}(Y|X_i)\delta\right),$$

for individual i and deductible j, with g^{-1} the logistic distribution function * and X_i exogeneous explanatory variables as opposed to endegeneous variables Y_i. The parameters of this model equation are the regression coefficients β and γ and the threshold parameter θ_k's.

1.5.3 Application on the large dataset of Subsection 1.3.2

We want to test for evidence of adverse selection on the full comprehensive (FC) coverage product. So, we study in this subsection only the three datasets relative to that coverage. First, we model the claim number, and then we test for the asymmetry of information.

Modelling the claim number

Modelling count data in the generalized linear model framework can be done by choosing an appropriate distribution: the Poisson and overdispersed Poisson distribution, cf. Table 1.1, where the canonical link function is the logarithm. Since for a Poisson distribution $\mathcal{P}(\lambda)$, $P(Y = 0) = e^{-\lambda}$, the GLM Poisson consists in assuming

$$E(Y/\ x_i) = e^{x_i^T\beta} \Leftrightarrow -\log P(Y = 0/\ x_i) = x_i^T\beta.$$

where x_i denotes the covariates. In practice, this models suffers a subparametrization of the Poisson distribution, one single parameter.

One could think that the Negative binomial in an extended GLM † framework will tackle this issue, but in practice the mass in zero is so high, that both Poisson and negative binomial distributions are inappropriate. As presented in Table 1.14, the high number of zero-claim will compromise the good fit of regular discrete distributions.

Claim number	0	1	2	3	4	5	5 <
Frequency	43687	5308	667	94	17	2	38

Table 1.14: Claim number for Full Comp. agent subset

As presented in Zeileis et al. (2008) and the references therein, the issue is solved by using a zero-inflated distribution, e.g., a zero-inflated Poisson distribution. The mass probability

∗. Note that in this form, it is easy to see that g^{-1} can be any distribution functions (e.g. normal or extreme value distributions).

†. The negative binomial distribution does not belong to the exponential family, except if the shape parameter is known. So, the trick is to use a maximum likelihood procedure for that shape parameter at outer iteration whereas each inner iteration use a GLM fit given the current value of the shape parameter.

function is given by

$$P(Y=k)=\begin{cases}\pi & \text{if } k=0,\\ (1-\pi)\frac{\lambda^k}{k!}e^{-\lambda} & \text{otherwise.}\end{cases}$$

Note that Y is a mixture of a Bernoulli distribution $\mathcal{B}(\pi)$ with a Poisson distribution $\mathcal{P}(\lambda)$. The mean of the zero-inflated Poisson distribution is $(1-\pi)\lambda$. Using the GLM framework and the canonical link functions, a zero-inflated GLM Poisson model is defined as

$$E(Y/\ x_i)=\frac{1}{1+e^{\ x_i^{1T}\gamma}}e^{\ x_i^{2T}\beta},$$

where the covariate vectors x_i^1, x_i^2 are parts of the vector x_i. Now there are two (vector) coefficients to estimate β and γ. The GLM is implemented in R base by the `glm` function. For the zero-inflated model, we need to use the **pscl** package, cf. Jackman (2011).

Still studying the FC agent dataset, we fit three distributions on the claim number: Poisson, zero-inflated Poisson and Negative binomial distributions. As shown in Table 1.19 in Appendix 1.8.2, the three models are similar in terms of log-likelihood or AIC. But, differences appear at the predictions.

Despite being equivalent for first probabilities $P(X=0,1,2)$, classic and zero-inflated Poisson distributions decrease too sharply compared to the observed number of claims. The negative Binomial distribution (fourth line) is far better. In Appendix 1.8.2, we give the regression summary for zero-inflated negative binomial distribution on the FC agent subset. We obtain the same conclusion for other FC subsets.

Claim number	0	1	2	3	4	5	6
Observed	43687	5308	667	94	17	2	2
Poisson	43337.9	5896.0	500.9	39.8	3.7	0.417	0.054
zeroinfl. Poisson	43677.6	5267.7	745.0	80.2	7.5	0.665	0.058
zeroinfl. NB	43704.6	5252.6	704.7	98.8	14.9	2.457	0.442

Table 1.15: Claim number prediction for Full Comp. agent subset

Testing adverse selection

Now that we have modelled the claim frequency, we turn to the modelling of the deductible choice as described in the previous section: an ordered logistic model. We test for evidence of adverse selection on three datasets: agent, broker and direct with Full. Comp. products. Let us note that we cannot test adverse selection on TPL covers, since there is no deductible for this cover. As reported in Subsection 1.5.1, adverse selection testing is done by a fit of a GLM to explain the deductible choice Z_i. In addition to the exogeneous variables X_i for ith individual, the regression will use the observed claim number Y_i (endogeneous) and its expected value coming from the zero-inflated negative binomial regression $\widehat{E}(Y|X_i)$ (exogeneous).

The numerical illustrations reveal that it is more relevant to cluster some deductible values which are too few in the dataset. Actually, the deductible is valued in {0, 150, 300, 500, 600, 1000, 2000, 2500}. As 300 euros is the standard deductible, very high deductibles are rarely chosen. So, we choose to regroup deductible values greater than 500 together. In Table 1.16,

we report the proportion of customers by deductible value for the first two datasets. Small deductible values might reveal high-risk individuals, so we decide to keep those values.

Deductible (€)	0	150	300	500+	0	150	300	500+
Proportion (%)	5.17	10.29	70.85	13.68	4.78	7.85	68.21	17.46
		Agent channel				Broker channel		

Table 1.16: Frequency table for Full Comp. deductibles values

As shown in Appendix 1.8.2 for FC agent subset, the endogeneous variable Y_i is not statistically significant despite being negative, i.e. the higher the loss number, the lower the deductible. But the expected value $\widehat{E}(Y|X_i)$ is significant. For the two other FC datasets, both coefficients for Y_i and $\widehat{E}(Y|X_i)$ are not significant, but these datasets are also smaller in size. We conclude that there is no adverse selection for FC datasets.

After removing insignificant variables in the deductible regression, we integrate the deductible choice predicted probabilities to the lapse regression. Let Z_i denotes the deductible for the ith individual, we incorporate fitted probabilities $\hat{P}(Z_i = 0)$, $\hat{P}(Z_i = 150)$ and $\hat{P}(Z_i = 500+)$. We choose to consider 300 euro as the baseline category, as 300-euro deductible is the standard "unchosen" deductible. For the FC agent dataset, the three probabilities, $\hat{P}(Z_i = 0)$, $\hat{P}(Z_i = 150)$ and $\hat{P}(Z_i = 500+)$, are significant, see Appendix 1.8.2, whereas for the two other FC datasets some probabilities are not significant. We perform the usual predictions for the lapse rate (-5%, 0% and +5% for the proposed premium). But we do not present here the lapse rate predictions since predictions are almost unchanged *.

This section shows how to use GLM modelling to test for evidence of adverse selection. In our dataset, no adverse selection is detected. The inclusion of deductible choice probability neither improves the lapse predictions nor helps in understanding the lapse decision at aggregate level. But we believe that the deductible choice (especially non standard ones) by a customer plays a major role in the propensy of lapse when renewing its policy. Low-risk invididuals, i.e. with high deductibles, are likely to be the most sensitive customers, unlike to high-risk individuals.

1.6 Other regression models

This section presents other regression models. There are mainly two (static) extensions to GLMs in two directions: (i) additive models where the linear predictor is composed of smooth terms and (ii) mixed models where we add a random term (as opposed to fixed term, i.e. deterministic). These two extensions are available for the exponential family distribution, leading to generalized additive models and generalized linear mixed models, respectively. In this paper, we discard mixed models as they are inefficient in our context. The first subsection introduces generalized additive models, and then the second subsection is devoted to an application. The last subsection details other regression models than generalized additive models.

∗. difference less than 0.1% pt.

1.6.1 Model presentation

The Generalized Additive Models (GAM) were introduced by Hastie and Tibshirani (1990) by unifying generalized linear models and additive models. So, GAMs combine two flexible and powerful methods: (i) the exponential family which can deal with many distribution for the response variable and (ii) additive models which relax the linearity assumption of the predictor.

Theoretical presentation

In this subsection, we present Generalized Additive Models in two steps: from linear to additive models and then from additive to generalized additive models. Fitting algorithms are then briefly presented, whereas smoothing techniques are detailed in Appendix 1.8.1. Finally, we apply GAMs on the large dataset of Subsection 1.3.2.

Assuming observations X_i and response variables Y_i are identically and independently distributed random variables having the same distribution of generic random variables X and Y, respectively. In a linear model, the model equation is

$$Y = X\Theta + \mathcal{E}$$

where Y as always stands for the response variable, X the design matrix and $\mathcal{E}$ the random noise. Linear models assume by definition a linear relationship motivated by mathematical tractability rather than empirical evidence. One candidate to extend the linear model is the additive model defined by

$$Y = \alpha + \sum_{j=1}^{p} f_j(X_j) + \mathcal{E},$$

with f_j smooth function of the jth explanatory variable X_j and $\mathcal{E}$ is assumed to be a centered random variable with variance σ^2.

A GAM is characterized by three components:

1. a random component: Y_i follows a distribution of the exponential family $\mathcal{F}_{exp}(\theta_i, \phi_i, a, b, c)$,
2. a systematic component: the covariate vector X_i provides a smooth predictor $\eta_i = \alpha + \sum_{j=1}^{p} f_j(X_{ij})$,
3. a link function $g : \mathbb{R} \mapsto S$ which is monotone, differentiable and invertible, such that $E(Y_i) = g^{-1}(\eta_i)$,

for $i \in \{1, \ldots, n\}$, where θ_i is the shape parameter, ϕ_i the dispersion parameter, a, b, c three functions (characterizing the distribution), f_j's smooth functions and S a set of possible values of the expectation $E(Y_i)$. Note that linear models (and GLMs) are special cases of additive models (and GAMs) with $f_j(x) = \beta_j x$.

We present here only the main idea of fitting algorithms and we do not go into details, see Appendix 1.8.1 for a list of smoothing procedures. All smoothers have a smoothing parameter λ, (the polynom degree, the bandwidth or the span). A first concern is how to choose a criterion on which to optimize λ (hence to have an automatic selection). Then, a second concern is to find a reliable estimate of the parameters α and smooths coefficients given a smoothing value λ.

We present the procedure in the reverse way. Assuming a value of λ, we present an algorithm to fit the model. Hastie and Tibshirani (1990) propose a local averaging generalized

Fisher scoring method. However, Wood (2008) proposes a recent and reliable method: the Penalized Iteratively Reweighted Least Square method (PIRLS). The PIRLS is (unsurprisingly) an iterative method aiming to minimize the penalized deviance

$$\widetilde{D} = D(f_1, \dots, f_p) + \sum_{j=1}^{p} \lambda_j \int f_j''(x_j)^2 dx_j,$$

where the second term penalizes the wiggly behavior of smooth functions.

Given a set of basis functions $(b_{jk})_{jk}$, we can express the smooth function f_j as $f_j(x) = \sum_{k=1}^{K_j} \beta_{jk} b_{jk}(x)$. So, in the end, the GAM can be represented as a GLM with $\eta_i = \widetilde{X}_i \beta$ with $\widetilde{X}_i$ containing the basis functions evaluated at the covariate values and β containing linear parameter α and coefficients β_{jk}'s. Thus, the first term is fully determined. Hence, the penalized deviance is given by

$$\widetilde{D}(\beta) = D(\beta) + \sum_{j} \lambda_j \beta^T S_j \beta,$$

where S_j contains known coefficients and zero's and $D(\beta)$ the GLM version of the deviance for the fixed-basis GAM model. In Appendix 1.8.1, we present in details the PIRLS algorithm to solve the problem $\min \widetilde{D}(\beta)$.

The PIRLS algorithm gives for any λ the corresponding fitted coefficient $\hat{\beta}(\lambda)$, i.e. smooth functions $\hat{f}_j$. Now, we must find a criterion to select the appropriate vector λ. We cannot choose the smoothing parameter λ as the parameter minimizing the deviance, because the model will overfit the data. In the literature, there are many criteria to select the smoothing parameter: likelihood measures such as Restricted Maximum Likelihood (REML), Maximum Likelihood (ML) and cross validation measures such as Generalized Cross Validation (GCV), Generalized Approximate Cross Validation (GACV). These methods differ whether the smoothing parameter is treated as a random effect or not. So, we either maximize a quantity linked to the likelihood (ML/REML) or minimize a prediction error (GCV/GACV).

Expressions of log-likelihood criterion (ML and REML) use the deviance of the model, the satured deviance and a third-term penalizing the wiggliness of the smooth function f_j. The optimization procedure consists in using a Newton method for the optimization of the parameter λ where in each iteration a PIRLS is used (to find $\beta(\lambda)$). So, this is a nested optimization where outer iteration optimizes over λ and the inner iterations optimized over β, see Wood (2010) for details.

An alternative approach seeks in minimizing the prediction error. The predictive error may seem difficult to assess, but the trick is to use a leave-one-out procedure. It consists in computing n deviances D_{-i} where D_{-i} is the deviance without the ith observation. The deviance cross validation is just a sum of the D_{-i}'s. In practice we do not fit n times the model (clearly too expensive!) but an approximation is used to compute the GCV or GACV. Then again, a nested optimization procedure using the PIRLS scheme is used.

In Appendix 1.8.1, we report an example of GAM fit, showing the criterion and the choice of polynomial basis have few impact on the final model. Thus, in the following, we use the REML criterion and thin plate regression: the default in the R function `gam`.

Binary regression and model selection

As for GLMs, the binary regression means we assume that Y_i follows a Bernoulli distribution $\mathcal{B}(\pi_i)$, π_i being linked to explanatory variables. So, the model equation is

$$\pi_i = g^{-1}(\eta_i),$$

where g is the link function and η_i the predictor. Unlike the GLM where the predictor was linear, for GAMs the predictor is a sum of smooth functions:

$$\alpha_0 + \sum_{j=1}^{p} f_j(X_j) \quad \text{or} \quad \alpha_0 + \sum_{i=1}^{p_1} \alpha_i X_i + \sum_{j=1}^{p_2} f_j(X_j),$$

the latter being a semi-parametric approach. As suggested in Hastie and Tibshirani (1995), the purpose to use linear terms can be motivated to avoid too much smooth terms and are longer to compute (than linear terms). For instance, if a covariate represents the date or the time of events, it is "often" better to consider the effect as an increasing or decreasing trend with a single parameter α_i.

As for GLMs, we are able to compute confidence intervals using the Gaussian asymptotic distribution of the estimators. The variable selection for GAMs is similar to those of GLMs. The true improvement is a higher degree of flexibility to model the effect of one explanatory variables on the response.

The procedure for variable selection is similar to the backward approach of GLMs, but a term is dropped only if no smooth function and no linear function with this term is relevant. That is to say, a poor significance of a variable modelled by a smooth function might be significant when modelled by a single linear term. We will use the following acceptance rules of Wood (2001) to drop an explanatory variable:

(a) Is the estimated degrees of freedom for the term close to 1?

(b) Does the plotted confidence interval band for the term include zero everywhere?

(c) Does the GCV score drop (or the REML score jump) when the term is dropped?

If the answer is "yes" to all questions (a, b, c), then we should drop the term. If only question (a) answer is "yes", then we should try a linear term. Otherwise there is no general rule to apply. For all the computation of GAMs, we use the recommended R package **mgcv** written by S. Wood.

1.6.2 Application to the large dataset

In Section 1.3.2, the GLM analysis of this large dataset reveals that the channel distribution strongly impacts the GLM outputs. Especially, the lapse gap between tied-agent and other channels is far stronger than what we could expect. Moreover, the price sensitivity gap measured by the lapse deltas is also high. Let us see this it still holds with GAM results.

On each channel and cover, we first estimate a GAM by modelling all the terms by a smooth function. And then we apply the Wood's rules to remove, to linearize or to categorize the explanatory variables. In Appendix 1.8.2, we provide the regression summary for one of the nine subsets.

Comments on regression summary

In this subsection, we briefly comment on the nine regression summaries. Let us start with the Third-Part Liability cover. For the agent subset, for which we have a market proxy, we

keep four non linear terms (premium difference variables and car class) all modelled jointly with the price ratio. We try to model these terms independently of price ratio, but this was worse in terms of REML scores. On the broker subset, we keep two non linear terms (difference to technical premium and vehicle age). Only the first term is modelled jointly with the price ratio, because the second term has a linear effect with the price ratio. Due to a small size, the direct subset was hard to handle with a GAM. We restrict the price ratio to be a smooth term of small order. This dataset also shows some strange results with a negative elasticity for small premium increase.

Studying Partial Comprehensive coverage is also challenging. For the agent subset, despite many attempts, only the price ratio (alone) has a real benefit to be modelled non linearly. This dataset is sufficiently big to make a lot of explanatory variables significant. And so, we believe a big part of price sensitivity is explained by linear terms. As for the TPL covers, the same variables are modelled non linearly for the broker subset, jointly with the price ratio. The high estimated degrees of freedoms emphasize this non linearity. Turning to the direct channel, only the difference to technical premium variable is modelled through a smooth function, jointly with the price ratio.

Finally, we study the Full Comprehensive coverage product. As always, the agent subset has many nonlinear terms. Three terms (driver age, difference to technical premium and car class) are smoothed together with the price ratio. Again, the estimated degrees of freedom are high, especially for the difference to technical premium variable. Regarding the broker subset, four terms (driver age, vehicle age, difference to technical premium and car class) are modelled non linearly. We retrieve the difference with technical premium and the vehicle age as non linear terms. There might be a process made by brokers to target old vehicles and/or to detect a strong difference with technical premium. So, the brokers have a major impact on the lapse decision. Ending with the direct subset, only two terms are modelled non linearly (the driver age, difference to technical premium): the estimated degree of freedom for the policyholder age variable is high. This may be linked to the close relationship between the motor (technical) premium and the policyholder age.

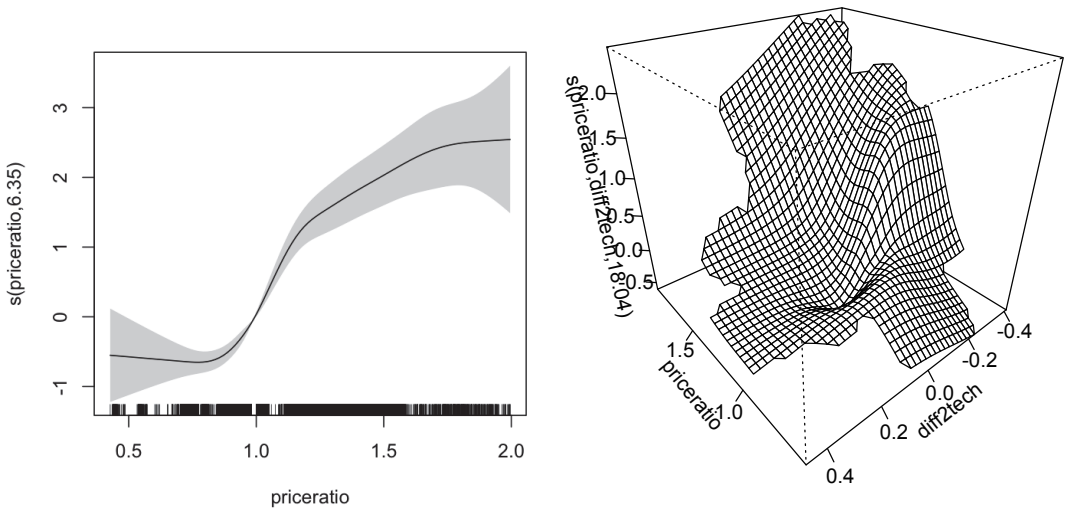
Examples of fitted smooth functions

In the preceding analysis, we observe some trends between channel distributions. Notably, the broker channel results are more sensitive to the difference with technical premium and the vehicle age variables than the other two channels. There is also a data size effect, since the data sets gradually increase in size from TPL and PC to FC covers. Of course, the more we have data, the more the regression is reliable.

On Figure 1.2, we plot two fitted smooth functions from two different GAM regressions [*]. Figure 1.2a represents the smooth function for the price ratio variable of the PC-agent regression. We observe that the smooth function is highly non linear, i.e. a high degree of freedom of 6.35. The smooth function features a very sharp increase of the price ratio around 1: such steep increase is not possible with a linear predictor.

Figure 1.2b is the plot of the bivariate smooth function of the price ratio and the difference to technical premium variable for FC broker dataset. There is a small hollow in the curve around the point (1, 0), a price ratio of 1 and a zero difference with technical premium. Locally,

∗. The grey area represents the standard error bandwidth around the smooth function. It is standard to use an area rather than two simples curves for the confidence interval: this suggests smooth functions lies in such area.



(a) PC agent - price ratio smooth function (b) FC broker - bivariate smooth function

Figure 1.2: GAM smooth functions

the price elasticity of the lapse decision is negative. Fortunately, this business inconsistency is small and located. If we had market variables for this dataset, it could be of interest to check whether this anomaly vanishes.

Discussion on predictions

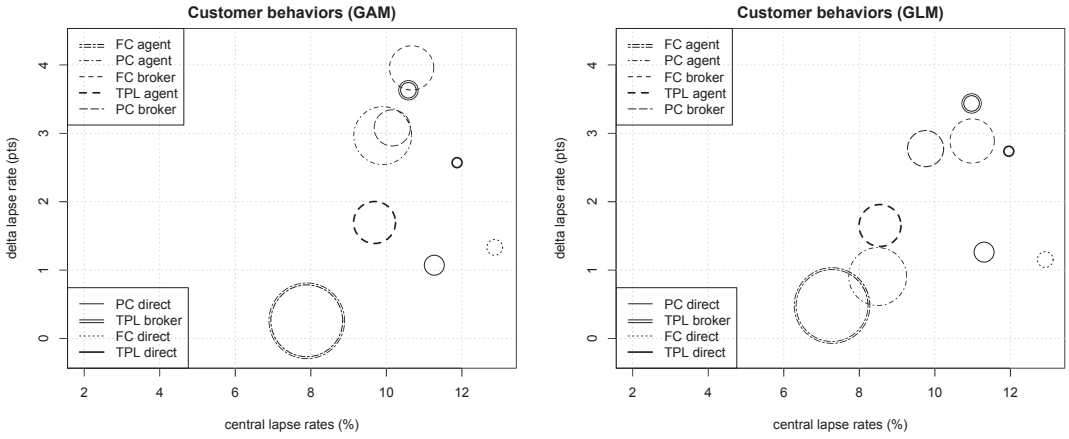
As for the GLM analysis, we turn to the analysis of the distribution channel and the coverage type by looking at the lapse rate predictions. We also consider an average lapse rate function defined as

$$\hat{\pi}_n(p) = \frac{1}{n}\sum_{i=1}^{n} g^{-1}\left(\hat{\mu} + x_i(p)^T\hat{\beta}_{-p} + z_i(p)^T\hat{\beta}_{+p} \times p + \sum_{j=1}^{p}\hat{f}_j(\tilde{z}_i(p), p)\right), \tag{1.5}$$

where $(\hat{\mu}, \hat{\beta}_{-p}, \hat{\beta}_{+p})$ are the fitted parameters, $\hat{f}_j$ are the fitted smooth functions, $(x_i, z_i, \tilde{z}_i)$ are parts of explanatory variables of the ith individual and g is the logit link function. What differentiates Equation (1.5) with Equation (1.1) is the inclusion of additive terms in the predictor.

On Figure 1.3, we plot the usual bubble plot to compare GAMs and GLMs. We observe that GAM delta lapse rate predictions are higher than GLM ones in most cases. This is especially true for PC agent or FC broker: there is a high jump upward. Only two channel-covers have a lower delta lapse rate $\Delta_{1+}(5\%)$ with GAMs: the FC direct case, a case where the dataset is small (so the GAM model selection was hard) and the FC agent case where the difference is limited.

In terms of central lapse rates, most of predictions $\hat{\pi}_n(1)$ are higher, i.e. shift to the right on Figure 1.3. It means that the customers in the portfolio are more price-sensitive even if we propose exactly the same premium as last year. On a private motor insurance, most people expect a better bonus-malus from year to another, hence a premium decrease.



(a) GAM (b) GLM

Figure 1.3: GAM vs. GLM - comparison of distribution channels and cover types

Now, we stop the GAM analysis and conclude on the pros and cons of GAMs. GAMs are less known tools than GLMs in actuarial science. But since their introduction in the 90's, GAMs are well studied and use state-of-the-art fitting procedures. There are two ways to perform model selections: prediction errors vs. likelihoods. In this paper, we follow the Wood's rule to select variables based on the restricted maximum likelihood. We tested other statistical quantities, but the impact remains limited.

As for GLMs, GAMs allow us to assess an overall estimated price elasticity (via $\hat{\pi}_n(1)$ and $\Delta_{1+}(5\%)$) taking into account the individual features of each policy. The additional complexity coming with additive modelling compared to GLMs permit to really fit the data. Especially for broker lines, we get a more cautious view of customer price sensitivity. For small datasets, GAM predictions may lead to irrelevant results. Furthermore, as already noticed for GLMs, GAMs predictions are reliable for with a small range of price change: extrapolating outside observed price ratio range leads to doubtful results.

Finally, GAMs need a longer time to fit than GLMs and require a better computing power. This is a limitation for GAMs to be used easily by everyone. In addition, some user judgement is needed to select, to linearize or to reject explanatory variables in order to get the final model for GAMs. Even with Wood's rules, newcomers may find it hard to choose between two GAM models with the same "score", i.e. with the same likelihood or prediction errors.

1.6.3 Other regression models

GLMs and GAMs are static models. One option to take into account for dynamics would be to use time serie models on regression coefficients of GLMs. But this was impossible with our datasets due to a limited number of years and it is rather a trick than an appropriate solution. Generalized Linear Mixed Models (GLMM), where the linear predictor becomes the sum of a (unknown deterministic) fixed term and a random term, are a natural extension of GLMs to deal with heterogeneity across time.

Among many others, Frees (2004) presents GLMMs in the context of longitudinal and

panel data. Since a panel data model cannot deal with right-censoring (that occurs when a policy is terminated), they are not appropriate to our policy termination problem, i.e. lapse. Despite discarding GLMMs for dynamic lapse modelling, we try to use the GLMMs on one period in order to model endogeneous effects such as dropping coverage with a random term. Unfortunately, this reveals inefficient.

The Survival Regression Model of Cox (1972) allow to remove the inherent limits of the static regression models previously presented. By nature, they take into account the dynamic aspects of the response variable. by a lifetime variable. In our context, we model the lifetime of a policy. As GLMs and GAMs demonstrate, renewing a policy for the first time is not motivated by the same factors as renewing one for the tenth time. An application of such models may be found in Brockett et al. (2008) and Chapter 4 of Dutang (2011).

The full power of survival models is not only to model one lapse reason. Other policy termination factors can be integrated so as to model the complete life cycle of a policy. With a full picture integrating other cash flows such as claims, and premiums, insurance risk could also be better assessed. Further advanced models than the Cox model regression exists, such as state-space models, e.g., Fahrmeir (1994) or stochastic counting processes, see, e.g., Andersen et al. (1995); Aalen et al. (2008). Some attempts have been done to use Fahrmeir (1994)'s state space model, but the fitting process was too heavy to be quickly used.

1.7 Conclusion

Fitting price-sensitivity is a complex topic. Being dependent on the market's environment, price elasticity forecasts require rigorous attention to details to prevent the risk of erroneous conclusions. Not surprisingly, a data cleaning process is essential prior to any regression fitting. In short, some supplied explanatory variables substantially affect the results. Omitting these variables in the data can, in itself, lead to unreliable findings.

These must-have variables include distribution channels, market premium proxies, rebate levels, coverage types, driver age, and cross-selling indicators. In Section 1.3, the small dataset only provides the driver age: this example leads to inconclusive results. On the large dataset, the coverage type, and the cross-selling indicators were added to the regression fit. This enabled us to refine our analysis. Having or not having a household policy with the same insurer was thus proven to be a driving factor in renewing or allowing a contract to lapse.

However, fully reliable predictions are only achieved when the rebate level and market premium proxies are used. In Section 1.4, the price sensitivity fit was considerably enhanced, along with our ability to fine tune the results, thanks to the inclusion of distribution channels, a market proxy, and a rebate level. With the gradual addition of explanatory variables, we have seen an increased accuracy of the lapse rate predictions. Disposing of market variables proved to make testing market scenarios possible (e.g. -5%, +5%). Being able to provide such forecasts is highly valuable in taking pricing actions. If those market proxies are no longer available, we are likely to get back to less meaningful results.

Adverse selection resulting from an asymmetry of information is a widely known risk in insurance. Section 1.5 investigates for empirical evidence of adverse selection and studies its relationship to the lapse decision of customers. On our large dataset, no adverse selection is detected. At aggregate level, adverse selection does not have a big influence. Nevertheless, at individual level, choosing a non-standard deductible when underwriting a new policy will certainly have consequences on the termination of this policy.

Generalized Linear Models are widely known and respected methods in non-life insurance. However, they have some inherent constraints with GLMs. Thus, in Section 1.6, we test Generalized Additive Models, which allow for non linear terms in the predictor. Like GLMs, the quality of the findings attained is directly related to the data provided. Using limited variables will produce approximate results, whereas, dealing with an extensive set of variables lead to proven results.

Applying GAMs, despite their additional complexity, can be justified in cases where GLMs fail to provide realistic lapse predictions and we have substantial datasets. Note that GAMs can model interactions between explanatory variables. Not restricted to linear terms, they consequently provide us with a more adaptive tool. Caution should however be exercised, as they may overfit the data when applied to limited datasets. This could then imply business inconsistency.

In this paper, we have explored the price elasticity topic from various viewpoints. Once again, our research has further demonstrated that the quality of data used in actuarial studies unequivocally affects the findings reached. In addition, the key role of the market proxies in estimating price sensitivity has been established. Market competition modelling, see, e.g., Demgne (2010), Dutang et al. (2012), is therefore relevant.

The conclusions drawn from customer price sensitivity studies should in any respect be weighed carefully. Charging higher premiums to loyal customers could seem unfair in light of the fact that those same customers usually have a better claims history. By the same token, relying on the market context with its inherent uncertainty to predict price sensitivity could be misleading. In summary, insurers must have a well informed overview of the market, the customer base, and a keen awareness of the pros and cons of potential pricing adjustments. The models presented herein serve as decision-making support tools and reinforce business acumen.

1.8 Appendix

1.8.1 Generalized linear and additive models

Univariate exponential family

Clark and Thayer (2004) defines the exponential family by the following density or mass probability function

$$f(x) = e^{d(\theta)e(x)+g(\theta)+h(x)},$$

where d, e, g and h are known functions and θ the vector of paremeters. Let us note that the support of the distribution can be $\mathbb{R}$ or $\mathbb{R}_+$ or $\mathbb{N}$. This form for the exponential family is called the natural form. When we deal with generalized linear models, we use the natural form of the exponential family, which is

$$f(x, \theta, \phi) = e^{\frac{\theta x - b(\theta)}{a(\phi)} + c(x,\phi)},$$

where a, b, c are known functions and θ, ϕ* denote the parameters. This form is derived from the previous by setting $d(\theta) = \theta$, $e(x) = x$ and adding a dispersion parameter ϕ. The exponential family of distributions in fact contains the most frequently used distributions.

∗. the canonic and the dispersion parameters.

For example, the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with $\theta = \mu$ and $\phi = \sigma^2$, see Clark and Thayer (2004) for details.

Fitting procedure

To determine the parameter vector β, we use the maximum likelihood estimation. For n observations, the log-likelihood of the model given a distribution from the exponential family is written as follows:

$$\ln(\mathcal{L}(\theta_1, \ldots, \theta_n, \phi, y_1, \ldots, y_n)) = \sum_{i=1}^{n} \left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]. \tag{1.6}$$

Let us define $\mu_i = E(Y_i)$ and $\eta_i = g(\mu_i) = X_i\beta$, the linear prediction where i is the number of the observation, n the total number of observations.

For all i and j,

$$\frac{\partial \ln(\mathcal{L}_i)}{\partial \beta_j} = \frac{\partial \ln(\mathcal{L}_i)}{\partial \mu_i} \times \frac{\partial \mu_i}{\partial \beta_j} = (g^{-1})'(g(\mu_i)) \times \frac{y_i - \mu_i}{Var(Y_i)} X_{ij}.$$

Maximum likelihood equations are then: $\sum_i \frac{\partial \ln(\mathcal{L}_i)}{\partial \beta_j} = \sum_i (g^{-1})'(g(\mu_i)) \times \frac{y_i - \mu_i}{Var(Y_i)} X_{ij} = 0$, for all j. Therefore, we get the equations, as a function of the β_i's:

$$\sum_i \frac{\partial \ln(\mathcal{L}_i)}{\partial \beta_j} = \sum_i (g^{-1})'(X_i\beta) \times \frac{y_i - g^{-1}(X_i\beta)}{(b')^{-1}(g^{-1}(X_i\beta))} X_{ij} = 0. \tag{1.7}$$

These equations are not linear with respect to the β_is, and cannot be solved easily. As always for complex equation, we use an iterative algorithm to find the solution. Most softwares, such as R, use an iterative weighted least-squares method, see Section 2.5 of McCullagh and Nelder (1989).

Link functions for binary regression

Log-likelihood for canonical link Using the expression of the variance function and the canonical logit function ($g^{-1}(x) = \frac{1}{1+e^{-x}}$ and $(b')^{-1}(x) = x(1-x)$), Equation (1.7) becomes

$$0 = \sum_i \frac{e^{-\eta_i}}{1+e^{-\eta_i}} \times \frac{y_i - \frac{1}{1+e^{-\eta_i}}}{\frac{1}{1+e^{-\eta_i}} \frac{e^{-\eta_i}}{1+e^{-\eta_i}}} X_{ij} = \sum_i (y_i(1+e^{-\eta_i}) - 1) X_{ij},$$

for $j = 1, \ldots, p$. These equations are called the likelihood equations. If we put it in a matrix version, we get the so-called score equation

$$X^T(Y - \mu(\beta)) = 0.$$

Thus, the Fisher information matrix for β in the case of logit link is

$$\mathcal{I}(\pi) \stackrel{\triangle}{=} -E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \beta_j \partial \beta_k} \right) = \text{diag}(\pi_i(1-\pi_i)).$$

Since we use the maximum likelihood estimator, the estimator $\hat{\beta}$ has the good property of being asymptotically unbiased and Gaussian with variance matrix approximated by Fisher information $\mathcal{I}(\pi(\hat{\beta}))$ *.

*. see subSection 4.4.4 of McCullagh and Nelder (1989).

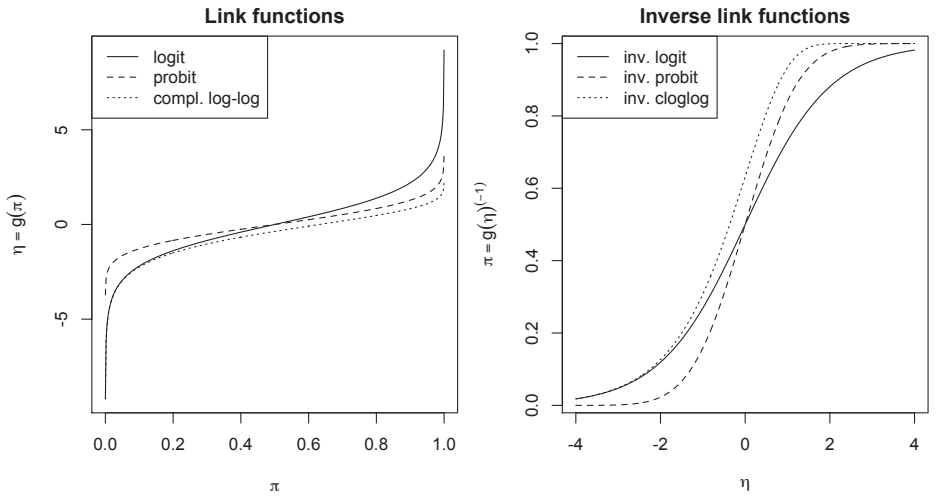


Figure 1.4: Link functions for binary regression

Univariate smoothing

In this paragraph, we present gradually some classic smoothing procedures, from the simplest to more complex methods. Probably the simplest method to get a smooth function is to regress a polynom on the whole data. Assuming observations are denoted by $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, a multiple regression model is appropriate with

$$Y = \alpha_0 + \alpha_1 X + \cdots + \alpha_p X^p.$$

Using $f(x) = \sum_i \alpha_i x^i$ is clearly not flexible and a better tool has to be found. One way to be more flexible in the smoothing is to subdivide the interval $[\min(x), \max(x)]$ into K segments. And then we can compute the average of the response variable Y on each segment $[c_k, c_{k+1}[$. This is called the bin smoother in the literature. As shown on Hastie and Tibshirani (1990) Figure 2.1, this smoother is rather unsmooth.

Another way to find a smooth value at x, we can use points about x, in a symmetric neighborhood $N_S(x)$. Typically, we use the k nearest point at the left and k nearest at the right of x to compute the average of y_i's. We have

$$s(y|x) = \frac{1}{\text{Card}N_S(x)} \sum_{i \in N_S(x)} y_i,$$

where the cardinal $\text{Card}N_S(x)$ does not necessarily equal to $2k+1$ if x is near the boundaries. Again we do not show the result and refers the reader to Hastie and Tibshirani (1990) Figure 2.1. This method, called the running mean, takes better into account the variability of the data. However we lose the smoothness of previous approches.

An extension of this approach is to fit the linear model $y = \mu + \alpha x$ on the points (x_i, y_i) in the neighborhood (for $i \in N_S(x)$). That is to say we have a serie of intercepts μ and slopes α for all observations. We called this method the running line, which generalizes the running mean, where α is forced to 0.

Another enhancement is to weight the points in the regression (for x_i) inversely relative to the distance to x_i. Generally we use the tricube weight function

$$w(z) = (1 - |z|^3)^3 \mathbb{1}_{|z|<1}.$$

So the weight for x_j when computing the smooth value of x_i is $w(z_j)$ with $z_j = \frac{|x_i - x_j|}{b}$ and b the bandwith. Introduced by Cleveland (1979), this method is known as LOcally WEigthed Smoothing Scatterplots (LOWESS). Other weight function can be used as long as it is a symmetric, decreasing from 0, strictly positive on $]-1, 1[$ and null elsewhere.

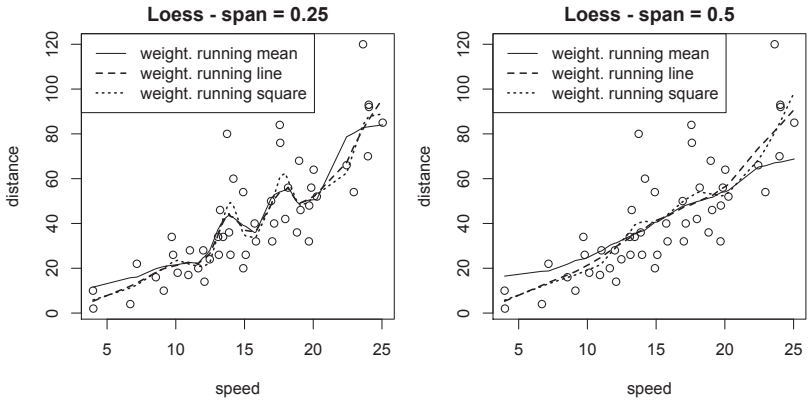


Figure 1.5: LOWESS

Summarising the LOWESS approach, we have w a weight function, d the degree of polynoms (1 for running line, 0 for running mean), f the span defined as the proportion of points to use in the regression*. On Figure 1.5, we plot different LOWESS methods for 2 spans. Clearly the righthand side plot is smoother than the lefthand one. For a given span, as one can expected the running mean method is more robust, unlike the running square which goes up and down easily.

A popular method for smoothing is the Kernel smoothing. Choosing a Kernel function (i.e. a symmetric density function) k, the associated smoother is

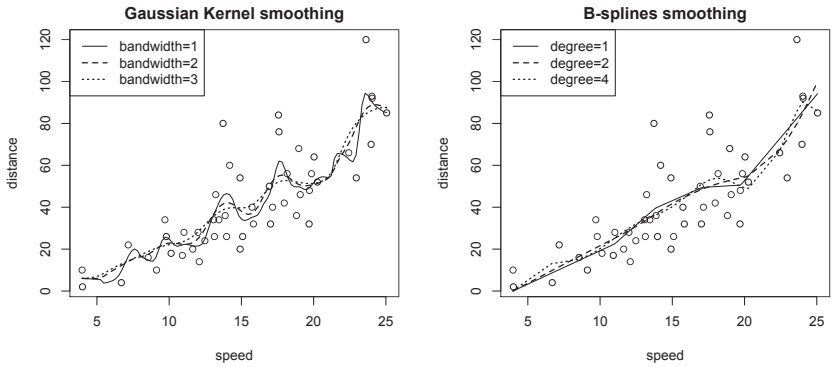
$$s(y|x) = \frac{\sum_{i=1}^n k\left(\frac{x-x_i}{b}\right) y_i}{\sum_{i=1}^n k\left(\frac{x-x_i}{b}\right)},$$

where b denoting the bandwidth is used to scale the data $(x - x_i)/b$.

As suggested in Venables and Ripley (2002), Kernel smoothing can be seen as a local weighted running mean approach. However, the power of the Kernel approach relies on the use of more complex Kernel functions. Common Kernel functions are the standard normal density function or the Epanechnikov function (a bisquare function). With Figure 1.6a, not surprisingly, increasing the bandwidth increases the smoothness of the fitted curve.

The last and recent tool to fit a smooth curve is to use spline functions. The approach consists in splitting the interval in K knots $(t_1, \ldots, t_K)$ and fit a polynom on each segment, while imposing smooth conditions at the knots. One intuitive spline function is the polynom of third degree, since the smooth conditions ($f' = f'' = 0$) are easily written done.

∗. closely linked to the bandwidth.



(a) Kernel smoothing (b) B-spline smoothing

Figure 1.6: Kernel and B-spline smoothing

A crucial fact is that using K polynomials can be represented by a banded matrix, a band for each segment. Using the matrix representation emphasizes that we use a basis of functions to approximate the function f. Many polynomial basis can be used (e.g. Bernstein polynoms for Bézier curves).

One popular basis is the B-spline basis. They are defined recursively starting polynoms of degree 0 defined by $B_{i,0}(t) = \mathbb{1}_{t_i \leq t < t_{i+1}}$ * and higher order $B_{i,d}$ obtained by convex combination (with t_i's increasing knots) of $(B_{i,d-1})_i$'s. To fit the data, we then minimize a penalized least square and use some quantiles as the knots (for instance quartiles).

The thin plate regression uses a basis of thin plate (also known as polyharmonic functions) functions $\phi_{md}(r) = \alpha_{md} r^{2m-d} \log(r)$ if d is even and $\alpha_{md} r^{2m-d}$ if d is odd. The smooth functions is computed as $s(y|x) = \sum_{i=1}^{n} \delta_i \phi_{md}(||x - x_i||_2)$. A low-rank approximation of this smooth function is then used to decrease the computational burden, see Wood (2003). This method avoids the knot placement problems of traditional regression spline modelling.

On Figure 1.7, we plot smooth functions for the `cars` data with different smoothing selection and basis functions (where default is the bottom-left figure). Both for the REML and GCV criterion, the function basis has no influence on the estimation of the smooth function f. Let us note that the estimated degrees of freedom are also very closed whatever the method is.

On Figure 1.6b, we use the quantiles 20%, 40%, 60% and 80% as interior knots and three different degrees. As one can expect, B-splines with high degrees better fit the data. To conclude with this smoother presentation, all smoothers presented here are linear smoothers, since they can be written by $\hat{y} = Sy$ with S the smoother matrix (depending on observations x).

PIRLS algorithm

The Penalized Iteratively Reweighted Least Square (PIRLS) has the following scheme

1. Initiate μ_i^0 typically with y_i.
2. Iterate while no change in deviance $\widetilde{D}(\beta^k)$

∗. See theorem 1.5 of Steihaug (2007).

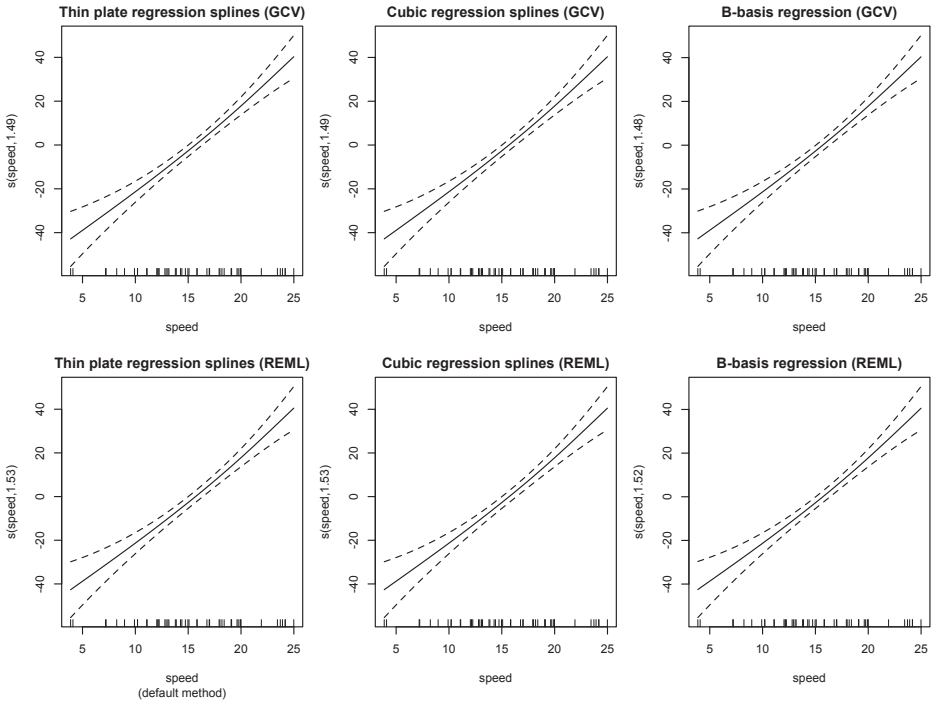


Figure 1.7: Additive model tests

(a) compute the weight $w_i^k = \frac{1}{g'(\mu_i^{k-1})}\sqrt{\frac{\omega_i}{V(\mu_i^{k-1})}}$,

(b) evaluate the pseudo data

$$z_i^k = g'(\mu_i^{k-1})(y_i - \mu_i^{k-1}) + \eta_i^{k-1},$$

with $\eta_i^{k-1} = g(\mu_i^{k-1})$,

(c) minimize with respect to β the least square objective

$$||W(z - \widetilde{X}\beta)||^2 + \sum_j \lambda_j \beta^T S_j \beta,$$

with $W = \text{diag}(w_1, \ldots, w_n)$. We get β^k.

(d) prepare next estimate with $\eta^k = \widetilde{X}\beta^k$ and $\mu_i = g^{-1}(\eta_i^k)$.

1.8.2 R outputs

Bronchitis dataset

Let us study the example of Bronchitis data of Turner (2008). The data consists of 212 patients, on which we measure the presence/absence of bronchitis B for `bron`, the air pollution level in the locality of residence P for `poll` and the number of cigarettes smoked per day C for `cigs`, see Appendix 1.8.2.

Let us first regress the bronchitis indicator on all variables

$$Y = \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} 1 & P_1 & C_1 \\ \vdots & \vdots & \vdots \\ 1 & P_n & C_n \end{pmatrix},$$

with a logit link function. The regression summary is given below

```
Call: glm(formula = bron ~ 1 + cigs + poll, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4023  -0.5606  -0.4260  -0.3155   2.3594

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.08491    2.95100  -3.417 0.000632 ***
cigs          0.21169    0.03813   5.552 2.83e-08 ***
poll          0.13176    0.04895   2.692 0.007113 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null deviance: 221.78  on 211  degrees of freedom
Residual deviance: 174.21  on 209  degrees of freedom - AIC: 180.21
```

So the GLM fit seems good because all variables (including intercept) are significant with a very low p-value. However the plot of residuals * (see Figure 1.8a) against fitted values † is quite puzzling. Two distinct curves are shown: one for ill patients and the other for healthy ones.

When categorizing the P variable, we lose information but we transform binary data into binomial data. This makes the fit better on this aspect, see Figure 1.8b. So for the same data, with the same (significant) variables, the two analyses of residuals lead to different conclusions. Hence, conclusions of residual analysis must be taken with great care.

GLM outputs of Section 1.3.1

See below the summary table with coefficients values, standard errors, z-statistics and p-value. For confidentiality reason, all the deviance and AIC statistics shown in this paper have been scaled by the same positive coefficient.

Here follows the regression summary when variables are categorical.

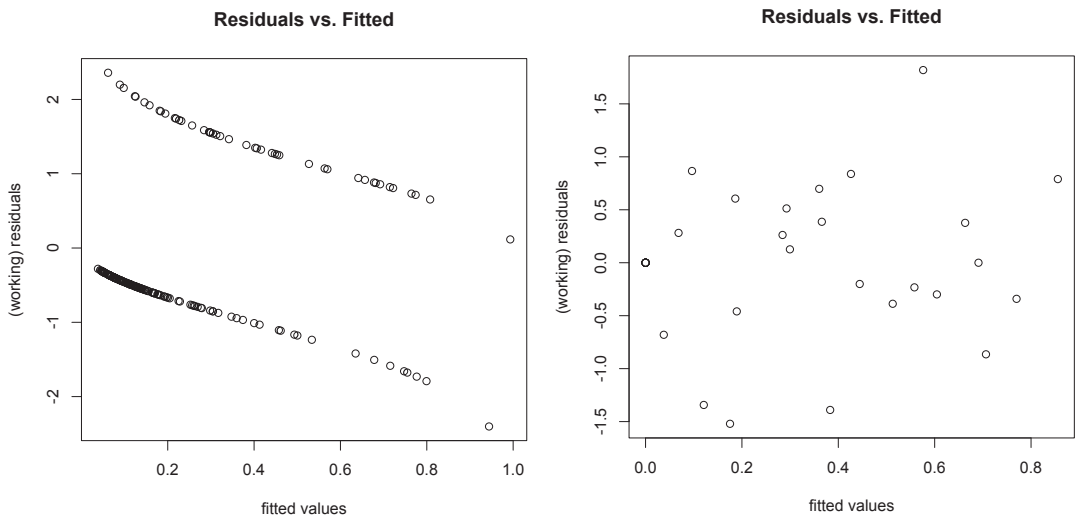
```
Call: glm(formula = did_lapse ~ agepolgroup2 + priceratio:agegroup4 +
    priceratio * (gender + agevehgroup2 + prembeforegroup2),
    family = binomial(), data = workdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1587  -0.6633  -0.6060  -0.5193   2.8747

Coefficients:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -2.522477   0.120852 -20.873  < 2e-16 ***
```

∗. Working residuals are $\hat{\epsilon}_i = Y_i - \hat{\pi}_i$. Note that using other residual types, Pearson, Studentized, do not change this behavior.

†. Fitted values are $\hat{\pi}_i$.



(a) Binary data (b) Binomial data

Figure 1.8: Analysis of residuals for binary regression

```
agepolgroup2(4,49]                          -0.153793   0.007270 -21.154  < 2e-16 ***
genderMALE                                   0.681454   0.117045   5.822 5.81e-09 ***
agevehgroup2(5,10]                          -0.684290   0.106741  -6.411 1.45e-10 ***
agevehgroup2(10,99]                         -0.262674   0.101038  -2.600  0.00933 **
prembeforegroup2(500,1e+03]                 -0.295837   0.137011  -2.159  0.03083 *
prembeforegroup2(1e+03,1e+04]               -0.923435   0.283603  -3.256  0.00113 **
priceratio                                   1.018771   0.120903   8.426  < 2e-16 ***
priceratio:agegroup4(35,60]                 -0.352247   0.008083 -43.579  < 2e-16 ***
priceratio:agegroup4(60,99]                 -0.674209   0.011248 -59.938  < 2e-16 ***
priceratio:genderMALE                       -0.607070   0.116885  -5.194 2.06e-07 ***
priceratio:agevehgroup2(5,10]                0.956935   0.106426   8.992  < 2e-16 ***
priceratio:agevehgroup2(10,99]               0.766736   0.100552   7.625 2.44e-14 ***
priceratio:prembeforegroup2(500,1e+03]       0.569856   0.138151   4.125 3.71e-05 ***
priceratio:prembeforegroup2(1e+03,1e+04]     1.340304   0.285123   4.701 2.59e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 53978  on 56026  degrees of freedom
Residual deviance: 53258  on 56012  degrees of freedom - AIC: 53261
```

Variable list for Subsection 1.3.2

The dataset is quite rich, therefore we have the detailed features of each policy. We write below a subset of the available variables:

- Policy: a dummy variable indicating the lapse, the policy age, the cover type (TPL, PC or FC) and the product, the bonus class for PC and FC covers and the bonus evolution,
- Policyholder: the policyholder age and the gender, the marital status and the job group,
- Premium: the last year premium, the technical premium and the proposed premium, the payment frequency, the market premium, i.e. the tenth lowest NB premium for a particular category,
- Car: the mileage, the vehicle age, the car usage, the car class,
- Cross-selling: the number of AXA contract in household, a dummy variable on household

policy,
- Claims: the claim amount, the claim number per year,
- Agent: the cumulative rebate, the technical rebate, the age difference between the agent and the policyholder.

GLM outputs for Subsection 1.3.2

The regression summary is given below

```
Call: glm(formula = lapse ~ lastprem_group2 + diff2tech + directdebit +
    product + nbclaim0708percust + vehiclage + householdNbPol +
    polholderage + maritalstatus2 + jobgroup2 + gender + polage +
    bonusevol2 + cover + priceratio:(lastprem_group2 + diff2tech +
    paymentfreq + glasscover + region2 + nbclaim08percust + householdNbPol +
    diffdriverPH7 + channel + typeclassTPL + bonusevol2), family = binomial("logit"),
    data = idata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1241  -0.4366  -0.3427  -0.2402   3.3497

Coefficients:
                                        Estimate Std. Error z value Pr(>|z|)
(Intercept)                           -2.6456876  0.1822517 -14.517  < 2e-16 ***
lastprem_group2(500,5e+03]             0.2008839  0.0952157   2.110 0.034878 *
diff2tech                              6.9600797  0.7949370   8.756  < 2e-16 ***
directdebit                           -0.0422104  0.0097823  -4.315 1.60e-05 ***
productT1                             -0.1060909  0.0185019  -5.734 9.80e-09 ***
productT2                             -1.0107703  0.0336376 -30.049  < 2e-16 ***
productT3                             -0.3869057  0.0193135 -20.033  < 2e-16 ***
nbclaim0708percust                     0.0802148  0.0061759  12.988  < 2e-16 ***
vehiclage                             -0.0172387  0.0010180 -16.934  < 2e-16 ***
householdNbPol                        -0.1638354  0.0156899 -10.442  < 2e-16 ***
polholderage                          -0.0106258  0.0003000 -35.417  < 2e-16 ***
maritalstatus2b                       -0.1455813  0.0266586  -5.461 4.74e-08 ***
maritalstatus2d                       -0.1088016  0.0119736  -9.087  < 2e-16 ***
jobgroup2public                       -0.1529926  0.0079183 -19.321  < 2e-16 ***
gender                                -0.0739520  0.0077666  -9.522  < 2e-16 ***
polage                                -0.0245842  0.0006806 -36.123  < 2e-16 ***
bonusevol2up-down                      1.9010618  0.1746998  10.882  < 2e-16 ***
coverpartial compr.                    0.0244814  0.0099107   2.470 0.013504 *
coverTPL                              -0.0349025  0.0131839  -2.647 0.008112 **
priceratio:lastprem_group2(0,500]      1.0418939  0.1840274   5.662 1.50e-08 ***
priceratio:lastprem_group2(500,5e+03]  1.0246974  0.2000580   5.122 3.02e-07 ***
priceratio:diff2tech                  -8.7933934  0.7867136 -11.177  < 2e-16 ***
priceratio:paymentfreq                -0.0136538  0.0010577 -12.909  < 2e-16 ***
priceratio:glasscover                 -0.0865708  0.0139001  -6.228 4.72e-10 ***
priceratio:region2_02-04-05-11         0.3608514  0.0207136  17.421  < 2e-16 ***
priceratio:region2_03-09-10            0.1368317  0.0109978  12.442  < 2e-16 ***
priceratio:region2_04-05-06-07         0.0935641  0.0103280   9.059  < 2e-16 ***
priceratio:region2_12-13               0.3938396  0.0166819  23.609  < 2e-16 ***
priceratio:region2_14-15-16            0.4424354  0.0160587  27.551  < 2e-16 ***
priceratio:region2_17_                 0.4812002  0.0243385  19.771  < 2e-16 ***
priceratio:nbclaim08percust           -0.0374916  0.0102707  -3.650 0.000262 ***
priceratio:householdNbPol              0.0794544  0.0157004   5.061 4.18e-07 ***
priceratio:diffdriverPH7learner 17     0.2768748  0.0578518   4.786 1.70e-06 ***
priceratio:diffdriverPH7only partner   0.0976821  0.0077879  12.543  < 2e-16 ***
priceratio:diffdriverPH7young drivers  0.1684370  0.0148135  11.371  < 2e-16 ***
priceratio:channelbroker               0.3954067  0.0089064  44.396  < 2e-16 ***
priceratio:channeldirect               0.3715832  0.0132034  28.143  < 2e-16 ***
priceratio:typeclassTPL                0.0108773  0.0016963   6.412 1.43e-10 ***
bonusevol2up-down:priceratio          -1.8295464  0.1740807 -10.510  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
    Null deviance: 62279  on 121813  degrees of freedom
Residual deviance: 58739  on 121809  degrees of freedom - AIC: 58747
```

Group j	Observed r_j	Fitted $\frac{1}{n_j}\sum_{i=1}^{n_j}\hat{\pi}_i(p_i)$	Group j	Observed r_j	Fitted $\frac{1}{n_j}\sum_{i=1}^{n_j}\hat{\pi}_i(p_i)$
Agent	8.840	7.714	FC	8.962	7.492
Broker	9.245	8.896	PC	9.464	8.846
Direct	11.837	9.005	TPL	10.222	12.522

Table 1.17: Lapse rates (%)

	$\Delta_{1-}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$	$\Delta_{1-}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$
Channel agent	-0.983	8.652	1.23	-0.759	8.732	0.75
Channel broker	-1.344	9.123	1.841	-1.255	9.422	1.299
Channel direct	-1.246	12.341	1.143	-1.18	11.597	1.268
Channel	One fit by channel			One fit for all channels		
	$\Delta_{1-}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$	$\Delta_{1-}(5\%)$	$\hat{\pi}_n(1)$	$\Delta_{1+}(5\%)$
Coverage FC	-0.926	8.297	1.01	-0.622	8.723	0.97
Coverage PC	-0.635	9.347	1.195	-0.714	9.244	1.063
Coverage TPL	-0.973	12.011	1.876	-0.899	10.179	1.178
Coverage	One fit by coverage			One fit for all coverages		

Table 1.18: Predicted lapse rates by channel and coverage

GLM outputs for Subsection 1.4.2

The regression summary without using the market proxy is given below.

```
Call: glm(formula = lapse ~ diff2tech + product2 + region2 + cumulrebate3 +
    nbclaim0608percust + isinsuredinhealth + isinsuredinlife +
    vehiclage + householdNbPol + polholderage + maritalstatus2 +
    jobgroup2 + gender + typeclassTPL + bonusevol2 + priceratio:(diff2tech +
    paymentfreq + nbclaim08percust + nbclaim0608percust + nbclaim0708percust +
    isinsuredinaccident + householdNbPol + gender + typeclassTPL +
    bonusevol2), family = binomial("logit"), data = idata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2613  -0.4104  -0.3482  -0.2792   3.1127

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -1.3513224  0.1034727 -13.060  < 2e-16 ***
diff2tech                    7.8972018  1.4461272   5.461 4.74e-08 ***
product2T1                  -0.1275087  0.0321359  -3.968 7.25e-05 ***
product2T2                  -0.2762145  0.0348857  -7.918 2.42e-15 ***
region2_02-04-11             0.2886433  0.0427885   6.746 1.52e-11 ***
region2_05                   0.1878357  0.0277600   6.766 1.32e-11 ***
region2_08-09                0.0661201  0.0259573   2.547 0.010857 *
region2_10                   0.4506006  0.0906820   4.969 6.73e-07 ***
region2_12-13                0.3729663  0.0404406   9.223  < 2e-16 ***
region2_14-15-16             0.4591227  0.0406760  11.287  < 2e-16 ***
```

```
region2_17                          0.4469127  0.0609890   7.328 2.34e-13 ***
cumulrebate3                        0.0131512  0.0220328   0.597 0.550581
nbclaim0608percust                  0.2538161  0.0861386   2.947 0.003213 **
isinsuredinhealth                  -0.2117021  0.0737189  -2.872 0.004082 **
isinsuredinlife                    -0.0904838  0.0403864  -2.240 0.025061 *
vehiclage                          -0.0418472  0.0024594 -17.015  < 2e-16 ***
householdNbPol                     -0.1608386  0.0347312  -4.631 3.64e-06 ***
polholderage                       -0.0142367  0.0007987 -17.824  < 2e-16 ***
maritalstatus2b                    -0.2473493  0.0756033  -3.272 0.001069 **
maritalstatus2d                    -0.1026557  0.0339761  -3.021 0.002516 **
jobgroup2public                    -0.1564253  0.0212887  -7.348 2.01e-13 ***
gender                             -0.8573031  0.1748974  -4.902 9.50e-07 ***
typeclassTPL                       -0.1127455  0.0320514  -3.518 0.000435 ***
bonusevol2up-down                   3.5129944  0.6064173   5.793 6.91e-09 ***
priceratio:diff2tech               -8.7833478  1.4474939  -6.068 1.30e-09 ***
priceratio:paymentfreq             -0.0314041  0.0025894 -12.128  < 2e-16 ***
priceratio:nbclaim08percust        -0.1047064  0.0383473  -2.730 0.006324 **
priceratio:nbclaim0608percust      -0.2269052  0.0913726  -2.483 0.013017 *
priceratio:nbclaim0708percust       0.1429228  0.0365854   3.907 9.36e-05 ***
priceratio:isinsuredinaccident     -0.1395317  0.0505194  -2.762 0.005746 **
priceratio:householdNbPol           0.0817417  0.0347087   2.355 0.018519 *
priceratio:gender                   0.7813407  0.1758044   4.444 8.81e-06 ***
priceratio:typeclassTPL             0.1300911  0.0320887   4.054 5.03e-05 ***
priceratio:bonusevol2up-down       -3.3300573  0.6048578  -5.506 3.68e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null deviance: 9151  on 18893  degrees of freedom
Residual deviance: 8866  on 18860  degrees of freedom - AIC: 8873
```

GLM outputs for Subsection 1.5.3

	Poisson	zeroinfl. Poisson	zeroinfl. NB
$\log \mathcal{L}$	-27571	-28372	-28105
AIC	45197	46797	46258
Deg. of free.	27	26	26

Table 1.19: Model adequacy for claim frequency of FC agent

Here follows the regression summary for zero-inflated NB distribution fit.

```
Call: zeroinfl(formula = nbclaim08FC ~ bonuspercentnew + bonusevol2 + lastprem_group2
+ isinsuredinhealth + isinsuredinlife + isinsuredinaccident + polage + vehiclage + polholderage
+ typeclassFC + diffdriverPH2 + gender |  lastprem_group2 + diff2tech
+ isinsuredinaccident + polage + polholderage, data = subdata, dist = "negbin")

Pearson residuals:
    Min      1Q  Median      3Q     Max
-0.6907 -0.3701 -0.3263 -0.2836 27.6615

Count model coefficients (negbin with log link):
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                -2.5053555  0.0463173 -54.091  < 2e-16 ***
bonuspercentnew            -0.0045481  0.0004473 -10.168  < 2e-16 ***
bonusevol2up-down           0.2814031  0.0108215  26.004  < 2e-16 ***
lastprem_group2(500,5e+03]  0.2867385  0.0125864  22.782  < 2e-16 ***
isinsuredinhealth           0.2536512  0.0129962  19.517  < 2e-16 ***
isinsuredinlife             0.1500995  0.0101994  14.716  < 2e-16 ***
isinsuredinaccident         0.1545091  0.0132603  11.652  < 2e-16 ***
polage                     -0.0045662  0.0008071  -5.657 1.54e-08 ***
vehiclage                  -0.0116381  0.0012641  -9.207  < 2e-16 ***
polholderage                0.0052154  0.0006398   8.152 3.59e-16 ***
```

```
typeclassFC                     0.0259947  0.0012908  20.139  < 2e-16 ***
diffdriverPH2all drivers > 24   0.1603390  0.0110572  14.501  < 2e-16 ***
diffdriverPH2commercial         0.5143316  0.0338102  15.212  < 2e-16 ***
diffdriverPH2learner 17         0.2501158  0.0642750   3.891 9.97e-05 ***
diffdriverPH2same              -0.1661160  0.0111876 -14.848  < 2e-16 ***
diffdriverPH2young drivers      0.2524112  0.0158128  15.962  < 2e-16 ***
gender                         -0.0593577  0.0088454  -6.711 1.94e-11 ***
Log(theta)                      0.2848294  0.0330418   8.620  < 2e-16 ***

Zero-inflation model coefficients (binomial with logit link):
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                -7.299505   0.367536 -19.861  < 2e-16 ***
lastprem_group2(500,5e+03] -0.484487   0.081025  -5.979 2.24e-09 ***
diff2tech                  -7.214606   0.562964 -12.815  < 2e-16 ***
isinsuredinaccident        -0.256634   0.098848  -2.596  0.00942 **
polage                     -0.011704   0.004260  -2.747  0.00601 **
polholderage                0.094674   0.004658  20.326  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 1.3295
Number of iterations in BFGS optimization: 77
Log-likelihood: -2.81e+04 on 24 Df
```

GLM outputs for Subsection 1.5.3

It follows the regression summary for ordered logistic regression for FC agent subset.

```
Call: polr(formula = deductibleFC3 ~ nbclaim08FC + ClaimNBhat + bonuspercentnew +
    lastprem_group2 + diff2tech + isinsuredinaccident + polage +
    vehiclage + polholderage + typeclassFC, data = subdata, Hess = TRUE,
    method = "logistic")

Coefficients:
                               Value Std. Error    t value pvalue
nbclaim08FC                -2.900e-02  8.425e-03 -3.442e+00  0.180
ClaimNBhat                  1.656e+00  9.401e-02  1.762e+01  0.036
bonuspercentnew             1.391e-02  3.357e-04  4.143e+01  0.015
lastprem_group2(500,5e+03] -3.026e-01  1.129e-02 -2.679e+01  0.024
diff2tech                  -1.720e+00  6.900e-02 -2.493e+01  0.026
isinsuredinaccident        -2.964e-01  9.988e-03 -2.968e+01  0.021
polage                     -2.789e-02  3.594e-04 -7.759e+01  0.008
vehiclage                   4.625e-02  1.056e-03  4.381e+01  0.015
polholderage               -9.538e-03  2.921e-04 -3.266e+01  0.019
typeclassFC                 1.169e-01  1.154e-03  1.013e+02  0.006

Intercepts:
        Value    Std. Error t value
0|150    -2.3565   0.0354   -66.5322
150|300  -0.4060   0.0334   -12.1655
300|500   4.1764   0.0341   122.4217

Residual Deviance: 664289.21
AIC: 664315.21
```

The GLM regression summary for lapse on the FC agent subset including deductible choice probabilities is available on request to the author.

GAM outputs for Subsection 1.6.2

Below we give the regression summary for the TPL agent dataset. Other summaries are available on request to the author.

```
Family: binomial - Link function: logit

Formula: lapse ~ product2 + region2 + cumulrebate3 + nbclaim0608percust +
    isinsuredinhealth + isinsuredinlife + vehiclage + householdNbPol +
    polholderage + maritalstatus2 + jobgroup2 + gender + bonusevol2 +
    priceratio:(paymentfreq + nbclaim08percust + nbclaim0608percust +
        nbclaim0708percust + isinsuredinaccident + bonusevol2) +
    s(priceratio, diff2tech) + s(priceratio, diff2top10agent) +
    s(priceratio, diff2top10direct) + s(priceratio, typeclassTPL)

Parametric coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -0.9881832  0.0744176 -13.279  < 2e-16 ***
product2T1                    -0.2957239  0.0365839  -8.083 6.30e-16 ***
product2T2                    -0.5888125  0.0439784 -13.389  < 2e-16 ***
region2_02-04-11               0.2474500  0.0432128   5.726 1.03e-08 ***
region2_05                     0.1820856  0.0279436   6.516 7.21e-11 ***
region2_08-09                  0.0627676  0.0260959   2.405 0.016161 *
region2_10                     0.4597820  0.0908178   5.063 4.13e-07 ***
region2_12-13                  0.3600178  0.0408722   8.808  < 2e-16 ***
region2_14-15-16               0.4440049  0.0377465  11.763  < 2e-16 ***
cumulrebate3                   0.1287561  0.0241245   5.337 9.44e-08 ***
nbclaim0608percust             0.2144964  0.0968126   2.216 0.026720 *
isinsuredinhealth             -0.2018414  0.0739308  -2.730 0.006331 **
isinsuredinlife               -0.0978298  0.0405763  -2.411 0.015908 *
vehiclage                     -0.0367641  0.0025963 -14.160  < 2e-16 ***
householdNbPol                -0.0783881  0.0048668 -16.107  < 2e-16 ***
polholderage                  -0.0150938  0.0008334 -18.111  < 2e-16 ***
maritalstatus2b               -0.2629597  0.0760885  -3.456 0.000548 ***
maritalstatus2d               -0.1017553  0.0341228  -2.982 0.002863 **
jobgroup2public               -0.1161175  0.0217312  -5.343 9.12e-08 ***
gender                        -0.0790535  0.0209269  -3.778 0.000158 ***
bonusevol2up-down              7.4827223  1.0625789   7.042 1.89e-12 ***
priceratio:paymentfreq        -0.0343715  0.0026481 -12.980  < 2e-16 ***
priceratio:nbclaim08percust   -0.0893319  0.0393116  -2.272 0.023062 *
priceratio:nbclaim0608percust -0.2010502  0.1016136  -1.979 0.047864 *
priceratio:nbclaim0708percust  0.1538349  0.0369590   4.162 3.15e-05 ***
priceratio:isinsuredinaccident -0.1409923  0.0508941  -2.770 0.005600 **
priceratio:bonusevol2up-down  -7.2677291  1.0573222  -6.874 6.26e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                edf Ref.df Chi.sq  p-value
s(priceratio,diff2tech)       12.440 16.687 113.56  < 2e-16 ***
s(priceratio,diff2top10agent)     8.901 12.069  29.36  0.00361 **
s(priceratio,diff2top10direct)  8.177 11.277  18.63  0.07569 .
s(priceratio,typeclassTPL)      4.160  5.687  43.91 5.43e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0176   Deviance explained = 3.46%
REML score =  44028  Scale est. = 1         n = 187733
```

Bibliography

Aalen, O., Borgan, O. and Gjessing, H. (2008), *Survival and Event History Analysis*, Springer.

Andersen, P., Borgan, O., Gill, R. and Keiding, N. (1995), *Statistical Models Based on Counting Processes*, Springer, Corrected Edition.

Atkins, D. C. and Gallop, R. J. (2007), 'Re-thinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models', *Journal of Family Psychology* **21**(4), 726–735.

Bella, M. and Barone, G. (2004), 'Price-elasticity based on customer segmentation in the italian auto insurance market', *Journal of Targeting, Measurement and Analysis for Marketing* **13**(1), 21–31.

Bland, R., Carter, T., Coughlan, D., Kelsey, R., Anderson, D., Cooper, S. and Jones, S. (1997), Workshop - customer selection and retention, *in* 'General Insurance Convention & ASTIN Colloquium'.

Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J. and Perez-Marin, A. M. (2008), 'Survival analysis of a household portfolio insurance policies: How much time do you have to stop total customer defection?', *Journal of Risk and Insurance* **75**(3), 713–737.

Chiappori, P.-A. and Salanié, B. (2000), 'Testing for asymmetric information in insurance markets', *Journal of Political Economy* **108**(1), 56–78.

Clark, D. R. and Thayer, C. A. (2004), 'A primer on the exponential family of distributions', *2004 call paper program on generalized linear models* .

Cleveland, W. S. (1979), 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association* **368**(829-836).

Cox, D. R. (1972), 'Regression models and life-tables', *Journal of the Royal Statistical Society: Series B* **34**(2), 187–200.

Cummins, J. D. and Venard, B. (2007), *Handbook of international insurance*, Springer.

Dardanoni, V. and Donni, P. L. (2008), Testing for asymmetric information in insurance markets with unobservable types. HEDG working paper.

Demgne, E. J. (2010), Etude des cycles de réassurance, Master's thesis, ENSAE.

Dionne, G., Gouriéroux, C. and Vanasse, C. (2001), 'Testing for evidence of adverse selection in the automobile insurance market: A comment', *Journal of Political Economy* **109**(2), 444–453.

Dreyer, V. (2000), Study the profitability of a customer, Master's thesis, ULP - magistère d'actuariat. Confidential memoir - AXA Insurance U.K.

Dutang, C. (2011), Regression models of price elasticity in non-life insurance, Master's thesis, ISFA.

Dutang, C. (2012), The customer, the insurer and the market. Working paper, ISFA.

Dutang, C., Albrecher, H. and Loisel, S. (2012), A game to model non-life insurance market cycles. Working paper, ISFA.

Fahrmeir, L. (1994), 'Dynamic modelling and penalized likelihood estimation for discrete time survival data', *Biometrika* **81**(2), 317–330.

Faraway, J. J. (2006), *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Parametric Regression Models*, CRC Taylor& Francis.

Frees, E. W. (2004), *Longitudinal and Panel Data*, Cambridge University Press.

Guillen, M., Parner, J., Densgsoe, C. and Perez-Marin, A. M. (2003), *Using Logistic Regression Models to Predict and Understand Why Customers Leave an Insurance Company*, Vol. 6 of *Innovative Intelligence* Shapiro and Jain (2003), chapter 13.

Hamel, S. (2007), Prédiction de l'acte de résiliation de l'assuré et optimisation de la performance en assurance automobile particulier, Master's thesis, ENSAE. Mémoire confidentiel - AXA France.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized Additive Models*, Chapman and Hall.

Hastie, T. J. and Tibshirani, R. J. (1995), 'Generalized additive models', *to appear in Encyclopedia of Statistical Sciences* .

Jackman, S. (2011), *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University*, Department of Political Science, Stanford University. R package version 1.04.1.

Kagraoka, Y. (2005), Modeling insurance surrenders by the negative binomial model. Working Paper 2005.

Kelsey, R., Anderson, D., Beauchamp, R., Black, S., Bland, R., Klauke, P. and Senator, I. (1998), Workshop - price/demand elasticity, *in* 'General Insurance Convention & ASTIN Colloquium'.

Kim, C. (2005), 'Modeling surrender and lapse rates with economic variables', *North American Actuarial Journal* **9**(4), 56–70.

Loisel, S. and Milhaud, X. (2011), 'From deterministic to stochastic surrender risk models: Impact of correlation crises on economic capital', *European Journal of Operational Research* **214**(2).

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, 2nd edn, Chapman and Hall.

McFadden, D. (1981), Econometric Models of Probabilistic Choice, *in* 'Structural Analysis of Discrete Data with Econometric Applications', The MIT Press, chapter 5.

Milhaud, X., Maume-Deschamps, V. and Loisel, S. (2011), 'Surrender triggers in Life Insurance: What main features affect the surrender behavior in a classical economic context?', *Bulletin Français d'Actuariat* **22**(11).

Nelder, J. A. and Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society* **135**(3), 370–384.

Ohlsson, E. and Johansson, B. (2010), *Non-Life Insurance Pricing with Generalized Linear Models*, Springer.

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org*

Rabehi, H. (2007), Study of multi-risk household, Master's thesis, ISUP. Mémoire confidentiel - AXA France.

Rothschild, M. and Stiglitz, J. E. (1976), 'Equilibrium in competitive insurance markets: An essay on the economics of imperfect information', *The Quarterly Journal of Economics* **90**(4), 630–649.

Sergent, V. (2004), Etude de la sensibilité de l'assuré au prix en assurance auto de particuliers, Master's thesis, ISUP. Mémoire confidentiel - AXA France.

Shapiro, A. F. and Jain, L. C. (2003), *Intelligent and Other Computational Techniques in Insurance*, World Scientific Publishing.

Steihaug, T. (2007), Splines and b-splines: an introduction, Technical report, University of Oslo.

Su, L. and White, H. (2003), Testing conditional independence via empirical likelihood. UCSD Department of Economics Discussion Paper.

Turner, H. (2008), Introduction to generalized linear models, Technical report, Vienna University of Economics and Business.

Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with S*, 4th edn, Springer.

Wood, S. N. (2001), 'mgcv: GAMs and Generalized Ridge Regression for R', *R News* **1**, 20–25.

Wood, S. N. (2003), 'Thin plate regression splines', *Journal of the Royal Statistical Society: Series B* **65**(1), 95–114.

Wood, S. N. (2008), 'Fast stable direct fitting and smoothness selection for generalized additive models', *Journal of the Royal Statistical Society: Series B* **70**(3).

Wood, S. N. (2010), 'Fast stable reml and ml estimation of semiparametric glms', *Journal of the Royal Statistical Society: Series B* **73**(1), 3–36.

Yeo, A. C. and Smith, K. A. (2003), *An integretated Data Mining Approach to Premium Pricing for the Automobile Insurance Industry*, Vol. 6 of *Innovative Intelligence* Shapiro and Jain (2003), chapter 5.

Zeileis, A., Kleiber, C. and Jackman, S. (2008), 'Regression models for count data in r', *Journal of Statistical Software* **27**(8).

Théorie des jeux

Chapitre 2

Theorie des jeux et cycles de marché

—

A game-theoretic approach to non-life insurance market cycles

Little by little, one travels far.
J.R.R. Tolkien (1892–1973)

Ce chapitre se base sur l'article Dutang et al. (2012) dont une partie est déjà soumise à l'*European Journal of Operation Research*.

2.1 Introduction

Insurance market cycles and the study of their causes have been puzzling actuaries for many years. Feldblum (2001) discusses four main causes that could explain the presence of underwriting through their aggregate effect. These causes are (i) actuarial pricing procedure, (ii) underwriting philosophy, (iii) interest rate fluctuations and (iv) competitive strategies. He compares contributions through out the 20^{th} century on the topic, see also Markham (2007) for an overview.

Actuarial pricing procedures are subject to claim cost uncertainty, information lag (due to accouting, regulatory and legal standards). Such effects are likely to generate fluctuations around an equilibrium price, when extrapolating premiums, see, e.g., Venezian (1985); Cummins and Outreville (1987). In addition, a hard behavior of underwriters combined with a lack of coordination is an extra recipe for underwriting cycles. In particular, price policies cannot be sold independently of the market premium, but neither can the market premium be driven by one's individual actions. This is called underwriting philosophy by Feldblum (2001), and is also noticed in Jablonowski (1985), who assumes (i) insurers do not make decisions in isolation from other firms in the market, and (ii) profit maximization is not the exclusive, or even the most important, motivation of insurers. Interest rate deviations further increase the frequency and the amplitude of market cycles, as they have an impact on the investment result and (indirectly) on the maximum rebate that underwriters can afford to keep presumably low-risk customers. Fields and Venezian (1989) were among the first to demonstrate this effect. Finally, the competition level on most mature insurance markets are such that any increase in market share can only be carried out by price decrease * (due to very little product differentiation). Coupled with capital constraints (e.g. Gron (1994)) and price inelasticity, insurers are forced not to deviate too much from market trends.

On a different level, basic economic models suggest that the equilibrium premium is the marginal cost, as any upward deviation from this equilibrium will result in losing all the policies in the next period. This theory would imply that all insurers price at the market premium However, in practice customers do not move from a insurer to a cheaper insurer as swiftly as economic models anticipate. There is an inertia of the insurance demand, preventing all insured to shop arround for the cheapest insurer when their premium is slightly higher than the market premium. So customer behavior is much more complicated. In addition to customer loyalty, Feldblum (2001) points out that it is difficult for a new insurer to enter successfully into the non-life insurance market.

More refined economic models focus on moral hazard and adverse selection. The celebrated model of Rothschild and Stiglitz (see Rothschild and Stiglitz (1976)) deals with a utility-based agent framework where insureds have private information on their own risk. Insurers provide a menu of contracts (a pair of premium and deductible) and high-risk individuals choose full coverage, whereas low-risk individuals are more attracted by partial coverage. Note that the equilibrium price may not exist if all insurers offer just one type of contract. Picard (2009) considers an extension by allowing insurers to offer participating contracts (such as mutual-type contracts). This feature guarantees the existence of an equilibrium, which forces (rational) insureds to reveal their risk level. An important source of applications of such models is health insurance where moral hazard and adverse selection play a major role, see,

∗. The hunger for market share is driven by the resulting reduction of claim uncertainty is increasing the policy number, which is motivated by the law of large numbers.

e.g., Geoffard et al. (1998), Wambach (2000); Mimra and Wambach (2010) and Picard (2009).

However, the economic models mentioned above can not address the insurance market cycle dynamics, so that one has to look for further alternatives. Taylor (1986, 1987) deals with discrete-time underwriting strategies of insurers and provides first attempts to model strategic responses to the market, see also, Kliger and Levikson (1998); Emms et al. (2007); Moreno-Codina and Gomez-Alvado (2008). The main pitfall of the optimal control approach is that it focuses on one single insurer and thus implicitly assumes that insurers are playing a game against an impersonal market player and the market price is independent of their own actions.

In this paper, we want to investigate the suitability of game theory for insurance market modelling. The use of game theory in actuarial science has a long history dating back to K. Borch and J. Lemaire, who mainly used cooperative games to model risk transfer between insurer and reinsurer, see, e.g., Borch (1960, 1975), Lemaire and Quairière (1986). Bühlmann (1984) and Golubin (2006) also studied risk transfer with cooperative tools. Among the articles using noncooperative game theory to model the non-life insurance market, Bertrand oligopoly models are studied by Polborn (1998), Rees et al. (1999), Hardelin and de Forge (2009). Powers and Shubik (1998, 2006) also study scale effects of the number of insurers and the optimal number of reinsurers in a market model having a central clearing house. More recently, Taksar and Zeng (2011) study non-proportional reinsurance with stochastic continuous-time games. Demgne (2010) seems to be the first study from a game theory point of view of (re)insurance market cycles. She uses well known economic models: pure monopoly, Cournot's oligopoly (i.e. war of quantity), Bertrand's oligopoly (i.e. war of price) and Stackelberg (leader/follower game). For all these models, she tests various scenarios and checks the consistency of model outputs with reinsurance reality.

Finally, in many ruin theory models, one assumes that the portfolio size remains constant over time (see, e.g., Asmussen and Albrecher (2010) for a recent survey). Non-homogeneous claim arrival processes have usually been studied in the context of modelling catastrophe events. More recently, non-constant portfolio size has been considered, see, e.g., Trufin et al. (2009) and the references therein. Malinovskii (2010) uses a ruin framework to analyze different situations for an insurer in its behavior against the market.

This paper aims to model competition and market cycles in non-life insurance markets with noncooperative game theory in order to extend the player-vs-market reasoning of Taylor (1986, 1987)'s models. A main contribution is to show that incorporating competition when setting premiums leads to a significant deviation from the actuarial premium and from a one-player optimized premium. Furthermore, the repeated game models a rational behavior of insurers in setting premium in a competitive environment, although the resulting market premium is cyclical. The rest of the paper is organized as follows. Section 2.2 introduces a one-period noncooperative game. Existence and uniqueness of premium equilibrium are established. Section 2.3 relaxes assumptions on objective and constraint components of the one-period model. The existence of premium equilibrium is still guaranteed, but uniqueness may not hold. A reasonable choice of an equilibrium is proposed in this situation. Section 2.4 then works out with the repeated version of the one-period model of Section 2.3. A conclusion and perspectives are given in Section 2.5.

2.2 A one-period model

In a first attempt to model the non-life insurance market cycle, we ignore for simplicity investment results, although they play a key role for third-part liability insurance product for which interest rate fluctuations have a big impact, as well as loss reserving. So, our framework is consistent only for short-tail business.

Consider I insurers competing in a market of n policyholders with one-year contracts (where n is considered constant). The "game" for insurers is to sell policies to this insured market by setting the premium. Let $(x_1, \dots, x_I) \in \mathbb{R}^I$ be a price vector, with x_j representing premium of insurer j. Once the premium is set by all insurers, the insureds choose to renew or to lapse from their current insurer. Then, insurers pay claims, according to their portfolio size, during the coverage year. At the end of the year, underwriting results are determined, and insurer capital is updated: some insurer may be bankrupt.

In the next subsections, we present the four components of the game: a lapse model, a loss model, an objective function and a solvency constraint function. In the sequel, a subscript $j \in \{1, \dots, I\}$ will always denote a player index whereas a subscript $i \in \{1, \dots, n\}$ denotes an insured index.

2.2.1 Lapse model

Being with current insurer j, the insurer choice C_i of insured i for the next period follows an I-dimensional multinomial distribution $\mathcal{M}_I(1, p_{j\to})$ with probability vector $p_{j\to} = (p_{j\to 1}, \dots, p_{j\to I})$ summing to 1. The probability mass function is given by $P(C_i = k \mid j) = p_{j\to k}$. It seems natural and it has been verified empirically that the probability to choose an insurer is highly influenced by the previous period choice. In other words, the probability to lapse $p_{j\to k}$ with $k \neq j$ is generally much lower than the probability to renew $p_{j\to j}$. To our knowledge, only the UK market shows lapse rates above 50%. Those probabilities have to depend on the premium x_j, x_k proposed by insurer j and k, respectively.

Assume at the beginning of the game that the insurer portfolio sizes are n_j (such that $\sum_{j=1}^{I} n_j = n$). The portfolio size $N_j(x)$ of insurer j for the next period is a random variable determined by the sum of renewed policies and businesses coming from other insurers. Hence,

$$N_j(x) = B_{jj}(x) + \sum_{k=1, k\neq j}^{I} B_{kj}(x).$$

$N_j(x)$ is a sum of I independent binomial variables $(B_{kj})_k$ where B_{kj} has parameters $\mathcal{B}(n_k, p_{k\to j}(x))$.

In the economics literature, $p_{j\to k}$ is considered in the framework of discrete choice models. In the random utility maximization setting, McFadden (1981) or Anderson et al. (1989) propose multinomial logit and probit probability choice models. In this paper, we choose a multinomial logit model, since the probit link function does not really enhance the choice model despite its additional complexity. Working with unordered choices, we arbitrarily set the insurer reference category for $p_{j\to k}$ to j, the current insurer. We define the probability for a customer to go from insurer j to k given the price vector x by the following multinomial logit model

$$p_{j\to k} = \lg_j^k(x) = \begin{cases} \dfrac{1}{1+\sum_{l\neq j} e^{f_j(x_j, x_l)}} & \text{if } \ j = k, \\ \dfrac{e^{f_j(x_j, x_k)}}{1+\sum_{l\neq j} e^{f_j(x_j, x_l)}} & \text{if } \ j \neq k, \end{cases} \tag{2.1}$$

where the sum is taken over the set $\{1, \ldots, I\}$ and f_j is a price sensitivity function. In the following, we consider two types of price functions

$$\bar{f}_j(x_j, x_l) = \mu_j + \alpha_j \frac{x_j}{x_l} \quad \text{and} \quad \tilde{f}_j(x_j, x_l) = \tilde{\mu}_j + \tilde{\alpha}_j(x_j - x_l).$$

The first function $\bar{f}_j$ assumes a price sensitivity with the ratio of the proposed premium x_j and competitor premium x_l, whereas $\tilde{f}_j$ works with the premium difference $x_j - x_l$. Parameters μ_j, α_j represent a base lapse level and price sensitivity. We assume that insurance products display positive price-elasiticity of demand $\alpha_j > 0$. One can check that $\sum_k \lg_j^k(x) = 1$.

The above expression can be rewritten as

$$\lg_j^k(x) = \lg_j^j(x) \left(\delta_{jk} + (1 - \delta_{jk}) e^{f_j(x_j, x_k)} \right),$$

with δ_{ij} denoting the Kronecker product. It is difficult to derive general properties of the distribution of a sum of binomial variables with different probabilities, except when the size parameters n_j are reasonably large, in which case the normal approximation is appropriate. With this insurer choice probability, the expected portfolio size of insurer j reduces to

$$\hat{N}_j(x) = n_j \times \lg_j^j(x) + \sum_{l \neq j} n_l \times \lg_l^j(x),$$

where n_j denotes the last year portfolio size of insurer j.

2.2.2 Loss model

Let Y_i be the aggregate loss of policy i during the coverage period. We assume no adverse selection among insured of any insurers, i.e. Y_i are independent and identically distributed (i.i.d.) random variables, $\forall i = 1, \ldots, n$. As already mentioned, we focus on short-tail business. Thus, we assume a simple frequency – average severity loss model

$$Y_i = \sum_{l=1}^{M_i} Z_{i,l},$$

where the claim number M_i is independent from the claim severities $Z_{i,l}$. Therefore, the aggregate claim amount for insurer j is

$$S_j(x) = \sum_{i=1}^{N_j(x)} Y_i = \sum_{i=1}^{N_j(x)} \sum_{l=1}^{M_i} Z_{i,l},$$

where $N_j(x)$ is the portfolio size of insurer j given the price vector x. We consider two main cases of the loss models: (i) Poisson-lognormal model: $M_i \overset{i.i.d.}{\sim} \mathcal{P}(\lambda)$ and $Z_{i,l} \overset{i.i.d.}{\sim} \mathcal{LN}(\mu_1, \sigma_1^2)$, (ii) negative binomial-lognormal model: $M_i \overset{i.i.d.}{\sim} \mathcal{NB}(r, p)$ and $Z_{i,l} \overset{i.i.d.}{\sim} \mathcal{LN}(\mu_2, \sigma_2^2)$. We choose a different parameter set for the claim severity distribution, because if we want a significant difference between the two loss models, changing only the claim number distribution does not reveal sufficient. These two instances of the frequency-average severity model are such the aggregate claim amount $S_j(x) = \sum_{i=1}^{N_j(x)} Y_i$ is still a compound distribution of the same kind, since Y_i are assumed i.i.d. random variables.

Hence, the insurer aggreggate claim amount $S_j(x)$ is a compound distribution $\sum_{l=1}^{\widetilde{M}_j(x)} Z_l$ such that the claim number M_j and claim severity Z_l follow

– a Poisson-lognormal with $\widetilde{M}_j(x) \sim \mathcal{P}(N_j(x)\lambda)$ and $Z_l \overset{i.i.d.}{\sim} \mathcal{LN}(\mu_1, \sigma_1^2)$,
– a negative binomial-lognormal with $\widetilde{M}_j(x) \sim \mathcal{NB}(N_j(x)r, p)$ and $Z_l \overset{i.i.d.}{\sim} \mathcal{LN}(\mu_2, \sigma_2^2)$.

In the numerical applications, these two loss models are denoted PLN and NBLN, respectively.

In addition to these two base loss models, we will also test a variation of the Negative binomial model, in which the claim numbers are correlated among insurers. Concretely,

– draw u from a uniform distribution $\mathcal{U}(0,1)$.
– set $\lambda_j = Q_j(u)$ where Q_j is the quantile function a random Gamma variable with shape parameter $N_j(x)r$ and rate parameter $\frac{p}{1-p}$.
– draw a compound variable S_j with claim frequency $\mathcal{P}(\lambda_j)$ and claim frequency $\mathcal{LN}(\mu_2, \sigma_2^2)$.

Since a Poisson-Gamma mixture follows a negative binomial distribution, the resulting margingal claim number distribution for insurer j is a Negative Binomial distribution with parameters $\mathcal{BN}(N_j(x)r, p)$. However, now the loss frequency among insurers is comonotonic. We will denote this model by PGLN in the numerical applications.

2.2.3 Objective function

In the two previous subsections, we presented two components of the insurance markets: the lapse model (how insureds react to premium changes) and the loss model (how insureds face claims). We now turn our attention to the underwriting strategy of insurers, i.e. on how they set premiums.

In Subsection 2.2.1, we assumed that price elasticity of demand for the insurance product is positive. Thus, if the whole market underwrites at a loss, any actions of a particular insurer to get back to profitability will result in a reduction of his business volume. This has two consequences for possible choice of objective functions: (i) it should use a decreasing demand function of price x_j given the competitors price $x_{-j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_I)$ and (ii) it should depend on an assessment of the insurer break-even premium per unit of exposure π_j.

We suppose that insurer j maximizes the expected profit of renewing policies defined as

$$O_j(x) = \frac{n_j}{n}\left(1 - \beta_j\left(\frac{x_j}{m_j(x)} - 1\right)\right)(x_j - \pi_j), \tag{2.2}$$

where π_j is the break-even premium j and $m_j(x)$ is a market premium proxy. The objective function O_j defined as the product of a demand function and an expected profit per policy represent a company-wide expected profit. O_j targets renewal business and does not take into account new business explicitly. In addition to focusing on renewal business only, the objective function locally approximates the true insurer choice probability $\lg_j^j$ presented in Subsection 2.2.1. However, since the demand function $D_j(x) = n_j/n(1 - \beta_j(x_j/m_j(x) - 1))$ is not restricted to [0,1], demand D_j can exceed the current market share n_j/n, but profit per policy will decline when the premium decreases. Thus, maximising the objective function O_j leads to a trade-off between increasing premium to favour higher projected profit margins and decreasing premium to defend the current market share. Note that O_j has the nice property to be infinitely differentiable.

Parameter π_j corresponds to the estimated mean loss of insurer j and is expressed as

$$\pi_j = \omega_j \overline{a}_{j,0} + (1 - \omega_j)\overline{m}_0$$

where $\overline{a}_{j,0}$ is the actuarial premium defined as empirical average loss per policy over a certain number of past years, the market premium $\overline{m}_0$ is defined as the past average of the market

premium weighted by the gross written premium and $\omega_j \in [0,1]$ is the credibility factor of insurer j. If insurer j is the market leader, then ω_j should be close to 1, whereas when insurer j is a follower, ω_j should be close to 0. Note that π_j takes into account expenses implicitly via the actuarial and the market premiums.

The market proxy used in Equation (2.2) is the mean price of the other competitors

$$m_j(x) = \frac{1}{I-1}\sum_{k\neq j} x_k.$$

The market proxy aims to assess other insurer premiums without specifically targeting one competitor. By excluding the price x_j to compute the market proxy $m_j(x)$, we suppose insurer j is not dominant in the market. If, for example, insurer j underwrites 80% of the total premium available in the market, $m_j(x)$ will not be appropriate, but in such cases the market competition is low. We could have used the minimum of the competitors' premium, but then $m_j(x)$ would not have been a continuous function of the price vector x. Furthermore, insurer j does not necessarily take into account to be the cheapest insurer.

2.2.4 Solvency constraint function

In addition to maximizing a certain objective function, insurers must satisfy a solvency constraint imposed by the regulator. Currently, European insurers report their solvency margin in the Solvency I framework, based on the maximum of a percentage of gross written premium and aggregate claim mean. According to Derien (2010), a non-life insurer computes its solvency margin as

$$\text{SM} = \max(18\% \times \text{GWP}, 26\% \times \text{AC}) \times \max(50\%, \text{AC net of reins/AC gross of reins}),$$

where GWP denotes the gross written premium and AC the aggregate claim mean [*]. Discarding reinsurance, the Solvency I framework leads to a solvency margin

$$\text{SM} = \max(9\% \times \text{GWP}, 13\% \times \text{AC}).$$

This approach is not really satisfactory, as it does not take into account the risk volality of underwritten business. Since 2005, actuaries are well busy with the upcoming Solvency II framework. In this new framework, the quantitative part leads to the computation of two capital values, both based on the difference between a certain quantile and the mean of the aggregate loss. The solvency capital requirement (SCR) is based on the 99.5%-quantile, whereas the minimum capital requirement (MCR) is based on the 85%-quantile.

In our game context, we want to avoid the simplistic Solvency I framework, but still want to keep the tractablity for the SCR computation rule. We recall that the aggregate claim amount is assumed to be a frequency - average severity model, i.e. Cat-losses are ignored. A simplification is to approximate a q-quantile $Q(n,q)$ of aggregate claim amount of n i.i.d. risks by a bilinear function of n and $\sqrt{n}$

$$Q(n,q) = E(Y)n + k_q\sigma(Y)\sqrt{n}, \tag{2.3}$$

∗. The percentages 18% and 26% are replaced respectively by 16% and 23% when the GWP exceeds 57.5 Meur or AC exceeds 40.3 Meur.

where the coefficient k_q has to be determined and Y is the generic individual claim severity variable. The first term corresponds to the mean of the aggregate claim amount, while the second term is related to standard deviation.

Three methods have been tested to compute the solvency coefficient k_q: (i) a normal approximation $k_q^N = \Phi^{-1}(q)$, where Φ is the distribution function of the standard normal distribution, (ii) a simulation procedure with 10^5 sample size to get k_q^S as the empirical quantile and (iii) the Panjer recursion to compute the aggregate claim quantile k_q^P *.

While the normal approximation is based on the first two moments of the distribution only, simulation and Panjer methods need to have assumptions on claim frequency and claim severity distributions: we use the PLN and NBLN models defined in Subsection 2.2.2. We also need a risk number n. In Table 2.1, we report solvency coefficients for $n = 1000$ risks. Panjer and simulation methods appear twice since two loss models (PLN and NBLN) are tested.

prob q	k_q^N	k_q^P-PLN	k_q^P-NBLN	k_q^S-PLN	k_q^S-NBLN
0.75	0.674	1.251	0.913	0.649	0.627
0.8	0.842	1.431	1.104	0.829	0.812
0.85	1.036	1.642	1.332	1.029	1.03
0.9	1.282	1.912	1.627	1.299	1.312
0.95	1.645	2.321	2.083	1.695	1.759
0.99	2.326	3.117	2.997	2.475	2.633
0.995	2.576	3.419	3.352	2.777	2.976

Table 2.1: Solvency coefficient k

Numerical experiments show that the normal approximation is less conservative for high quantiles (i.e. $k_q^N < k_q^P$) when the claim number follows a negative binomial distribution, and the reverse for the Poisson distribution. Based on this study, we choose to approximate quantiles at 85% and 99.5% levels with coefficients $k_{85} = 1$ and $k_{995} = 3$.

Thus, using the approximation (2.3), the solvency capital requirement SCR is deduced as

$$\mathrm{SCR}_q \approx k_q \sigma(Y)\sqrt{n},$$

which is more complex than the Solvency I framework. Numerical investigations show that the Solvency I requirement corresponds to a 75% quantile. Therefore, we decide to choose the adapted solvency constraint function

$$g_j^1(x_j) = \frac{K_j + n_j(x_j - \pi_j)(1 - e_j)}{k_{995}\sigma(Y)\sqrt{n_j}} - 1, \tag{2.4}$$

where k_{995} is the solvency coefficient and e_j denotes the expense rate as a percentage of gross written premium. The numerator corresponds to the sum of current capital K_j and expected profit on the in-force portfolio (without taking into account new business). It is easy to see that the constraint $g_j^1(x) \geq 0$, is equivalent to $K_j + n_j(x_j - \pi_j)(1 - e_j) \geq k_{995}\sigma(Y)\sqrt{n_j}$, but g_j^1 is normalized with respect to capital, providing a better numerical stability.

In addition to the solvency constraint, we need to impose bounds on the possible premium. A first choice could be simple linear constraints as $x_j - \underline{x} \geq 0$ and $\overline{x} - x_j \geq 0$, where $\underline{x}$

∗. See, e.g., Theorem 12.4.3 of Bowers et al. (1997). Panjer recursion requires that the claim distribution is discrete. So before using Panjer algorithm, we use a lower discretization of the lognormal claim distribution.

and $\overline{x}$ represent the minimum and the maximum premium, respectively. But the following reformulation is equivalent and numerically more stable:

$$g_j^2(x_j) = 1 - e^{-(x_j - \underline{x})} \geq 0 \quad \text{and} \quad g_j^3(x_j) = 1 - e^{-(\overline{x} - x_j)} \geq 0.$$

The minimum premium $\underline{x}$ could be justified by a prudent point of view by regulators while the maximum premium $\overline{x}$ could be set, e.g., by a consumer right defense association. In the sequel, we set $\underline{x} = E(Y)/(1 - e_{min}) < \overline{x} = 3E(Y)$, where e_{min} is the minimum expense rate.

Overall, the constraint function $g_j(x_j) \geq 0$ is equivalent to

$$\{x_j, g_j(x_j) \geq 0\} = \left\{x_j \in [\underline{x}, \overline{x}], \quad K_j + n_j(x_j - \pi_j)(1 - e_j) \geq k_{995}\sigma(Y)\sqrt{n_j}\right\}. \tag{2.5}$$

2.2.5 Game sequence

For noncooperative games, there are two main solution concepts, there are two main solution concepts: Nash equilibrium and Stackelberg equilibrium. The Nash equilibrium assumes player actions are taken simultaneously while for the Stackelberg equilibrium actions take place sequentially, see, e.g., Fudenberg and Tirole (1991); Osborne and Rubinstein (2006). In our setting, we consider the Nash equilibrium as the most appropriate concept. We give below the definition of a generalized Nash equilibrium extending the Nash equilibrium with constraint functions.

Definition. *For a game with I players, with payoff functions O_j and constraint function g_j, a generalized Nash equilibrium is a vector $x^\star = (x_1^\star, \ldots, x_I^\star)$ such that for all $j = 1, \ldots, I$, $x_j^\star$ solves the subproblem*

$$\max_{x_j} O_j(x_j, x_{-j}^\star) \quad s.t. \quad g_j(x_j, x_{-j}^\star) \geq 0.$$

where x_j and x_{-j} denote action of player j and the other players' action, respectively.

A (generalized) Nash equilibrium is interpreted as a point at which no player can profitably deviate, given the actions of the other players. When each player's strategy set does not depend on the other players' strategies, a generalized Nash equilibrium reduces to a standard Nash equilibrium. Our game is a Nash equilibrium problem since our constraint functions g_j defined in Equation (2.4) depend on the price x_j only.

The game sequence is given as follows:

(i) Insurers set their premium according to a generalized Nash equilibrium $x^\star$, solving for all $j \in \{1, \ldots, I\}$

$$x_{-j} \mapsto \underset{x_j, g_j(x_j) \geq 0}{\arg\max} \; O_j(x_j, x_{-j}).$$

(ii) Insureds randomly choose their new insurer according to probabilities $p_{k \to j}(x^\star)$: we get $N_j(x)$.

(iii) For the one-year coverage, claims are random according to a frequency-average severity model relative to the portfolio size $N_j(x^\star)$.

(iv) Finally the underwriting result is determined by $UW_j(x^\star) = N_j(x^\star)x_j^\star(1-e_j) - S_j(x^\star)$, where e_j denotes the expense rate.

If the solvency requirement is not fullfilled, in Solvency I, the regulator response is immediate: depending on the insolvency severity, regulators can withdraw the authorisation to underwrite new business or even force the company to go run-off or to sell part of its portfolio. In Solvency II, this happens only when the MCR level is not met. There is a buffer between

MCR and SCR where regulators impose some specific actions to help returning to the SCR level.

In our game, we choose to remove players which have a capital below MCR and to authorize players to continue underwriting when capital is between the MCR and the SCR. Note that the constraint function will be active when computing the Nash equilibrium, if the capital is between the MCR and SCR.

2.2.6 Properties of the premium equilibrium

In this subsection, we investigate the properties of premium equilibrium. We start by showing existence and uniqueness of a Nash equilibrium. Then, we focus on the sensitivity analysis on model parameters of such equilibrium.

Proposition 2.2.1. *The I-player insurance game with objective function and solvency constraint function defined in Equations (2.2) and (2.5), respectively, admits a unique (Nash) premium equilibrium.*

Proof. The strategy set is $R = [\underline{x}, \overline{x}]^I$, which is nonempty, convex and compact. Given $x_{-j} \in [\underline{x}, \overline{x}]$, the function $x_j \mapsto O_j(x)$ is a quadratic function with second-degree term $-\beta_j x_j^2/m_j(x) < 0$ up to a constant n_j/n. Thus, this function is (strictly) concave. Moreover, for all players, the constraint functions g_j^1 are linear functions, hence also concave. By Theorem 1 of Rosen (1965), the game admits a Nash equilibrium, i.e. existence is guaranteed. By Theorem 2 of Rosen (1965), uniqueness is verified if we have the following inequality for all $x, y \in R$,

$$\sum_{j=1}^{I} r_j(x_j - y_j)\nabla_{x_j} O_j(y) + \sum_{j=1}^{I} r_j(y_j - x_j)\nabla_{x_j} O_j(x) > 0, \tag{2.6}$$

for some $r \in \mathbb{R}^I$ with strictly positive components $r_i > 0$. As the function $x_j \mapsto O_j(x)$ is a strictly concave and differentiable function for all x_{-j}, we have $\nabla_{x_j} O_j(x)(y_j - x_j) > O_j(y) - O_j(x)$ and equivalently $\nabla_{x_j} O_j(y)(x_j - y_j) > O_j(x) - O_j(y)$. Thus,

$$(x_j - y_j)\nabla_{x_j} O_j(y) + (y_j - x_j)\nabla_{x_j} O_j(x) > O_j(y) - O_j(x) + O_j(x) - O_j(y) = 0.$$

Taking $r = 1$, equation (2.6) is verified. □

Proposition 2.2.2. *Let $x^\star$ be the premium equilibrium of the I-player insurance game. For each player j, if $x_j^\star \in]\underline{x}, \overline{x}[$, the player equilibrium $x_j^\star$ depends on the parameters in the following way: it increases with break even premium π_j, solvency coefficient k_{995}, loss volatility $\sigma(Y)$, expense rate e_j and decreases with sensitivity parameter β_j and capital K_j. Otherwise when $x_j^\star = \underline{x}$ or $\overline{x}$, the premium equilibrium is independent of any parameters.*

Proof. The premium equilibrium $x_j^\star$ of insurer j solves the necessary Karush-Kuhn-Tucker conditions:

$$\begin{aligned} &\nabla_{x_j} O_j(x^\star) + \sum_{1 \le l \le 3} \lambda_l^{j\star} \nabla_{x_j} g_j^l(x_j^\star) = 0, \\ &0 \le \lambda^{j\star},\ g_j(x_j^\star) \ge 0,\ g_j(x_j^\star)^T \lambda^{j\star} = 0, \end{aligned} \tag{2.7}$$

where $\lambda^{j\star} \in \mathbb{R}^3$ are Lagrange multipliers, see, e.g., Facchinei and Kanzow (2009). In the last part of equation (2.7), $g_j(x_j^\star)^T \lambda^{j\star} = 0$ is the complementarity equation implying that the l constraint g_j^l is either active ($g_j^l(x_j^\star) = 0$) or inactive ($g_j^l(x_j^\star) > 0$), but $\lambda_l^{j\star} = 0$.

We suppose that $x_j^\star \in]\underline{x}, \overline{x}[$. Hence, $\lambda_2^{j\star} = \lambda_3^{j\star} = 0$. There are two cases: either the solvency constraint g_j^1 is active or not. Let us assume the solvency constraint is inactive. Insurer j's premium equilibrium verifies $\nabla_{x_j} O_j(x^\star) = 0$, i.e.

$$\frac{n_j}{n}\left(1 - 2\beta_j \frac{x_j^\star}{m_j(x^\star)} + \beta_j + \beta_j \frac{\pi_j}{m_j(x^\star)}\right) = 0. \tag{2.8}$$

Let x_y^j be the premium vector with the j component being y, i.e. $x_y^j = (x_1, \ldots, x_{j-1}, y, x_{j+1}, \ldots, x_I)$. We denote by z a parameter of interest and define the function F as

$$F_x^j(z, y) = \frac{\partial O_j}{\partial x_j}(x_y^j, z),$$

where the objective function depends (also) on the interest parameter z. Equation (2.8) can be rewritten as $F_{x^\star}^j(z, x_j^\star) = 0$.

By the continuous differentiability of F with respect to z and y and the fact that $F_x^j(z, y) = 0$ has at least one solution (z_0, y_0), we can invoke the implicit function theorem, see Appendix 2.6.1. So there exists a function φ defined in a neighborhood of (z_0, y_0) such that $F_x^j(z, \varphi(z)) = 0$ and $\varphi(z_0) = y_0$. Furthermore, if $\frac{\partial F_x^j}{\partial y}(z_0, y_0) \neq 0$, the derivative of φ is given by

$$\varphi'(z) = -\left.\frac{\frac{\partial F_x^j}{\partial z}(z, y)}{\frac{\partial F_x^j}{\partial y}(z, y)}\right|_{y=\varphi(z)}.$$

In our case, we have

$$\frac{\partial F_x^j}{\partial y}(z, y) = \frac{\partial^2 O_j}{\partial x_j^2}(x_y^j, z) = -2\alpha_j \frac{n_j}{n m_j(x)} < 0.$$

As a consequence, the sign of φ' is simply

$$\operatorname{sign}\big(\varphi'(z)\big) = \operatorname{sign}\left(\frac{\partial F_x^j}{\partial z}(z, \varphi(z))\right).$$

Let us consider $z = \pi_j$. We have

$$\frac{\partial F_x^j}{\partial z}(z, y) = \frac{n_j \beta_j}{n m_j(x)} > 0.$$

Thus, the function $\pi_j \mapsto x_j^\star(\pi_j)$ is increasing.

Let z be the sensitivity coefficient β_j. We have

$$\frac{\partial F_x^j}{\partial z}(z, y) = \frac{n_j}{n}\left(-2\beta_j \frac{y}{m_j(x)} + 1 + \frac{\pi_j}{m_j(x)}\right).$$

Using $F_x^j(z, \varphi(z)) = 0$, it leads to

$$\frac{\partial F_x^j}{\partial z}(z, \varphi(z)) = \frac{n_j}{n}\frac{-1}{z} < 0.$$

Thus, the function $\beta_j \mapsto x_j^\star(\beta_j)$ is decreasing. In such a case of an inactive constraint, the premium equilibrium is independent of the initial portfolio size n_j.

When the solvency constraint is active, the premium equilibrium $x_j^\star$ verifies $g_j^1(x_j^\star) = 0$, i.e.

$$x_j^\star = \pi_j + \frac{k_{995}\sigma(Y)\sqrt{n_j} - K_j}{n_j(1 - e_j)}. \tag{2.9}$$

Here, the implicit function theorem is not necessary since $x_j^\star$ does not depend on $x_{-j}^\star$. We deduce that $x_j^\star$ is an increasing function of $\pi_j, k_{995}, \sigma(Y), e_j$ and a decreasing function K_j.

The function $n_j \mapsto x_j^\star(n_j)$ is not necessarily monotone. Let z be n_j. Differentiating Equation (2.9) with respect to z, we get

$$\varphi'(z) = \frac{1}{z^{3/2}(1 - e_j)}\left(-\frac{k\sigma(Y)}{2} + \frac{K_j}{\sqrt{z}}\right),$$

whose sign depends on the value of the other parameters.

□

2.2.7 Numerical illustration

All numerical applications are carried out with the R software, R Core Team (2012). Please refer to Appendix 2.6.1 for computation details.

Base parameters

We consider a three-player game operating a 10 000-customer insurance market, i.e. $n = 10000$, $I = 3$. To ensure that insurers already have underwritten business, we provide d-year history for portfolio size, where $d = 3$. Although we provide a 3-year history in this section, we only consider the one-period equilibrium. So only the value at year 0 matters.

Insurer portfolio size $n_j(t)$'s are given in Table 2.2. The portfolio size is chosen such that player 1 is the leader, player 2 the challenger and player 3 the outsider with 45%, 32% and 23% market shares, respectively.

time	P1	P2	P3
-2	4200	3800	2000
-1	4700	3200	2100
0	4500	3200	2300

Table 2.2: Insurer portfolio sizes

We consider two types of loss model: (i) loss $E(Y) = 1, \sigma(Y) = 4.472$, Poisson-Lognormal model, (ii) loss $E(Y) = 1, \sigma(Y) = 10$, Negative Binomial-Lognormal model. The loss history is such that the actuarially based premiums $\bar{a}_{j,0}$'s are given in Table 2.3 and the market premium $\bar{m}_0$ is 1.190, 1.299 for PLN and NBLN, respectively.

	P1	P2	P3
PLN	1.066	1.159	0.972
NBLN	1.079	1.189	1.035

Table 2.3: Actuarially based premium $\bar{a}_{j,0}$

The weight parameters $(\omega_j)_j$ used in the computation of the insurer break-even premium are $\omega = (1/3, 1/3, 1/3)$. Before giving the sensitivity parameters β_j's, we present the lapse models. For customer behavior, we have two parameters μ_j, α_j per player given a price sensitivity function. At first, we consider the price function based on the premium ratio

$$\bar{f}_j(x_j, x_l) = \mu_j + \alpha_j \frac{x_j}{x_l}.$$

The central lapse rate parameters (i.e. lapse rate when every insurers use the same premium) are set to 10%, 14% and 18% for $j = 1, 2$ or 3, respectively. In addition to this first constraint, we also impose that on increase of 5% compared to other players increases the total lapse rate by 5 percentage points.

Let $x^1 = (1, 1, 1)$ and $x^{1.05} = (1.05, 1, 1)$. The two constraints are equivalent to

$$\mathrm{lg}_1^2(x^1) + \mathrm{lg}_1^3(x^1) = 10\% \quad \text{and} \quad \mathrm{lg}_1^2(x^{1.05}) + \mathrm{lg}_1^3(x^{1.05}) = 15\%$$

for Insurer 1. We get $\mu_1 = -12.14284$ and $\alpha_1 = 9.25247$. With this central lapse rate parameters, the expected numbers of lost policies when all insurers propose the same premium are 450.1, 448.0 and 414.0.

Secondly, we consider the price function based on the premium difference

$$\tilde{f}_j(x_j, x_l) = \tilde{\mu}_j + \tilde{\alpha}_j(x_j - x_l).$$

Calibration is done similarly as for f_j. In Figure 2.5 in Appendix 2.6.1, we plot the total lapse ratio function of each player for the two different price function f_j (left graph) and $\tilde{f}_j$ (right graph). In a grey dot-dash horizontal line, we highlight the central rates at 10%, 14% and 18% (the premium of other players is set to 1.4). In the central graph, we plot the total lapse rate function of player 1 with the two different price functions.

Price sensitivity parameters β_j of objective functions are fitted in the following way $1 - \beta_j \left(\frac{x_j}{m_j(x)} - 1\right) \approx \mathrm{lg}_j^j(x)$. Using $x = (1.05, 1, 1)$, we get

$$\beta_j = \frac{1 - \mathrm{lg}_j^j(x)}{0.05}.$$

Using the premium ratio function $\bar{f}_j$, we have $(\beta_1, \beta_2, \beta_3) = (3.0, 3.8, 4.6)$.

Last parameters are capital values and the expense rates. Capital values (K_1, K_2, K_3) are set such that the solvency coverage ratio is 133%. Expense rates are $(e_1, e_2, e_3) = (10\%, 10\%, 10\%)$.

Results and sensitivity analysis

Since we consider two loss models (PLN, NBLN) and two price sensitivity functions $\bar{f}_j, \tilde{f}_j$ (denoted by 'ratio' and 'diff', respectively), we implicitly define four sets of parameters, which

differ on loss model and price sensitivity functions. In Table 2.4, we report premium equilibria of the four models (PLN-ratio, PLN-diff, NBLN-ratio and NBLN-diff), differences between equilibrium vector $x^\star$ and actuarial and average market premium, and expected difference in portfolio size ($\Delta\hat{N}_1$ negative means insurer 1 expects to lose customers).

	$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - \bar{a}\|\|^2$	$\|\|x^\star - \bar{m}\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$
PLN-ratio	1.612	1.583	1.531	11.064	13.199	-258.510	-43.479	301.989
PLN-diff	1.659	1.621	1.566	11.191	13.297	-382.879	-38.401	421.280
NBLN-ratio	1.727	1.697	1.648	11.994	15.645	-239.465	-35.301	274.766
NBLN-diff	1.777	1.738	1.685	12.152	15.752	-385.404	-29.645	415.049

Table 2.4: Base premium equilibrium

The premium equilibrium vector $x^\star$ is quite similar between the four different tested models. The change between price sensitivity functions $\bar{f}_j, \tilde{f}_j$ from an insurer point of view is a change in sensitivity parameter β_j in its objective function. The change between $\bar{f}_j, \tilde{f}_j$ results in a slight increase of premium equilibrium whereas the change between PLN or NBLN loss models is significantly higher. Unlike the sensitivity function change, a change in loss models does not impact the objective function but the constraint function (an increase in $\sigma(Y)$).

In Tables 2.5, 2.6, we perform a sensitivity analysis considering the NBLN-ratio model as the base model. Table 2.5 reports the analysis with respect to capital (K_3 decreases) and sensitvity parameter (β_j increases). Table 2.6 focuses on actuarially based premiums ($\bar{a}_{j,0}$ increases), average market premium ($\bar{m}_0$ increases) and credibility factors (ω_j increases).

	$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - \bar{a}\|\|^2$	$\|\|x^\star - \bar{m}\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$
base	1.727	1.697	1.648	11.994	15.645	-239.465	-35.301	274.766
capital down	1.797	1.764	1.79	12.185	15.678	-96.943	87.126	9.817
	$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - \bar{a}\|\|^2$	$\|\|x^\star - \bar{m}\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$
base	1.727	1.697	1.648	11.994	15.645	-239.465	-35.301	274.766
sensitivity up	1.643	1.62	1.575	11.736	15.479	-207.466	-42.836	250.302

Table 2.5: Base premium equilibrium

	$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - \bar{a}\|\|^2$	$\|\|x^\star - \bar{m}\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$
base	1.727	1.697	1.648	11.994	15.645	-239.465	-35.301	274.766
actuarial up	1.766	1.714	1.665	13.015	15.706	-325.31	4.752	320.558
market up	1.91	1.874	1.822	12.72	20.503	-240.803	-29.484	270.287
	$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - \bar{a}\|\|^2$	$\|\|x^\star - \bar{m}\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$
base	1.727	1.697	1.648	11.994	15.645	-239.465	-35.301	274.766
credibility up	1.68	1.657	1.599	11.839	15.545	-232.155	-68.27	300.425

Table 2.6: Base premium equilibrium

2.3 Refinements of the one-period model

In this section, we propose refinements on the objective and constraint functions of the previous section.

2.3.1 Objective function

The objective function given in Subsection 2.2.3 is based on an approximation of the true demand function. For insurer j, the expected portfolio size is given by

$$\hat{N}_j(x) = n_j \times \mathrm{lg}_j^j(x) + \sum_{l \neq j} n_l \times \mathrm{lg}_l^j(x),$$

where lg_j^l's are lapse functions and lg_j^j the "renew" function. Note that the expected size $\hat{N}_j(x)$ contains both renewal and new businesses. So, a new objective function could be

$$\widetilde{O}_j(x) = \frac{\hat{N}_j(x)}{n}(x_j - \pi_j),$$

where π_j is the break-even premium as defined in Subsection 2.2.3. However, we do not consider this function, since the function $x_j \mapsto \widetilde{O}_j(x)$ does not verify some generalized convexity properties, which we will explain in Subsection 2.3.3. And also, the implicit assumption is that insurer j targets the whole market: this may not be true in most competitive insurance markets.

Instead, we will test the following objective function

$$\widetilde{O}_j(x) = \frac{n_j \, \mathrm{lg}_j^j(x)}{n}(x_j - \pi_j), \tag{2.10}$$

taking into account only renewal business. This function has the good property to be infinitely differentiable. Using the definition lg_j^j in equation (2.1), one can show that the function $x_j \mapsto \mathrm{lg}_j^j(x)$ is a strictly decreasing function, see Appendix 2.6.1. As for the objective function O_j, maximising $\widetilde{O}_j$ is a trade-off between increasing premium for better expected profit and decreasing premium for better market share.

2.3.2 Constraint function

We also change the solvency constraint function $x_j \mapsto g_j^1(x_j)$ defined in equation (2.4), which is a basic linear function of the premium x_j. We also integrate other insurer premium x_{-j} in the new constraint function, i.e. $x_j \mapsto \tilde{g}_j^1(x)$. We could use the following constraint function

$$\tilde{g}_j^1(x) = \frac{K_j + \hat{N}_j(x)(x_j - \pi_j)(1 - e_j)}{k_{995}\sigma(Y)\sqrt{\hat{N}_j(x)}} - 1,$$

the ratio of the expected capital and the required solvency capital. Unfortunately, this function does not respect a generalized convexity property, that we will define in the Subsection 2.3.3. So instead, we consider a simpler version

$$\tilde{g}_j^1(x) = \frac{K_j + n_j(x_j - \pi_j)(1 - e_j)}{k_{995}\sigma(Y)\sqrt{\hat{N}_j(x)}} - 1, \tag{2.11}$$

by removing the expected portfolio size $\hat{N}_j$ in the numerator. This function also has the good property to be infinitely differentiable. The other two constraint functions g_j^2, g_j^3 are identical as in Subsection 2.2.4.

2.3.3 Properties of premium equilibrium

Conditions on the existence of a generalized Nash equilibrium can be found in Facchinei and Kanzow (2009) or Dutang (2012*b*). In our setting, we need to show (i) the objective function $O_j(x)$ is quasiconcave with respect to x_j, (ii) the constraint function $g_j(x)$ is quasiconcave with respect to x_j, (iii) the action set $\{x_j \in X_j, g_j(x_j, x_{-j}) \geq 0\}$ is nonempty.

Recall that a function $f : X \mapsto Y$ is concave if $\forall x, y \in X, \forall \lambda \in [0,1]$, we have $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$. Note that a convex and concave function is linear. If inequalities are strict, we speak about concavity. A function $f : X \mapsto Y$ is quasiconcave if $\forall x, y \in X, \forall \lambda \in]0,1[$, we have

$$f(\lambda x + (1-\lambda)y) \geq \min(f(x), f(y)).$$

Again, if inequalities are strict, we speak about strict quasiconcavity. As for concavity, there exist special characterizations when f is C^2.

Proposition. *When f is a differentiable function on an open convex $O \subset \mathbb{R}^n$, then f is quasiconcave if and only if $\forall x, y \in O, f(x) \geq f(y) \Rightarrow \nabla f(y)^T(x-y) \geq 0$.*
When f is a C^2 function on an open convex $O \subset \mathbb{R}^n$, then f is quasiconcave if and only if $\forall x \in O, \forall d \in \mathbb{R}^n, d^T \nabla f(x) = 0 \Rightarrow d^T \nabla^2 f(x) d \leq 0$.

Proof. See Theorems 2 and 5 of Diewert et al. (1981). □

From the last proposition, it is easy to see that for a C^2 univariate function, quasiconcavity implies unimodality. Furthermore, f is pseudoconcave if and only if $\forall x, y$, we have $f(x) > f(y) \Rightarrow \nabla f(y)^T(x-y) > 0$.

Proposition. *When f is a C^2 function on an open convex $O \subset \mathbb{R}^n$, then if $\forall x \in O, \forall d \in \mathbb{R}^n, d^T \nabla f(x) = 0 \Rightarrow d^T \nabla^2 f(x) d < 0$, then f is pseudoconcave, which in turn implies strict quasiconcavity.*

Proof. See Corollary 10.1 of Diewert et al. (1981). □

Examples of quasiconcave functions include monotone, concave or log-concave functions. A univariate quasiconcave function is either monotone or unimodal. More properties can be found in Diewert et al. (1981). Figure 2.6 in Appendix 2.6.1 relates on the different concepts of convexity.

Proposition 2.3.1. *The I-player insurance game with objective function and solvency constraint function defined in Equations (2.10) and (2.11), respectively, admits a generalized Nash premium equilibrium, if for all $j = 1, \ldots, I$, $\tilde{g}_j^1(\overline{x}) > 0$.*

Proof. Properties of the expected portfolio size function have been established in Appendix 2.6.1. The objective function can be rewritten as

$$\widetilde{O}_j(x) = \mathrm{lg}_j^j(x, f)(x_j - \pi_j),$$

up to a constant n_j/n. $\widetilde{O}_j$ has been built to be continuous on $\mathbb{R}^I_+$. Note that we stress the dependence on the price sensitivity function f. Using Appendix 2.6.1, the gradient of the objective function is proportional to

$$\frac{\partial \widetilde{O}_j(x)}{\partial x_j} = \lg^j_j(x,f)(1 - S_j(x)(x_j - \pi_j)), \quad \text{where} \quad S_j(x) = \sum_{l \neq j} f'_{j1}(x_j, x_l) \lg^l_j(x,f).$$

The gradients cancel at $1 = S_j(x^{j\star})(x^\star_j - \pi_j)$, where $x^{j\star} = (x_1, \ldots, x_{j-1}, x^\star_j, x_{j+1}, \ldots, x_I)$.

The second-order derivative is given by

$$\begin{aligned}\frac{\partial^2 \widetilde{O}_j(x)}{\partial x_j^2} &= \lg^j_j(x,f) \left((x_j - \pi_j) 2S_j^2(x) - 2S_j(x) - (x_j - \pi_j) \sum_{l \neq j} f'_{j1}(x_j, x_l)^2 \lg^l_j(x,f) \right) \\ &= \lg^j_j(x,f) 2S_j(x) \left[(x_j - \pi_j) S_j(x) - 1 \right] - \lg^j_j(x,f)(x_j - \pi_j) \sum_{l \neq j} f'_{j1}(x_j, x_l)^2 \lg^l_j(x,f).\end{aligned}$$

The sign of the second order derivative at $x^{j\star}$ is

$$\operatorname{sign}\left(\frac{\partial^2 \widetilde{O}_j(x^{j\star})}{\partial x_j^2} \right) = - \lg^j_j(x^{j\star}, f)(x^\star_j - \pi_j) \sum_{l \neq j} f'_{j1}(x^\star_j, x_l)^2 \lg^l_j(x^{j\star}, f).$$

However, the root of the gradient is such that $x^\star_j - \pi_j = 1/S_j(x^{j\star}) > 0$. So we have

$$\operatorname{sign}\left(\frac{\partial^2 O_j(x^{j\star})}{\partial x_j^2} \right) < 0.$$

Hence, the function $x_j \mapsto O_j(x)$ is pseudoconcave, and thus strictly quasiconcave.

Functions g^2_j, g^3_j are strictly concave since second-order derivatives are

$$\frac{\partial^2 g^2_j(x)}{\partial^2 x_j} = -e^{-(x_j - \underline{x})} < 0 \quad \text{and} \quad \frac{\partial^2 g^3_j(x)}{\partial^2 x_j} = -e^{-(\overline{x} - x_j)} < 0.$$

We verify quasiconcavity of the function $\tilde{g}^1_j$ with respect to x_j. The function $x_j \mapsto \tilde{g}^1_j(x)$ is monotone since its gradient

$$\frac{\partial \tilde{g}^1_j(x)}{\partial x_j} = \frac{K_j + n_j(x_j - \pi_j)(1 - e_j)}{2k_{995}\sigma(Y)\hat{N}^{3/2}_j(x)} \left(-\frac{\partial \hat{N}_j(x)}{\partial x_j} \right) + \frac{n_j(1 - e_j)}{k_{995}\sigma(Y)\sqrt{\hat{N}_j(x)}}$$

is positive for all $x \in \mathbb{R}^I_+$. Thus, function $x_j \mapsto \tilde{g}^1_j(x)$ is (strictly) quasiconcave.

Let $X_j = [\underline{x}, \overline{x}]$. The constraint set is $C_j(x_{-j}) = \{x_j \in X_j, \tilde{g}^1_j(x_j, x_{-j}) \geq 0\}$ where $x_j \mapsto \tilde{g}^1_j(x)$ is strictly increasing, continuous and by assumption, for all $j = 1, \ldots, I$, $\tilde{g}^1_j(\overline{x}) > 0$. Thus, $C_j(x_{-j})$ is a nonempty convex closed set. Furthermore, the point-to-set mapping C_j is upper semi-continuous by using Example 5.10 of Rockafellar and Wets (1997). Using Theorem 13 of Hogan (1973) and the continuity of $\tilde{g}^1_j$, the point-to-set mapping is also lower semi-continuous. By Theorem 4.1 of Facchinei and Kanzow (2009), there exists a generalized Nash equilibrium. □

Non-uniqueness issues

Uniqueness of a generalized Nash equilibrium is not guaranteed in general. Furthermore, there is no particular reason for a player to choose a certain Nash equilibrium rather than another one. Rosen (1965) studied uniqueness of such an equilibrium in a jointly convex game (i.e. where objective functions are convex and the constraint function is common and convex). To deal with non-uniqueness, he studies a subset of generalized Nash equilibrium, where Lagrange multipliers resulting from the Karush-Kuhn-Tucker (KKT) conditions are normalized. Such a normalized equilibrium is unique given a scale of the Lagrange multiplier when the constraint function verifies additional assumptions. Other authors such as von Heusinger and Kanzow (2009) or Facchinei et al. (2007) define normalized equilibrium when Lagrange multipliers are set equal.

Another way is to look for generalized Nash equilibria having some specific properties, such as Pareto optimality. The selection of the equilibrium is particularly developed for games with finite action sets. In that setting, one can also use a mixed strategy, by playing ramdomly one among many equilibrium strategies.

Parameter sensitivity

Proposition 2.3.2. *Let $x^\star$ be a premium equilibrium of the I-player insurance game. For each player j, if $x_j^\star \in]\underline{x}, \overline{x}[$, player equilibrium $x_j^\star$ depends on parameter in the following way: it increases with break-even premium π_j, solvency coefficient k_{995}, loss volatility $\sigma(Y)$, expense rate e_j and decreases with lapse parameter μ_j, α_j and capital K_j. Otherwise when $x_j^\star = \underline{x}$ or $\overline{x}$, premium equilibrium is independent of any parameters.*

Proof. As explained in Appendix 2.6.1, the KKT conditions at a premium equilibrium $x^\star$ are such there exist Lagrange multipliers $\lambda^{j\star}$,

$$\frac{\partial \widetilde{O}_j}{\partial x_j}(x) - \lambda_1^{j\star} \frac{\partial \tilde{g}_j^1}{\partial x_j}(x) = 0,$$

when assuming g_j^2, g_j^3 functions are not active. And the complementarity constraint is such that $\lambda_1^{j\star} \times \tilde{g}_j^1(x^\star) = 0$.

If the solvency constraint $\tilde{g}_j^1$ is inactive, then we necessarily have $\lambda_{j1}^\star = 0$. Let x_y^j be the premium vector with the j component being y, i.e. $x_y^j = (x_1, \ldots, x_{j-1}, y, x_{j+1}, \ldots, x_I)$. We denote by z a parameter of interest, say for example e_j. We define the function F as

$$F_x^j(z, y) = \frac{\partial O_j}{\partial x_j}(x_y^j, z),$$

where the objective function depends on the interest parameter z. By the continuous differentiability of F with respect to z and y, we can invoke the implicit function theorem, see Appendix 2.6.1. So there exists a function φ such that $F_x^j(z, \varphi(z)) = 0$, and the derivative is given by

$$\varphi'(x) = - \left. \frac{\frac{\partial F_x^j}{\partial z}(z, y)}{\frac{\partial F_x^j}{\partial y}(z, y)} \right|_{y=\varphi(z)}.$$

In our case *, we have

$$F_x^j(z,y) = \frac{n_j}{n}\,\mathrm{lg}_j^j(x_y^j)[1 - S_j(x_y^j)(y-\pi_j)],$$

and

$$\frac{\partial F_x^j}{\partial z}(z,y) = \frac{\partial^2 O_j}{\partial z \partial x_j}(x_y^j, z), \quad \text{and} \quad \frac{\partial F_x^j}{\partial y}(z,y) = \frac{\partial^2 O_j}{\partial x_j^2}(x_y^j, z).$$

The first-order derivative is given by

$$\frac{\partial F_x^j}{\partial y}(z,y) = 2\frac{n_j}{n}\,\mathrm{lg}_j^j(x_y^j)S_j(x_y^j)\left[(y-\pi_j)S_j(x_y^j) - 1\right] - \frac{n_j}{n}\,\mathrm{lg}_j^j(x_y^j)(y-\pi_j)\sum_{l\neq j} f'_{j1}(y,x_l)^2\,\mathrm{lg}_j^l(x_y^j).$$

Using $F_x^j(z,\varphi(z)) = 0$ whatever z represents, it simplifies to

$$\frac{\partial F_x^j}{\partial y}(z,\varphi(z)) = -\frac{n_j}{n}\,\mathrm{lg}_j^j(x_{\varphi(z)}^j)(\varphi(z)-\pi_j)\sum_{l\neq j} f'_{j1}(\varphi(z),x_l)^2\,\mathrm{lg}_j^l(x_{\varphi(z)}^j).$$

Let z now be the insurer's break-even premium $z = \pi_j$. We have

$$\frac{\partial F_x^j}{\partial z}(z,y) = n_j\,\mathrm{lg}_j^j(x_y^j)S_j(x_y^j).$$

Thus, the derivative of φ is

$$\varphi'(z) = \frac{S_j\left(x_{\varphi(z)}^j\right)}{(\varphi(z)-z)\sum_{l\neq j} f'_{j1}\left(\varphi(z),x_l\right)^2\,\mathrm{lg}_j^l\left(x_{\varphi(z)}^j\right)}.$$

By definition, $F_x^j(z,\varphi(z)) = 0$ is equivalent to

$$1 = S_j\left(x_{\varphi(z)}^j\right)(\varphi(z)-z).$$

Thus $\varphi(z) - z > 0$. We conclude that $\varphi'(z) > 0$, i.e. the function $\pi_j \mapsto x_j^\star(\pi_j)$ is increasing.

Let z be the intercept lapse parameter $z = \mu_j$. By differentiating the lapse probability, we have

$$\frac{\partial\,\mathrm{lg}_j^j}{\partial z}(x_y^j) = -\,\mathrm{lg}_j^j(x_y^j)\sum_{l\neq j}\mathrm{lg}_j^l(x_y^j) \quad \text{and} \quad \left.\frac{\partial\,\mathrm{lg}_j^k}{\partial z}(x_y^j)\right|_{j\neq k} = -\,\mathrm{lg}_j^k(x_y^j)\sum_{l\neq j}\mathrm{lg}_j^l(x_y^j) + \mathrm{lg}_j^k(x_y^j).$$

We get

$$\frac{\partial F_x^j}{\partial z}(z,y) = -n_j\,\mathrm{lg}_j^j(x_y^j)(1-\mathrm{lg}_j^j(x_y^j))\left[1 - S_j(x_y^j)(y-\pi_j)\right] - n_j\,\mathrm{lg}_j^j(x_y^j)^2 S_j(x_y^j).$$

Note the first term when $y = \varphi(z)$ since $F_x^j(z,\varphi(z)) = 0$. We finally obtain

$$\varphi'(x) = -\frac{S_j\left(x_{\varphi(z)}^j\right)\mathrm{lg}_j^j\left(x_{\varphi(z)}^j\right)}{(\varphi(z)-z)\sum_{l\neq j} f'_{j1}\left(\varphi(z),x_l\right)^2\,\mathrm{lg}_j^l\left(x_{\varphi(z)}^j\right)}.$$

∗. To simplify, we do not stress the dependence of lg_j^k and S_j on f.

Using $1 = S_j\left(x^j_{\varphi(z)}\right)(\varphi(z) - z)$, we have

$$\varphi'(x) = -\frac{S_j\left(x^j_{\varphi(z)}\right)^2 \lg^j_j\left(x^j_{\varphi(z)}\right)}{\sum\limits_{l\neq j} f'_{j1}\left(\varphi(z), x_l\right)^2 \lg^l_j\left(x^j_{\varphi(z)}\right)} < 0.$$

Thus, the function $\mu_j \mapsto x^\star_j(\mu_j)$ is decreasing.

Let z be the slope lapse parameter $z = \alpha_j$.

$$\frac{\partial \lg^j_j}{\partial z}(x^j_y) = -\lg^j_j(x^j_y)\sum_{l\neq j}\Delta_{j,l}(x^j_y)\lg^l_j(x^j_y)$$

and

$$\left.\frac{\partial \lg^k_j}{\partial z}(x^j_y)\right|_{j\neq k} = -\lg^k_j(x^j_y)\sum_{l\neq j}\Delta_{j,l}(x^j_y)\lg^l_j(x^j_y) + \lg^k_j(x^j_y)\Delta_{j,k}(x^j_y),$$

where $\Delta_{j,l}(x^j_y) = x_j/x_l$ if we use the premium ratio function f_j and $x_j - x_l$ if we use the premium difference function $\tilde{f}_j$. We get

$$\begin{aligned}\frac{\partial F^j_x}{\partial z}(z,y) = &-n_j \lg^j_j(x^j_y) S^\Delta_j(x^j_y)[1 - S_j(x^j_y)(y - \pi_j)] \\ &- n_j \lg^j_j(x^j_y)^2 S^\Delta_j(x^j_y) - n_j \lg^j_j(x^j_y)\sum_{l\neq j} f'_{j1}(y, x_l)\Delta_{j,l}(x^j_y)\lg^l_j(x^j_y),\end{aligned}$$

where $S^\Delta_j(x^j_y) = \sum_{l\neq j}\Delta_{j,l}(x^j_y)\lg^l_j(x^j_y)$. Again the first term cancels when $y = \varphi(z)$. Hence, we have

$$\varphi'(z) = -\frac{\lg^j_j\left(x^j_{\varphi(z)}\right) S^\Delta_j\left(x^j_{\varphi(z)}\right) + \sum_{l\neq j} f'_{j1}(\varphi(z), x_l)\Delta_{j,l}\lg^l_j\left(x^j_{\varphi(z)}\right)}{(\varphi(z) - z)\sum\limits_{l\neq j} f'_{j1}\left(\varphi(z), x_l\right)^2 \lg^l_j\left(x^j_{\varphi(z)}\right)}.$$

Using $1 = S_j\left(x^j_{\varphi(z)}\right)(\varphi(z) - z)$, we have

$$\varphi'(z) = -S_j\left(x^j_{\varphi(z)}\right)\frac{\lg^j_j\left(x^j_{\varphi(z)}\right) S^\Delta_j\left(x^j_{\varphi(z)}\right) + \sum_{l\neq j} f'_{j1}(\varphi(z), x_l)\Delta_{j,l}\left(x^j_{\varphi(z)}\right)\lg^l_j\left(x^j_{\varphi(z)}\right)}{\sum\limits_{l\neq j} f'_{j1}\left(\varphi(z), x_l\right)^2 \lg^l_j\left(x^j_{\varphi(z)}\right)}.$$

If we use the premium ratio function, we have $\Delta_{j,l}(.) = \varphi(z)/x_l > 0$ as well as $f'_{j1}\left(\varphi(z), x_l\right) > 0$. It is immediate that $\varphi'(z) < 0$. Otherwise when we use the premium difference function $(\Delta_{j,l}(.) = \varphi(z) - x_l)$, we cannot guarantee that the numerator is positive.

If the solvency constraint $\tilde{g}^1_j$ is active, then we necessarily have $\lambda^\star_{j1} > 0$, $g^1_j(x^\star) = 0$. Let x^j_y be the premium vector with the j component being y as above. We denote by z a parameter of interest, then we define the function G as

$$G^j_x(z,y) = \tilde{g}^1_j(x^j_y, z),$$

where the objective function depends on the interest parameter z. Again, we apply the implicit function theorem with a function ϕ such that $G_x^j(z, \phi(z)) = 0$. The first-order derivative is given by

$$\frac{\partial G_x^j}{\partial y}(z, y) = \frac{\partial g_j^1}{\partial x_j}(z, y) > 0,$$

since $x_j \mapsto \tilde{g}_j^1$ is a strictly increasing function. Therefore, the sign of ϕ' is

$$\operatorname{sign}\big(\phi'(z)\big) = -\operatorname{sign}\left(\frac{\partial G_x^j}{\partial z}(z, \phi(z))\right).$$

Let $z = \pi_j$ be the actuarial premium. We have

$$\frac{\partial G_x^j}{\partial z}(z, y) = -\frac{n_j(1 - e_j)}{k_{995}\sigma(Y)\sqrt{\hat{N}_j(x_y^j)}} < 0,$$

independently of y or z. So, $\operatorname{sign}(\phi'(z)) > 0$, i.e. the function $\pi_j \mapsto x_j^\star(\pi_j)$ is increasing as in the previous case.

Let $z = K_j$ be the capital. We have

$$\frac{\partial G_x^j}{\partial z}(z, y) = \frac{1}{k_{995}\sigma(Y)\sqrt{\hat{N}_j(x_y^j)}} > 0.$$

So $\operatorname{sign}(\phi'(z)) < 0$, i.e. the function $K_j \mapsto x_j^\star(K_j)$ is decreasing.

Let $z = \sigma(Y)$ be the actuarial premium. We have

$$\frac{\partial G_x^j}{\partial z}(z, y) = -\frac{1}{z^2} \times \frac{K_j + n_j(y - \pi_j)(1 - e_j)}{k_{995}\sqrt{\hat{N}_j(x_y^j)}},$$

which simplifies to $\frac{\partial G_x^j}{\partial z}(z, \phi(z)) = -1/z < 0$ using the definition of G^j. Thus, the function $\sigma(Y) \mapsto x_j^\star(\sigma(Y))$ is decreasing. By a similar reasoning, we have for $z = k_{995}$, that ϕ is decreasing. □

2.3.4 Numerical application

We use the same set of parameters as in Subsection 2.2.7. As discussed above, a generalized premium equilibrium is not necessarily unique: in fact there are many of them. In Tables 2.7 and 2.8, we report generalized Nash equilibria found with different starting points (2^{10} feasible points randomly drawn in the hypercube $[\underline{x}, \overline{x}]^I$). Premium equilibrium are sorted according to the difference with average market premium $\bar{m}$.

In Table 2.8, this computation is done for the Negative Binomial-Lognormal loss model (NBLN), whereas Table 2.7 reports the computation for Poisson-Lognormal model (PLN). Both tables use the price ratio function $\bar{f}_j$. The last column of those tables reports the number of optimization sequences converging to a given equilibrium.

Most of the time, other equilibriums found hit one of the barriers $\underline{x}, \overline{x}$. It may appear awkward that such points are optimal in a sense, but one must not forget the Lagrange

$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - \bar{a}\|\|^2$	$\|\|x^\star - \bar{m}\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$	Nb
1	1	1	0.1132	0.0084	-19	-16	35	1
1.3041	1.025	1.0283	0.0497	0.0645	-2479	1264	1216	13
1	1.3183	1.0065	0.0964	0.0754	1001	-1899	898	4
1	1.0001	1.3427	0.1162	0.0896	722	701	-1423	3
1.0185	1.3694	1.3993	0.133	0.2215	2646	-1507	-1139	114
1.3856	1.0844	1.4501	0.1144	0.2696	-1729	2758	-1029	142
1.419	1.4541	1.1247	0.1121	0.3004	-1564	-1233	2797	111
1.0449	1.3931	3	3.4379	3.9075	3787	-1490	-2297	1
3	1.1738	1.5381	3.2767	4.0418	-4490	5412	-922	3

Table 2.7: Premium equilibria - PLN price ratio function

$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - a\|\|^2$	$\|\|x^\star - m\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$	Nb
1.3644	1.0574	1.0661	0.1239	0.0611	-2635	1397	1239	10
1	1.3942	1.0208	0.1201	0.1003	1315	-2258	943	1
1	1.001	1.4206	0.1398	0.1192	851	818	-1670	3
1.0044	1.4216	1.4569	0.0887	0.1781	3333	-1923	-1411	109
1.4875	1.1726	1.5792	0.1836	0.2781	-1622	2696	-1075	116
1.555	1.6092	1.2508	0.2323	0.3598	-1369	-1210	2579	97
1.561	1.2526	3	3.0865	3.5394	-1405	3695	-2291	4
1.7346	3	1.4348	3.2546	3.7733	-955	-3174	4129	5
3	1.4664	3	6.226	6.8384	-4485	6746	-2261	2
3	1.3699	1.7658	3.4794	3.7789	-4482	5299	-817	4
3	1.9041	1.5497	3.6941	4.0712	-4462	-743	5205	12
3	3	1.7542	6.407	7.0956	-4354	-2970	7324	4

Table 2.8: Premium equilibria - NBLN price ratio function

multipliers (not reported here). Those are not zero when a constraint g_j^i is active, (where $i = 1, 2, 3$ and $j = 1, \ldots, I$).

Tables 2.12 and 2.13 in Appendix 2.6.1 report the computation when we use the price difference function $\tilde{f}_j$. The number of different premium equilibria is similar as in the previous case.

This numerical application reveals that in our refined game, we have many generalized premium equilibria. In our insurance context, a possible way to deal with multiple equilibria is to choose as a premium equilibrium, the generalized Nash equilibrium $x^\star$ that is closest to the average market premium $\bar{m}$. This option is motivated by the high level of competition present in most mature insurance markets (e.g. Europe and North America) where each insurer sets the premium with a view towards the market premium.

However, this solution has drawbacks: while a single Nash equilibrium may be seen as a self-enforcing solution, multiple generalized Nash equilibria cannot be self-enforcing. We will not pursue this refined one-shot game further and focus on the simple insurance game of Section 2.2.

2.4 Dynamic framework

In practice, insurers play an insurance game over several years, gather new information on incurred losses, available capital and competition level. We present in this section the dynamic framework based on the one-shot game of Subsection 2.2. The first subsection gives a justification for the chosen dynamic model, compared to other possible dynamic game models. The following subsections present the dynamic game, some properties and numerical illustrations.

2.4.1 Dynamic game models

Defining dynamic games is quite complex. Basar and Olsder (1999) is a major reference on noncooperative dynamic game theory. Extending a static game to a dynamic game consists not only of adding a time dimension t for the control variable. It also requires the definition of a state equation ($x_{t+1} = f(x_t, \dots)$) and a state variable x_t, "linking" all the information together, see Definition 5.1 of Basar and Olsder (1999) *.

Depending on which information the players have about the state variable, different classes of games are defined: open-loop (knowing only the first state x_1), closed-loop (all states x_t up to time t), feedback (only the current state x_t). Computational methods for dynamic equilibrium generally use backward equations, e.g. Theorem 6.6 of Basar and Olsder (1999) for feedback strategies and Theorem 6.10 in a stochastic setting. This method does not correspond to the insurance market reality: (i) premium is not set backwardly, the claim uncertainty is a key element in insurance pricing, (ii) the time horizon is infinite rather than finite.

A class of discrete-time games, first introduced by Shapley (1953), use a finite state space where a transition probability models the evolution of the current state depending on player actions. As the set of possible strategies (a serie of pure or mixed actions) is huge, Shapley (1953) focuses on strategies depending on the current state only. These games are also referred to Markov games. Despite our game has a Markovian property, we do neither limit our strategy space to a finite set, nor use a finite state space.

Another kind of dynamic games is evolutionary games, e.g. Sigmund and Hofbauer (1998). Evolutionary games are different in spirit to the classical game theory since they try to model non-rational behavior of players meeting randomly. The different types of individuals represent the different type of strategies. Again a recurrence equation is used to model the average proportion of individuals of type i at time t. In the actuarial literature, Ania et al. (2002) use an evolutionary setting to extend the Rotschild-Stiglitz framework on optimal insurance coverage by individuals. Non-rational behavioral game theory does not seem the appropriate tool for insurance market cycles.

Finally, repeated games study long-term interactions between players during the repetition of one-shot finite games. The horizon either infinite or finite plays a major role in the analysis of such games, in particular whether punishment strategies and threats are relevant. Most of the theory (Folk theorems) focuses on the set of achievable payoffs rather than the characterization of the equilibrium. Folk theorems demonstrate that wellfare outcomes can be attained when players have a long-term horizon, even if it is not possible in the one-shot game, see, e.g., Osborne and Rubinstein (2006). Our game does not belong to this framework,

∗. We deal with discrete-time games, if working continuous-time games, the state equation is replaced by a differential equation. Such games are thus called differential games.

since strategic environments (action sets) evolve over a time, the action set is not finite and stochastic pertubations complete the picture.

We choose a repeated game but with infinite action space, such that at each period, insurers set new premiums depending on past observed losses. A generalized Nash equilibrium is computed at each period. Our repeated game does not enter the framework of dynamic games as presented in Basar and Olsder (1999), but it shares some properties of Markov games and classical repeated games.

2.4.2 Deriving a dynamic model

In this subsection, we describe the repeated game framework. Now, insurers have a past history: past premium $x^\star_{j,t}$ gross written premium $\text{GWP}_{j,t}$, portfolio size $n_{j,t}$, capital $K_{j,t}$ at the beginning of year t. Let d be the history depth for which economic variables (e.g. market premium) will be computed. In this setting, objective $O_{j,t}$ and constraint functions $g_{j,t}$ are also time-dependent.

At the beginning of each time period, the average market premium is determined as

$$\bar{m}_{t-1} = \frac{1}{d}\sum_{u=1}^{d} \underbrace{\frac{\sum_{j=1}^{N} \text{GWP}_{j,t-u} \times x^\star_{j,t-u}}{\text{GWP}_{.,t-u}}}_{\text{market premium for year } t-u},$$

which is the mean of last d market premiums. With current portfolio size $n_{j,t-1}$ and initial capital $K_{j,t-1}$, each insurer computes its actuarially based premium as

$$\bar{a}_{j,t} = \frac{1}{1-e_{j,t}} \frac{1}{d} \sum_{u=1}^{d} \underbrace{\frac{s_{j,t-u}}{n_{j,t-u}}}_{\text{avg ind loss}},$$

where $s_{j,t}$ denotes the observed aggregate loss of insurer j during year t. Thus, break-even premiums are $\pi_{j,t} = \omega_j \bar{a}_{j,t} + (1-\omega_j)\bar{m}_{t-1}$. Thus, the objective function in the dynamic model is given by

$$O_{j,t}(x) = \frac{n_{j,t}}{n}\left(1 - \beta_{j,t}\left(\frac{x_j}{m_j(x)} - 1\right)\right)(x_j - \pi_{j,t}),$$

and the solvency constraint function by

$$g^1_{j,t}(x_j) = \frac{K_{j,t} + n_{j,t}(x_j - \pi_{j,t})(1-e_{j,t})}{k_{995}\sigma(Y)\sqrt{n_{j,t}}} - 1.$$

It is important to note that the characteristics of insurers evolve over time, notably the break-even premium $\pi_{j,t}$, the expense rate $e_{j,t}$ and the sentivity parameter $\beta_{j,t}$.

The game sequence for period t is as follows

1. The insurers maximize their objective function subject to the solvency constraint

$$\sup_{x_{j,t}} O_{j,t}(x_{j,t}, x_{-j,t}) \text{ such that } g_{j,t}(x_{j,t}) \geq 0.$$

2. Once the premium equilibrium vector $x^\star_t$ is determined, customers randomly lapse or renew, so we get a realization $n^\star_{j,t}$ of the random variable $N_{j,t}(x^\star)$.

3. Aggregate claim amounts $S_{j,t}$ are randomly drawn according to the chosen loss model (either PLN, NBLN or PGLN) with the frequency shape parameter multiplied by $n^{\star}_{j,t}$. So we get a new aggregate claim amount $s_{j,t}$ for period t per insurer j.
4. The underwriting result for insurer j is computed by $UW_{j,t} = n^{\star}_{j,t} \times x^{\star}_{j,t} \times (1 - e_j) - s_{j,t}$.
5. Finally, we update the capital by the following equation $K_{j,t} = K_{j,t-1} + UW_{j,t}$.

This game sequence is repeated over T years, but insurers are pulled out of the market when they have either a tiny market share ($< 0.1\%$) or a negative capital. Furthermore, we remove players from the game when the capital is below the minimum capital requirement (MCR), whereas we keep them if capital is between MCR and solvency capital requirement (SCR).

Let $I_t \subset \{1, \dots, I\}$ be the set of insurers at the beginning of year t and $R_t \subset \{1, \dots, I\}$ the set of removed insurers at the end of year t. If some insurers are removed, i.e. $\text{Card}(R_t) > 0$, then corresponding policyholders randomly move to other insurers according to a I_{t+1}-dimensional multinomial distribution. Say from $l \in R_t$ to $j \in I_{t+1}$, insured randomly move with multinomial distribution $\mathcal{M}_{I_{t+1}}(n_{l,t}, p^{-}_{l\to}(x^{\star}_t))$, where the probability vector $p^{-}_{l\to}(x^{\star}_t)$ has jth component given by

$$p^{-}_{l\to j}(x^{\star}_t) = \frac{\lg^j_l(x^{\star}_t)}{1 - \sum_{k\in R_t} \lg^k_l(x^{\star}_t)}.$$

When there are no more insurers, i.e. $\text{Card}(I_{t+1}) = 0$, the game ends, while if there is a single insurer, i.e. $\text{Card}(I_{t+1}) = 1$, the game continues and the survivor insurer set the highest premium.

In the current framework, we make the following implicit simplifying assumptions: (i) the pricing procedure is done (only) once a year (on January 1), (ii) all policies start at the beginning of the year, (iii) all premium are collected on January 1, (iv) every claim is (fully) paid on December 31 and (v) there is no inflation and no stock/bond market to invest premium.

In practice, these assumptions do not hold: (i) pricing by actuarial and marketing departments can be done more frequently, e.g. every 6 months, (ii) policies start and are renewed all over the year, (iii) premium is collected all over the year, (iv) claims are settled every day and there are reserves for incurred-but-not-reported claims and (v) there are inflation on both claims and premiums, and the time between the premium payment and a possible claim payment is used to invest in stock/bond markets. However, we need the above simplifications to have a sufficiently simple model.

2.4.3 Properties of premium equilibrium

Proposition 2.4.1. *For the repeated $I-$player insurance game defined in the previous subsection, the probability that there is at least two non-bankrupt insurers at time t decreases geometrically as t increases.*

Proof. As reported in Appendix 2.6.1, insurer choice probability functions $x_j \mapsto \lg^j_l(x)$ are (strictly) decreasing functions from 1 to 0. Note that $\lg^j_l(x) = 0$ (respectively $\lg^j_l(x) = 1$) is only attained when x_j tends to $+\infty$ ($-\infty$). When $i \neq j$ functions $x_i \mapsto \lg^j_l(x)$ are strictly increasing. Let $\underline{x}^{j-} = (\underline{x}, \dots, \underline{x}, \overline{x}, \underline{x}, \dots, \underline{x})$ and $\overline{x}^j_- = (\overline{x}, \dots, \overline{x}, \underline{x}, \overline{x}, \dots, \overline{x})$. We have

$$0 < \lg^j_l(\underline{x}^{j-}) < \lg^j_l(x) < \lg^j_l(\overline{x}^j_-) < 1,$$

for all $x \in [\underline{x}, \overline{x}]^I$. Taking supremum and infimum on player j, we get

$$0 < \underline{p}_l = \inf_j \lg_l^j(\underline{x}^{j-}) \quad \text{and} \quad \sup_j \lg_l^j(\overline{x}_-^j) = \overline{p}_l < 1.$$

Using the definition of portfolio size $N_{j,t}(x)$ given in Subsection 2.2.1 as a sum of binomial random variables $B_{lj,t}(x)$, we have

$$\begin{aligned}
&P\left(N_{j,t}(x) = m_j | N_{j,t-1} > 0, \mathrm{Card}(I_{t-1}) > 1\right) \\
&\qquad = P\left(\sum_{l \in I_{t-1}} B_{lj,t}(x) = m_j \,\middle|\, N_{j,t-1} > 0, \mathrm{Card}(I_{t-1}) > 1\right) \\
&\qquad\qquad = \sum_{\substack{\tilde{m}_1, \ldots, \tilde{m}_{I_{t-1}} \geq 0 \\ \text{s.t. } \sum_l \tilde{m}_l = m_j}} \prod_{l \in I_{t-1}} P(B_{lj,t}(x) = \tilde{m}_l) \\
&\qquad = \sum_{\substack{\tilde{m}_1, \ldots, \tilde{m}_{I_{t-1}} \geq 0 \\ \text{s.t. } \sum_l \tilde{m}_l = m_j}} \prod_{l \in I_{t-1}} \binom{n_{l,t-1}}{\tilde{m}_l} \lg_l^j(x)^{\tilde{m}_l} \left(1 - \lg_l^j(x)\right)^{n_{l,t-1} - \tilde{m}_j} \\
&\qquad\qquad > \sum_{\substack{\tilde{m}_1, \ldots, \tilde{m}_{I_{t-1}} \geq 0 \\ \text{s.t. } \sum_l \tilde{m}_l = m_j}} \prod_{l \in I_{t-1}} \binom{n_{l,t-1}}{\tilde{m}_l} \underline{p}_l^{\tilde{m}_l} (1 - \overline{p}_l)^{n_{l,t-1} - \tilde{m}_j} = \xi > 0
\end{aligned}$$

Therefore,

$$\begin{aligned}
&P(\mathrm{Card}(I_t) = 0 | \mathrm{Card}(I_{t-1}) > 1) = \\
&\qquad P\left(\forall j \in I_{t-1}, N_{j,t}(x) \geq 0, K_{j,t-1} + N_{j,t}(x) x_{j,t}^\star (1 - e_j) < \sum_{i=1}^{N_{j,t}(x)} Y_i\right) \\
&\qquad \geq P\left(\forall j \in I_{t-1}, N_{j,t}(x) > 0, K_{j,t-1} + N_{j,t}(x) x_{j,t}^\star (1 - e_j) < \sum_{i=1}^{N_{j,t}(x)} Y_i\right) \\
&\quad = \sum_{\substack{m_1, \ldots, m_{I_{t-1}} \geq 0 \\ \text{s.t. } \sum_l m_l = n}} \prod_{l \in I_{t-1}} P_t\left(N_{j,t}(x) = m_j\right) P\left(K_{j,t-1} + m_j x_{j,t}^\star (1 - e_j) < \sum_{i=1}^{m_j} Y_i\right) \\
&\quad \geq \sum_{\substack{m_1, \ldots, m_{I_{t-1}} > 0 \\ \text{s.t. } \sum_l m_l = n}} \prod_{l \in I_{t-1}} P_t\left(N_{j,t}(x) = m_j\right) P\left(K_{j,t-1} + m_j x_{j,t}^\star (1 - e_j) < \sum_{i=1}^{m_j} Y_i\right) \\
&\qquad > \xi \sum_{\substack{m_1, \ldots, m_{I_{t-1}} > 0 \\ \text{s.t. } \sum_l m_l = n}} \prod_{l \in I_{t-1}} P\left(K_{j,t-1} + m_j x_{j,t}^\star (1 - e_j) < \sum_{i=1}^{m_j} Y_i\right) > 0
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&P(\mathrm{Card}(I_t) > 1 | \mathrm{Card}(I_{t-1}) > 1) = \\
&\qquad 1 - P(\mathrm{Card}(I_t) = 0 | \mathrm{Card}(I_{t-1}) > 1) - P(\mathrm{Card}(I_t) 1 | \mathrm{Card}(I_{t-1}) > 1) \\
&\qquad\qquad \leq 1 - P(\mathrm{Card}(I_t) = 0 | \mathrm{Card}(I_{t-1}) > 1) < 1 - \tilde{\xi} < 1.
\end{aligned}$$

By successive conditioning, we get

$$P(\text{Card}(I_t) > 1) = P(\text{Card}(I_0) > 1) \prod_{s=1}^{t} P(\text{Card}(I_s) > 1 | \text{Card}(I_{s-1}) > 1) < \left(1 - \tilde{\xi}\right)^t.$$

So, the probability $P(\text{Card}(I_t) > 1)$ decreases geometrically as t increases. □

Proposition 2.4.2. *For the repeated $I-$player insurance game defined in the previous subsection, if for all $k \neq j$, $x_j \leq x_k$ and $x_j(1-e_j) \leq x_k(1-e_k)$, then the underwriting result by policy is ordered $UW_j \leq_{icx} UW_k$ where UW_j is the random variable*

$$UW_j = x_j(1-e_j) - \frac{1}{N_j(x)} \sum_{i=1}^{N_j(x)} Y_i.$$

Proof. Let us consider a price vector x such that $x_j < x_k$ for all $k \neq j$. Since the change probability $p_{k\to j}$ (for $k \neq j$) is a decreasing function (see Appendix 2.6.1), $p_{k\to j}(x) > p_{k\to l}(x)$ for $l \neq j$ given the initial portfolio sizes n_j's are constant.

Below we use the stochastic orders ($\leq_{\text{st}}$, $\leq_{\text{cx}}$) and the majorization order ($\leq_{\text{m}}$) whose definitions and main properties are recalled in the Appendices 2.6.2 and 2.6.2 respectively. Using the convolution property of the stochastic order I times, we can show a stochastic order of the portfolio size

$$N_k(x) \leq_{\text{st}} N_j(x), \forall k \neq j.$$

Let us consider the underwriting result per policy

$$uw_j(x,n) = \frac{1}{n}\left(nx_j(1-e_j) - \sum_{i=1}^{n} Y_i\right) = x_j(1-e_j) - \sum_{i=1}^{n} \frac{1}{n} Y_i,$$

for insurer j having n policies, where Y_i denotes the total claim amount per policy.

Let $n < \tilde{n}$ be two policy numbers and $a_{\tilde{n}}, a_n \in \mathbb{R}^{\tilde{n}}$ be defined as

$$a_{\tilde{n}} = \left(\frac{1}{\tilde{n}}, \ldots, \frac{1}{\tilde{n}}\right) \quad \text{and} \quad a_n = \left(\underbrace{\frac{1}{n}, \ldots, \frac{1}{n}}_{\text{size } n}, \underbrace{0, \ldots, 0}_{\text{size } \tilde{n}-n}\right).$$

Since $a_{\tilde{n}} \leq_{\text{m}} a_n$ and $(Y_i)_i$'s are i.i.d. random variables, we have $\sum_i a_{\tilde{n},i} Y_i \leq_{\text{cx}} \sum_i a_{n,i} Y_i$ i.e.

$$\sum_{i=1}^{\tilde{n}} \frac{1}{\tilde{n}} Y_i \leq_{\text{cx}} \sum_{i=1}^{n} \frac{1}{n} Y_i.$$

For all increasing convex functions ϕ, the function $x \mapsto \phi(x+a)$ is still increasing and convex. Thus for all random variables X, Y such that $X \leq_{\text{icx}} Y$ and real numbers a, b, $a \leq b$, we have

$$E(\phi(X+a)) \leq E(\phi(X+b)) \leq E(\phi(Y+b)),$$

i.e. $a + X \leq_{\text{icx}} b + Y$.

As $x_j(1-e_j) \leq x_k(1-e_k)$ and using the fact that $X \leq_{\text{cx}} Y$ is equivalent to $-X \leq_{\text{cx}} -Y$, we have

$$uw_j(x,\tilde{n}) \leq_{\text{icx}} uw_k(x,n), \forall k \neq j.$$

Using Theorem 3.A.23 of Shaked and Shanthikumar (2007), except that for all ϕ convex, $E(\phi(uw_j(x,n)))$ is a decreasing function of n and $N_k(x) \leq_{\text{st}} N_j(x)$, we can show

$$UW_j = uw_j(x, N_j(x)) \leq_{\text{icx}} uw_k(x, N_k(x)) = UW_k.$$

□

2.4.4 Numerical illustration

In this subsection, we present numerical illustrations of the repeated game. As explained at the beginning of this section, objective and solvency constraint functions depend on parameters evolving over time: the portfolio size $n_{j,t}$, the capital $K_{j,t}$, the break-even premium $\pi_{j,t}$. Doing this, we want to mimic the real economic actions of insurers on a true market: in private motor insurance, each year insurers and mutuals update their tariff depending on last year experience of the whole company and the claim experience of each particular customer through bonus-malus. In our game, we are only able to catch the first aspect.

In addition to this parameter update, we want to take into account the portfolio size evolution over time. As $n_{j,t}$ will increase or decrease, the insurer j may become a leader or lose leadership. Hence, depending on market share (in terms of gross written premium), we update the lapse, the expense and the sensitivity parameters $\alpha_{j,t}, \mu_{j,t}, e_{j,t}$ and $\beta_{j,t}$, respectively. Before the game starts, we define three sets of parameters for the leader, the outsider and the challenger, respectively. At the beginning of each period t, each insurer has its parameter updated according to its market share $\text{GWP}_{j,t-1}/\text{GWP}_{t-1}$. When the market share is above 40% (respectively 25%), insurer j uses the "leader" parameter set (respectively the "outsider" parameter set), otherwise insurer j uses the "challenger" parameter set. There is only one parameter not evolving over time: the credibility factor ω_j which is set to a common value of $\omega_j = 9/10$ in our numerical experiments.

For the following numerical application, the parameters are summarized in Table 2.9. Lapse parameters * are identical as in Subsection 2.2.7, but we change the expense and sensitivity parameters to get realistic outputs.

	α_j	μ_j	e_j	β_j
leader	-12.143	9.252	0.2	3.465
challenger	-9.814	7.306	0.18	4.099
outsider	-8.37	6.161	0.16	4.6

Table 2.9: Parameter sets

Random paths

In order to compare the two different loss models (PLN, NBLN) and the sensitivity functions $\bar{f}_j, \tilde{f}_j$, we fix the seed for losses and insured moves. That's why we observe similar patterns for the four situations. On Figures 2.1 and 2.2, we plot the individual premium $x^{\star}_{j,t}$, the gross written premium $\text{GWP}_{j,t}$, the loss ratio $\text{LR}_{j,t}$ and the solvency coverage ratio $K_{j,t}/\text{SCR}_{j,t}$.

∗. We give here only the lapse parameter for the price sensitivity function $\bar{f}_j$, but there are also three parameter sets for $\tilde{f}_j$.

An interesting feature of these random paths is that a cyclic pattern is observed for the market individual premium m_t, strong correlations of gross written premiums $\text{GWP}_{j,t}$ and loss ratio $\text{LR}_{j,t}$ for each insurer j. We fit a basic second-order autoregressive model on market premium (i.e. $X_t - m = a_1(X_{t-1} - m) + a_2(X_{t-2} - m) + \epsilon_t$) *. Estimation on the serie $(m_t)_t$ leads to period of 11.01 and 9.82 years, respectively for Figures 2.1 and 2.2.

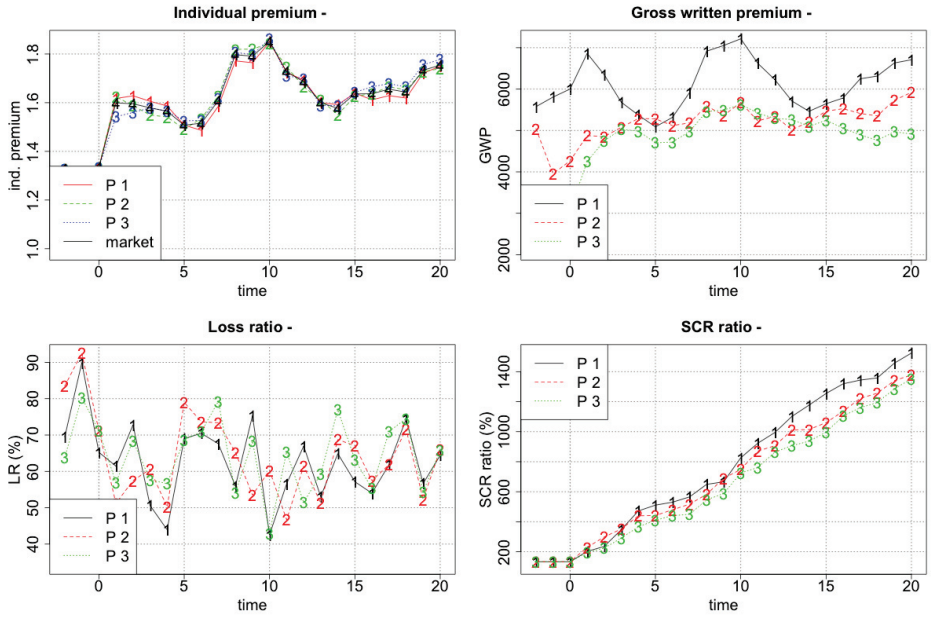


Figure 2.1: A random path for NBLN loss model and $\tilde{f}_j$

Furthermore, this numerical application shows that insurers set premium well above the pure premium $E(Y) = 1$. Thus, the insurer capitals tend to infinite as we observe on the (bottom right) plot of the solvency coverage ratio. We also do the computation of PLN/NBLN loss models with the sensitivity function $\bar{f}_j$. Similar comments apply, see Appendix 2.6.2.

Some Monte-Carlo estimates

In this subsection, we run the repeated game a certain of times with the NBLN loss model and the price sensitivity function $\tilde{f}_j$ in order to assess certain indicators by a Monte-Carlo method. We choose a sample size of $2^{14} \approx 16000$ and a time horizon $T = 20$. Our indicators are: (i) the ruin probability of insurer j before time $\widetilde{T}$ and (ii) the probability a leader at time $\widetilde{T}$, where $\widetilde{T} = T/2$ or T.

Results are given in Table 2.10. Estimates of ruin probabilities are extremely low, because the safety loadings of equilibrium premium are very high, see previous Figures. Leadership probabilities are more interesting. Recalling that Insurer 1 is the leader at time 0, the probability for Insurer 1 to be leader after t periods decreases quickly as t increases. After only 20 periods, Insurer 1 has losed its initial advantage.

Then, we look at the underwriting result by policy to see if some insurers underwrite a deliberate loss. As the first quartile is above zero, we observe that negative underwriting

∗. When $a_2 < 0$ and $a_1^2 + 4a_2 < 0$, the AR(2) is p-periodic with $p = 2\pi \arccos\left(\frac{a_1}{2\sqrt{-a_2}}\right)$.

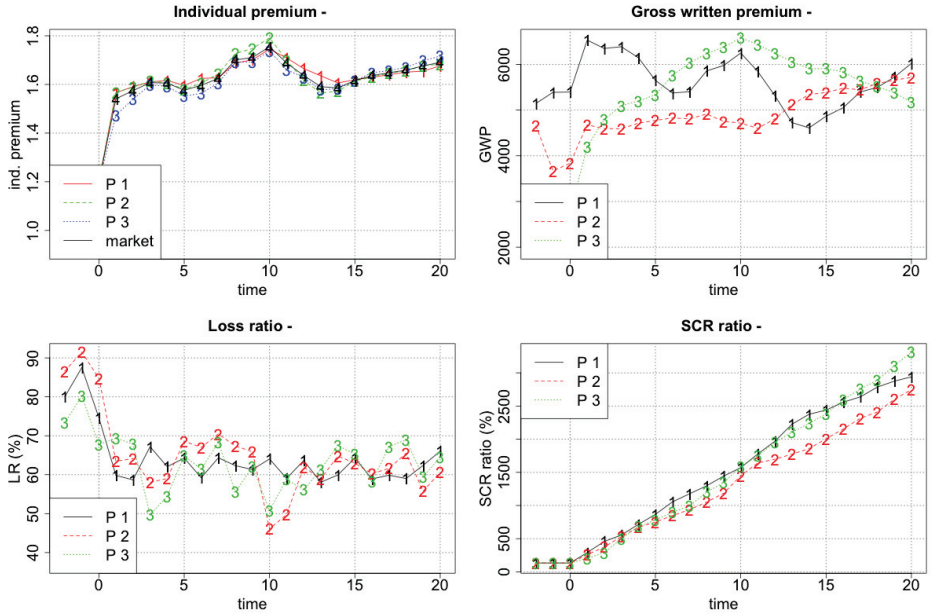


Figure 2.2: A random path for PLN loss model and $\tilde{f}_j$

	Ruin before $t = 10$	Ruin before $t = 20$	Leader at $t = 5$	Leader at $t = 10$	Leader at $t = 20$
Insurer 1	6.1e-05	6.1e-05	0.593	0.381	0.331
Insurer 2	0	0	0.197	0.308	0.329
Insurer 3	0.000244	0.000244	0.21	0.312	0.34

Table 2.10: Ruin and leadership probabilities

results are rather marginal. In fact, the probability of negative underwriting results are $(0.0352, 0.0378, 0.0358)$ for Insurers 1, 2 and 3, respectively.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Insurer 1	-0.7905	0.2309	0.3617	0.3563	0.4869	1.2140
Insurer 2	-0.4340	0.2279	0.3600	0.3555	0.4869	1.1490
Insurer 3	-0.4730	0.2308	0.3627	0.3563	0.4871	1.0950

Table 2.11: Summary statistics of underwriting result by policy at $t = 20$

On the left-hand plot of Figure 2.3, we analyze the (individual) market premium m_t. We plot quantiles at 5%, 50% and 95% as well as two random paths. The two plotted random paths show a cyclic behavior, whereas the three quantiles are stable in time. On each random path, we can fit an AR(2) model and estimate the cycle period. Only in 240 random paths, the fitted AR(2) is not periodic, i.e. $a_2 \geq 0$ or $a_1^2 + 4a_2 \geq 0$. Otherwise, the period is computed as $p = 2\pi \arccos\left(\frac{a_1}{2\sqrt{-a_2}}\right)$. On the right-hand plot of Figure 2.3, we plot the histogram of

estimated periods: average period is around 10.

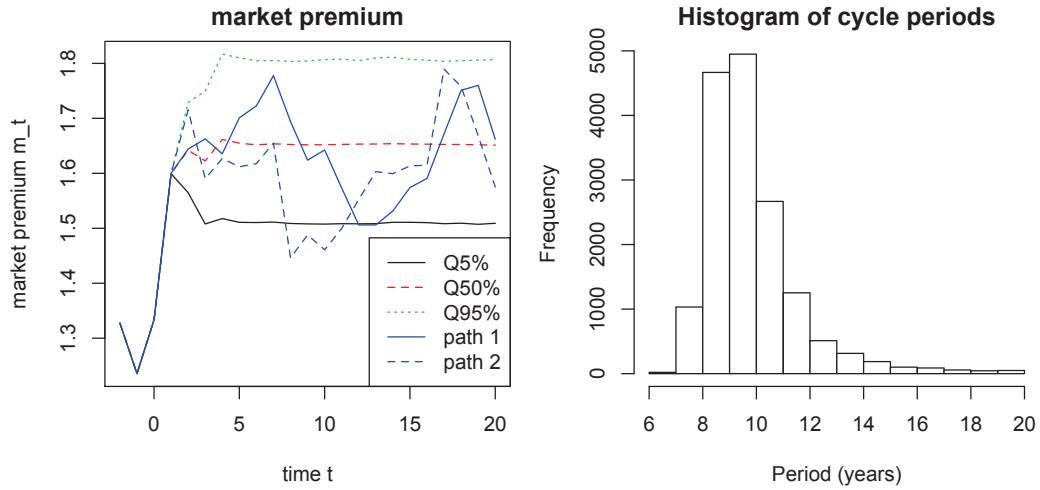


Figure 2.3: Market premium

Finally, on a short time horizon, $T = 5$, we want to assess the impact of initial portfolio size $n_{j,0}$ and capital value $K_{j,0}$ on the probability to be leader at time T. We consider Insurer 1 as the leader and Insurers 2 and 3 as identical competitors. We take different values of $K_{1,0}$ and $n_{1,0}$ for Insurer 1 and deduce capital values and portfolio sizes for Insurers 2 and 3 as $K_{2,0} = K_{3,0} = k_0\sigma(Y)\sqrt{n_{2,0}}$ and $n_{2,0} = n_{3,0} = (n - n_{1,0})/2$, where k_0 is a fixed solvency coverage ratio. The sensitivity analysis consists in increasing both market shares and capital values of Insurer 1 while other competitors have a decreasing market share and a constant coverage ratio.

We look at the probability for insurer i to be a leader in terms of gross written premium at period $T = 5$, i.e.

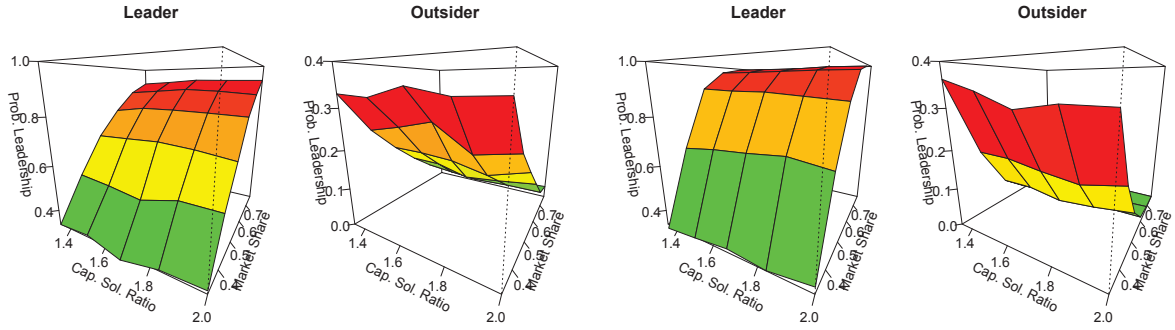
$$p_i = P(\forall k \neq i, \mathrm{GWP}_{i,T} > \mathrm{GWP}_{k,T} | N_{i,0} = n_i, K_{i,0} = k_i).$$

We test two loss models NBLN and PGLN, for which the margingal claim distribution is a compound negative binomial distribution with lognormal distribution, but for PGLN, the loss frequency among insurers is comonotonic, see Subsection 2.2.2.

On Figures 2.4, we observe the probability to a leader after five periods is an increasing function of the initial market share. The initial capital does not seem to have any influence, which can be explained by the high profit per policy. As one could expect, the comonotonic loss model (Figure 2.4b) is favorable to Insurer 1 than the independent case (Figure 2.4a).

2.5 Conclusion

This paper assesses the suitability of noncooperative game theory for insurance market modelling. The game-theoretic approach proposed in this paper gives first answers of the effect of competition on the insurer solvency whose a significant part is linked to the ability of insurers to sell contracts. The proposed game models a rational behavior of insurers in setting premium taken into account other insurers. The ability of an insurer to sell contracts



(a) NBLN (b) PGLN

Figure 2.4: Leadership probabilities of Leader (Insurer 1) and Outsiders (Insurers 2 and 3)

is essential for its survival. Unlike the classic risk theory where the collection of premiums is fixed per unit of time, the main source of risk for an insurance company in our game is a premium risk. We extends one-player model of Taylor (1986, 1987) and subsequent extensions based optimal control theory. To our knowledge, the use of a repeated of noncooperative game to model non-life insurance markets is new.

The game can be extended in various directions. A natural next step is to consider adverse selection among policyholders, since insurers do not propose the same premium to all customers. In practice, insurers do not propose the same premium to all customers. Considering two risk classes of individuals would be an interesting extension of the game, but we would also have to find a way to differentiate premium between individuals. A second extension is to model investment results as well as loss reserves. We could also consider reinsurance treaties for players in addition a catastrophe generator. However, we must avoid not to complexify too much the game as the game is already complex and deriving theoretical properties is an issue.

2.6 Appendix

2.6.1 On the one-period game

Implicit function theorem

Below the implicit function theorem, see, e.g., (Zorich, 2000, Chap. 8).

Theorem. *Let F be a bivariate C^1 function on some open disk with center in (a,b), such that $F(a,b)=0$. If $\frac{\partial F}{\partial y}(a,b)\neq 0$, then there exists an $h>0$, and a unique function φ defined for $]a-h,a+h[$, such that*

$$\varphi(a)=b \quad \text{and} \quad \forall |x-a|<h, F(x,\varphi(x))=0.$$

Moreover on $|x-a|<h$, the function φ is C^1 and

$$\varphi'(x)=-\left.\frac{\frac{\partial F}{\partial x}(x,y)}{\frac{\partial F}{\partial y}(x,y)}\right|_{y=\varphi(x)}.$$

Computation details

Computation is based on a Karush-Kuhn-Tucker (KKT) reformulation of the generalized Nash equilibrium problem (GNEP). We present briefly the problem reformulation and refer the interested readers to e.g. Facchinei and Kanzow (2009), Dreves et al. (2011) or Dutang (2012*a*). In our setting we have I players and three constraints for each player. For each j of the I subproblems, the KKT conditions are

$$\nabla_{x_j} O_j(x) - \sum_{1\leq m\leq 3} \lambda_m^j \nabla_{x_j} g_j^m(x) = 0,$$
$$0 \leq \lambda^j \perp g_j(x) \geq 0.$$

The inequality part is called the complementarity constraint. The reformulation proposed uses a complementarity function $\phi(a,b)$ to reformulate the inequality constraints $\lambda^j, g_j(x) \geq 0$ and $\lambda^{jT} g_j(x) = 0$.

A point satisfying the KKT conditions is also a generalized Nash equilibrium if the objective functions are pseudoconcave and a constraint qualification holds. We have seen that objective functions are either strictly concave or pseudoconcave. Whereas constraint qualifications are always verified for linear constraints, or strictly monotone functions, see Theorem 2 of Arrow and Enthoven (1961), which is also verified.

By definition, a complementarity function is such that $\phi(a,b) = 0$ is equivalent to $a, b \geq 0$ and $ab = 0$. A typical example is $\phi(a,b) = \min(a,b)$ or $\phi(a,b) = \sqrt{a^2+b^2} - (a+b)$ called the Fischer-Burmeister function. With this tool, the KKT condition can be rewritten as

$$\begin{matrix} \nabla_{x_j} L_j(x, \lambda^j) = 0 \\ \phi_\cdot(\lambda^j, g_j(x)) = 0 \end{matrix},$$

where L_j is the Lagrangian function for the subproblem j and $\phi_\cdot$ denotes the component wise version of ϕ. So, subproblem j reduces to solving a so-called nonsmooth equation. In this paper, we use the Fischer-Burmeister complementarity function. This method is implemented in the R package **GNE***.

Graphs of lapse functions

Properties of multinomial logit function

We recall that the choice probability function is defined as

$$\mathrm{lg}_j^k(x) = \mathrm{lg}_j^j(x) \left(\delta_{jk} + (1-\delta_{jk}) e^{f_j(x_j, x_k)} \right),$$

and

$$\mathrm{lg}_j^j(x) = \frac{1}{1 + \sum_{l\neq j} e^{f_j(x_j, x_l)}},$$

where the summation is over $l \in \{1, \ldots, I\} - \{j\}$ and f_j is the price function. The price function f_j goes from $(t,u) \in \mathbb{R}^2 \mapsto f_j(t,u) \in \mathbb{R}$. Partial derivatives are denoted by

$$\frac{\partial f_j(t,u)}{\partial t} = f'_{j1}(t,u) \quad \text{and} \quad \frac{\partial f_j(t,u)}{\partial u} = f'_{j2}(t,u).$$

$*$. Dutang (2012*c*).

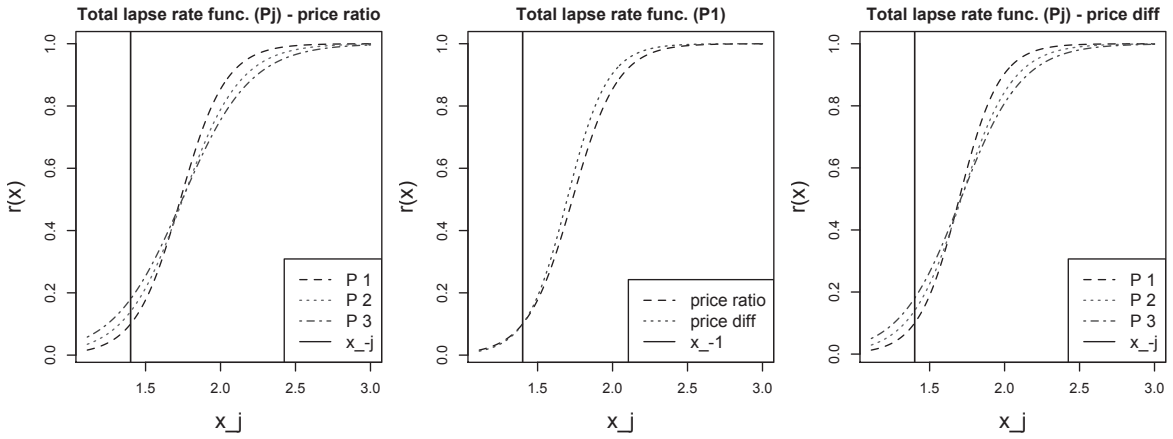


Figure 2.5: Total lapse rate functions

Derivatives of higher order use the same notation principle.

The lg function has the good property to be infinitely differentiable. We have

$$\frac{\partial \lg_j^j(x)}{\partial x_i} = -\frac{\partial}{\partial x_i}\left(\sum_{l\neq j} e^{f_j(x_j,x_l)}\right)\frac{1}{\left(1+\sum\limits_{l\neq j} e^{f_j(x_j,x_l)}\right)^2}.$$

Since we have

$$\frac{\partial}{\partial x_i}\sum_{l\neq j} e^{f_j(x_j,x_l)} = \delta_{ji}\sum_{l\neq j} f'_{j1}(x_j,x_l)e^{f_j(x_j,x_l)} + (1-\delta_{ji})f'_{j2}(x_j,x_l)e^{f_j(x_j,x_i)},$$

we deduce

$$\frac{\partial \lg_j^j(x)}{\partial x_i} = -\delta_{ji}\sum_{l\neq j}\frac{f'_{j1}(x_j,x_l)e^{f_j(x_j,x_l)}}{\left(1+\sum\limits_{l\neq j} e^{f_j(x_j,x_l)}\right)^2} - (1-\delta_{ji})f'_{j2}(x_j,x_l)\frac{f'_{j1}(x_j,x_i)}{\left(1+\sum\limits_{l\neq j} e^{f'_{j1}(x_j,x_l)}\right)^2}.$$

This is equivalent to

$$\frac{\partial \lg_j^j(x)}{\partial x_i} = -\left(\sum_{l\neq j} f'_{j1}(x_j,x_l)\lg_j^l(x)\right)\lg_j^j(x)\delta_{ij} - f'_{j2}(x_j,x_l)\lg_j^i(x)\lg_j^j(x)(1-\delta_{ij}).$$

Furthermore,

$$\frac{\partial \lg_j^j(x)}{\partial x_i}\left(\delta_{jk}+(1-\delta_{jk})e^{f_j(x_j,x_k)}\right) = -\left(\sum_{l\neq j} f'_{j1}(x_j,x_l)\lg_j^l(x)\right)\lg_j^k(x)\delta_{ij} - f'_{j2}(x_j,x_i)\lg_j^i(x)\lg_j^k(x)(1-\delta_{ij}).$$

and also

$$\lg_j^j(x)\frac{\partial}{\partial x_i}\left(\delta_{jk}+(1-\delta_{jk})e^{f_j(x_j,x_k)}\right) = \lg_j^j(x)(1-\delta_{jk})\left(\delta_{ik}f'_{j2}(x_j,x_k)e^{f_j(x_j,x_k)} + \delta_{ij}f'_{j1}(x_j,x_k)e^{f_j(x_j,x_k)}\right).$$

Hence, we get

$$\frac{\partial \lg_j^k(x)}{\partial x_i} = -\delta_{ij}\left(\sum_{l\neq j} f'_{j1}(x_j,x_l)\lg_j^l(x)\right)\lg_j^k(x) - (1-\delta_{ij})f'_{j2}(x_j,x_i)\lg_j^i(x)\lg_j^k(x) + (1-\delta_{jk})\left[\delta_{ij}f'_{j1}(x_j,x_k)\lg_j^k(x) + \delta_{ik}f'_{j2}(x_j,x_k)\lg_j^k(x)\right].$$

Similarly, the second order derivative is given by *

$$\begin{aligned}\frac{\partial^2 \lg_j^k(x)}{\partial x_m \partial x_i} = &-\delta_{ij}\left(\delta_{jm}\sum_{l\neq j} f''_{j11}(x_j,x_l)\lg_j^l + (1-\delta_{jm})f''_{j12}(x_j,x_m)\lg_j^m + \sum_{l\neq j} f'_{j1}(x_j,x_l)\frac{\partial \lg_j^l}{\partial x_m}\right)\lg_j^k \\ &- \delta_{ij}\left(\sum_{l\neq j} f'_{j1}(x_j,x_l)\lg_j^l\right)\frac{\partial \lg_j^k}{\partial x_m} \\ -(1-\delta_{ij})&\left(\left(\delta_{jm}f''_{j21}(x_j,x_i) + \delta_{im}f''_{j22}(x_j,x_i)\right)\lg_j^i\lg_j^k + f'_{j2}(x_j,x_i)\frac{\partial \lg_j^i}{\partial x_m}\lg_j^k + f'_{j2}(x_j,x_i)\lg_j^i\frac{\partial \lg_j^k}{\partial x_m}\right) \\ &+ (1-\delta_{jk})\delta_{ij}\left(\left(f''_{j11}(x_j,x_k)\delta_{jm} + f''_{j12}(x_j,x_k)\delta_{km}\right)\lg_j^k + f'_{j1}(x_j,x_k)\frac{\partial \lg_j^k}{\partial x_m}\right) \\ &+ (1-\delta_{jk})\delta_{ik}\left(\left(f''_{j21}(x_j,x_k)\delta_{jm} + f''_{j22}(x_j,x_k)\delta_{im}\right)\lg_j^k + f'_{j2}(x_j,x_k)\frac{\partial \lg_j^k}{\partial x_m}\right).\end{aligned}$$

Portfolio size function

We recall that the expected portfolio size of insurer j is defined as

$$\hat{N}_j(x) = n_j \times \lg_j^j(x) + \sum_{l\neq j} n_l \times \lg_l^j(x),$$

where n_j's denotes last year portfolio size of insurer j and $\lg_j^k$ is defined in equation (2.1).

The function $\phi_j : x_j \mapsto \lg_j^j(x)$ has the following derivative

$$\phi'_j(x_j) = \frac{\partial \lg_j^j(x)}{\partial x_j} = -\left(\sum_{l\neq j} f'_{j1}(x_j,x_l)\lg_j^l(x)\right)\lg_j^j(x).$$

For the two considered price function, we have

$$f'_{j1}(x_j,x_l) = \alpha_j\frac{1}{x_l} \quad \text{and} \quad \tilde{f}'_{j1}(x_j,x_l) = \tilde{\alpha}_j,$$

which are positive. So, the function ϕ_j will be a decreasing function.

For $l \neq j$, the function $\phi_l : x_j \mapsto \lg_l^j(x)$ has the following derivative

$$\phi'_l(x_j) = \frac{\partial \lg_l^j(x)}{\partial x_j} = -f'_{j2}(x_l,x_j)\lg_l^j(x)\lg_l^j(x) + f'_{j2}(x_l,x_j)\lg_l^j(x) = f'_{j2}(x_l,x_j)\lg_l^j(x)(1-\lg_l^j(x)).$$

∗. We remove the variable x when possible.

For the two considered price function, we have

$$f'_{j2}(x_j, x_l) = -\alpha_j \frac{x_j}{x_l^2} \text{ and } \tilde{f}'_{j2}(x_j, x_l) = -\tilde{\alpha}_j,$$

which are negative. So, the function ϕ_l will also be a decreasing function.

Therefore the portfolio size $x_j \mapsto \hat{N}_j(x)$ function has the following derivative

$$\frac{\partial \hat{N}_j(x)}{\partial x_j} = -n_j \left(\sum_{l \neq j} f'_{j1}(x_j, x_l) \lg_j^l(x) \right) \lg_j^j(x) + \sum_{l \neq j} n_l f'_{j2}(x_l, x_j) \lg_l^j(x)(1 - \lg_l^j(x)).$$

Hence, it is decreasing from the total market size $\sum_l n_l$ to 0. So the function $x_j \mapsto \hat{N}_j$ is both a quasiconcave and a quasiconvex function.

Therefore, using the C^2 characterization of quasiconcave and quasiconvex functions, we have that

$$\frac{\partial \hat{N}_j(x)}{\partial x_j} = 0 \Rightarrow \frac{\partial^2 \hat{N}_j(x)}{\partial x_j^2} = 0.$$

Note that the function $\hat{N}_j(x)$ is horizontal (i.e. has gradient of 0) when $x_j \to 0$ and $x_j \to +\infty$ for fixed x_{-j}.

Finally, we also need

$$\frac{\partial^2 \lg_j^j(x)}{\partial x_j^2} = -\sum_{l \neq j} f'_{j1}(x_j, x_l) \frac{\partial \lg_j^l}{\partial x_j} \lg_j^j - \left(\sum_{l \neq j} f'_{j1}(x_j, x_l) \lg_j^l \right) \frac{\partial \lg_j^j}{\partial x_j},$$

as f''_{j11} is 0 for the two considered functions. Since,

$$\left. \frac{\partial \lg_j^l}{\partial x_j} \right|_{l \neq j} = -\lg_j^l \sum_{n \neq j} f'_{j1}(x_j, x_n) \lg_j^n + \lg_j^l f'_{j1}(x_j, x_l) \text{ and } \frac{\partial \lg_j^j}{\partial x_j} = -\left(\sum_{l \neq j} f'_{j1}(x_j, x_l) \lg_j^l \right) \lg_j^j,$$

then we get

$$\frac{\partial^2 \lg_j^j(x)}{\partial x_j^2} = -\lg_j^j \sum_{l \neq j} \left(f'_{j1}(x_j, x_l) \right)^2 \lg_j^l + 2 \left(\sum_{l \neq j} f'_{j1}(x_j, x_l) \lg_j^l \right)^2 \lg_j^j .$$

Convexity concepts

Numerical applications for refined one-period game

2.6.2 On the dynamic game

Borel-Cantelli lemma and almost sure convergence

A random variable sequence $(X_n)_n$ is said to converge almost surely to X, if $P(X_n \to X) = 1$. A simple characterization of almost sure convergence is

$$X_n \xrightarrow{\text{p.s.}} X \Leftrightarrow P\left(\cap_{n_0 \geq 0} \cup_{n \geq 0} |X_n - X| \geq \epsilon \right) = 0, \forall \epsilon > 0.$$

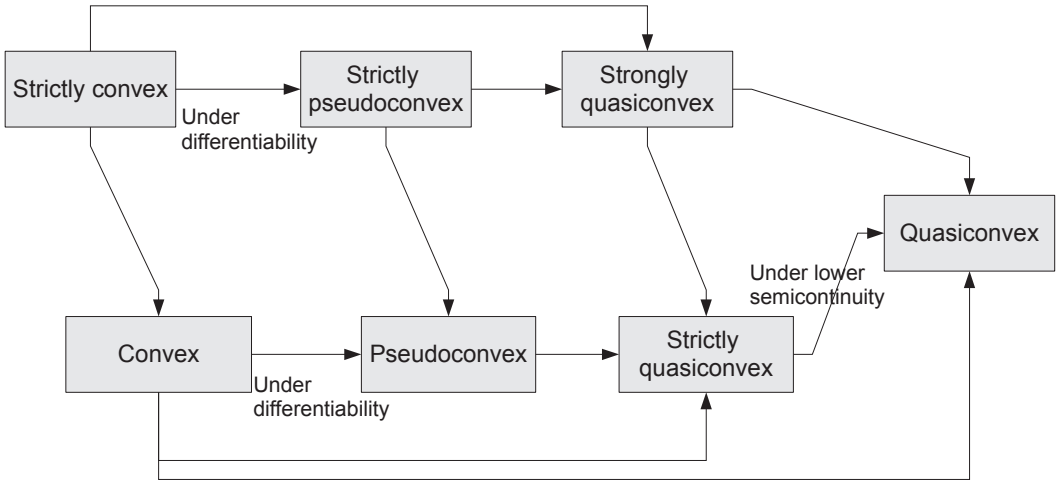


Figure 2.6: Convexity and generalized convexity, from Bazaraa et al. (2006)

$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - a\|\|^2$	$\|\|x^\star - m\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$	Nb
1.4048	1.0537	1.0642	0.1413	0.0802	-2583	1357	1226	11
1.507	1.1725	1.5968	0.2059	0.3061	-1609	2695	-1086	128
1.5527	1.6031	1.2504	0.2267	0.3525	-1359	-1217	2576	106
1	1.4896	1.5299	0.1449	0.2675	3127	-1790	-1338	56
2.4177	2.4107	2.3994	4.0024	4.6571	-83	-21	104	542
3	2.8239	2.8087	7.9441	8.838	-1191	509	683	38
1.5495	1.2359	3	3.0797	3.5277	-1182	3479	-2297	2
1.6611	3	1.3993	3.172	3.6768	-616	-3197	3813	6
3	3	2.9022	8.812	9.7701	-305	-238	543	2
3	2.9326	3	8.912	9.877	-219	321	-102	6
2.8877	2.874	3	8.3306	9.273	175	264	-439	11
2.8559	3	2.8279	8.0789	9.0088	259	-759	500	20
1	1.0214	3	3.065	3.4201	1277	1023	-2299	1
3	1.7906	1.5134	3.5479	3.8892	-4498	-357	4856	9
3	3	1.7839	6.436	7.1316	-4487	-3167	7654	1

Table 2.12: Premium equilibria - PLN with price difference function

Lemma (Borel–Cantelli)**.** *Let B_n be a sequence of events on a probability space. If the serie $\sum_n P(B_n)$ is finite $\sum_n P(B_n) < +\infty$, then*

$$P\left(\cap_{n_0 \geq 0} \cup_{n \geq 0} B_n\right) = 0.$$

An application of the Borel-Cantelli lemma to almost sure convergence is

$$\forall \epsilon > 0, \sum_n P(|X_n - X| \geq \epsilon) < +\infty \Rightarrow X_n \xrightarrow{\text{p.s.}} X.$$

Notation and definition of classic stochastic orders

Using the notation of Shaked and Shanthikumar (2007), we denote by $\leq_{\text{st}}$ the stochastic order, which is characterised as $X \leq_{\text{st}} Y$ if $\forall x \in \mathbb{R}, P(X > x) \leq P(Y > x)$. They are

$x_1^\star$	$x_2^\star$	$x_3^\star$	$\|\|x^\star - a\|\|^2$	$\|\|x^\star - m\|\|^2$	$\Delta\hat{N}_1$	$\Delta\hat{N}_2$	$\Delta\hat{N}_3$	Nb
1.4048	1.0537	1.0642	0.1413	0.0802	-2583	1357	1226	11
1.507	1.1725	1.5968	0.2059	0.3061	-1609	2695	-1086	128
1.5527	1.6031	1.2504	0.2267	0.3525	-1359	-1217	2576	106
1	1.4896	1.5299	0.1449	0.2675	3127	-1790	-1338	56
2.4177	2.4107	2.3994	4.0024	4.6571	-83	-21	104	542
3	2.8239	2.8087	7.9441	8.838	-1191	509	683	38
1.5495	1.2359	3	3.0797	3.5277	-1182	3479	-2297	2
1.6611	3	1.3993	3.172	3.6768	-616	-3197	3813	6
3	3	2.9022	8.812	9.7701	-305	-238	543	2
3	2.9326	3	8.912	9.877	-219	321	-102	6
2.8877	2.874	3	8.3306	9.273	175	264	-439	11
2.8559	3	2.8279	8.0789	9.0088	259	-759	500	20
1	1.0214	3	3.065	3.4201	1277	1023	-2299	1
3	1.7906	1.5134	3.5479	3.8892	-4498	-357	4856	9
3	3	1.7839	6.436	7.1316	-4487	-3167	7654	1

Table 2.13: Premium equilibria - NBLN with price difference function

various other *equivalent* characterizations, including expectectation order for all increasing function $E(\phi(X)) \leq E(\phi(Y))$, quantile order $F_X^{-1}(u) \leq F_Y^{-1}(u)$, distribution function order $F_X(x) \geq F_Y(x)$.

One important property of this stochastic order is the stability under convolutions, i.e. if $X \leq_{\text{st}} Y$ and $\tilde{X} \leq_{\text{st}} \tilde{Y}$, then $X+\tilde{X} \leq_{\text{st}} Y+\tilde{Y}$, see theorem 1.A.3 of Shaked and Shanthikumar (2007). By this mean, we can show that an ordering of binomial distributions. If $X \sim \mathcal{B}(n,p)$ and $Y \sim \mathcal{B}(n,q)$, such that $p \leq q$, then $X \leq_{\text{st}} Y$. Theorem 1.A.3 of Shaked and Shanthikumar (2007) also shows that the stochastic order is closed under mixtures.

The stochastic order is sometimes denoted by $\leq_1$ since $X \leq_1 Y$ requires that for all differentiable functions ϕ such that $\forall x, \phi^{(1)}(x) \geq 0$, we have $E(\phi(X)) \leq E(\phi(Y))$. With this reformulation in mind, we define the convex order denoted by $X \leq_2 Y$ or $X \leq_{\text{cx}} Y$ as $E(\phi(X)) \leq E(\phi(Y))$ for all convex functions ϕ. If we restrict to differentiable functions, it means $\forall x, \phi^{(2)}(x) \geq 0$. This explains the relation between notations $\leq_1$ and $\leq_2$.

Note that if expectations exist, then $X \leq_{\text{cx}} Y$ implies that $E(X) = E(Y)$, $Var(X) \leq Var(Y)$ and $E((X-a)_+) \leq E((Y-a)_+)$. By theorem 3.A.12 of Shaked and Shanthikumar (2007), the convex order is closed under mixtures and convolutions. We also have that $X \leq_{\text{cx}} Y$ is equivalent to $-X \leq_{\text{cx}} -Y$.

A third stochastic order is the increasing convex order: $X \leq_{\text{icx}} Y$ if for all increasing convex functions ϕ, $E(\phi(X)) \leq E(\phi(Y))$. For ϕ differentiable, it means that $\phi^{(1)}(x) \geq 0$, $\phi^{(2)}(x) \geq 0$.

Notation and definition of majorization orders

Using the book of Marshall and Olkin (1979), the majorization order $\leq_{\text{m}}$ is defined as $a \leq_{\text{m}} \tilde{a}$ if

$$\forall 1 \leq k \leq n-1, \sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} \tilde{a}_i \quad \text{and} \quad \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \tilde{a}_i.$$

A direct consequence of property B.2.c of Marshall and Olkin (1979) is that if $X_1, \ldots, X_n$ are exchangeable and $a, \tilde{a} \in \mathbb{R}^n$, $a \leq_{\mathrm{m}} \tilde{a}$ implies $\sum_i a_i X_i \leq_{\mathrm{cx}} \sum_i \tilde{a}_i X_i$.

Numerical experiment

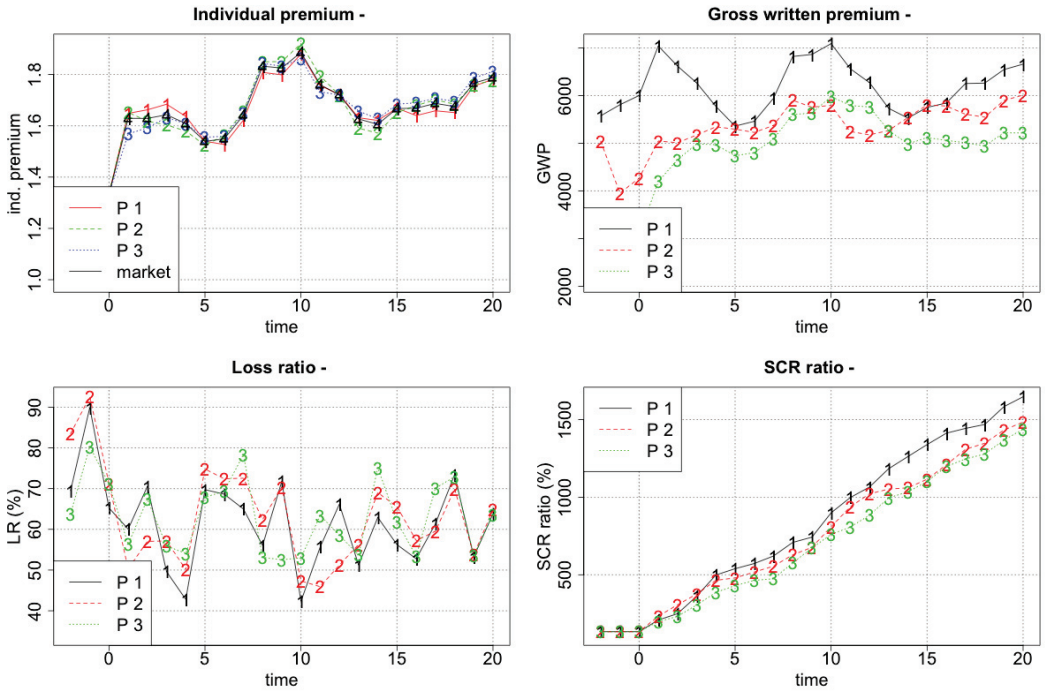


Figure 2.7: A random path for NBLN loss model and $\bar{f}_j$

Bibliography

Anderson, S. P., Palma, A. D. and Thisse, J.-F. (1989), 'Demand for differentiated products, discrete choice models, and the characteristics approach', *The Review of Economic Studies* **56**(1), 21–35.

Ania, A., Troeger, T. and Wambach, A. (2002), 'An evolutionary analysis of insurance markets with adverse selection', *Games and economic behavior* **40**(2), 153–184.

Arrow, K. J. and Enthoven, A. C. (1961), 'Quasiconcave programming', *Econometrica* **29**(4), 779–800.

Asmussen, S. and Albrecher, H. (2010), *Ruin Probabilities*, 2nd edition edn, World Scientific Publishing Co. Ltd. London.

Basar, T. and Olsder, G. J. (1999), *Dynamic Noncooperative Game Theory*, SIAM.

Bazaraa, M. S., Sherali, H. D. and Shetty, C. M. (2006), *Nonlinear Programming: Theory and Algorithms*, Wiley interscience.

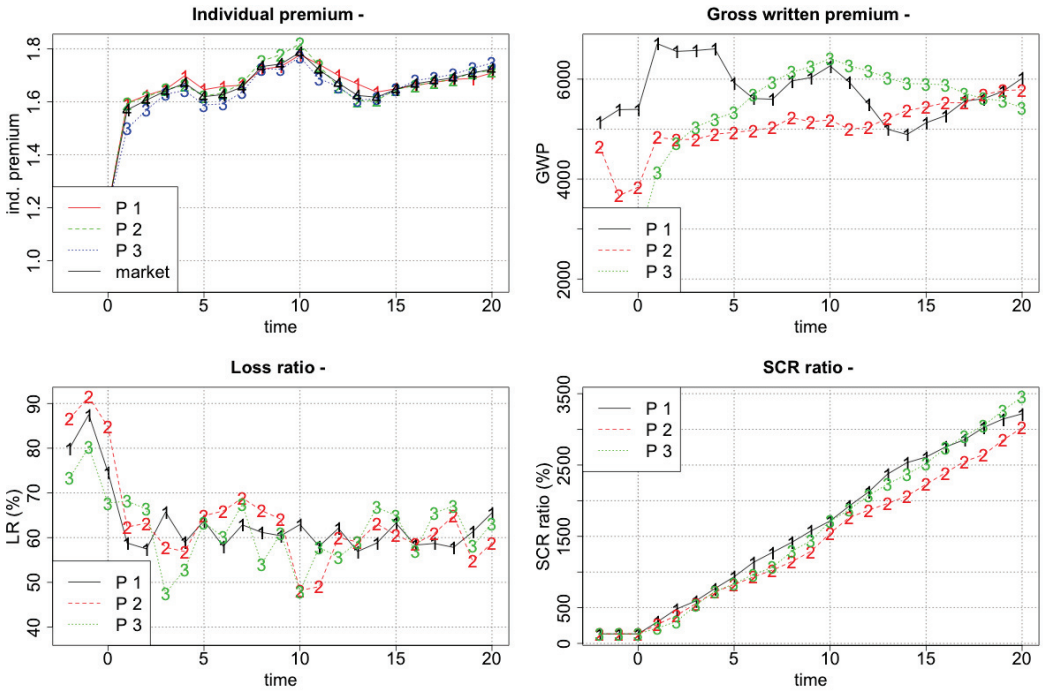


Figure 2.8: A random path for PLN loss model and $\bar{f}_j$

Borch, K. (1960), 'Reciprocal reinsurance treaties seen as a two-person cooperative game', *Skandinavisk Aktuarietidskrift* **1960**(1-2), 29–58.

Borch, K. (1975), 'Optimal insurance arrangements', *ASTIN Bulletin* **8**(3), 284–290.

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A. and Nesbitt, C. J. (1997), *Actuarial Mathematics*, The Society of Actuaries.

Bühlmann, H. (1984), 'The general economic premium principle', *ASTIN Bulletin* **14**(1), 13–21.

Cummins, J. D. and Outreville, J. F. (1987), 'An international analysis of underwriting cycles in property-liability insurance', *The Journal of Risk Finance* **54**(2), 246–262.

Demgne, E. J. (2010), Etude des cycles de réassurance, Master's thesis, ENSAE.

Derien, A. (2010), Solvabilité 2: une réelle avancée?, PhD thesis, ISFA.

Diewert, W. E., Avriel, M. and Zang, I. (1981), 'Nine kinds of quasiconcavity and concavity', *Journal of Economic Theory* **25**(3).

Dreves, A., Facchinei, F., Kanzow, C. and Sagratella, S. (2011), 'On the solutions of the KKT conditions of generalized Nash equilibrium problems', *SIAM Journal on Optimization* **21**(3), 1082–1108.

Dutang, C. (2012*a*), A survey of GNE computation methods: theory and algorithms. Working paper, ISFA.

Dutang, C. (2012*b*), Fixed-point-based theorems to show the existence of generalized Nash equilibria. Working paper, ISFA.

Dutang, C. (2012*c*), *GNE: computation of Generalized Nash Equilibria*. R package version 0.9.

Dutang, C., Albrecher, H. and Loisel, S. (2012), A game to model non-life insurance market cycles. Working paper, ISFA.

Emms, P., Haberman, S. and Savoulli, I. (2007), 'Optimal strategies for pricing general insurance', *Insurance: Mathematics and Economics* **40**(1), 15–34.

Facchinei, F., Fischer, A. and Piccialli, V. (2007), 'On generalized Nash games and variational inequalities', *Operations Research Letters* **35**(2), 159–164.

Facchinei, F. and Kanzow, C. (2009), Generalized Nash equilibrium problems. Updated version of the 'quaterly journal of operations research' version.

Feldblum, S. (2001), Underwriting cycles and business strategies, *in* 'CAS proceedings'.

Fields, J. A. and Venezian, E. C. (1989), 'Interest rates and profit cycles: A disaggregated approach', *Journal of Risk and Insurance* **56**(2), 312–319.

Fudenberg, D. and Tirole, J. (1991), *Game Theory*, The MIT Press.

Geoffard, P. Y., Chiappori, P.-A. and Durand, F. (1998), 'Moral hazard and the demand for physician services: First lessons from a French natural experiment', *European Economic Review* **42**(3-5), 499–511.

Golubin, A. Y. (2006), 'Pareto-Optimal Insurance Policies in the Models with a Premium Based on the Actuarial Value', *Journal of Risk and Insurance* **73**(3), 469–487.

Gron, A. (1994), 'Capacity constraints and cycles in property-casualty insurance markets', *RAND Journal of Economics* **25**(1).

Hardelin, J. and de Forge, S. L. (2009), Raising capital in an insurance oligopoly market. Working paper.

Hogan, W. W. (1973), 'Point-to-set maps in mathematical programming', *SIAM Review* **15**(3), 591–603.

Jablonowski, M. (1985), 'Earnings cycles in property/casualty insurance: A behavioral theory', *CPCU Journal* **38**(3), 143–150.

Kliger, D. and Levikson, B. (1998), 'Pricing insurance contracts - an economic viewpoint', *Insurance: Mathematics and Economics* **22**(3), 243–249.

Lemaire, J. and Quairière, J.-P. (1986), 'Chains of reinsurance revisited', *ASTIN Bulletin* **16**(2), 77–88.

Malinovskii, V. K. (2010), Competition-originated cycles and insurance companies. work presented at ASTIN 2009.

Markham, F. J. (2007), An investigation into underwriting cycles in the South African short-term insurance market for the period 1975 to 2006, Technical report, University of the Witwatersrand.

Marshall, A. W. and Olkin, I. (1979), *Inequalities: theory of majorization and its applications*, Academic Press.

McFadden, D. (1981), Econometric Models of Probabilistic Choice, *in* 'Structural Analysis of Discrete Data with Econometric Applications', The MIT Press, chapter 5.

Mimra, W. and Wambach, A. (2010), A Game-Theoretic Foundation for the Wilson Equilibrium in Competitive Insurance Markets with Adverse Selection. CESifo Working Paper No. 3412.

Moreno-Codina, J. and Gomez-Alvado, F. (2008), 'Price optimisation for profit and growth', *Towers Perrin Emphasis* **4**, 18–21.

Osborne, M. and Rubinstein, A. (2006), *A Course in Game Theory*, Massachusetts Institute of Technology.

Picard, P. (2009), Participating insurance contracts and the Rothschild-Stiglitz equilibrium puzzle. working paper, Ecole Polytechnique.

Polborn, M. K. (1998), 'A model of an oligopoly in an insurance market', *The Geneva Paper on Risk and Insurance Theory* **23**(1), 41–48.

Powers, M. R. and Shubik, M. (1998), 'On the tradeoff between the law of large numbers and oligopoly in insurance', *Insurance: Mathematics and Economics* **23**(2), 141–156.

Powers, M. R. and Shubik, M. (2006), 'A "square- root rule" for reinsurance', *Cowles Foundation Discussion Paper No. 1521.* .

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org*

Rees, R., Gravelle, H. and Wambach, A. (1999), 'Regulation of insurance markets', *The Geneva Paper on Risk and Insurance Theory* **24**(1), 55–68.

Rockafellar, R. T. and Wets, R. J.-V. (1997), *Variational Analysis*, Springer-Verlag.

Rosen, J. B. (1965), 'Existence and Uniqueness of Equilibrium Points for Concave N-person Games', *Econometrica* **33**(3), 520–534.

Rothschild, M. and Stiglitz, J. E. (1976), 'Equilibrium in competitive insurance markets: An essay on the economics of imperfect information', *The Quarterly Journal of Economics* **90**(4), 630–649.

Shaked, M. and Shanthikumar, J. G. (2007), *Stochastic Orders*, Springer.

Shapley, L. (1953), 'Stochastic games', *Proc. Nat. Acad. Sci. U.S.A.* **39**(10), 1095–1100.

Sigmund, K. and Hofbauer, J. (1998), *Evolutionary Games and Population Dynamics*, Cambridge University Press.

Taksar, M. and Zeng, X. (2011), 'Optimal non-proportional reinsurance control and stochastic differential games', *Insurance: Mathematics and Economics* **48**(1), 64–71.

Taylor, G. C. (1986), 'Underwriting strategy in a competitive insurance environment', *Insurance: Mathematics and Economics* **5**(1), 59–77.

Taylor, G. C. (1987), 'Expenses and underwriting strategy in competition', *Insurance: Mathematics and Economics* **6**(4), 275–287.

Trufin, J., Albrecher, H. and Denuit, M. (2009), 'Impact of underwriting cycles on the solvency of an insurance company', *North American Actuarial Journal* **13**(3), 385–403.

Venezian, E. C. (1985), 'Ratemaking methods and profit cycles in property and liability insurance', *Journal of Risk and Insurance* **52**(3), 477–500.

von Heusinger, A. and Kanzow, C. (2009), 'Optimization reformulations of the generalized Nash equilibrium problem using the Nikaido-Isoda type functions', *Computational Optimization and Applications* **43**(3).

Wambach, A. (2000), 'Introducing heterogeneity in the Rothschild-Stiglitz model', *Journal of Risk and Insurance* **67**(4), 579–591.

Zorich, V. (2000), *Mathematical Analysis I*, Vol. 1, Universitext, Springer.

Chapitre 3

Synthèse des méthodes de calcul d'équilibre de Nash généralisé

—

A survey of GNE computation methods: theory and algorithms

Theory attracts practice as the magnet attracts iron.
Carl Friedrich Gauss (1777–1855)

We consider a generalized game of N players defined by their objective function $\theta_i : \mathbb{R}^n \mapsto \mathbb{R}$ and their constraint function $g^i : \mathbb{R}^n \mapsto \mathbb{R}^{m_i}$. The generalized Nash equilibrium problem (GNEP for short) extends standard Nash equilibrium, since ones' player strategy depends on the rival players' strategies. Thus, when each player's strategy set does not depend on the other players' strategies, the GNEP reduces to standard Nash equilibrium problem. The GNEP consists in finding $x^\star$ such that for all $i = 1, \ldots, N$, $x_i^\star$ solves the subproblem

$$\min_{x_i \in \mathbb{R}^{n_i}} \theta_i(x_i, x_{-i}^\star) \quad \text{s.t.} \quad g^i(x_i, x_{-i}^\star) \leq 0, \tag{3.1}$$

where (x_i, x_{-i}) denotes the vector $(x_1, \ldots, x_i, \ldots, x_N) \in \mathbb{R}^n$ with $n = \sum_i n_i$ the total number of variables and $m = \sum_i m_i$ the total number of constraints.

GNEP arises from many practical problems, including telecommunications, engineering and economics applications, see Facchinei and Kanzow (2009) and the references therein for an overview of GNEPs. This paper aims to make a survey of computational methods to solve general GNEPs defined in Equation (3.1).

The paper is organized as follows: Section 3.1 present the different reformulations of GNEPs. Section 3.2 describes the numerous optimization methods that can solved a nonlinear reformulation of the GNEP. Finally, Section 3.3 carries out a numerical comparison of all algorithms presented in the previous section, before Section 3.4 concludes.

3.1 Problem reformulations

As presented in Equation (3.1), the generalized Nash equilibrium problem is not directly solvable. This section aims to present the different reformulations of the GNEP. On Figure 3.1, we present a basic flow-chart of the relationship among the different reformulations that we present below.

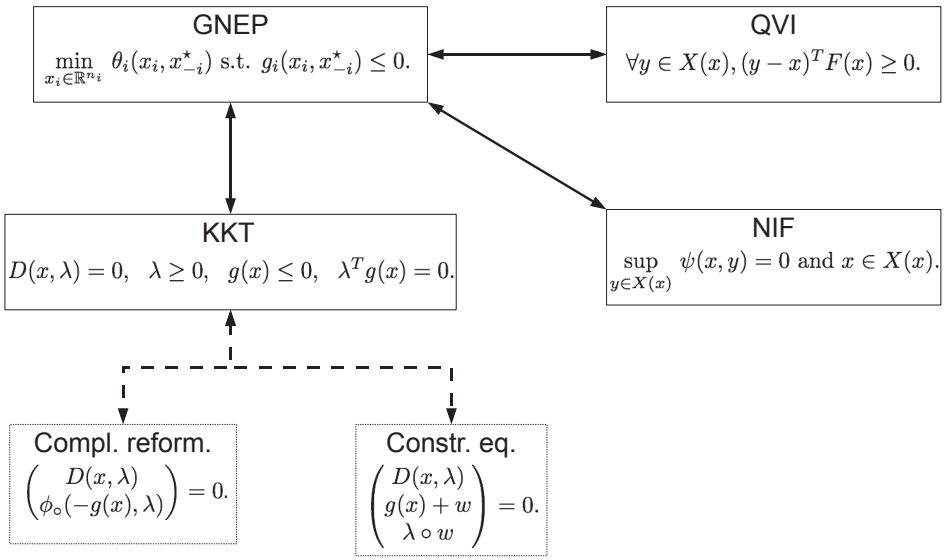


Figure 3.1: Map of GNE reformulations

3.1.1 The KKT conditions

The first reformulation uses the Karush-Kuhn-Tucker (KKT) conditions of the N optimization subproblems. We assume that both constraints and objective functions are twice continuously differentiable C^2. Let $x^\star$ be a solution of the GNEP. If a constraint qualification holds for all players, then for all player i, there exists a Lagrange multiplier $\lambda^{i\star} \in \mathbb{R}^{m_i}$ such that

$$\nabla_{x_i}\theta_i(x^\star) + \sum_{1\leq j\leq m_i} \lambda_j^{i\star}\nabla_{x_i} g_j^i(x^\star) = 0 \qquad (\in \mathbb{R}^{n_i}).$$
$$0 \leq \lambda^{i\star},\ -g^i(x^\star) \geq 0,\ g^i(x^\star)^T\lambda^{i\star} = 0 \qquad (\in \mathbb{R}^{m_i}).$$

Concatening the N subproblems, we get the following "extended" KKT system

$$D(x,\lambda) = 0, \quad \lambda \geq 0, \quad g(x) \leq 0, \quad \lambda^T g(x) = 0, \tag{3.2}$$

where the functions D, g are defined as

$$D(x,\lambda) = \begin{pmatrix} \nabla_{x_1} L_1(x,\lambda^1) \\ \vdots \\ \nabla_{x_N} L_N(x,\lambda^N) \end{pmatrix} \in \mathbb{R}^n, \lambda = \begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^N \end{pmatrix} \in \mathbb{R}^m, g(x) = \begin{pmatrix} g^1(x) \\ \vdots \\ g^N(x) \end{pmatrix} \in \mathbb{R}^m,$$

and L_i is the Lagrangian function $L_i(x,\lambda^i) = \theta_i(x) + g^i(x)^T\lambda^i$. The following theorem precises the necessary and sufficient condition between the original GNEP in Equation (3.1) and the KKT system in Equation (3.2).

Theorem. *Let a GNEP with twice continuity and differentiability of objective and constraint functions.*

(i) If $x^\star$ solves the GNEP at which all the player's subproblems satisfy a constraint qualification, then there exists $\lambda^\star \in \mathbb{R}^m$ such that $x^\star, \lambda^\star$ solve equation 3.2.

(ii) If $x^\star, \lambda^\star$ solve equation 3.2 and that the functions θ_i's are player convex and $\{y_i, g^i(y_i, x_{-i}) \leq 0\}$ are closed convex sets, then $x^\star$ solves the original GNEP.

Facchinei and Kanzow (2009) and Facchinei et al. (2009) report the previous theorem, respectively in Theorem 4.6 and Proposition 1. Using Fritz John conditions, see, e.g., Simon (2011) or Bazaraa et al. (2006), the player convexity of θ_i in item (ii) can be relaxed to player pseudoconvexity, i.e. $x_i \mapsto \theta_i(x)$ is pseudoconvexe.

The complementarity reformulation

A complementarity function $\phi : \mathbb{R}^2 \to \mathbb{R}$ is a function verifying the following property

$$\phi(a,b) = 0 \Leftrightarrow a \geq 0, b \geq 0, ab = 0.$$

Examples are $\phi_\wedge(a,b) = \min(a,b)$, $\phi_{FB}(a,b) = \sqrt{a^2+b^2} - (a+b)$, see, e.g., Facchinei and Pang (2003).

The complementarity reformulation of the KKT conditions is

$$\Phi(x,\lambda) = 0 \ \text{ where } \ \Phi(x,\lambda) = \begin{pmatrix} D(x,\lambda) \\ \phi_\circ(-g(x),\lambda) \end{pmatrix}, \tag{3.3}$$

where $\phi_\circ$ is the component-wise version of the complementarity function ϕ and D defined from the extended system. This reformulation of the GNEP is given in Facchinei et al. (2009), Facchinei and Kanzow (2009) and Dreves et al. (2011). For a general discussion of semismooth reformulations of optimization problems, Fukushima and Qi (1999).

The constrained equation reformulation

Dreves et al. (2011) also propose a constrained equation of the KKT system. Let $\circ$ denote the component-wise product, i.e. $x \circ y = (x_1 y_1, \ldots, x_N y_N)$. Let $w \in \mathbb{R}^m$ be slack variables (to transform inequality into equality). The KKT conditions are equivalent to

$$H(x, \lambda, w) = 0, \quad (x, \lambda, w) \in Z, \quad \text{and} \quad H(x, \lambda, w) = \begin{pmatrix} D(x, \lambda) \\ g(x) + w \\ \lambda \circ w \end{pmatrix}, \tag{3.4}$$

where Z is the set $Z = \{(x, \lambda, w) \in \mathbb{R}^{n+2m}, \lambda \geq 0, w \geq 0\}$.

3.1.2 The QVI reformulation

Let us first define the Variational Inequality (VI) and Quasi-Variational Inequality (QVI) problems. Variational inequality problems $\text{VI}(K, F(x))$ consist in finding $x \in K$ such that

$$\forall y \in K, (y - x)^T F(x) \geq 0,$$

where $F : K \mapsto \mathbb{R}^n$. VI problems typically arise in minimization problems. Quasi-variational inequality problems are an extension of VI problems where the constraint set K depends on x. The QVI problem is defined as

$$\forall y \in K(x), (y - x)^T F(x) \geq 0,$$

which is denoted by $\text{QVI}(K(x), F(x))$. Note that a QVI has very complex structure since y must satisfy $y \in K(x)$ for a vector x we are looking for.

The GNEP in Equation (3.1) can be reformulated as a QVI problem

$$\forall y \in X(x), (y - x)^T F(x) \geq 0, \quad \text{with} \quad F(x) = \begin{pmatrix} \nabla_{x_1} \theta_1(x) \\ \vdots \\ \nabla_{x_N} \theta_N(x) \end{pmatrix}, \tag{3.5}$$

and a constrained set $X(x) = \{y \in \mathbb{R}^n, \forall i, g^i(y_i, x_{-i}) \leq 0\}$. We have the following theorem precising the equivalence between the GNEP and the QVI.

Theorem. *If objective functions are C^1 and player-convex, the action sets $X_\nu(x_{-\nu})$ are closed convex, then we have $x^\star$ solves the GNEP if and only if $x^\star$ solves the QVI$(X(x), F(x))$ defined in Equation (3.5).*

This is Theorem 3.3 of Facchinei and Kanzow (2009) or Equation (3) Kubota and Fukushima (2010).

Penalized sequence of VI

We can express the KKT conditions for the QVI problem, but we naturally get back to Equation (3.2). According to Facchinei and Kanzow (2009), methods to solve general QVI

problem arising from GNEP are still missing. However, Fukushima and Pang (2005) propose to solve the QVI($X(x), F(x)$) as a sequence of penalized variational inequality problems VI($\tilde{X}, \tilde{F}_k$), where $\tilde{F}_k$ is defined as

$$\tilde{F}_k(x) = \begin{pmatrix} \nabla_{x_1}\theta_1(x) \\ \vdots \\ \nabla_{x_N}\theta_N(x) \end{pmatrix} + \begin{pmatrix} P_1(x) \\ \vdots \\ P_N(x) \end{pmatrix}, \tag{3.6}$$

with

$$P_\nu(x) = \sum_{i=1}^{m_\nu} \left(u_i^k + \rho_k g_i^\nu(x) \right)_+ \nabla_{x_\nu} g_i^\nu(x).$$

The set $\tilde{X}$ is either $\mathbb{R}^n$ or a box constraint set $[l, u] \subset \mathbb{R}^n$, (ρ_k) an increasing sequence of penalty parameters and (u_k) a bounded sequence. Theorem 3 of Fukushima and Pang (2005) shows the convergence of the VIP solutions $x_k^\star$ to a solution of the QVI under some smoothness conditions. We will see later that the QVI reformulation for a certain class of generalized games reduces to a standard VI problem. In that case, it makes sense to use that reformulation.

3.1.3 Nikaido-Isoda reformulation

We present a last reformulation of the GNEP, which was originally introduced in the context of standard Nash equilibrium problem. We define the Nikaido-Isoda function as the function ψ from $\mathbb{R}^{2n}$ to $\mathbb{R}$ by

$$\psi(x, y) = \sum_{\nu=1}^{N} [\theta(x_\nu, x_{-\nu}) - \theta(y_\nu, x_{-\nu})]. \tag{3.7}$$

This function represents the unilateral player improvement of the objective function between actions x and y. Let $\hat{V}$ be the gap function

$$\hat{V}(x) = \sup_{y \in X(x)} \psi(x, y).$$

Theorem 3.2 of Facchinei and Kanzow (2009) shows the relation between GNEPs and the Nikaido-Isoda function.

Theorem. *If objective functions θ_i are continuous, then $x^\star$ solves the GNEP if and only if $x^\star$ solves the equation*

$$\hat{V}(x) = 0 \quad \textit{and} \quad x \in X(x), \tag{3.8}$$

where the set $X(x) = \{y \in \mathbb{R}^n, \forall i, g^i(y_i, x_{-i}) \leq 0\}$ and $\hat{V}$ defined in (3.7). Furthermore, the function $\hat{V}$ is such that $\forall x \in X(x), \hat{V}(x) \geq 0$.

As for the QVI reformulation, Equation (3.8) has a very complex structure. There is no particular algorithm able to solve this problem for a general constrained set $X(x)$. But a simplification will occur in a special case.

3.1.4 Jointly convex case

In this subsection, we present reformulations for a subclass of GNEP called jointly convex case. Firstly, the jointly convex setting requires that the constraint function is common to all players $g^1 = \cdots = g^N = g$. Then, we assume, there exists a closed convex subset $X \subset \mathbb{R}^n$ such that for all player i,

$$\{y_i \in \mathbb{R}^{n_i}, g(y_i, x_{-i}) \leq 0\} = \{y_i \in \mathbb{R}^{n_i}, (y_i, x_{-i}) \in X\}.$$

The convexity of X implies that the constraint function g is quasiconvex with respect to all variables. However, we generally assume that g is convex with respect to all variables.

KKT conditions for the jointly convex case

In the jointly convex case, the KKT conditions (3.2) become

$$\nabla_{x_i}\theta_i(x^\star) + \nabla_{x_i}g(x^\star)\lambda^{i\star} = 0,\ 0 \leq \lambda^{i\star},\ -g(x^\star) \geq 0,\ g(x^\star)^T\lambda^{i\star} = 0, \tag{3.9}$$

since the constraint function is common to all players. But, there are still N Lagrange multipliers λ^i. Under the same condition as Subsection 3.1.1, a solution of this KKT system is also a solution of the original GNEP.

VI formulation for the jointly convex case

In the jointly convex case, the QVI reformulation (3.5) simplifies to a variational inequality problem VI(X, F)

$$\forall y \in X, (y - x)^T F(x) \geq 0, \quad \text{with} \quad F(x) = \begin{pmatrix} \nabla_{x_1}\theta_1(x) \\ \vdots \\ \nabla_{x_N}\theta_N(x) \end{pmatrix}, \tag{3.10}$$

under certain conditions with $X = \{y \in \mathbb{R}^n, \forall i, g(y_i, x_{-i}) \leq 0\}$. To understand that VI problem solutions are a subclass of GNEs, we just compare the KKT conditions of the VIP (3.10) and Equation 3.9. This is given in the following theorem, see, e.g., Theorem 3.9 of Facchinei and Kanzow (2009), Theorem 3.1 of Facchinei et al. (2007).

Theorem. *Assuming θ_i and g are C^1 functions and g is convex and θ_i player-convex. The subset of variational equilibrium verifying Equation (3.10) are the solution of the KKT system (3.9) with a common multiplier $\lambda^1 = \cdots = \lambda_N = \lambda^\star$.*

GNEs verifying the VI problem in Equation (3.10) are called variational or normalized equilibrium, see also Kulkarni and Shanbhag (2010) for a detailed discussion of the VI representation of the QVI reformulation of GNEPs.

NIF formulation for the jointly convex case

Recalling that for the Nikaido-Isoda function (3.7), the gap function is

$$\hat{V}(x) = \sup_{y \in X(x)} \psi(x, y).$$

In the jointly convex case, we get

$$\hat{V}(x) = 0 \quad \text{and} \quad x \in X(x), \tag{3.11}$$

where the set $X(x) = \{y \in \mathbb{R}^n, g(y_i, x_{-i}) \leq 0\}$. Still the computation of $\hat{V}$ is a complex optimization over a constrained set $X(x)$. As in the previous subsection, the class of GNE called variational equilibrium can be characterized by the NI formulation. We have the folllowing theorem.

Theorem. *Assuming θ_i and g are C^1 functions and g is convex and θ_i player-convex. $x^\star$ is a variational equilibrium if and only if $x^\star \in X$ and $V(x^\star) = 0$ with V defined as*

$$V(x) = \sup_{y \in X} \psi(x, y).$$

In the rest of the paper, we do not study all algorithms but rather focus on the most promising ones. We restrict our attention to general GNEPs and algorithms to solve the KKT system presented in Subsection 3.1.1. So, we do not study jointly convex GNEPs for which special methods have been proposed in the literature. These two situations differs widely, since in the general GNEP, we have to solve a nonlinear equation, while for the jointly convex case, we solve a fixed point equation or a minimization problem.

3.2 Methods to solve nonlinear equations

As introduced in many optimization books, see, e.g., Dennis and Schnabel (1996); Nocedal and Wright (2006); Bonnans et al. (2006), an optimization method to solve a nonlinear equation or more generally to find the minimum of a function is made of two components: a local method and a globalization scheme. Assuming the initial point is not "far" from the root or the optimal point, local methods use a local approximation of the function, generally linear or quadratic approximation based on the Taylor expansion, that is easier to solve. The globalization studies adjustments to be carried out, so that the iterate sequence still converges when algorithms are badly initialized.

To emphasize the prominent role of the globalization, we first look at a simple example of a nonlinear equation. Let $F : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be defined as

$$F(x) = \begin{pmatrix} x_1^2 + x_2^2 - 2 \\ e^{x_1 - 1} + x_2^3 - 2 \end{pmatrix}.$$

This function only has two roots $x^\star = (1, 1)$ and $\bar{x} = (-0.7137474, 1.2208868)$. We notice that the second component of F explodes as x_1 tends to infinity.

On Figure 3.2, we plot the contour level of the norm $||F(x)||_2$, as well as two iterate sequences $(x_n), (y_n)$ (see numbers 0, 1, 2,...) starting from the point $(x_0, y_0) = (-1, -3/2)$. The first sequence (x_n) corresponds to a "pure" Newton method, which we will present after, whereas the second sequence (y_n) combine the Newton method with a line search (LS). We can observe the sequence (y_n) converges less abruptly to the solution $x^\star$ than the sequence (x_n).

On Figure 3.3, we plot the contour level of the norm $||F(x)||_2$ with two iterate sequences $(x_n), (y_n)$, for pure and line-search Newton, respectively. But this time, sequences are initiated at $(x_0, y_0) = (2, 1/2)$. Despite being close the solution $\bar{x}$, the pure sequence (x_n) wanders in

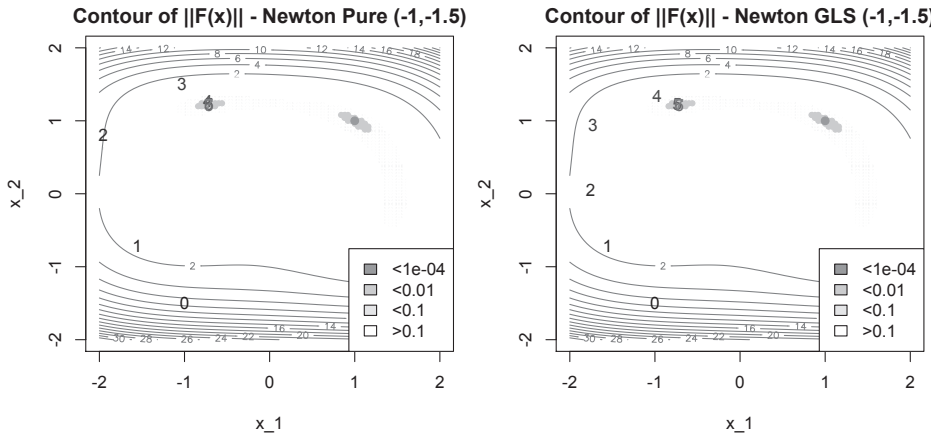


Figure 3.2: Pure and LS Newton methods, convergence to $x^\star$

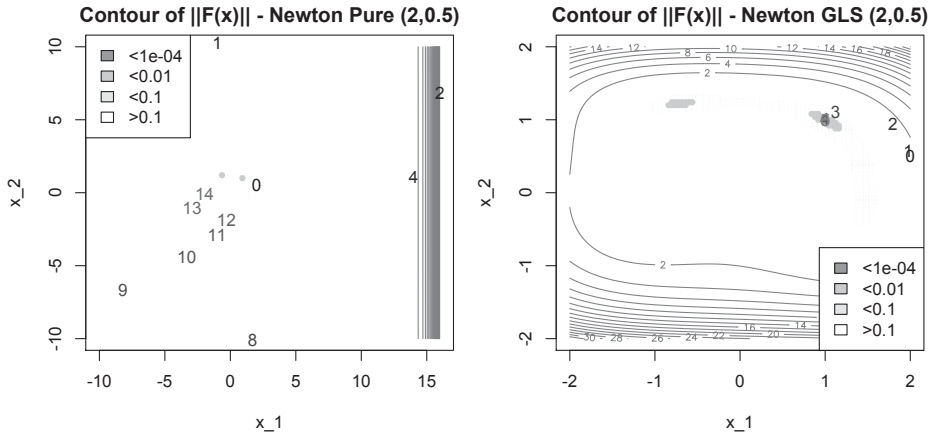


Figure 3.3: Pure and LS Newton methods, convergence to $\bar{x}$

the whole space $\mathbb{R}^2$ whereas the sequence (y_n) converges very quickly to $\bar{x}$. The sequence of (x_n) diverges because we use a pure local method.

This small example illustrates how essential a globalization scheme is. The globalization step stabilizes the sequence and improves its convergence. In the sequel, we first consider a smooth equation $F(z) = 0$. Subsection 3.2.1 presents local methods whereas Subsection 3.2.2 describes globalization schemes in a smooth framework. Then in Subsection 3.2.3, we present necessary extensions to deal with semismooth equations as encountered in Equation (3.3). Finally, Subsection 3.2.4 details contrained equation methods.

3.2.1 Local methods

Local methods are sequences $z_{k+1} = z_k + d_k$ where d_k is a root of a certain equation based on a local approximation of the root function $F : \mathbb{R}^n \mapsto \mathbb{R}^n$. The first and most simple method, the Newton method, uses a first-order Taylor expansion

$$F(z + h) = F(z) + J(z)h + o(h)\,,$$

where J denotes the Jacobian* $\operatorname{Jac} F$. Let $M_k^N(h) = F(z_k) + J(z_k)h$ be the model function. At the kth iterate, the Newton step consists in solving the system $M_k^N(d_k) = 0$. We get the following equation

$$J(z_k)d_k = -F(z_k). \tag{3.12}$$

Note that there is no guarantee that $J(z_k)$ is nonsingular.

Another method consists in replacing the Jacobian by an approximate matrix, which will be always invertible. The direction d_k solves

$$H_k d_k = -F(z_k), \tag{3.13}$$

where H_k is updated by a so-called quasi-Newton scheme. This is equivalent to a model function $M_k^{QN}(h) = F(z_k) + H_k h$. A quasi-Newton scheme needs an iterative process to update the approximate Jacobian from H_k to H_{k+1}. The choice of the matrix is large, since there are n^2 terms. We can set first that

$$F(z_k) = M_{k+1}^{QN}(z_{k+1} - z_k) \Leftrightarrow H_{k+1}s_k = y_k,$$

where $s_k = z_{k+1} - z_k$ and $y_k = F(z_{k+1}) - F(z_k)$. The latter equation is called the secant equation. We could have considered a scheme to approximate directly the inverse of the Jacobian by W_k with the secant equation $s_k = W_{k+1}y_k$. But still, we have an underdetermined system n equations for $n \times n$ matrix.

Having no other property on the Jacobian of F (e.g. symmetry or positiveness), we generally consider the matrix that makes the smallest possible change to the preceding Jacobian according to the Frobenius norm†, see (Nocedal and Wright, 2006, Chap. 11). For an implementation point of view, the smallest possible change feature reduces the linear algebra computation work from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ cost.

If we consider minimizing $\min_H ||H - H_k||_F$ for all matrix H verifying the secant equation, we get the (good) Broyden scheme

$$H_{k+1} = H_k + \frac{y_k - H_k s_k}{s_k^T s_k} s_k^T \Leftrightarrow W_{k+1} = W_k + \frac{(s_k - W_k y_k) y_k^T W_k}{s_k^T W_k y_k},$$

using the Sherman-Morrison formula. Otherwise, if we minimize $\min_W ||W - W_k||_F$ for all matrix W verifying the secant equation, then we will obtain the (bad) Broyden scheme

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k) y_k^T H_k}{y_k^T H_k s_k} \Leftrightarrow W_{k+1} = W_k + \frac{s_k - W_k y_k}{y_k^T y_k} y_k^T.$$

According to Broyden (1965), this method appears often unsatisfactory in practice, so it will be discused no further. For a general discussion of quasi-Newton methods, we refer to Dennis

∗. The Jacobian of a function f is defined as usual by

$$\operatorname{Jac} f(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_a} \\ \vdots & \frac{\partial f_i}{\partial x_j} & \vdots \\ \frac{\partial f_b}{\partial x_1} & \cdots & \frac{\partial f_b}{\partial x_a} \end{pmatrix}(x).$$

And the ∇f denotes the transpose of the Jacobian.

†. The Frobenius norm (also called the Euclidean norm) for matrix A is defined as $||A||_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$.

and Morée (1977) and for details on quasi-Newton methods specialized to nonlinear equations, see Broyden (1965) or (Nocedal and Wright, 2006, Chap. 11).

Another way to solve the equation $F(z) = 0$ requires minimizing the norm $f(z) = \frac{1}{2}||F(z)||_2^2$. But not all minima of the function f are roots of F because we must have $f(z) = 0$. A widely known method for this least square problem is the Gauss-Newton method. We minimize the model function $\frac{1}{2}M_k^N(d_k)^T M_k^N(d_k)$. We get

$$J(z_k)^T J(z_k) d_k = -J(z_k)^T F(z_k). \tag{3.14}$$

To prevent the right-hand side matrix to be nonsingular, the Levenberg-Marquardt method modifies Equation (3.14) to

$$\left[J(z_k)^T J(z_k) + \lambda_k I\right] d_k = -J(z_k)^T F(z_k), \tag{3.15}$$

where I denotes the $n \times n$ identity matrix and $\lambda_k \geq 0$. Various choices of Levenberg-Marquart parameter are possible: Fan and Yuan (2005) consider terms $\lambda_k = ||F(z_k)||_2^\delta$ with $\delta \in [1, 2]$. In practice, $\delta = 1$, i.e. $\lambda = ||F(z_k)||_2$, works much better than other δ's. We may also use $||J_k||$ or $\min(||F(z_k)||^2, ||J_k^T F(z_k)||^2)$. The Levenberg-Marquart is sometimes referred to as the modified Gauss-Newton method for its relation to Gauss-Newton method.

Summarizing this first subsection, local methods consider sequences of the form $z_{k+1} = z_k + d_k$ where the direction d_k is a solution of one of the above equations. Newton direction uses Equation (3.12), quasi-Newton Equation (3.13) and Levenberg-Marquardt Equation (3.15).

3.2.2 Globalization schemes

Now, we focus on globalization schemes, as Figures 3.2 and 3.3 show their relevancy. There are mainly two frameworks: line search and trust-region methods. To our knowledge, only two other methods exist continuation/homotopy, e.g., Allgower and Georg (2003) or Chapter 8 of Facchinei and Pang (2003)), or successive approximations, see, e.g., Qi and Jiang (1997) or Qi and Chen (1995).

Line-search techniques

Line-search techniques are a refinement of the local sequence by considering the sequence $z_{k+1} = z_k + t_k d_k$ where $t_k \in]0, 1]$ is the stepsize in direction d_k at the current iterate z_k. Note that the direction d_k is not unitary, i.e. $||d_k|| >> 1$ generally. Line-search techniques propose criteria to choose t_k. As the stepsize may reduces the (full) step from z_k to z_{k+1}, line-search version of an algorithm is sometimes called the damped version of that algorithm.

Let f be a merit function. We define the function $t \mapsto \phi_k(t) = f(z_k + t d_k)$. We want to find a good minimizer of ϕ_k. However, it is useless to find the global minimizer $\arg\min\ \phi_k(t)$, because we want to solve the outer problem $F(z) = 0$, and not the inner problem $\min\ \phi_k(t)$.

In the following, we assume we have a descent direction d_k for the merit function f, as a minimal condtion to choose t_k is $f(z_{k+1}) < f(z_k)$. This descent direction condition translates to $\phi_k'(0) < 0$. We are focused on two things, t_k should be big enougth to ensure a sufficient decrease of ϕ, and also t_k should not be too small to guarantee a sufficient big step.

One could think that requiring $f(z_{k+1}) < f(z_k)$ is enough to show convergence, but unfortunately not. In literature, see, e.g., Dennis and Schnabel (1996); Nocedal and Wright (2006);

Bonnans et al. (2006), two typical conditions are used to determine an appropriate stepsize. Let $0 < c_1 < 1/2 < c_2 < 1$. The Armijo condition ensures a decrease of f

$$\phi_k(t) \leq \phi_k(0) + tc_1\phi_k'(0) \Leftrightarrow f(x_k + td_k) \leq f(x_k) + tc_1\nabla f(x_k)^T d_k.$$

The curvature condition ensures an increase of $\nabla\phi$, implying a decrease of f,

$$\phi_k'(t) \geq c_2\phi_k'(0) \Leftrightarrow \nabla f(x_k + td_k)^T d_k \geq c_2\nabla f(x_k)^T d_k.$$

These two conditions are referred to the Wolfe conditions. In this paper, we use a backtracking algorithm, for which the curvature condition is always satisfied. Let $t_{k,0} = 1$ be the initial guess of the stepsize. The backtracking algorithm is defined as follows

Repeat until $f(x_k + td_k) \leq f(x_k) + t_{k,i}c_1\nabla f(x_k)^T d_k$ satisfied,
– propose a new $t_{k,i+1}$ using $t_{k,i}, \dots t_{k,0}$.
end Repeat

This algorithm always tests a full step with $t_{k,0} = 1$, otherwise the above algorithm tries a new stepsize. For the backtracking line search, a classic result shows that the full step will be eventually satisfied as z_k tends to a solution. We test two stepsize proposal algorithms. The geometric line search uses

$$t_{k,i+1} = \rho \times t_{k,i},$$

with $0 < \rho < 1$, whereas the quadratic line search uses a quadratic approximation of ϕ using the information $\phi_k(t_{k,i})$, $\phi_k(t_{k,i-1})$, $\phi_k'(t_{k,i-1})$. We get

$$t_{k,i+1} = -\frac{1}{2}\frac{\phi_k'(t_{k,i-1})t_{k,i}^2}{\phi_k(t_{k,i}) - \phi_k(t_{k,i-1}) - \phi_k'(t_{k,i-1})t_{k,i}}.$$

Other proposal, such as cubic approximation, are given in Chapter 3 of Bonnans et al. (2006) or Chapter 6 of Dennis and Schnabel (1996).

Until now, we do not specify the merit function f. For nonlinear equation, we generally choose $f(z) = \frac{1}{2}||F(z)||_2^2$, sometimes referred to the residual function. This merit function has some defficiencies, since a local minimum is not necessarily a root of the function F. We will see later in the GNEP context, that f has still some interesting properties.

Line-search methods require to a tractable formula for the gradient $\nabla f(z) = \operatorname{Jac} F(z)^T F(z)$, when testing the Armijo condition. However, in a quasi-Newton framework, we do not necessarily have a tractable Jacobian. One way to deal with this is to use a numerical Jacobian, e.g., based on the forward difference. We use Dennis and Schnabel (1996)'s algorithm A5.4.1 defined by

$$D(F)(z) = (D_1, \dots, D_n), \quad \text{with} \quad D_j = \frac{F(z + h_j e_j) - F(z)}{h_j} \in \mathbb{R}^n,$$

where e_j is the jth unit vector and h_j a small step, typically, $h_j = \sqrt{\epsilon} z_j$ where ϵ is the epsilon machine ($\epsilon = 1e^{-16}$).

Trust-region approach

Trust-region strategies relaxe the constraint that d_k is a descent direction. Line search assumes the "best" point from z_k lies on the half-line $z_k + \mathbb{R}_+ d_n$. Quoting Bonnans et al. (2006), "what is magic in this half line? answer: nothing". Trust-region approach will look for

appropriate steps h_k in a "small" region around z_k. Such regions are not the half-line as in line search.

To find the root of $F(x) = 0$, trust-region methods minimizes a local quadratic approximation m_k of a merit function $f : \mathbb{R}^n \mapsto \mathbb{R}$ on a bounded subset: the trust region $\{z, ||z - z_k|| \leq \Delta_k\}$. The name comes from viewing Δ_k as providing a region in which we can trust m_k to adequately model f. Classic trust region methods consider the following model function

$$m_k(h) = f(z_k) + g(x_k)^T h + \frac{1}{2} h^T \tilde{H}_k h,$$

with $g(x_k)$ the approximate gradient of f and $\tilde{H}_k$ a (ideally) positive definite matrix approximating the Hessian of f.

To adapt the trust region radius Δ_k, we define the following ratio

$$\rho_k(h) = \frac{f(x_k) - f(x_k + h)}{m_k(0) - m_k(h)}.$$

$\rho_k(h)$ is the ratio of the actual reduction and the predicted reduction of the merit function for a step h. The higher is $\rho_k(h)$, the higher is the reduction of the merit function f for a given h. Now, we can define a generic algorithm for a model function m_k, see, e.g., Nocedal and Wright (2006).

Init $\Delta_0 = 1, \eta_0 > 0, 0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 < 1 < \gamma_2$.

Iterate until a termination criterion is satisfied,

- get $h_k = \underset{||h|| < \Delta_k}{\arg\min}\ m_k(h)$ (approximately),
- compute $\rho_k(h_k)$,
  - if $\rho_k(h_k) < \eta_1$ then $\Delta_{k+1} = \gamma_1 \Delta_k$ (unsuccessful),
  - else if $\rho_k(h_k) > \eta_2$ and $||h_k|| = \Delta_k$ then $\Delta_{k+1} = \min(\gamma_2 \Delta_k, \Delta_{max})$ (very successful),
  - else $\Delta_{k+1} = \Delta_k$.
- next iterate
  - if $\rho_k(h_k) > \eta_0$ then $x_{k+1} = x_k + h_k$,
  - else $x_{k+1} = x_k$.

end Iterate

Typical values of parameters are $\Delta_0 = 1$ or $\frac{||g_0||}{10}$, $\Delta_{max} = 10^{10}$ for radius bounds, $\eta_0 = 10^{-4}$, $\eta_1 = \frac{1}{4}$, $\eta_2 = \frac{3}{4}$ for ratio threshold, $\gamma_1 = \frac{1}{2}$ and $\gamma_2 = 2$ for radius expansion coefficients.

If readers have been attentive, then they should have noticed that the algorithm cannot be used directly. In fact, we have to determine how to compute the solution h_k of the following minimization problem

$$\min_{||h|| < \Delta_k} m_k(h).$$

As for line search techniques, this problem has to be solved approximately as this problem is not our primary concern. There are two main methods to achieve this: Powell's dogleg and double dogleg methods. These two methods are heuristics to compute in one step an approximate solution h_k.

Let p_k^S be the scaled steepest descent direction and p_k^N be the Newton point defined as

$$p_k^S = -\frac{g_k^T g_k}{g_k^T \tilde{H}_k g_k} g_k \quad \text{and} \quad p_k^N = -\tilde{H}_k^{-1} g_k.$$

The Powell dogleg method uses a linear approximation of the model function $m_k(h)$. The Powell dogleg method is as follows, see, e.g., Chapter 6 of Powell (1970).

- If $||p^N|| \leq \Delta_k$, then $h^\star = p^N$.
- Else if $||p^S|| \geq \Delta_k$, then $h^\star = \Delta_k / ||p^S|| \times p^S$.
- Otherwise, we choose a convex combination between the two points p^S and p^N. That is we find a $\lambda \in [0,1]$ such that $||p^S + \lambda(p^N - p^S)|| = \Delta_k$. We get $h^\star = \lambda^\star p^N + (1-\lambda^\star)p^S$ with

$$\lambda^\star = \frac{- < p^S, p^N - p^S > + \sqrt{< p^S, p^N - p^S >^2 - ||p^N - p^S||^2 (||p^S||^2 - \Delta_k)}}{||p^N - p^S||^2},$$

  where $< ., . >$ denotes the scalar product and $||.||$ denotes the Euclidean norm.

Nocedal and Wright (2006) also propose a "simple" dogleg method which remove the step $||p^S|| \geq \Delta_k$, see Algorithm 11.6.

The double dogleg method finds an approximate solution of the model function $m_k(h)$ assuming the Hessian matrix is $\tilde{H}_k = L^T L$. The double dogleg method is a variant Powell dogleg method. Dennis and Schnabel (1996) propose the following procedure.

- If $||p^N|| \leq \Delta_k$, then $h^\star = p^N$.
- Else if $\eta_k ||p^N|| \leq \Delta_k$, then $h^\star = p^N \times \eta_k / \Delta_k$.
- Else if $||p^S|| \geq \Delta_k$ then $h^\star = p^S \times \Delta_k / ||p^S||$.
- Otherwise, we choose λ such that $||p^S + \lambda(\eta_k p^N - p^S)|| = \Delta_k$. We get back to the Powell dogleg case with $\eta_k p^N$ instead of p^N.

where the parameter $\eta_k \leq 1$ is defined as

$$\eta_k = 0.2 + 0.8 \frac{\alpha^2}{\beta |g_k^T d^N|},$$

with $\alpha = ||g_k||^2, \beta = ||Lg_k||^2$.

As for line-search techniques, we use the merit function $f(z) = \frac{1}{2}||F(z)||_2^2$. We recall that the gradient is given by

$$g(z) = \mathrm{Jac}\, F(z)^T F(z).$$

Therefore, the approximated Hessian $\tilde{H}_k$ is $\mathrm{Jac}\, F(z_k)^T \mathrm{Jac}\, F(z_k)$ as in the Gauss-Newton model. Hence, the steepest descent point and the Newton point have the following expression

$$p_k^S = -\frac{g_k^T g_k}{g_k^T J_k^T J_k g_k} g_k \quad \text{and} \quad p_k^N = -J_k^T g_k.$$

As in the previous subsection, when working a quasi-Newton method, the Jacobian is numerically approximated by a forward difference.

The special case of the Levenberg-Marquardt method

Until now, all globalization methods are adapted for the Newton or the Broyden direction defined in Equations (3.12) and (3.13). We need to precise how to globalize the Levenberg-Marquardt direction. This method was introduced in the context of the least-square problem $\min \frac{1}{2}||F(z)||_2^2$. In fact, there is a relation between the trust-region approach and the Levenberg-Marquardt method. The solution to the quadratic problem

$$\min_{||h|| < \Delta_k} f(z_k) + g(x_k)^T h + \frac{1}{2} h^T \tilde{H}_k h$$

is equivalent to the problem of finding $\lambda^\star, h^\star$ such that

$$(\tilde{H}_k + \lambda^\star I)h^\star = -g(x_k) \quad \text{and} \quad \lambda^\star(\Delta_k - ||h^\star||) = 0, \tag{3.16}$$

with the condition that $\tilde{H}_k + \lambda^\star I$ is a positive semidefinite matrix. We note in Equation (3.16) that variable $\lambda^\star$ has the same role as in the Levenberg-Marquardt method.

We can easily interpret Equation (3.16). If $h^\star$ lies strictly inside the trust-region then parameter $\lambda^\star$ is zero. Otherwise, $h^\star$ hits the radius Δ_k, and then parameter $\lambda^\star$ is set to a positive value. With this interpretation, the use a trust-region approach with the Levenberg-Marquardt method is redundant.

However, we still consider a globalization strategy for the Levenberg-Marquardt method. Firstly, we test the geometric line-search strategy, defined in Subsection 3.2.2, which is proposed in Fan and Yuan (2005), Yamashita and Fukushima (2000). Secondly, we use the adaptive Levenberg-Marquardt method discussed in Fan (2003). The method consists in adjusting the parameter λ_k based on $\lambda_k = \mu_k||F(z_k)||_2$, where μ_k is udpated at each iteration depending on the value of the ratio ρ_k. Their Algorithm 2.2 updates μ_k as in the generic algorithm of the previous Subsection 3.2.2. So, we do not restate here the updating scheme.

3.2.3 Extension to a semismooth setting

In the two previous subsections, a blanket assumption of differentiability of the root function F is assumed. In the GNEP context, the root function is defined as

$$\Phi(x, \lambda) = \begin{pmatrix} D(x, \lambda) \\ \phi_\circ(-g(x), \lambda) \end{pmatrix},$$

where the bottom part has some specificities: the function Φ is not everywhere differentiable. Hence, some adjustments have to be made to use local methods and globalization strategies, previously presented.

To emphasize the pitfalls between the assumed conditions of the previous methods and our current setting, we present local/global convergence results in the differentiable setting. Then, we describe the necessary adjustment for the GNEP context.

Differentiable setting

By definition, the Newton and Levenberg-Marquardt methods require the root function F to be differentiable, when the Broyden method does not. Let us analyze the convergence conditions for these methods. We concatenate all theorems in one, see, e.g., Theorems 11.2 and 11.5 of Nocedal and Wright (2006) for Newton and Broyden, respectively, Theorem 2.1 of Yamashita and Fukushima (2000) or Theorem 10 of Fischer (2002) for Levenberg-Marquardt.

Theorem. *Suppose that F is continuously differentiable in a open convex set $O \subset\in \mathbb{R}^n$. Let $z^\star$ be a solution of $F(z) = 0$ and let $(z_k^N)_k$ be the sequence generated by the Newton method. If the Jacobian at the solution $J(z^\star)$ is nonsingular, then $(z_k^N)_k$ converges superlinearly to $z^\star$. If in addition, F is Lipschitz continuously differentiable* near $z^\star$, then the convergence rate is quadratic.*

$*$. F is continuously differentiable C^1 and the Jacobian is Lipschitzian.

Let $(z_k^B)_k$ be the sequence generated by the Broyden method. If for the starting points, there exist $\delta, \epsilon > 0$, such that

$$||z_0^B - z^\star|| \leq \delta, ||B_0 - J(z^\star)|| \leq \epsilon,$$

then $(z_k^B)_k$ converges superlinearly to $z^\star$.

Let $(z_k^{LM})_k$ be the sequence generated by the Levenberg-Marquardt method. If F is Lipschitz continuously differentiable on O and $||F(x)||$ provides a local error bound, then $(z_k^{LM})_k$ converges superlinearly to $z^\star$.

In summary, convergence theorems require (i) F is Lipschitz continuously differentiable (LC) [1] in an open convex set around a solution $z^\star$ and (ii) the Jacobian at the solution $J(z^\star)$ is nonsingular for Newton and Broyden method or the function F verifies a local error bound.

Let f be the merit function. The global convergence of line-search techniques is guaranteed when we have the following conditions: (i) the set $\mathcal{L} = \{z, f(z) \leq f(z_0)\}$ is bounded, (ii) the Jacobian is bounded in an open convex set around a solution $z^\star$ and (iii) line-search always satisfies the Wolfe conditions with a descent direction, see, e.g. Theorem 11.6 of Nocedal and Wright (2006). We have seen that the backtracking algorithm satisfies the Wolfe conditions at each step.

The convergence of trust-region strategies is proved with similar conditions and requires also that the set $\mathcal{L}$ and the Jacobian are bounded. Furthermore, the approximated solution of the quadratic local problem $\min\ m_k(h)$ such that $||h|| < \Delta_k$ verifies two conditions: (i) $m_k(0) - m_k(h^\star) \geq c_1 ||J_k^T F_k|| \min(\Delta_k, ||J_k^T F_k||/||J_k^T J_k||)$ for some $c_1 > 0$ and (ii) $||h^\star|| \leq \gamma \Delta_k$ for some $\gamma \geq 1$.

Overall, local methods and globalization strategies need differentiability by definition and Lipschitz continuity and some additional conditions for convergence. In the next subsection, we will see how these conditions can be weakened.

Toward the non-differentiable setting

Getting back to our original GNEP, we want to solve the KKT conditions (3.3) using the complementarity reformulation of Subsection 3.1.1. Thus, the root function $\Phi : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n \times \mathbb{R}^m$ is defined as

$$\Phi(x, \lambda) = \begin{pmatrix} D(x, \lambda) \\ \phi_\circ(-g(x), \lambda) \end{pmatrix},$$

where the first component $D(x, \lambda)$ is composed of N derivatives of the Lagrangian function $L_i(x, \lambda^i)$ and the second component $\phi_\circ(-g(x), \lambda)$ is the component-wise application of the complementarity function ϕ on the overall constraint function $g : \mathbb{R}^n \mapsto \mathbb{R}^m$. Firstly, only the bottom component has some non-differentiability problem, because most complementarity functions $\phi(.,.)$ are non-differentiable at $(0, 0)$. In this paper, we use the minimum function $\phi_\wedge(a, b) = \min(a, b)$ and the Fischer-Burmeister function $\phi_{FB}(a, b) = \sqrt{a^2 + b^2} - (a + b)$.

To deal with non-differentiability, Clarke (1975) introduced the generalized gradient. Few laters, its famous book, Clarke (1990), provides a comprehensive presentation of the mathematical fundamentals of nonsmooth analysis. We briefly present here some elements necessary to our paper, refer to Appendix 3.5.1 for details.

Let $G : \mathbb{R}^n \mapsto \mathbb{R}^m$ a function with component G_j. By the Rademacher theorem, a locally Lipschitzian function is almost everywhere differentiable. We define first the limiting Jacobian, also called B(ouligand) subdifferential, denoted by $\partial_B G(x)$.

Definition (limiting Jacobian)**.** *The limiting Jacobian of G at x is defined as*

$$\partial_B G(x) = \{V_x, \exists (x_k)_k \in D_G, x_k \to x, JacG(x_k) \to V_x\},$$

where D_G is the differentiability set of G.

For his desirable properties, the Clarke's generalized Jacobian, based on the limiting Jacobian, is commonly used.

Definition (generalized Jacobian)**.** *The generalized Jacobian $\partial G(x)$ of G at x is the convex hull of the limiting Jacobian $\partial_B G(x)$ *.*

Let us start with an example. Using the Example 7.1.2 of Facchinei and Pang (2003), we define a function $G : \mathbb{R}^2 \mapsto \mathbb{R}^2$ with components $G_1(x) = \min(x_1, x_2)$ and $G_2(x) = |x_1|^3 - x_2$. G is not differentiable at (0,0). By splitting $\mathbb{R}^2$ into 4 parts, we can compute the limiting Jacobian

$$\partial_B G((0,0)) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} \right\}.$$

Then, the generalized Jacobian is $\partial G((0,0)) = \text{co } \partial_B G((0,0))$. From a computational point of view, using a component-wise version of the generalized Jacobian $\partial G(x) \subset \partial G_1(x) \times \cdots \times \partial G_m(x)$ is sometimes useful, see Appendix 3.5.1.

Now, we focus on the properties of the generalized Jacobian and more flexible criterion than continuous differentiability. Before introducing the semismoothness, we present directional derivative of a function.

Definition (directional derivative)**.** *For a locally Lipschitzian function G, the (classical) directional derivative at x along $h \in \mathbb{R}^n$ is defined as*

$$G'(x;h) = \lim_{t \downarrow 0} \frac{G(x+th) - G(x)}{t}.$$

The Hadamard directional derivative is

$$G'_H(x;h) = \lim_{h' \to h, t \downarrow 0} \frac{G(x+th') - G(x)}{t}.$$

The difference between classical and Hadamard directional derivatives lies the fact that we look at all directions $h' \to h$ and not only at h. Now, we can define the semismoothness.

Definition (semismooth)**.** *A locally Lipschitzian function G is semismooth at x, if for all $d \in \mathbb{R}^n$, the following limit exists*

$$\lim_{\substack{V \in \partial G(x+t\tilde{d}) \\ \tilde{d} \to d, t \downarrow 0}} V\tilde{d}.$$

In the case, this limit exists and equals to $G'_H(x;h)$.

∗. Note that if $m = 1$, the generalized Jacobian reduces to the generalized gradient, see Theorem 2.5.1 of Clarke (1990) for this characterization of the generalized gradient.

Therefore, the semismoothness of function H at x requires the Hadamard directional derivative at x to exists along any direction converging to h and not only along h.

Examples of semismooth functions are smooth, convex and piecewise linear functions. An important property is that composite, scalar products, sums, minimum, maximum preserve semismoothness. Indeed for the minimum function, the breakpoint is at $(0,0)$. By standard calculations, the directional derivate of $\phi_\wedge$ at this point is $\phi_\wedge((0,0);(a,b)) = \min(a,b)$. Futhermore, for all nonzero vector (a,b), $\phi_\wedge$ is differentiable and is uniquely given by

$$\nabla\phi_\wedge(a,b) = \begin{pmatrix} \mathbb{1}_{a\leq b} \\ \mathbb{1}_{a>b} \end{pmatrix}.$$

We deduce

$$\partial_B\phi_\wedge(0,0) = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \text{ and } \partial\phi_\wedge(0,0) = \left\{ \begin{pmatrix} \lambda \\ 1-\lambda \end{pmatrix}, \lambda \in [0,1] \right\}. \tag{3.17}$$

Furthermore, for all $V \in \partial\phi_\wedge(c,d)$, such that $(c,d) \to (a,b)$, we have

$$V \begin{pmatrix} a \\ b \end{pmatrix} - \phi_\wedge((0,0);(a,b)) = a\mathbb{1}_{c\leq d} + b\mathbb{1}_{c>d} - \min(a,b) = o((a,b)). \tag{3.18}$$

By Appendix 3.5.1, we conclude that $\phi_\wedge$ is semismooth at (0,0).

Finally, we introduce the strong semismoothness, also called 1-order semismoothness, that will be used in most convergence theorems.

Definition (strongly semismooth)**.** *A locally Lipschitzian function G is strongly semismooth at x if for all $d \to 0, \forall V \in \partial G(x+d)$, we have*

$$Vd - G'(x,d) = \mathcal{O}\big(||d||^2\big).$$

Based on Equation (3.18), we conclude that the minimum function $\phi_\wedge$ is not strongly semismooth at $(0,0)$. But, the Fischer-Burmeister function is strongly semismooth at $(0,0)$. In fact, by standard calculations, for all nonzero vector (a,b), we have

$$\nabla\phi_{FB}(a,b) = \begin{pmatrix} \frac{a}{\sqrt{a^2+b^2}} - 1 \\ \frac{b}{\sqrt{a^2+b^2}} - 1 \end{pmatrix}.$$

We deduce

$$\partial_B\phi_{FB}(0,0) = \{\nabla\phi_{FB}(a,b), (a,b) \neq (0,0)\} \text{ and } \partial\phi_{FB}(0,0) = \bar{B}((-1,-1),1), \tag{3.19}$$

where $\bar{B}$ denotes the closed ball.

Furthermore, for all $V \in \partial\phi_{FB}(c,d)$, such that $(c,d) \to (a,b)$, we have

$$V \begin{pmatrix} a \\ b \end{pmatrix} = \phi_{FB}(a,b) \text{ and } \phi_{FB}((0,0);(a,b)) = \phi_{FB}(a,b).$$

Hence, we have

$$V \begin{pmatrix} a \\ b \end{pmatrix} - \phi_{FB}((0,0);(a,b)) = 0.$$

We conclude that ϕ_{FB} is strongly semismooth at $(0,0)$, as proved in Kanzow and Kleinmichel (1998).

Now, we can express appropriately the generalized Jacobian of the GNEP. We denote by $J_\Phi(z)$ elements of the generalized Jacobian $\partial\Phi(z)$. Using chain rules and previous definitions, we have

$$J_\Phi(z) = \begin{pmatrix} \mathrm{Jac}_x D(x,\lambda) & \mathrm{diag}\left[\left(\nabla_{x_i} g^i(x)\right)_i\right] \\ -D_a(z)\mathrm{Jac}_x g(x) & D_b(z) \end{pmatrix}, \tag{3.20}$$

where Jac_x denotes the Jacobian with respect to x and diag[...] represents a block diagonal matrix, see Appendix 3.5.1 for a extended representation of the generalized Jacobian J_Φ.

The diagonal matrices D_a and D_b are given by

$$D_a(z) = \mathrm{diag}[a^1(x,\lambda^1),\ldots,a^N(x,\lambda^N)] \text{ and } D_b(z) = \mathrm{diag}[b^1(x,\lambda^1),\ldots,b^N(x,\lambda^N)],$$

with $a^i(x,\lambda^i), b^i(x,\lambda^i) \in \mathbb{R}^{m_i}$ defined as

$$(a^i_j(x,\lambda^i_j), b^i_j(x,\lambda^i_j)) = \begin{cases} \left(\phi'_a(-g^i_j(x),\lambda^i_j), \phi'_b(-g^i_j(x),\lambda^i_j)\right) & \text{if } (-g^i_j(x),\lambda^i_j) \neq (0,0), \\ (\xi_{ij},\zeta_{ij}) & \text{if } (-g^i_j(x),\lambda^i_j) = (0,0), \end{cases}$$

where ϕ'_a (resp. ϕ'_b) denotes the derivative of ϕ with respect to the first (second) argument a (b) and $(\xi_{ij},\zeta_{ij}) \in \bar{B}(p_\phi, c_\phi)$, the closed ball at p_ϕ of radius c_ϕ.

Let us precise the top-left part.

$$\mathrm{Jac}_x D(x,\lambda) = \begin{pmatrix} \mathrm{Jac}_{x_1} L_1(x,\lambda^1) & \ldots & \mathrm{Jac}_{x_N} L_1(x,\lambda^1) \\ \vdots & & \vdots \\ \mathrm{Jac}_{x_1} L_N(x,\lambda^N) & \ldots & \mathrm{Jac}_{x_N} L_N(x,\lambda^N) \end{pmatrix},$$

where $L_i(x,\lambda^i) = \nabla_{x_i}\theta_i(x) + \sum_{j=1}^{m_i} \nabla_{x_i} g^i_j(x)\lambda^i_j$. The top-right part is block diagonal and given by

$$\begin{pmatrix} \mathrm{Jac}_{x_1} g^1(x)^T & & 0 \\ & \ddots & \\ 0 & & \mathrm{Jac}_{x_N} g^N(x)^T \end{pmatrix}.$$

Let us specify the parameters p_ϕ and c_ϕ for the two considered complementarity functions. For the minimum function, using Equation (3.17), we have $p_\phi = (1/2, 1/2)$ and $c_\phi = 1/2$. For the Fischer-Burmeister function, using Equation (3.19), we have $p_\phi = (-1,-1)$ and $c_\phi = 1$. We refer to Kanzow and Kleinmichel (1998) and Facchinei and Pang (2003) for other complementarity functions.

If functions θ_i and g_i are $\mathrm{C}^{\mathrm{k}+1}$, by the chain rule, the root function Φ defined in Equation (3.20) is C^{k} except at points z such that $g^i_j(x) = \lambda^i_j = 0$. At these points, when the complementarity function is $\phi_\wedge$, Φ is semismooth, while for ϕ_{FB}, Φ is strongly semismooth.

Extension and local convergence in the semismooth framework

As the Jacobian of the root function is not available, the direction computation of local methods presented in Subsection 3.2.1 must be adapted. The solution consists in replacing the Jacobian by an element of the generalized Jacobian.

Considering the Newton method (3.12), the direction solves

$$J_k d_k = -F(z_k), \tag{3.21}$$

whereas for the Levenberg-Marquardt method (3.15), the direction solves

$$\left[J_k^T J_k + \lambda_k I\right] d_k = -J_k^T F(z_k), \tag{3.22}$$

with $J_k \in \partial F(z_k)$. Corresponding sequences are called generalized Newton and generalized Levenberg-Marquardt methods, respectively. For the quasi-Newton direction does not require any modification. Some authors also use the B-subdifferential $\partial_B F(z_k)$ or the component wise B-subdifferential $\partial_B F_1(z_k) \times \cdots \times \partial_B F_n(z_k)$ instead of the generalized Jacobian in Equations (3.21) and (3.22).

Now, we present theorems for local convergence. Local convergence theorems for smooth functions have been extended to nonsmooth functions by Qi and Sun (1993), cf. Theorem 3.2. We give below a slightly more general version than the original version.

Theorem. *Let $z^\star$ a solution of $F(z^\star) = 0$. If F is locally Lipschitzian and semismooth at $z^\star$ and all elements* $J^\star \in \partial_B F(z^\star)$ are nonsingular, then the generalized Newton method is well defined and converges superlinearly to $z^\star$. If in addition, F is strongly semismooth at $z^\star$, then the convergence rate is quadratic.*

A version of the previous theorem exists when the generalized Newton method use the limiting Jacobian $J_k \in \partial_B F(z_k)$ (instead of the generalized Jacobian) in Equation (3.21), see, e.g., Sun and Han (1997), Jiang and Ralph (1998) or Facchinei et al. (2009). On a similar idea, Jeyakumar (1998) presents a convergence theorem for a generalized Newton method when working with approximate Jacobian, which reduces to a single valued function under certain assumptions.

For the quasi-Newton approach, extensions have been proposed in the literature, e.g. Ip and Kyparisis (1992) and Qi (1997), where the differentiability is needed at a solution rather than in an open convex. Lopes and Martinez (1999) give a minimal condition (lesser than semismoothness) for a general quasi-Newton method to converge linearly.

As in the differentiable setting, e.g., Dennis and Schnabel (1996), the convergence of quasi-Newton methods for semismooth functions is done in two steps: (i) a theorem gives conditions of local linear convergence based on the limited difference between approximate Jacobian and elements in the generalized Jacobian, and (ii) another theorem gives an additional condition for a general quasi-Newton method to converge superlinearly. We report here Theorems 4.1 and 4.2 of Sun and Han (1997).

Theorem. *Let $z^\star$ a solution of $F(z^\star) = 0$. If F is locally Lipschitzian in the open convex $D \subset \mathbb{R}^n$ such as $z^\star \in D$. Consider the sequence $z_0 \in D$ and $z_{k+1} = z_k - V_k^{-1} F(z_k)$ with V_k a $n \times n$ matrix updated by a quasi-Newton scheme.*
Suppose F is semismooth at $z^\star$ and for all $J^\star \in \partial_b F(x^\star)$ are nonsingular. There exist constant $\epsilon, \Delta > 0$ such that if $||z_0 - z^\star|| \leq \epsilon$ and there exists $W_k \in \partial_b F(z_k)$ such that $||V_k - W_k|| \leq \Delta$, then the quasi-Newton sequence is well defined and converges linearly to $z^\star$.

Theorem. *Let F be locally Lipschitzian in a open convex $D \subset \mathbb{R}^n$. Assume F is semismooth and $\forall J^\star \in \partial_b F(z^\star)$ are nonsingular. Consider a sequence of nonsingular matrices V_k and*

∗. Originally, Qi and Sun (1993) use the generalized Jacobian and not the limiting Jacobian. But as mentioned in Qi and Jiang (1997) and Qi (1993), there is a weaker condition for superlinear convergence to hold, that is all elements in the limiting Jacobian are nonsingular.

points $z_{k+1} = z_k - V_k^{-1}F(z_k)$. *If* $(z_k)_k$ *converges to* $z^\star$ *then* $(z_k)_k$ *converges superlinearly to* $z^\star$ *and* $F(z^\star) = 0$ *is equivalent to* $\exists W_k \in \partial_b F(z_k)$,

$$\lim_{k\to+\infty} \frac{||(V_k - W_k)s_k||}{||s_k||} = 0,$$

with $s_k = z_{k+1} - z_k$.

Local convergence of the generalized Levenberg-Marquardt method is studied by Facchinei and Kanzow (1997), cf. Theorem 6.

Theorem. *Let* $z^\star$ *a solution of* $F(z^\star) = 0$, *with* F *a locally Lipschitzian, semismooth at* $z^\star$ *and* $\forall J^\star \in \partial_B F(z^\star)$ *are nonsingular. If the Levenberg-Marquardt parameters* (λ_k) *converge to 0, then the (generalized) Levenberg-Marquardt converges superlinearly to* $z^\star$. *If in addition,* F *is strongly semismooth and locally directionnally differentiable at* $z^\star$, *and* $\lambda_k = O(||J_k g_k||)$ *or* $\lambda_k = O(||g_k||)$, *then the sequence converges quadratically.*

We turn our attention to the assumption analysis of preceding theorems. All theorems require the root function to be semismooth at a solution $z^\star$. This is verified for our function Φ defined in Equation (3.3) as long as the complementarity function ϕ is semismooth. Furthermore, strong semismoothness improves the convergence rate. In the previous subsection, we have seen this requires for ϕ to be strongly semismooth, as e.g., the Fischer-Burmeister function.

The nonsingularity condition is required only for elements $J^\star$ of the limiting Jacobian at a solution $z^\star$. As analyzed in Facchinei et al. (2009), the limiting Jacobian (3.20) might have some identical rows at the bottom part. Let us investigate first this issue. We recall first the expression of generalized Jacobian

$$J_\Phi(z) = \begin{pmatrix} \mathrm{Jac}_x D(x,\lambda) & \mathrm{diag}\left[\left(\nabla_{x_i} g^i(x)\right)_i\right] \\ -D_a(z)\mathrm{Jac}_x g(x) & D_b(z) \end{pmatrix}.$$

As only terms (ξ_{ij}, ζ_{ij}) in diagonal matrices D^a and D^b change between the generalized and the limiting Jacobian of Φ, we study the nonsingularity condition directly on the generalized Jacobian. In a detailed form, the bottom part has the following structure

$$\left(\begin{array}{ccc|ccc} -D_1^a(x,\lambda^1)\mathrm{Jac}_{x_1} g^1(x) & \dots & -D_1^a(x,\lambda^1)\mathrm{Jac}_{x_N} g^1(x) & D_1^b(x,\lambda^1) & & 0 \\ \vdots & & \vdots & & \ddots & \\ -D_N^a(x,\lambda^N)\mathrm{Jac}_{x_1} g^N(x) & \dots & -D_N^a(x,\lambda^N)\mathrm{Jac}_{x_N} g^N(x) & 0 & & D_N^b(x,\lambda^N) \end{array}\right), \tag{3.23}$$

where D_i^a and D_i^b are $m_i \times m_i$ diagonal matrices. In the following, we denote by D^a-part and D^b-part the left and right parts of Equation (3.23).

Assume the generalized Jacobian has two identical rows, say for players i and $\tilde{i}$ and components j_i and $j_{\tilde{i}}$. The, the D^b-part requires that the j_ith row of D_i^b and the $j_{\tilde{i}}$th row of $D_{\tilde{i}}^b$ equals zero

$$b_j^i(x,\lambda_j^i) = b_{j_{\tilde{i}}}^{\tilde{i}}(x,\lambda_{j_{\tilde{i}}}^{\tilde{i}}) = 0, \tag{3.24}$$

with $b_j^i(x,\lambda_j^i) = \phi_b'(-g_j^i(x),\lambda_j^i)$. Identical rows in the D^a-part is equivalent to the n dimensional equation

$$a_j^i(x,\lambda_j^i)\mathrm{Jac}_x g_j^i(x) = a_{j_{\tilde{i}}}^{\tilde{i}}(x,\lambda_{j_{\tilde{i}}}^{\tilde{i}})\mathrm{Jac}_x g_{j_{\tilde{i}}}^{\tilde{i}}(x). \tag{3.25}$$

If $\phi = \phi_\wedge$ for which $\phi'_{\wedge b}(a,b) = \mathbb{1}_{b<a}$, then Equation (3.24) leads to $g_j^i(x) < \lambda_j^i$ and $g_{j_{\tilde{i}}}^{\tilde{i}}(x) < \lambda_{j_{\tilde{i}}}^{\tilde{i}}$. Since at a solution $z^\star$ both $g_j^i(x)$ and λ_j^i are nonnegative and (at least) one is zero. Hence, $g_j^i(x) = g_{j_{\tilde{i}}}^{\tilde{i}}(x) = 0$, i.e. active constraints. For the D^a-part using $\phi'_{\wedge a}(a,b) = \mathbb{1}_{a<b}$, Equation (3.25) leads to $a_j^i(x,\lambda_j^i) = a_{j_{\tilde{i}}}^{\tilde{i}}(x,\lambda_{j_{\tilde{i}}}^{\tilde{i}}) = 1$. Furthermore, $\mathrm{Jac}_x g_j^i$ and $\mathrm{Jac}_x g_{j_{\tilde{i}}}^{\tilde{i}}$ are identical. So if the constraints $g_j^i, g_{j_{\tilde{i}}}^{\tilde{i}}$ are active and shared, then both Equations (3.24) and (3.25) are satisfied, i.e. the generalized Jacobian has two identical (null) rows.

If we use the Fischer-Burmeister function, Equation (3.24) is equivalent to

$$g_j^i(x) = 0 \quad \text{and} \quad \lambda_j^i > 0,$$

and the same for $\tilde{j}$. This is equivalent to $g_j^i, g_{j_{\tilde{i}}}^{\tilde{i}}$ are strongly active constraints. These conditions are such that $a_j^i(x,\lambda_j^i)$ and $a_{j_{\tilde{i}}}^{\tilde{i}}(x,\lambda_{j_{\tilde{i}}}^{\tilde{i}})$ are non-zero and expressed as $a_j^i(x,\lambda_j^i) = a_{j_{\tilde{i}}}^{\tilde{i}}(x,\lambda_{j_{\tilde{i}}}^{\tilde{i}}) = -1$. Equation (3.25) again requires that the constraint functions g_j^i and $g_{j_{\tilde{i}}}^{\tilde{i}}$ have identical derivatives.

We conclude that for these two complementarity functions, the bottom part of generalized Jacobian Φ has two identical rows if and only if there are two *shared* and *active* constraints. For the top part, there are less problem in general. As shown in Appendix 3.5.1, matrices are less sparse: identical rows implies that the constraint functions are linear and two objective derivatives are identical.

If we exclude the pathological case described above, the nonsingularity of limiting Jacobian * (3.20) requires that $J^\star$ has a non-zero determinant. Using the determinant formula for partitionned matrix and the Schur complement, the nonsingularity condition can be expressed in two different formulations

(i) that $\mathrm{Jac}_x D(x,\lambda)$ and

$$D_b(z) + D_a(z)\mathrm{Jac}_x g(x)\mathrm{Jac}_x D(x,\lambda)^{-1}\mathrm{diag}\left[\left(\nabla_{x_i} g^i(x)\right)_i\right] \tag{3.26}$$

are nonsingular,

(ii) $D_b(z)$ and

$$\mathrm{Jac}_x D(x,\lambda) + \mathrm{diag}\left[\left(\nabla_{x_i} g^i(x)\right)_i\right] D_b(z)^{-1} D_a(z)\mathrm{Jac}_x g(x) \tag{3.27}$$

are nonsingular.

Note, since $D_b(z)$ is a diagonal matrix, its nonsingularity is equivalent to have nonnull terms. We will see in subsection 3.2.3 that (3.26) is typical. Following the above discussion, we conclude that both the generalized Newton and the generalized Levenberg-Marquardt method converges superlinearly to a solution $z^\star$ if one of the two conditions (3.26) and (3.27) are satisfied and no constraint functions are shared.

Finally, we focus on the convergence of the Broyden method. Theorems of Sun and Han (1997) give minimal conditions for a quasi-Newton method to converge (superlinearly). In Sun and Han (1997) or Qi and Jiang (1997), the Broyden method is used on smooth parts of the root function F. As emphasized in Jiang et al. (1996), when applying the Broyden method directly on F, it is hard to show the convergence for a general semismooth function. The weakest condition seems to be the Lipschitz continuity of B-derivative of F proved in Ip and Kyparisis (1992). Let us note that the function $F(x,\lambda)$ is not differentiable only when

∗. We do not have to require all elements of ∂F but only elements of $\partial_B F$.

$\lambda_j^i = g_j^i(x) = 0$. Otherwise, F is differentiable for which the convergence is established. In this paper, we test the Broyden method without proving the convergence. To our knowledge, the Broyden method is not used in the GNEP context.

Global convergence

Let us study the global convergence starting by line-search techniques. The merit function used is the same function as for the differentiable setting the residual norm.

$$f(z) = \frac{1}{2}|| \ F(z)||^2.$$

The gradient is given by $\nabla f(z) = V^T F(z)$, where $V \in \partial F(z)$. As mentioned in Qi and Sun (1998), the merit function may be C^1, even if F is only semismooth.

Jiang and Ralph (1998) and Qi and Sun (1998) show the convergence of the Newton method globalized with a backtracking (geometric) line search. When using the generalized Jacobian, Jiang (1999) shows the convergence of the corresponding algorithm. All proofs rely on the fact that after a large number of iteration, the full Newton step is accepted, i.e. we get back to local convergence.

Theorem. *Suppose that the root function F is a semismooth function and the merit function f is C^1. Then any accumulation points $z^\star$ of the line-search generalized Newton method is a stationary point of f, i.e. $\nabla f(z^\star) = 0$. If $z^\star$ is a solution of $F(z) = 0$ and all matrices in $\partial_B F(z^\star)$ are nonsingular, then the whole sequence converges superlinearly (resp. quadratically) to $z^\star$ (if F is strongly semismooth at $z^\star$).*

To our knowledge, the global convergence of general quasi-Newton methods for nonsmooth equation is not established. However, for the Levenberg-Marquardt with a backtracking line-search technique, Jiang and Ralph (1998) and Jiang (1999) show the convergence. We present below a light version of their theorem.

Theorem. *Let $(z_k)_k$ be a sequence of the generalized LM method globalized with a backtracking line search to solve $F(x) = 0$ for a semismooth function F and a C^1 merit function f. Assuming the direction step is solvable at each iteration, we denote by $z^\star$ an accumulation point. If $\lambda_k = \min(f(z_k), ||\nabla f(z_k)||)$ and elements in $\partial_B F(z^\star)$ are nonsingular, then $(z_k)_k$ converges superlinearly. If in addition F is strongly semismooth at $z^\star$, then it converges quadratically.*

The attentive reader may have noticed that in global convergence not all stationary points $z^\star$ of the merit function f are not necessarily a solution of the equation $F(z) = 0$. This is already the case in the smooth setting, e.g. $F(x) = \sin(5x) - x$ has only three roots (0 and $\pm 0.519148..$) but $(sin(5x) - x)^2/2$ has many local minima, see Figures 11.1 and 11.2 of Nocedal and Wright (2006).

Now, we focus on the second globalization strategy: the trust-region approach. We present first a convergence result of Jiang and Ralph (1998), based on a result of Jiang et al. (1998) in the context of complementarity problems.

Theorem. *Consider $F(z) = 0$ with F semismooth and a C^1 merit function f? Let (z_k) be generated by the trust-region algorithm with the submodel given $m_k(h) = J_k^T F(x_k)h + \frac{1}{2}h^T J_k^T J_k h$. If $z^\star$ is an accumulation point of the sequence, then $z^\star$ is a stationary point of f. If for all element of $\partial_B F(z^\star)$ are nonsingular, then the entire sequence converges to $x^\star$ superlinearly. If in addition, F is strongly semismooth at $z^\star$, then the convergence rate is quadratic.*

In our GNEP context, the root function Φ is given in Equation (3.3) and the merit function is defined as

$$f(z) = \frac{1}{2} \left\| \begin{pmatrix} D(x,\lambda) \\ \phi_\circ(-g(x),\lambda) \end{pmatrix} \right\|_2^2.$$

The gradient of f can be expressed as

$$\nabla f(z) = \begin{pmatrix} \mathrm{Jac}_x D(x,\lambda) & \mathrm{diag}\left[\left(\nabla_{x_i} g^i(x)\right)_i\right] \\ -D_a(z)\mathrm{Jac}_x g(x) & D_b(z) \end{pmatrix}^T \begin{pmatrix} D(x,\lambda) \\ \phi_\circ(-g(x),\lambda) \end{pmatrix}.$$

using Equation (3.20). As mentioned in Dreves et al. (2011), the gradient ∇f is single-valued since only the bottom part of the generalized Jacobian $\partial\Phi(z)$ contains multi-valued expressions (D_a and D_b when $g_j^i(x) = \lambda_j^i = 0$) and the bottom part of $\Phi(x,\lambda)$ has zero entries in that case. Hence, f is C^1 as long as the objective and constraint functions are C^2. Therefore, the line-search is well defined for our GNEP reformulation.

The nonsingularity assumption of elements in the limiting Jacobian $\partial_B\Phi$ was already studied in the previous subsection for local convergence. Furthermore, Dreves et al. (2011)'s theorem 3 analyzes the additional conditions for a stationary point of f to be a solution of the nonsmooth equation, hence of the original GNEP. The condition is that $\mathrm{Jac}_x D(x,\lambda)$ is nonsingular and the matrix

$$M = \mathrm{Jac}_x g(x)\mathrm{Jac}_x D(x,\lambda)^{-1}\mathrm{diag}\left[\left(\nabla_{x_i} g^i(x)\right)_i\right] \tag{3.28}$$

is a P_0-matrix*. The nonsingularity condition (3.28) for local convergence requires that $D_b(z) + MD_a(z)$ is nonsingular. If diagonal matrices D_a and D_b have non zero terms, then M is P_0-matrix implies that $D_b(z) + MD_a(z)$ is nonsingular. Thus, Conditions (3.28) and (3.26) are closely related.

3.2.4 Specific methods for constrained equations

This subsection aims to present methods specific to solve constrained (nonlinear) equations, first proposed by Dreves et al. (2011) in the GNEP context. The KKT system can be reformulated as a constrained equation, see Equation (3.4) of Subsection 3.1.1. Techniques to solve such equation may provide good alternatives to standard optimization procedures. In a VI problem context, Facchinei and Pang (2003) devotes a chapter to interior-point methods for solving such constrained equations (CE). Here, we focus on the method of Monteiro and Pang (1999) providing a general framework for CE problems.

A constrained equation is defined as

$$H(z) = 0, \quad z \in \Omega, \tag{3.29}$$

where Ω is a closed subset of $\mathbb{R}^n$. Generally, the constraint set Ω has a simple structure, e.g. the nonnegative orthant $\Omega = \mathbb{R}^n_+$ or a hypercube $\Omega = [l,u]$. As mentioned in Wang et al. (1996), in practice, the root function H has also a structured form

$$H(z) = \begin{pmatrix} F(z) \\ G(z) \end{pmatrix}.$$

*. A $m \times m$ square matrix M is a P_0-matrix if for all subscript set $\alpha \subset \{1,\ldots,m\}$ the determinant $\det(M_{\alpha\alpha}) \geq 0$.

The KKT reformulation of the GNEP (3.2) falls within this framework

Let $\mathring{\Omega}$ denote the interior of Ω. In the contrained equation literature, we assume that (i) Ω is a closed subset with nonempty interior $\mathring{\Omega}$, (ii) there exists a closed convex set $S \subset \mathbb{R}^n$, such that $0 \in S$, $H^{-1}(\mathring{S}) \cap \mathring{\Omega}$ is nonempty but $H^{-1}(\mathring{S}) \cap \text{bd } \Omega$ is empty and (iii) H is C^1 function. The set S contains zero, so that the set $H^{-1}(\mathring{S})$ contains a solution to Equation (3.29) and local points around. The second assumption requires that such points should not be on the boundary of Ω. The third assumption is just differentiability. In the following, these three assumptions will be referenced as the constrained equation blanket assumptions.

Potential reduction algorithms with line-search

Monteiro and Pang (1999) build a framework for potential functions, where these functions play a major role. A potential function $\mathring{S} \mapsto \mathbb{R}$ satisfies the following properties:

1. For all sequences $(u_k)_k$ in $\mathring{S}$ such that either $||u_k||$ tends to infinity or (u_k) tends to a point on the boundary bd $S \setminus \{0\}$, then $p(u_k)$ tends to infinity.
2. p is C^1 function on its domain and the curvature condition $u^T \nabla p(u) > 0$ for all nonzero vectors.
3. There exists a pair $(a, \bar{\sigma})$ in $\mathbb{R}^n \times]0, 1]$, such that for all $u \in \mathring{S}$, we have $||a||^2 u^T \nabla p(u) \geq \bar{\sigma}(a^T u)(a^T \nabla p(u))$.

The potential function has the dual objective to keep the sequences $(H(x_k))_k$ away from the set bd $S \setminus \{0\}$ and to help the convergence to the zero vector. The parameter a, known as the central vector, will play a crucial role to generate iterates in the constrained set Ω.

For example, if the subset S is the nonnegative orthant $\mathbb{R}^n_+$, then a typical potential function is

$$p(u) = \zeta \log ||u||_2^2 - \sum_{i=1}^{n} \log u_i \quad \text{for} \quad u > 0.$$

Monteiro and Pang (1999) prove that this function verifies the above conditions when $\zeta > n/2$ and with the pair $(a, \bar{\sigma}) = (\mathbb{1}_n, 1)$, $\mathbb{1}_n$ being the n-dimensional one vector.

In the GNEP context, the subset S is $\mathbb{R}^{n_\cdot} \times \mathbb{R}^{2m_\cdot}_+$ where $n_\cdot = \sum_i n_i$ is the total number of player variables and $m_\cdot = \sum_i m_i$ is the total number of constraints, i.e. $n = n_\cdot + m_\cdot$. The function H has components F and G given by

$$F(z) = D(x, \lambda) \quad \text{and} \quad G(z) = \begin{pmatrix} g(x) + w \\ \lambda \circ w \end{pmatrix},$$

where $z = (x, \lambda, w)$, see, e.g., Wang et al. (1996). Dreves et al. (2011) propose the following potential function

$$p\,(u) = \zeta \log \left(||u_1||_2^2 + ||u_2||_2^2 \right) - \sum_{i=1}^{2m_\cdot} \log(u_{2i}),$$

where $u = (u_1, u_2) \in \mathbb{R}^{n_\cdot} \times \mathbb{R}^{2m_\cdot}_+$ and $\zeta > m_\cdot$ in order to enter the potential framework. The pair of constants is $(a, \bar{\sigma}) = ((0_{n_\cdot}, \mathbb{1}_{m_\cdot}), 1)$.

The difficulty, compared to a classical nonlinear equation, is to ensure that all the iterates remains in the constrained set Ω. In order to solve Equation (3.29), Monteiro and Pang (1999) use a modified Newton globalized with a backtracking line-search. We report below their potential reduction Newton algorithm. The algorithm is divided into two parts: (i)

compute the direction using the central vector a and (ii) find an appropriate stepsize with a geometric line-search for which the merit function is $\psi(u) = p(H(u))$. Note that $H(u)$ is valued in $\mathbb{R}^{n\cdot} \times \mathbb{R}^{2m\cdot}$.

Init $z_0 \in \mathring{\Omega}, 0 < \rho, \alpha < 1$ and choose $\sigma_0 \in [0, \bar{\sigma}[$

Iterate until a termination criterion is satisfied,

– Solve the system * to get d_k

$$H(z_k) + \operatorname{Jac} H(z_k)d = \sigma_k \frac{a^T H(z_k)}{||a||_2^2} a. \tag{3.30}$$

– Find the smallest integer m_k such that

$$\psi(z_k + \rho^{m_k} d_k) \leq \psi(z_k) + \alpha \rho^{m_k} \nabla\psi(z_k)^T d_k, \quad \text{and} \quad z_k + \rho^{m_k} d_k \in \mathring{\Omega}.$$

– $z_{k+1} = z_k + \rho^{m_k} d_k$.

end Iterate

Due to the special structure H and a might have, the computation of d_k in Equation (3.30), a modified Newton direction because of the right-hand side term, may be further simplified by decomposing into its components F and G. In this form, the algorithm is defined when the Jacobian $\operatorname{Jac} H$ is nonsingular at $z_k \in \mathring{\Omega}$. Lemma 2 of Monteiro and Pang (1999) shows that the direction computed in the first step is a descent direction for the merit function ψ. So, the algorithm is well-defined. Their Theorem 3 shows the convergence of the potential reduction algorithm.

Theorem. *Assume p is a potential function, the constrained Equation (3.29) satisfies the constrained equation blanket assumptions, the Jacobian $\operatorname{Jac} H(z)$ is nonsingular for all $z \in \mathring{\Omega}$ and we have $\limsup_k \sigma_k < \bar{\sigma}$. Let (z_k) be a sequence generated by the potential reduction Newton algorithm. We have (i) the sequence $(H(z_k))$ is bounded and (ii) any accumulation point, if there exists, solves the constrained Equation (3.29). In particular, if (z_k) is bounded, the constrained equation has a solution.*

Application to GNEP

As already mentioned, Equation (3.4) of the GNEP can be reformulated as a constrained equation. The root function $H : \mathbb{R}^n \times \mathbb{R}^{2m} \mapsto \mathbb{R}^n \times \mathbb{R}^{2m}$ is defined as

$$H(x, \lambda, w) = \begin{pmatrix} D(x, \lambda) \\ g(x) + w \\ \lambda \circ w \end{pmatrix},$$

where the dimensions n, m correspond to the GNEP notation and $(a, \bar{\sigma})$ is given by $((0_n, \mathbb{1}_m), 1)$. The potential function is given by

$$p\,(u) = \zeta \log\left(||x||_2^2 + ||\lambda||_2^2 + ||w||_2^2\right) - \sum_{k=1}^{m} \log(\lambda_k) - \sum_{k=1}^{m} \log(w_k),$$

∗. In Monteiro and Pang (1999), they use the directional derivative along d in the left-hand side of Equation (3.30), which is equivalent to this formulation since H is C^1 under the blanket assumptions.

where $u = (x,\ \lambda, w) \in \mathbb{R}^n \times \mathbb{R}^m_{\star+} \times \mathbb{R}^m_{\star+}$ and $\zeta > m$. This reformulation of the potential function emphasizes the three components $u = (x,\ \lambda, w)$. For the line-search, the gradient ∇p is given by

$$\nabla p(x,\lambda,w) = \begin{pmatrix} \frac{2\zeta}{||x||_2^2+||\lambda||_2^2+||w||_2^2}x \\ \frac{2\zeta}{||x||_2^2+||\lambda||_2^2+||w||_2^2}\lambda - \lambda^{-1} \\ \frac{2\zeta}{||x||_2^2+||\lambda||_2^2+||w||_2^2}w - w^{-1} \end{pmatrix},$$

where λ and w have positive components and terms λ^{-1} and w^{-1} correspond to the component-wise inverse vector. Compared to the semismooth reformulation, the root function H is now C^1. The Jacobian is given by

$$\mathrm{Jac}\, H(x,\lambda,w) = \begin{pmatrix} \mathrm{Jac}_x D(x,\lambda) & \mathrm{diag}\left[\left(\nabla_{x_i} g^i(x)\right)_i\right] & 0 \\ \mathrm{Jac}_x g(x) & 0 & I \\ 0 & \mathrm{diag}[w] & \mathrm{diag}[\lambda] \end{pmatrix}.$$

As reported in Dreves et al. (2011), the computation of the direction $d_k = (d_{x,k}, d_{\lambda,k}, d_{w,k})$ in Equation (3.30) can be simplified due to the special structure of the above Jacobian matrix. The system reduces to a linear system of n equations to find $d_{x,k}$ and the $2m$ components $d_{\lambda,k}, d_{w,k}$ are simple linear algebra. Using the classic chain rule, the gradient of the merit function is given by

$$\nabla\psi(x,\lambda,w) = \mathrm{Jac}\, H(x,\lambda,w)^T \nabla p(H(x,\lambda,w)).$$

Again the computation of this gradient can be simplified due to the sparse structure of $\mathrm{Jac}\, H$. Theorem 4.3 of Dreves et al. (2011) is the direct application of the previous theorem in the GNEP context. We do not restate here their theorem, but present their nonsingularity result given in Theorem 4.6. The Jacobian matrix is nonsingular, if the matrix $\mathrm{Jac}_x D(x,\lambda)$ is nonsingular and

$$M = \mathrm{Jac}_x g(x) \mathrm{Jac}_x D(x,\lambda)^{-1} \mathrm{diag}\left[\left(\nabla_{x_i} g^i(x)\right)_i\right] \tag{3.31}$$

is a P_0-matrix. This is exactly Equation (3.28) given in the semismooth setting.

3.3 Numerical results

In this section, we perform a numerical illustration to compare the different methods presented in this paper. The implementation is done in the R statistical software and the package **GNE**, freely available on internet.

Our test problem is a simple two-player polynomial-objective game for which there are four generalized Nash equilibria. The objective functions (to be minimized) are given by

$$\theta_1(x) = (x_1 - 2)^2(x_2 - 4)^4 \quad \text{and} \quad \theta_2(x) = (x_2 - 3)^2(x_1)^4,$$

for $x \in \mathbb{R}^2$, while the constraint functions are

$$g_1(x) = x_1 + x_2 - 1 \leq 0 \quad \text{and} \quad g_2(x) = 2x_1 + x_2 - 2 \leq 0.$$

Objective functions are player strictly convave. This problem is simple but not simplistic, since second-order partial derivatives of objective functions are not constant, as for other

	$x_1^\star$	$x_2^\star$	$\lambda_1^\star$	$\lambda_1^\star$
$z^{\star 1}$	2	-2	0	5×2^5
$z^{\star 2}$	-2	3	8	0
$z^{\star 3}$	0	1	4×3^4	0
$z^{\star 4}$	1	0	2^9	6

Table 3.1: Four GNEs

test problem such as the river basin pollution game of Krawczyk and Uryasev (2000) or the example of Rosen (1965).

Due to the simple form of the objective function, we can solve the KKT system for this GNEP, manually. The solutions are listed in Table 3.1.

Before discussing the results, we detail the stopping criteria used in our optimization procedure. They are based on Chapter 7 of Dennis and Schnabel (1996). Algorithms always stop after a finite number of iterations with an exit code specifying whether the sequence converges or not: (1) convergence is achieved $||F(z)||_\infty < ftol$ with $ftol = 10^{-8}$, (2) algorithm is stuck because two consecutive iterates are too close $||(z_k - z_{k-1})/z_k||_\infty < xtol$ with $xtol = 10^{-8}$, (3) stepsize is too small $t_k < xtol$ or radius is too small $\Delta_k < xtol$, (4) the iteration limit is exceeded $k > k_{max}$ with $k_{max} = 300$ or generalized Jacobian is too ill-conditionned (5) or singular (6).

On this example, we compare the following methods: the Newton and Broyden methods with a globalization scheme (line search or trust-region), the Levenberg-Marquardt (LM) method with line search, the Levenberg-Marquardt with adaptive parameter, the constrained-equation modified Newton. In the following tables, the results are labelled as follows: GLS stands for geometric line search, QLS quadratic line search, PTR Powell trust-region, DTR double dogleg trust-region. We report the number of calls to the root function and the generalized Jacobian, the time (sec), the final iterate $z^\star$ (when successful convergence), the value of Euclidean norm at $z^\star$ and the exit code. Meanings of exit code are 1 for successful convergence, 2 or 3 consecutive iterates too small, 4 iteration limit exceeded, 5 or 6 ill-conditionned Jacobian.

In Table 3.2, we report the result with the complementarity function ϕ_{FB} and the starting point $z_0 = (5, 5, 0, 0)$, while for the constrained equation method, the starting point is $z_0 = (5, 5, 2, 2, 2, 2)$. Most methods converge to different equilibria. Surprisingly, the Newton method with geometric line search converges to $z^{\star 1}$, whereas all Broyden methods converge to $z^{\star 2}$ and Newton trust-region methods converge to $z^{\star 3}$, despite using the *same* initial points. There is only one method diverging to a local minimum of the merit function $1/2||F(z^\star)||_2^2$ which is not a root of F: the Newton method with quadratic line search. The LM method with adaptive parameter and modified Newton of constrained equation are stuck on singular matrices. In overall, there is a clear advantage for classic semismooth methods solving the extended KKT system on this example.

With the minimum complementarity function $\phi_\wedge$, we get similar results, see Table 3.4 in Appendix 3.5.2. Newton methods converges to a different GNE than convergent Broyden methods. This time, Levenberg-Marquardt method with adaptive parameter is convergent in relative few iterations. But again, the constrained equation Newton method is divergent because of a singular Jacobian.

	Fct. call	Jac. call	Time	$x_1^\star$	$x_2^\star$	$\lambda_1^\star$	$\lambda_1^\star$	$\|\|F(z^\star)\|\|$	Code
Newton - GLS	96	24	0.008	2	-2	-2.8e-12	160	2.8e-12	1
Newton - QLS	67	20	0.007					11	3
Newton - PTR	322	217	0.102	7e-04	1	324	-2.3e-17	8.6e-09	1
Newton - DTR	317	217	0.056	0.00066	1	324	-4.6e-17	6.9e-09	1
Broyden - GLS	78	4	0.005	-2	3	8	2.4e-09	3e-09	1
Broyden - QLS	52	3	0.005	-2	3	8	4.8e-13	1.2e-09	1
Broyden - PTR	91	3	0.006	-2	3	8	8.5e-09	1.2e-08	1
Broyden - DTR	127	3	0.008	-2	3	8	6.6e-13	1.1e-09	1
LM min - GLS	29	29	0.02	-2	3	8	-1.4e-09	3.7e-09	1
LM adaptive	368	184	0.111					0.00019	6
Mod. CE Newton	1782	158	0.295					2500	6

Table 3.2: Results with starting point $(5, 5, 0, 0)$ and ϕ_{FB}

	$z^{\star 1}$	$z^{\star 2}$	$z^{\star 3}$	$z^{\star 4}$	∞
min Newton GLS	58	213	280	394	55
FB Newton GLS	183	198	211	238	170
min Broyden PTR	106	362	45	385	102
FB Broyden PTR	104	381	35	248	232

Table 3.3: Number of GNEs found for 1000 random initial points

To further compare these methods and the complementarity function, we draw uniformly 1000 random initial points such that $z_0 \in [-10, 10] \times [-10, 10] \times \{1\} \times \{1\}$ and run algorithms on each of them. For simplicity, we restrict out comparison to Newton GLS and Broyden PTR methods test the Newton GLS method both with the minimum and Fischer-Burmeister complementarity functions. Results are summarized in Table 3.3, the first four columns store the number of sequences converging to a particular GNE, while the last column contains the number of diverging sequences (termination criteria remain the same as in the previous example.). With this comparison, the best method seems to be the Newton GLS method combined with the minimum function. We observe that using the minimum function tends to get only two GNEs, namely $z^{\star 2}$ and $z^{\star 4}$. The method finding almost equally all GNEs is the Newton GLS method with the Fischer-Burmeister function. Finally, the Broyden PTR method with the Fischer-Burmeister function seems very poor on this example.

3.4 Conclusion

The generalized Nash equilibrium problem (GNEP) is a useful tool for modelling many concrete applications in economics, computer science and biology, just to name a few. The demand for computational methods of the GNEP in general form is increasing. This survey paper aims to present and to compare the current optimization methods available for the GNEP. Our numerical experiments show an advantage for the KKT reformulation of the GNEP compared to the constrained equation reformulation. But, in Dreves et al. (2011), the

constrained equation reformulation was better. A method working for any general GNEP has yet to be found and its convergence to be proved.

3.5 Appendix

3.5.1 Analysis

Nonsmooth analysis

Definition (locally Lipschitzian)**.** *G is locally Lipschitzian (on $\mathbb{R}^n$) if* $\forall x \in \mathbb{R}^n, \exists U \in \mathcal{N}(x), \forall y, z \in U, \exists k_x > 0, ||G(y) - G(z)|| \leq k_x ||y - z||$.

From Clarke and Bessis (1999), the Rademacher theorem is

Theorem. *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a locally Lipschitz function. Then f is almost everywhere differentiable.*

From (Clarke, 1990, Cor 2.2.4, Chap. 2), for a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ locally Lipschitzian at x, we have that the generalized gradient $\partial f(y)$ is a singleton for all $y \in B(x, \epsilon)$ is equivalent to f is C^1 on $B(x, \epsilon)$.

From (Clarke, 1990, Prop 2.6.2, Chap. 2), we have the following properties of the generalized Jacobian.

Proposition.
- *$\partial G(x)$ is a nonempty, convex, compact subset of $\mathbb{R}^{m \times n}$, while $\partial_B G(x)$ is nonempty and compact.*
- *∂G is upper semicontinuous and closed at x and $\partial_B G$ is upper semicontinuous.*
- *$\partial G(x) \subset \partial G_1(x) \times \cdots \times \partial G_m(x)$, where the right-hand side is a matrix set where the ith row is the generalized gradient.*

The term $\partial G_1(x) \times \cdots \times \partial G_m(x)$ is sometimes denoted by $\partial_C G(x)$. But it is not Clarke's subdifferential, which seems to refer only to real-valued function, i.e. $\partial G(x) = \partial_C G(x)$.

From Theorem 2.3 of Qi and Sun (1993), we have the following equivalences

Proposition.
- *G is semismooth at x.*
- $\forall V \in \partial G(x + h), h \to 0, Vh - G'(x; h) = o(||h||)$.
- $\forall x \in D_G, G'(x + h; h) - G'(x; h) = o(||h||)$.

From Lemma 2.2 of Qi and Sun (1993) and Lemma 2.1 of Sun and Han (1997), we have the following properties

Proposition. *If G is semismooth at x, then $d \mapsto G'(x; d)$ is a Lipschitz function; $\forall h, \exists V \in G(x), Vh = G'(x; h)$ and $\forall h \to 0, G(x + h) - G(x) - G'(x; h) = o(||h||)$.*

The KKT system

The generalized Jacobian of the complementarity formulation has the following form

$$J(z) = \left(\begin{array}{ccc|ccc} \mathrm{Jac}_{x_1} L_1(x, \lambda^1) & \dots & \mathrm{Jac}_{x_N} L_1(x, \lambda^1) & \mathrm{Jac}_{x_1} g^1(x)^T & & 0 \\ \vdots & & \vdots & & \ddots & \\ \mathrm{Jac}_{x_1} L_N(x, \lambda^N) & \dots & \mathrm{Jac}_{x_N} L_N(x, \lambda^N) & 0 & & \mathrm{Jac}_{x_N} g^N(x)^T \\ \hline -D_1^a(x, \lambda^1)\mathrm{Jac}_{x_1} g^1(x) & \dots & -D_1^a(x, \lambda^1)\mathrm{Jac}_{x_N} g^1(x) & D_1^b(x, \lambda^1) & & 0 \\ \vdots & & \vdots & & \ddots & \\ -D_N^a(x, \lambda^N)\mathrm{Jac}_{x_1} g^N(x) & \dots & -D_N^a(x, \lambda^N)\mathrm{Jac}_{x_N} g^N(x) & 0 & & D_N^b(x, \lambda^N) \end{array}\right).$$

3.5.2 Numerical results

	Fct. call	Jac. call	Time	$x_1^\star$	$x_2^\star$	$\lambda_1^\star$	$\lambda_1^\star$	$\|\|F(z^\star)\|\|$	Code
Newton - GLS	14	6	0.003					71	5
Newton - QLS	9	6	0.003					71	6
Newton - PTR	38	17	0.005	2	-2	1e-25	160	1.4e-11	1
Newton - DTR	34	16	0.052	2	-2	-3.4e-29	160	4.7e-29	1
Broyden - GLS	1866	4	0.079	1	4.1e-18	512	6	7e-13	1
Broyden - QLS	93	4	0.005					71	5
Broyden - PTR	21	2	0.002	1	-7.6e-15	512	6	1.3e-11	1
Broyden - DTR	21	2	0.003	1	-4.6e-15	512	6	2.3e-12	1
LM min - GLS	33	33	0.023					4.9e-07	3
LM adaptive	18	9	0.006	-2	3	8	-3.9e-14	3.3e-11	1
Mod. CE Newton	1782	158	0.295					2500	6

Table 3.4: Results with starting point $(5, 5, 0, 0)$ and $\phi_\wedge$

Bibliography

Allgower, E. L. and Georg, K. (2003), *Introduction to Numerical Continuation Methods*, SIAM.

Bazaraa, M. S., Sherali, H. D. and Shetty, C. M. (2006), *Nonlinear Programming: Theory and Algorithms*, Wiley interscience.

Bonnans, J. F., Gilbert, J. C., Lemaréchal, C. and Sagastizábal, C. A. (2006), *Numerical Optimization: Theoretical and Practical Aspects, Second edition*, Springer-Verlag.

Broyden, C. G. (1965), 'A class of methods for solving nonlinear simultaneous equations', *Mathematics of Computation* **19**(92), 577–593.

Clarke, F. H. (1975), 'Generalized gradients and applications', *Transactions of the American Mathematical Society* **205**(1), 247–262.

Clarke, F. H. (1990), *Optimization and Nonsmooth Analysis*, SIAM.

Clarke, F. H. and Bessis, D. N. (1999), 'Partial subdifferentials, derivates and Rademacher's theorem', *Transactions of the American Mathematical Society* **351**(7), 2899–2926.

Dennis, J. E. and Morée, J. J. (1977), 'Quasi-newton methods, motivation and theory', *SIAM Review* **19**(1).

Dennis, J. E. and Schnabel, R. B. (1996), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM.

Dreves, A., Facchinei, F., Kanzow, C. and Sagratella, S. (2011), 'On the solutions of the KKT conditions of generalized Nash equilibrium problems', *SIAM Journal on Optimization* **21**(3), 1082–1108.

Facchinei, F., Fischer, A. and Piccialli, V. (2007), 'On generalized Nash games and variational inequalities', *Operations Research Letters* **35**(2), 159–164.

Facchinei, F., Fischer, A. and Piccialli, V. (2009), 'Generalized Nash equilibrium problems and Newton methods', *Math. Program., Ser. B* **117**(1-2), 163–194.

Facchinei, F. and Kanzow, C. (1997), 'A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems', *Mathematical Programming* **76**(3), 493–512.

Facchinei, F. and Kanzow, C. (2009), Generalized Nash equilibrium problems. Updated version of the 'quaterly journal of operations research' version.

Facchinei, F. and Pang, J.-S. (2003), *Finite-Dimensional Variational Inequalities and Complementary Problems. Volume II*, Springer-Verlag New York, Inc.

Fan, J.-Y. (2003), 'A modified Levenberg-Marquardt algorithm for singular system of nonlinear equations', *Journal of Computational Mathematics* **21**(5), 625–636.

Fan, J.-Y. and Yuan, Y.-X. (2005), 'On the quadratic convergence of the Levenberg-Marquardt Method without nonsingularity assumption', *Computing* **74**(1), 23–39.

Fischer, A. (2002), 'Local behavior of an iterative framework for generalized equations with nonisolated solutions', *Math. Program., Ser. A* **94**(1), 91–124.

Fukushima, M. and Pang, J.-S. (2005), 'Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games', *Comput. Manag. Sci.* **2**, 21–56.

Fukushima, M. and Qi, L., eds (1999), *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Academic Publishers.

Ip, C. and Kyparisis, J. (1992), 'Local convergence of quasi-Newton methods for B-differentiable equations', *Mathematical Programming* **56**(1-3), 71–89.

Jeyakumar, V. (1998), Simple Characterizations of Superlinear Convergence for Semismooth Equations via Approximate Jacobians, Technical report, School of Mathematics, University of New South Wales.

Jiang, H. (1999), 'Global convergence analysis of the generalized Newton and Gauss-Newton methods for the Fischer-Burmeister equation for the complementarity problem', *Mathematics of Operations Research* **24**(3), 529–543.

Jiang, H., Fukushima, M., Qi, L. and Sun, D. (1998), 'A Trust Region Method for Solving Generalized Complementarity Problem', *SIAM Journal on Optimization* **8**(1).

Jiang, H., Qi, L., Chen, X. and Sun, D. (1996), Semismoothness and superlinear convergence in nonsmooth optimization and nonsmooth equations, *in* 'Nonlinear Optimization and Applications', Plenum Press.

Jiang, H. and Ralph, D. (1998), Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares, *in* M. Fukushima and L. Qi, eds, 'Reformulation - nonsmooth, piecewise smooth, semismooth and smoothing methods', Boston MA: Kluwer Academic Publishers.

Kanzow, C. and Kleinmichel, H. (1998), 'A new class of semismooth Newton-type methods for nonlinear complementarity problems', *Computational Optimization and Applications* **11**(3), 227–251.

Krawczyk, J. and Uryasev, S. (2000), 'Relaxation algorithms to find Nash equilibria with economic applications', *Environmental Modeling and Assessment* **5**(1), 63–73.

Kubota, K. and Fukushima, M. (2010), 'Gap function approach to the generalized Nash equilibrium problem', *Journal of optimization theory and applications* **144**(3), 511–531.

Kulkarni, A. A. and Shanbhag, U. V. (2010), Revisiting generalized Nash games and variational inequalities. preprint.

Lopes, V. L. R. and Martinez, J. M. (1999), On the convergence of quasi-Newton methods for nonsmooth problems. preprint.

Monteiro, R. and Pang, J.-S. (1999), 'A Potential Reduction Newton Method for Constrained equations', *SIAM Journal on Optimization* **9**(3), 729–754.

Nocedal, J. and Wright, S. J. (2006), *Numerical Optimization*, Springer Science+Business Media.

Powell, M. (1970), A hybrid method for nonlinear algebraic equations, *in* P. Rabinowitz, ed., 'Numerical Methods for Nonlinear Algebraic Equations', Gordon & Breach, chapter 6.

Qi, L. (1993), 'Convergence analysis of some algorithms for solving nonsmooth equations', *Mathematics of Operations Research* **18**(1), 227–244.

Qi, L. (1997), 'On superlinear convergence of quasi-Newton methods for nonsmooth equations', *Operations Researchs Letters* **20**(5), 223–228.

Qi, L. and Chen, X. (1995), 'A globally convergent successive approximation method for severely nonsmooth equations', *SIAM Journal on control and Optimization* **33**(2), 402–418.

Qi, L. and Jiang, H. (1997), 'Semismooth KKT equations and convergence analysis of Newton and Quasi-Newton methods for solving these equations', *Mathematics of Operations Research* **22**(2), 301–325.

Qi, L. and Sun, D. (1998), A Survey of Some Nonsmooth Equations and Smoothing Newton Methods, Technical report, Applied Mathematics Report AMR 98/10, School of Mathematics, the University of New South Wales.

Qi, L. and Sun, J. (1993), 'A nonsmooth version of Newton's method', *Mathematical Programming* **58**(1-3), 353–367.

Rosen, J. B. (1965), 'Existence and Uniqueness of Equilibrium Points for Concave N-person Games', *Econometrica* **33**(3), 520–534.

Simon, L. (2011), Mathematical methods, Technical report, Berkeley, Lecture notes.

Sun, D. and Han, J. (1997), 'Newton and Quasi-Newton methods for a class of nonsmooth equations and related problems', *SIAM Journal on Optimization* **7**(2), 463–480.

Wang, T., Monteiro, R. and Pang, J.-S. (1996), 'An interior point potential reduction method for constrained equations', *Mathematical Programming* **74**(2), 159–195.

Yamashita, N. and Fukushima, M. (2000), On the rate of convergence of the Levenberg-Marquardt method, Technical report, Kyoto University.

Théorie de la ruine

Chapitre 4

Asymptotiques de la probabilité de ruine dans un modèle de risque avec dépendance

---

The $A + B/u$ rule for discrete and continuous time ruin models with dependence

A teacher can never truly teach unless he is still learning himself. A lamp can never light another lamp unless it continues to burn its own flame. The teacher who has come to the end of his subject, who has no living traffic with his knowledge but merely repeats his lessons to his students, can only load their minds ; he cannot quicken them.
Rabindranath Tagore (1861-1941)

Ce chapitre se base sur l'article Dutang et al. (2012) soumis à l'*Insurance: Mathematics and Economics*.

4.1 Introduction

Traditionally, the free surplus $(U_t)_t$ of an insurance company at time t is represented by

$$U_t = u + ct - \sum_{i=1}^{N_t} X_i,$$

where u is the initial surplus, c is the premium rate, $(X_i)_i$ are the successive claim amounts and $(N_t)_t$ is the claim arrival process (the claim waiting times are denoted by $(T_i)_i$). In the Cramér-Lundberg model, $(N_t)_t$ is modelled by a Poisson process, $(X_i)_i$ are independent and identically distributed (i.i.d.) random variables and claim severity $(X_i)_i$ are independent of the claim waiting times $(T_i)_i$. Andersen (1957) generalized the Cramér-Lundberg model by proposing a general renewal process for the claim arrival process $(N_t)_t$.

Since then, extensions have been proposed in many directions. Asmussen and Rolski (1991) studied ruin models with phase-type distributions for both claim severities X_i and claim waiting times T_i. Gerber and Shiu (1998) unified the analysis of ruin measures in the Cramér-Lundberg model, including the deficit at ruin, the claim causing the ruin or the ruin probability, by introducing a so-called discounted penalty function. Gerber and Shiu (2005), Song et al. (2010) and many others extended the Gerber-Shiu approach to a wider class of risk models. Various generalizations of the Sparre Andersen model have been proposed, such as for non-homogeneous claim arrivals (e.g. Lu and Garrido (2005), Albrecher and Asmussen (2006)), reinsurance treaties (e.g. Centeno (2002a,b)), multivariate risks (e.g. Cai and Li (2005),Collamore (1996)) and dependent risks (e.g. Albrecher and Boxma (2004),Boudreault et al. (2006),Albrecher and Teugels (2006)).

The ultimate ruin probability, i.e. $\psi(u) = P(\exists t > 0 : U_t < 0 | U_0 = u)$, is a major ruin measure and has received a considerable attention in the literature. For the Sparre Andersen model, with light-tailed claim amounts, $\psi(u) \sim Ce^{-\gamma u}$ as $u \to \infty$, where γ is the positive root of a simple equation involving the moment generating function of X_i (see, e.g., Asmussen and Albrecher (2010)). For heavy-tailed claim amounts, the ruin probability decreases at a slower polynomial rate since $\psi(u) \sim C/u^{\alpha}$ as $u \to \infty$ (e.g., Embrechts and Veraverbeke (1982); Klueppelberg and Stadtmueller (1998)). Concerning models with dependence, Albrecher and Teugels (2006), e.g., studied the ruin probability when claim size and claim waiting times, $(X_i, T_i)_i$, are correlated; they obtained again an exponential decrease for $\psi(u)$ in the case of light-tailed claim sizes. In a recent paper, Albrecher et al. (2011) investigated study the ruin probability when there is dependence by mixing among the claim sizes $(X_i)_i$ or the claim waiting times $(T_i)_i$, see also Constantinescu et al. (2011). They derived here an asymptotic formula $\psi(u) - A \sim B/u$ for Pareto correlated claims or inter-occurence times.

The main purpose of the present work is to show that the asymptotic rule $A+B/u$ applies to a wide class of dependent risk models in discrete and continuous time. That dependence will be incorporated through a mixing approach among claim amounts $(X_i)_i$ or claim interarrival times $(T_i)_i$. This translates a systemic risk behavior; by comparison, a dependence between claim sizes and waiting times would correspond to risks of catastrophes. Sufficient conditions are also given under which the ruin probability can be expanded as a series of terms $1/u^k$, $k \geq 0$.

Much care is paid on risk models that are formulated in discrete time. In fact, such models are often more appropriate in insurance because the surplus of the company is usually examined after regular time periods. Li et al. (2009) provided a review of standard risk models in discrete time. Our starting point is when claim amounts have a geometric distribution,

which implies an exponential decrease for $\psi(u)$. Adopting a mixing approach, we will focus on three particular cases of special interest. We also obtain asymptotics for the tail of the resulting claim distributions and then discuss the dependence structure involved.

The paper is organized as follows. Section 4.2 describes the mixing approach for both continuous and discrete time models. Section 4.3 establishes the asymptotic rule $A + B/u$ and some variants. Section 4.4 focuses on special features of the discrete time model. Except mentioned otherwise, all numerical illustrations are done with the R statistical software (R Core Team (2012)).

4.2 Model formulation

This section is devoted to the presentation of dependent risk models, first in the continuous time framework and then in the discrete time framework. In addition to a general formula of the ruin probability under the mixing approach, we present two and three special cases of mixing distributions for both time scales.

4.2.1 Continuous time framework

In this subsection, we present the continuous time framework based on the classic Cramér-Lundberg model.

Surplus process

The free surplus of an insurance company at time t is modeled by

$$U_t = u + ct - \sum_{i=1}^{N_t} X_i,$$

where u is the initial surplus, c is the premium rate, $(X_i)_i$ are the claim amounts and $(N_t)_{t\geq 0}$ is the Poisson claim arrival process with intensity λ. We assume that the $(X_i)_i$ are i.i.d. conditionally on a latent variable Θ (distributed as $X|\Theta = \theta$, say); they are independent of the claim arrival process. Θ can be interpreted as the heterogeneity in the claim process. In such setting, the claim sizes $(X_1, \ldots, X_n)$ are dependent random variables.

Ruin probabilities

Ruin occurs as soon as the surplus process becomes negative. Conditionally on $\Theta = \theta$, the ruin probability is thus defined as

$$\psi(u, \theta) = P(\exists t > 0 : U_t < 0 | U_0 = u, \Theta = \theta).$$

To determine such a probability, a standard method consists in looking at the state of the surplus after the first claim arrival. This leads to an integro-differential equation that can be solved by using Laplace-Stieltjes transforms, see, e.g., Asmussen and Albrecher (2010). In the case of exponentially distributed claims $(X_i)_i \sim \mathcal{E}(\theta)$, we have the well-known following formula

$$\psi(u, \theta) = \min\left(\frac{\lambda}{\theta c} e^{-u(\theta - \frac{\lambda}{c})}, 1\right),$$

where the min is equivalent to the net profit condition $\theta > \lambda/c$. Integrating over the parameter θ yields the ruin probability,

$$\psi(u) = F_\Theta(\theta_0) + I(u, \theta_0), \tag{4.1}$$

where $\theta_0 = \lambda/c$ and

$$I(u, \theta_0) = \int_{\theta_0}^{+\infty} \frac{\theta_0}{\theta} e^{-u(\theta-\theta_0)} dF_\Theta(\theta). \tag{4.2}$$

(4.1) is nothing else than Equation (5) of Albrecher et al. (2011).

Two special cases

Now, we briefly present the results for two particular distributions of the latent variable Θ, reported in Albrecher et al. (2011). Firstly, we consider for Θ a gamma distribution $\mathcal{G}a(\alpha, \lambda)$ with density

$$f_\Theta(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta}, \text{thus } F_\Theta(\theta) = \frac{\gamma(\alpha, \lambda\theta)}{\Gamma(\alpha)}, \theta > 0,$$

where $\gamma(.,.)$ (resp. $\Gamma(.)$) denotes the incomplete lower gamma function (the gamma function), see, Olver et al. (2010). The resulting claim generic variable X has a Pareto distribution with parameter $\mathcal{P}a(\alpha, \lambda)$, whose survival function is

$$P(X > x) = \frac{1}{\left(1 + \frac{x}{\lambda}\right)^\alpha}, x \geq 0.$$

Using the change of variable $y = \theta(\lambda + u)$, the integral $I(u, \theta_0)$ can be expressed in terms of the incomplete upper gamma function $\Gamma(.,.)$, see Appendix 4.6.1 for the definition of $\Gamma(.,.)$. We get

$$\psi(u) = \frac{\gamma(\alpha, \theta_0\lambda)}{\Gamma(\alpha)} + \frac{\lambda^\alpha \theta_0}{\Gamma(\alpha)} e^{\theta_0 u} \times \frac{\Gamma(\alpha - 1, \theta_0(\lambda + u))}{(\lambda + u)^{\alpha-1}}.$$

Note that the formula is only valid when the shape parameter verifies $\alpha > 1$, i.e. the density of $X/\Theta = \theta$ is log-concave.

Secondly, consider for Θ a Lévy distribution with density

$$f_\Theta(\theta) = \frac{\alpha}{2\sqrt{\pi\theta^3}} e^{-\alpha^2/4\theta}, \text{thus } F_\Theta(\theta) = \text{erfc}\left(\frac{\alpha}{2\sqrt{\theta}}\right), \theta > 0,$$

where erfc(.) denotes the complementary error function, see, Olver et al. (2010). The resulting claim distribution is a Weibull distribution $\mathcal{W}e(1/2, 1/\alpha^2)$ for which the distribution tail is

$$P(X > x) = e^{-\alpha\sqrt{x}}, x \geq 0.$$

Unlike the previous case, the computation of $I(u, \theta)$ in the Lévy case is more complicated. Using this time the change of variable $x = u\theta$, we get

$$I(u, \theta_0) = \frac{\theta_0 \alpha \sqrt{u^3}}{2\sqrt{\pi}} e^{u\theta_0} \int_{u\theta_0}^{+\infty} \frac{1}{\sqrt{x^5}} e^{-x - \alpha^2 u/(4x)} dx. \tag{4.3}$$

The latter integral is related to the generalized error function, a particular case of the generalized incomplete upper gamma function, which is defined as

$$\Gamma(a, x; b) = \int_x^{+\infty} t^{a-1} e^{-t-b/t} dt,$$

see, e.g., Chaudry and Zubair (1994, 2002). In Equation (4.3), we use $\Gamma(-3/2, \theta_0 u; \alpha^2 u/4)$. As for the classic incomplete gamma function, the function $\Gamma(.,.,.)$ satisfies a recurrence equation on the parameter a,

$$\Gamma(a+1, x; b) = a\Gamma(a, x; b) + b\Gamma(a-1, x; b) + x^a e^{-x-b/x},$$

see Theorem 2.2 of Chaudry and Zubair (2002). Using this equation, we are able to compute $\Gamma(-3/2, x; b)$ in terms of $\Gamma(-1/2, x; b)$ and $\Gamma(1/2, x; b)$, which can be both expressed in terms of the (classic) error function, see Appendix 4.6.2 for details. We get

$$I(u, \theta_0) = \frac{\theta_0\sqrt{u}}{\alpha} e^{u\theta_0} \left[\left(1 - \frac{1}{\alpha\sqrt{u}}\right) e^{\alpha\sqrt{u}} \mathrm{erfc}\left(d_+\right) \right.$$
$$\left. + \left(1 + \frac{1}{\alpha\sqrt{u}}\right) e^{-\alpha\sqrt{u}} \mathrm{erfc}\left(d_-\right) - \frac{2}{\sqrt{\pi u\theta_0}} e^{-u\theta_0 - \alpha^2/(4\theta_0)} \right],$$

where $d_+ = \sqrt{u\theta_0} + \alpha/(2\sqrt{\theta_0})$ and $d_- = \sqrt{u\theta_0} - \alpha/(2\sqrt{\theta_0})$. The constant term for the ruin probability appearing in Equation (4.3) is $F_\Theta(\theta_0) = \mathrm{erfc}(\lambda/2\sqrt{\theta_0})$.

4.2.2 Discrete time framework

The compound binomial risk model, introduced by Gerber (1988), is the discrete time analog of the Cramér-Lundberg model. Here too, we construct an extended version of this model by using a mixing approach. We are going to derive the ruin probability, for this risk process, as well as explicit formulas for three special cases.

Surplus process

The insurance portfolio is now examined at times $t \in \mathbb{N}$. Here too, the successive claim amounts form a sequence of i.i.d. random variables conditionally on $\Theta = \theta$ (distributed as $X|\Theta = \theta$).

The units of time and money are chosen such that the premium for each time unit is equal to one. The surplus of the insurance company at time t is then given by

$$U_t = u + t - \sum_{i=1}^{t} X_i,$$

where u is the initial surplus. When the claims are independent, this model is named compound binomial, because the number of strictly positive claims has a binomial distribution $\mathcal{B}(t, q)$ where $q = P(X > 0)$. The net profit condition is $E(X) < 1$ in order to avoid the certain ruin.

Ruin probability in infinite time

The definition of ruin probability has to be made precise since there is a non-zero probability for the surplus to be zero. In other words, we must specify if the ruin of the insurance company occurs when $U_t < 0$ or $U_t \leq 0$. Gerber (1988) considers the ruin as the first time the process U reaches 0, i.e.

$$\psi_G(u) = P(\exists t \in \mathbb{N}^+ : U_t \leq 0 | U_0 = u).$$

Shiu (1989) considers the ruin as the first time the process U becomes strictly negative:

$$\psi_S(u) = P(\exists t \in \mathbb{N}^+ : U_t < 0 | U_0 = u).$$

Graphically, ψ_G is the probability that the surplus process crosses the level 0 while ψ_S is the probability that the surplus crosses the level -1. We can switch from one formula to the other using the relation $\psi_G(u) = \psi_S(u-1)$. For the rest of the paper, we consider the ruin probability ψ_S.

Closed formulas for the ruin probability ψ_S are available (see, e.g., Willmot (1993), Sundt and dos Reis (2007)). Sundt and dos Reis (2007) derived the ruin probability when X is geometrically distributed. More precisely, assuming a geometric decreasing tail for the ruin probability, they proved that the claim amount distribution is of geometric type (see proof* of Theorem 1 of Sundt and dos Reis (2007)).

In the Sundt and dos Reis (2007) framework, when the claim distribution is geometric $\mathcal{G}e(q, \rho, 1-\alpha)$, see Appendix 4.6.3 for details, then the ultimate ruin probability is given by

$$\psi_S(u) = \min\left(\frac{(1-q)(1-\rho)}{q(1-\alpha)} \left(\frac{1-q}{q}(1-\rho) + \alpha \right)^u, 1 \right),$$

where the minimum is equivalent to the net profit condition $\frac{1-q}{q}(1-\rho)+\alpha < 1$. The net profit condition ensures the term in power of u does not explode. From this result, we can easily deduce the 0-modified geometric case, when $\rho = 1-\alpha$. When X is geometrically distributed $\mathcal{G}e(q, \rho)$, we have

$$\psi_S(u) = \min\left(\frac{1-q}{\rho} \left(\frac{1-\rho}{q} \right)^{u+1}, 1 \right). \tag{4.4}$$

Again the net profit condition (i.e. $\rho > 1-q$) ensures that the term $((1-\rho)/q)^{u+1}$ does not explode.

At our disposal, we have two closed formulas for the infinite time ruin probability. Now, let us extend the formula (4.4) by using again a mixing approach. We choose this formula rather than the previous one because of its tractability. Specifically, we suppose that $X_i/\Theta = \theta \sim \mathcal{G}e(q, e^{-\theta})$, then the overall ruin probability is

$$\psi(u) = \bar{F}_\Theta(\theta_0) + I(u, \theta_0), \tag{4.5}$$

where $\theta_0 = -\log(1-q)$ and

$$I(u, \theta_0) = \int_0^{\theta_0} \frac{1-q}{e^{-\theta}} \left(\frac{1-e^{-\theta}}{q} \right)^{u+1} dF_\Theta(\theta). \tag{4.6}$$

Compared to the continuous setting, (4.1) and (4.2), the integral in (4.6) is done over the interval $[0, \theta_0]$ for $I(u, \theta)$ rather than the interval $[\theta_0, +\infty[$. This is due to the fact that $\psi_S(u, \theta)$ is decreasing function of θ in the considered parametrization.

We do not choose the classic geometric distribution $\mathcal{G}e(\rho)$, because the net profit condition ($\rho > 1/2$) is restrictive on the type of parametrization for the parameter ρ. However, in that

*. Sundt and dos Reis (2007)'s reasoning works because substracting the recurrence equation at u and $u+1$ cancels the terms in the sum $\sum_{x=1}^{u+1}$. If we assume another type for the ruin probability, such as $\alpha/(u+\beta)$, it is far more difficult to get back to the claim distribution.

case, one could consider, for example, a geometric distribution $X/\Theta = \theta \sim \mathcal{G}e(1/(1+\theta))$. This leads to a ruin probability

$$\psi(u) = \int_0^1 \theta^{u+2} dF_\Theta(1) + \bar{F}_\Theta(1).$$

Choosing a uniform distribution $\Theta \sim \mathcal{U}(0,p)$ with $p \leq 1$ yields the surprisingly simple formula $\psi(u) = p^{u+2}/(u+3)$. This simple ruin probability is particularly interesting, because whether $p < 1$ or $p = 1$, the decrease of the ruin probability switches a geometric speed to a polynomial speed. In this special setting, the ruin probability is also explicit when Θ is beta distributed.

Three special cases

We present here results for three different distributions of Θ. Firstly, we consider an exponential distribution $\Theta \sim \mathcal{E}(\lambda)$. We use the following definite integral

$$I_1(a,b,x) = \int_0^x \left(e^{-\theta}\right)^a \left(1-e^{-\theta}\right)^b d\theta = \int_{e^{-x}}^1 p^a (1-p)^b \frac{dp}{p} = \bar{\beta}(a, b+1, e^{-x}),$$

for $x > 0$. $I_1(a,b,x)$ reduces to the beta function $\beta(a,b+1)$ when x tends to infinity. Using $I_1(\lambda+1,k,+\infty)$, the mass probability function of the claim distribution is given

$$P(X=k) = q\delta_{k0} + (1-\delta_{k0})\lambda(1-q)\beta(\lambda+1,k),$$

where δ_{ij} denotes the Kronecker product. With the presence of a beta function in this mass probability function, one can recognize the zero-modified Yule-Simon distribution, see, e.g., Simon (1955). This distribution appears in the study of word frequency. The survival function is given by

$$P(X>k) = \lambda(1-q)\beta(\lambda, k+1).$$

Using $I(u,\theta_0) = I_1(\lambda-1, u+1, \theta_0)$, the ruin probability can be derived.

Proposition 4.2.1. *Let us consider the discrete time framework of Subsection 4.2.2 with a latent variable Θ exponentially distributed $\mathcal{E}(\lambda)$.*

$$\psi(u) = (1-q)^\lambda + \frac{\lambda(1-q)}{q^{u+1}} \bar{\beta}(\lambda, u+1, 1-q), \forall u \geq 0.$$

Secondly, consider Θ follows a gamma distribution $\mathcal{G}a(\alpha,\lambda)$. We use the following integral

$$I_2(a,n,b,x) = \int_0^x \left(e^{-\theta}\right)^a \left(1-e^{-\theta}\right)^n \theta^b d\theta = \sum_{j=0}^n \binom{n}{j} (-1)^{n-j} \int_0^x e^{-\theta(a+n-j)} \theta^b d\theta,$$

yielding

$$I_2(a,n,b,x) = \sum_{j=0}^n \binom{n}{j} (-1)^{n-j} \int_0^{\tilde{x}} \frac{y^b}{(a+n-j)^{b+1}} e^{-y} dy.$$

with $\tilde{x} = x(a+j)$. Substituting $n-j$ to j gives

$$I_2(a,n,b,x) = \sum_{j=0}^n \binom{n}{j} (-1)^j \frac{\gamma(b+1,\tilde{x})}{(a+j)^{b+1}},$$

where $\gamma(.,.)$ denotes the incomplete lower gamma function. When x tends to infinity, only the term $\gamma(b+1,\tilde{x})$ changes and tends to $\Gamma(b+1)$. With the integral $I_2(\lambda,k-1,\alpha-1,+\infty)$, the resulting claim distribution has mass probability function

$$P(X=k)=q\delta_{k0}+(1-\delta_{k0})(1-q)\sum_{j=0}^{k-1}C_{k-1}^{j}(-1)^j\frac{\lambda^\alpha}{(\lambda+j)^\alpha}.$$

Similarly with $I_2(\lambda,k,\alpha-1,+\infty)$, the survival function is given by

$$P(X>k)=(1-q)\sum_{j=0}^{k}\binom{k}{j}(-1)^j\frac{\lambda^\alpha}{(\lambda+j)^\alpha}.$$

Using $I_2(\lambda-1,u+1,\alpha-1,\theta_0)$, the ruin probability can be deduced.

Proposition 4.2.2. *Let us consider the discrete time framework of Subsection 4.2.2 with a latent variable Θ gamma distributed $\mathcal{G}a(\alpha,\lambda)$.*

$$\psi(u)=\frac{\Gamma(\alpha,\lambda\theta_0)}{\Gamma(\alpha)}+\frac{1-q}{q^{u+1}}\sum_{j=0}^{u+1}\binom{u+1}{j}(-1)^j\frac{\gamma(\alpha,\theta_0(\lambda+j-1))}{\Gamma(\alpha)}\left(\frac{\lambda}{\lambda+j-1}\right)^\alpha,$$

with $\lambda>1$, $\theta_0=-\log(1-q)$ and for $u\geq 0$.

Finally, consider Θ is Lévy distributed $\mathcal{L}e(\alpha)$. We use the integral

$$I_3(a,n,b,x)=\int_0^x\left(e^{-\theta}\right)^a\left(1-e^{-\theta}\right)^n\theta^{-3/2}e^{-\frac{b}{\theta}}d\theta.$$

Using the change of variable, we have

$$I_3(a,n,b,x)=\sum_{j=0}^{n}\binom{n}{j}(-1)^{n-j}2\int_{\tilde{x}}^{\infty}e^{-\frac{a+n-j}{y^2}}e^{-by^2}dy.$$

with $\tilde{x}=x^{-1/2}$. This integral is linked to the generalized incomplete upper gamma function. Using Appendix 4.6.4, we get

$$I_3(a,n,b,x)=\sum_{j=0}^{n}\binom{n}{j}(-1)^j\frac{\sqrt{\pi}}{2\sqrt{b}}\left[e^{2\sqrt{b(a+j)}}\text{erfc}\left(\frac{\sqrt{b}}{\sqrt{x}}+\sqrt{a+j}\sqrt{x}\right)\right.$$
$$\left.+e^{-2\sqrt{b(a+j)}}\text{erfc}\left(\frac{\sqrt{b}}{\sqrt{x}}-\sqrt{a+j}\sqrt{x}\right)\right].$$

When x tends to infinity, we have

$$I_3(a,n,b)=\sum_{j=0}^{n}\binom{n}{j}(-1)^j\frac{\sqrt{\pi}}{\sqrt{b}}e^{-2\sqrt{b}\sqrt{a+j}}.$$

Using $I_3(0,k-1,\alpha^2/4)$ and $I_3(0,k,\alpha^2/4)$, the mass probability and survival functions are given by

$$P(X=k)=(1-q)\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^je^{-\alpha\sqrt{j}}\quad\text{and}\quad P(X>k)=(1-q)\sum_{j=0}^{k}\binom{k}{j}(-1)^je^{-\alpha\sqrt{j}}.$$

The expressions derived when Θ is Lévy distributed, are much more complex than in the continuous time framework. In Subsection 4.3.3, we study asymptotics for the survival function. The ruin probability can be computed using $I_3(-1, u+1, \alpha^2/4, \theta_0)$.

Proposition 4.2.3. *Let us consider the discrete time framework of Subsection 4.2.2 with a latent variable Θ Lévy distributed $\mathcal{L}e(\alpha)$.*

$$\psi(u) = erfc\left(\frac{\alpha}{2\sqrt{\theta_0}}\right) + \frac{1-q}{4q^{u+1}} \sum_{j=0}^{u+1} \binom{u+1}{j} (-1)^j \left[e^{\alpha\sqrt{j-1}} erfc\left(\frac{\alpha}{2\sqrt{\theta_0}} + \sqrt{j-1}\sqrt{\theta_0}\right) \right.$$
$$\left. + e^{-\alpha\sqrt{j-1}} erfc\left(\frac{\alpha}{2\sqrt{\theta_0}} - \sqrt{j-1}\sqrt{\theta_0}\right) \right],$$

with the convention* $\sqrt{-1} = i$, $\theta_0 = -\log(1-q)$ *and for* $u \geq 0$.

4.3 Asymptotics – the $A + B/u$ rule

This section is the core of the paper, where we establish the $A + B/u$ asymptotic rule for the ultimate ruin probability for both continuous and discrete time models. We also obtain an expansion of the ruin probability as a power series of $1/u$. Finally, we investigate the asymptotic behavior of the resulting claim distribution, which requires a special treatment with complex analysis.

4.3.1 Notation

We recall basics and notation of the asymptotic analysis; see e.g. Jones (1997), Olver et al. (2010). We introduce the standard Landau notation $\mathcal{O}()$, $o()$ and $\sim$. One says that f is asymptotically bounded by g as $x \to x_0$, denoted by

$$f(x) \underset{x_0}{=} \mathcal{O}(g(x)),$$

if there exists $K, \delta > 0$, such that for all $0 < |x - x_0| < \delta$, we have $|f(x)| \leq K|g(x)|$. In other words, in a neighborhood of x_0 excluding x_0, $|f(x)/g(x)|$ is bounded.

Then, f is said to be asymptotically smaller than g as $x \to x_0$, denoted by

$$f(x) \underset{x_0}{=} o(g(x)),$$

if for all $\epsilon > 0$, there exists $\delta > 0$, such that for all $0 < |x - x_0| < \delta$, we have $|f(x)| \leq \epsilon|g(x)|$. That is to say, in a neighborhood of x_0 excluding x_0, $|f(x)/g(x)|$ tends to 0.

And finally, f is asymptotically equivalent to g around x_0, if the ratio of f over g tends to 1, i.e.,

$$f(x) \underset{x_0}{\sim} g(x) \;\Leftrightarrow\; \frac{f(x)}{g(x)} \underset{x \to x_0}{\longrightarrow} 1.$$

This is equivalent to $f(x) = g(x) + o(g(x))$.

The asymptotic idea aims to approximate a complicated function f at x_0 by a sum of known and tractable terms $g(x)$, controlling the error by $o(g(x))$. Note that x_0 can be $+\infty$.

$*$. One can check that the term $j = 0$ is still a real number.

The usual way to achieve this is to take a series expansion of f around x_0 as g. f is said to take a series expansion at x_0 if for all $N \in \mathbb{N}$

$$f(x) \underset{x_0}{=} \sum_{n=1}^{N} a_n \phi_n(x) + o(\phi_N(x)),$$

where $(\phi_n)_n$ is a sequence of so-called gauge functions such that $\forall n \in \mathbb{N}$, $\phi_{n+1}(x) = o(\phi_n(x))$ around x_0. It can also be denoted by

$$f(x) \underset{x_0}{=} \sum_{n=1}^{N-1} a_n \phi_n(x) + \mathcal{O}(\phi_N(x)) \quad \text{or} \quad f(x) \underset{x_0}{\sim} \sum_{n=1}^{+\infty} a_n \phi_n(x).$$

Integration by part is a standard tool to study integral asymptotics and derive asymptotics, as pointed in Olver et al. (2010). Integration by part will be extensively used in the next two subsections. In Appendix 4.6.5, we recall two integration by part theorems.

4.3.2 Continuous time framework

In this subsection, we present and show the $A + B/u$ rule for the continuous time model.

Theorem 4.3.1. *Let us consider the continuous time framework of Subsection 4.2.1 with a positive latent variable Θ and $\theta_0 = \lambda/c$.*

(i) For all $u > 0$, the ruin probability is bounded

$$\psi(u) \leq F_\Theta(\theta_0) + \frac{1}{u} \times \frac{F_\Theta(\theta_0)}{\theta_0}.$$

(ii) If Θ has a continuous distribution with density f_Θ such that f_Θ is almost everywhere differentiable on $[\theta_0, +\infty[$ and f'_Θ being a Lebesgue-integrable, then we have

$$\psi(u) = F_\Theta(\theta_0) + \frac{f_\Theta(\theta_0)}{u} + o\left(\frac{1}{u}\right).$$

(iii) If in addition f_Θ is C^{k-1} almost everywhere on $[\theta_0, +\infty[$ and $f_\Theta^{(k)}$ is Lebesgue integrable and bounded on $[\theta_0, +\infty[$, then we have

$$\psi(u) = F_\Theta(\theta_0) + \sum_{i=0}^{k-1} \frac{h^{(i)}(0)}{u^{i+1}} + o\left(\frac{1}{u^k}\right),$$

where $h(x) = \theta_0 f_\Theta(x + \theta_0)/(x + \theta_0)$, so that

$$h^{(i)}(0) = \sum_{j=0}^{i} (-1)^j \frac{i!}{(i-j)!\theta_0^j} f_\Theta^{(i-j)}(\theta_0).$$

(iv) If f_Θ is C^∞ on $[\theta_0, +\infty[$, then we have

$$\psi(u) \underset{u \to +\infty}{\sim} F_\Theta(\theta_0) + \sum_{i=0}^{+\infty} \frac{h^{(i)}(0)}{u^{i+1}}.$$

Proof. (i) From (4.1) and (4.2), the ruin probability is given by

$$\psi(u) = F_\Theta(\theta_0) + \int_{\theta_0}^{+\infty} \psi_u(\theta) dF_\Theta(\theta), \quad \text{with} \quad \psi_u(\theta) = \frac{\theta_0}{\theta} e^{-u(\theta-\theta_0)},$$

where $\theta_0 = \lambda/c$.

Both ψ_u and F_Θ are bounded functions on $[\theta_0, +\infty[$. They also have bounded variations since they are monotone. In addition, ψ_u is continuous. So by Corollary 7.1.23 of Silvia (1999) or Theorem 12.1 of (Hildebrandt, 1971, Chap. 2), F_Θ is Stieltjes integrable with respect to the function ψ_u.

Then, we apply the integration by part theorem on $\int \psi_u dF_\Theta$ reported in Appendix 4.6.5. We get

$$\psi(u) = F_\Theta(\theta_0) + \lim_{b\to+\infty} \psi_u(b)F_\Theta(b) - \psi_u(\theta_0)F_\Theta(\theta_0) - \int_{\theta_0}^{+\infty} F_\Theta(t) d\psi_u(t).$$

Since ψ_u is continuously differentiable, the Stieltjes integral $\int F_\Theta d\psi_u$ reduces to a Riemann integral. We have

$$\psi'_u(\theta) = \frac{-1}{\theta^2}\theta_0 e^{-u(\theta-\theta_0)} + \frac{\theta_0}{\theta}(-u)e^{-u(\theta-\theta_0)} = -\theta_0 \left(\frac{1}{\theta^2} + \frac{u}{\theta}\right) e^{-u(\theta-\theta_0)}.$$

Furthermore, $\psi_u(\theta_0) = 1$ and

$$\lim_{b\to+\infty} \psi_u(b)F_\Theta(b) = 0.$$

Therefore, we obtain

$$\psi(u) = \theta_0 \int_{\theta_0}^{+\infty} F_\Theta(t) \left(\frac{1}{t^2} + \frac{u}{t}\right) e^{-u(t-\theta_0)} dt \leq \theta_0 \max_{t\in[\theta_0,+\infty[} F_\Theta(t) \left(\frac{1}{t^2} + \frac{u}{t}\right) \times \int_{\theta_0}^{+\infty} e^{-u(t-\theta_0)} dt.$$

We get

$$\psi(u) \leq F_\Theta(\theta_0) + \frac{1}{u} \times \frac{F_\Theta(\theta_0)}{\theta_0}.$$

(ii) Let $I(u, \theta_0) = \int_{\theta_0}^{+\infty} \psi_u(\theta) dF_\Theta(\theta)$. We assume a continuous distribution for the mixing variable Θ and make the change of variable $t = \theta - \theta_0$, we get

$$I(u, \theta_0) = \int_0^{+\infty} \frac{\theta_0}{\theta_0 + t} f_\Theta(t + \theta_0) e^{-ut} dt.$$

We easily recognize a Laplace transform of the function h defined as

$$h(t) = \frac{\theta_0}{\theta_0 + t} f_\Theta(t + \theta_0).$$

The minimum condition to apply an integration by part theorem is to require h to be absolutely continuous, see Appendix 4.6.5. Heil (2007) reports a version of the Fundamental Theorem of Calculus for absolutely continuous functions. So, absolute continuity of h on $[a, b]$ is equivalent to h is almost everywhere differentiable $[a, b]$ with h' being Lebesgue integrable on $[a, b]$.

Since $t \mapsto \theta_0/(\theta_0+t)$ is C^∞ on $[0,b]$ for $b>0$, h is absolutely continuous on $[0,b]$ if and only if f_Θ is. By assumption, f_Θ is almost everywhere differentiable on $\mathbb{R}_+$ with f'_Θ being Lebesgue integrable, hence h is absolutely continuous. Thus we have

$$\int_0^b h(t)e^{-ut}dt = \left[h(t)\frac{e^{-ut}}{-u}\right]_0^b + \frac{1}{u}\int_0^b h'(t)e^{-ut}dt = \frac{h(0)}{u} - \frac{h(b)e^{-bu}}{u} + \frac{1}{u}\int_0^b h'(t)e^{-ut}dt.$$

As b tends to infinity, we get

$$I(u,\theta_0) = \frac{h(0)}{u} + \frac{1}{u}\int_0^{+\infty} h'(t)e^{-ut}dt.$$

Using a property of the Laplace transform, see, e.g., Chapter 19 of Jeffrey and Dai (2008), we have

$$\int_0^{+\infty} h'(t)e^{-ut}dt \underset{u\to+\infty}{\longrightarrow} 0.$$

Finally, we conclude

$$\psi(u) = F_\Theta(\theta_0) + \frac{f_\Theta(0)}{u} + o\left(\frac{1}{u}\right).$$

(iii) As f_Θ is C^{k-1} almost everywhere on $[\theta_0,+\infty[$ and $f_\Theta^{(k)}$ is Lebesgue integrable, then $h^{(i)}$ is absolute continous for all $i \le k$. Applying k times the integration by part theorem, we get

$$I(u,\theta_0) = \sum_{i=0}^{k-1}\frac{h^{(i)}(0)}{u^{i+1}} + \frac{1}{u^k}\int_0^{+\infty} h^{(k)}(t)e^{-ut}dt.$$

Similarly if $h^{(k)}(t)$ is bounded on $[\theta_0,+\infty[$, then the latter term is controlled by $o\big(1/u^k\big)$. Let g be the function $t \mapsto \frac{\theta_0}{\theta_0+t}$. The ith-order derivative of h, if it exists, can be derived by the Leibniz formula

$$h^{(i)}(t) = \sum_{j=0}^{i}\binom{i}{j}g^{(j)}(t)f_\Theta^{(i-j)}(t+\theta_0) \quad \text{with} \quad g^{(j)}(t) = \frac{(-1)^j j!\theta_0}{(\theta_0+t)^{j+1}}.$$

Thus, we have

$$\psi(u) = F_\Theta(\theta_0) + \sum_{i=0}^{k-1}\frac{h^{(i)}(0)}{u^{i+1}} + o\left(\frac{1}{u^k}\right) \quad \text{with} \quad h^{(i)}(0) = \sum_{j=0}^{i}(-1)^j\frac{i!}{(i-j)!\theta_0^j}f_\Theta^{(i-j)}(\theta_0).$$

(iv) if f_Θ is C^∞, we have

$$\psi(u) \underset{u\to+\infty}{\sim} F_\Theta(\theta_0) + \sum_{i=0}^{+\infty}\frac{h^{(i)}(0)}{u^{i+1}}.$$

Unsurprisingly, we get back to asymptotic result (3.2.1) of (Olver et al., 2010, Chapter 3), since $I(u,\theta)$ is a Laplace transform. □

Remark 4.3.2. *A sufficient condition for f_Θ to be almost everywhere differentiable is local Lipchitzness. This is a consequence of the Rademacher theorem, see Appendix 4.6.5.*

Remark 4.3.3. *A similar approach can be done when mixing the waiting times* $(T_1, T_2, \dots)$. *Using Albrecher et al. (2011)'s Section 3, we have*

$$\psi(u) = \bar{F}_\Lambda(\lambda_0) + \int_0^{\lambda_0} \psi_u(\lambda) dF_\Lambda(\lambda), \quad \textit{with} \quad \psi_u(\lambda) = \frac{\lambda}{\lambda_0} e^{-u/\theta(1-\lambda/\lambda_0)}, \lambda_0 = \theta c.$$

We give here only the first terms of the series expansion assuming Λ *has a continuous distribution*

$$\psi(u) = \bar{F}_\Lambda(\lambda_0) + \frac{1}{cu} f_\Lambda(\lambda_0) + o\left(\frac{1}{u}\right).$$

Below, we present asymptotics for the two special cases analyzed in Subsection 4.2.1, based on known asymptotics listed in Appendix 4.6.1. When Θ is gamma distributed, we have

$$\psi(u) = \frac{\gamma(\alpha, \theta_0\lambda)}{\Gamma(\alpha)} + \frac{\lambda^\alpha \theta_0^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda\theta_0} \left(\frac{1}{\lambda + u} + \frac{\alpha - 1}{(\lambda + u)^2 \theta_0} + o\left(\frac{1}{u^2}\right)\right).$$

If we use Theorem 4.3.1, we get

$$\psi(u) = \frac{\gamma(\alpha, \lambda\theta_0)}{\Gamma(\alpha)} + \frac{\lambda^\alpha \theta_0^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda\theta_0} \left(\frac{1}{u} + \frac{1}{u^2} \left(\frac{\alpha - 1}{\theta_0} - \lambda \right) + o\left(\frac{1}{u^2}\right)\right).$$

These two expressions are similar with only different denominators $1/u$ against $1/(\lambda+u)$, but it does not matter for large values of u.

When Θ is Lévy distributed, the term $I(u, \theta_0)$ contains two terms linked with the error complementarity function. There exists expansion formula for the error function, cf. Appendix 4.6.1, but unfortunately the asymptotic of $I(u, \theta_0)$ leads to an explosive term $e^{\alpha\sqrt{u}}$. We conclude that a term-by-term asymptotic is not appropriate, a uniform expansion of the original function $\Gamma(3/2, x, b)$ is needed, when both x and b are large. But, we can still use Theorem 4.3.1 to get

$$\psi(u) = \operatorname{erfc}\left(\frac{\alpha}{2\sqrt{\theta_0}}\right) + \frac{\alpha}{2\sqrt{\pi\theta_0^3}} e^{-\alpha^2/4\theta_0} \left(\frac{1}{u} + \frac{1}{u^2} \left(\frac{\alpha^2}{4\sqrt{\theta_0}} - \frac{3}{2\theta_0} \right) + o\left(\frac{1}{u^2}\right)\right).$$

4.3.3 Discrete time framework

Now, let us turn our attention to the discrete-time framework, where the approach of this subsection shares strong similarities with the previous subsection.

Theorem 4.3.4. *Let us consider the discrete time framework of Subsection 4.2.2 with a positive latent variable* Θ *and* $\theta_0 = -\log(1 - q)$.

(i) For all $u \geq 0$, *the ruin probabiliy is lower bounded*

$$\bar{F}_\Theta(\theta_0) + F_\Theta(\theta_0) \frac{q}{u + 2} \leq \psi(u).$$

(ii) If Θ *has a continuous distribution with density* f_Θ *such that* f_Θ *is almost everywhere differentiable on* $[0, \theta_0]$ *with* f_Θ, f'_Θ *being bounded, then we have*

$$\psi(u) = \bar{F}_\Theta(\theta_0) + \frac{1}{u + 2} \times \frac{q f_\Theta(\theta_0)}{1 - q} + o\left(\frac{1}{u + 2}\right).$$

(iii) If in addition f_Θ is C^{k-1} almost everywhere on $[0,\theta_0]$ and successive derivatives of f_Θ are bounded on $[0,\theta_0]$, then we have

$$\psi(u) = \bar{F}_\Theta(\theta_0) + \sum_{i=0}^{k-1} \frac{\tilde{h}^{(i)}(0)}{(u+2)\dots(u+2+i)} + o\left(\frac{1}{(u+2)\dots(u+2+k-1)}\right),$$

with $\tilde{h}(x) = f_\Theta(-\log(1-xq))/(1-xq)^2$.

(iv) If f_Θ is C^∞ on $[0,\theta_0]$, then we have

$$\psi(u) \underset{u\to+\infty}{\sim} \bar{F}_\Theta(\theta_0) + \sum_{i=0}^{+\infty} \frac{\tilde{h}^{(i)}(0)}{(u+2)\dots(u+2+i)}.$$

Proof. (i) From (4.5) and (4.6), the ruin probability is given by

$$\psi(u) = \bar{F}_\Theta(\theta_0) + \int_0^{\theta_0} \psi_u(\theta) dF_\Theta(\theta), \quad \text{with} \quad \psi_u(\theta) = \frac{1-q}{e^{-\theta}} \left(\frac{1-e^{-\theta}}{q}\right)^{u+1},$$

where $\theta_0 = -\log(1-q)$.

First, we change the right-hand side Stieltjes integral by using the survival function $\bar{F}_\Theta$ rather than the cumulative distribution function. We get

$$\psi(u) = \bar{F}_\Theta(\theta_0) - \int_0^{\theta_0} \psi_u(\theta) d\bar{F}_\Theta(\theta).$$

Then, it is easy to see that both ψ_u and $\bar{F}_\Theta$ are also of bounded variation on $[0,\theta_0]$. They also have bounded variations since they are monotone. In addition, ψ_u is continuous. So by Corollary 7.1.23 of Silvia (1999), $\bar{F}_\Theta$ is Stieltjes integrable with respect to the function ψ_u.

Then we apply the integration by part theorem on $\int \psi_u d\bar{F}_\Theta$ reported in Appendix 4.6.5. We get

$$\psi(u) = \bar{F}_\Theta(\theta_0) - \psi_u(\theta_0)\bar{F}_\Theta(\theta_0) + \psi_u(0)\bar{F}_\Theta(0) + \int_0^{\theta_0} \bar{F}_\Theta(t) d\psi_u(t) = \int_0^{\theta_0} \bar{F}_\Theta(t) d\psi_u(t),$$

using $\psi_u(\theta_0) = 1$ and $\psi_u(0) = 0$.

Since ψ_u is continuously differentiable, the Stieltjes integral $\int \bar{F}_\Theta d\psi_u$ reduces to a Riemann integral. We have

$$\psi_u'(\theta) = (1-q)e^\theta \left(\frac{1-e^{-\theta}}{q}\right)^{u+1} + \frac{1-q}{q}(u+1)\left(\frac{1-e^{-\theta}}{q}\right)^u.$$

Therefore, we obtain

$$\psi(u) = \int_0^{\theta_0} (1-q)e^t \bar{F}_\Theta(t) \left(\frac{1-e^{-t}}{q}\right)^{u+1} dt + \int_0^{\theta_0} \frac{1-q}{q}(u+1)\bar{F}_\Theta(t)\left(\frac{1-e^{-t}}{q}\right)^u dt.$$

Let $J(u) = \int_0^{\theta_0} ((1-e^{-t})/q)^u d\theta$. Making the change of variable $qx = 1-e^{-t}$, we have

$$J(u) = q\int_0^1 \frac{1}{1-xq} x^u dx \quad \text{and} \quad q \times \frac{1}{u+1} \le J(u) \le \frac{q}{1-q} \times \frac{1}{u+1}.$$

Furthermore, we have

$$\max_{\theta\in[0,\theta_0]} \bar{F}_\Theta(\theta)e^\theta = \frac{1}{1-q}, \min_{\theta\in[0,\theta_0]} \bar{F}_\Theta(\theta)e^\theta = \bar{F}_\Theta(\theta_0).$$

Therefore, the ruin probability is bounded as

$$(1-q)\bar{F}_\Theta(\theta_0)J(u+1) + (u+1)\frac{1-q}{q}\bar{F}_\Theta(\theta_0)J(u) \leq \psi(u) \leq J(u+1) + (u+1)\frac{1-q}{q}J(u).$$

This yields

$$F_\Theta(\theta_0)\frac{q}{u+2} + F_\Theta(\theta_0) \leq \psi(u) \leq \frac{q}{1-q} \times \frac{1}{u+2} + 1.$$

(ii) Let $I(u,\theta_0) = \int_0^{\theta_0} \psi_u(\theta) dF_\Theta(\theta)$. We assume a continuous distribution for the mixing variable Θ and make the change of variable $x = (1-e^{-\theta})/q$, for which $qdx = e^{-\theta}d\theta$, we get

$$I(u,\theta_0) = q(1-q)\int_0^1 \frac{f_\Theta\left(-\log(1-xq)\right)}{(1-xq)^2} x^{u+1} dx.$$

Let h be $f_\Theta \circ g$ with $g(x) = -\log(1-xq)$. The minimum condition to apply an integration by part theorem is to require the integrand function $(h(x)/(1-xq)^2)$ to be absolutely continuous, see Appendix 4.6.5. As $x \mapsto 1/(1-xq)^2$ are C^∞, we must show h is absolutely continuous.

But $h = f_\Theta \circ g$ is not necessarily continuous if both f_Θ and h are absolutely continuous. According to Merentes (1991), if g is absolutely continous, then $f_\Theta \circ g$ is absolutely continuous if and only if f_Θ is locally Lipschitzian. Using the Rademacher theorem, see Appendix 4.6.5, we deduce that f_Θ is locally Lipschitizan, so h is absolutely continuous.

We obtain

$$I(u,\theta_0) = q(1-q)\left[\frac{h(x)}{(1-xq)^2}\frac{x^{u+2}}{u+2}\right]_0^1 - q(1-q)\underbrace{\int_0^1 \left(\frac{h'(x)}{(1-xq)^2} + \frac{2qh(x)}{(1-xq)^3}\right)\frac{x^{u+2}}{u+2}dx}_{J(u)}.$$

The first term equals to

$$\frac{qf_\Theta(\theta_0)}{(1-q)(u+2)}$$

while the integral term is controlled as

$$|J(u)| \leq \sup_{x\in[0,1]} \left|\frac{qf'_\Theta(g(x))}{(1-xq)^3} + \frac{2qf_\Theta(g(x))}{(1-xq)^3}\right| \int_0^1 \frac{x^{u+2}}{u+2}dx = C\frac{1}{(u+2)(u+3)} = o\left(\frac{1}{u+2}\right),$$

since f_Θ and f'_Θ are bounded on $[0,\theta_0]$. Combining the two preceding results, we get to

$$\psi(u) = \bar{F}_\Theta(\theta_0) + \frac{qf_\Theta(\theta_0)}{(1-q)(u+2)} + o\left(\frac{1}{u+2}\right).$$

(iii) As f_Θ is C^{k-1} almost everywhere on $[0,\theta_0]$ and $f_\Theta^{(k)}$ is Lebesgue integrable, then $h^{(i)}$ is absolute continous for all $i \leq k$. Applying k times the integration by part theorem, we get

$$I(u,\theta_0) = \sum_{i=0}^{k-1} \frac{\tilde{h}^{(i)}(0)}{(u+2)\ldots(u+2+i)} + \int_0^{\theta_0} \tilde{h}^{(k)}(t)\frac{x^{u+2+k-1}}{(u+2)\ldots(u+2+k-1)}dx,$$

where $\tilde{h}(x) = h(x)/(1-xq)^2$. Since successive derivatives $f_\Theta^{(i)}$ are bounded on $[0,\theta_0]$, the integral term is controlled by $o\big(1/u^k\big)$. The expression of the ith order derivative for a composition $f \circ g$ is complex, see Huang et al. (2006).

(iv) If f_Θ is C^∞ on $[0,\theta_0]$, we have

$$\psi(u) \underset{u\to+\infty}{\sim} \bar{F}_\Theta(\theta_0) + \sum_{i=0}^{+\infty} \frac{\tilde{h}^{(i)}(0)}{(u+2)\dots(u+2+i)}.$$

□

We examine below the three special cases studied in Subsection 4.2.2. Only one asymptotic is available via known asymptotics of the incomplete beta function asymptotic, see Appendix 4.6.1. Indeed, when Θ is exponentially distributed, we have

$$\psi(u) = (1-q)^\lambda + \lambda(1-q)^\lambda \frac{1}{u+2} + o\left(\frac{1}{u+2}\right).$$

Using Theorem 4.3.4, with $\tilde{h}(x) = \lambda(1-xq)^\lambda/(1-xq)^2$, leads to the same expansion.

For the two other distributions, gamma and Lévy, we have to use Theorem 4.3.4, as no asymptotic is available. When Θ is gamma distribution, the function $\tilde{h}$ is

$$\tilde{h}(x) = \frac{1}{(1-xq)^2}(1-xq)^\lambda \left(\log\left(\frac{1}{1-xq}\right)\right)^{\alpha-1} \frac{\lambda^\alpha}{\Gamma(\alpha)}.$$

Thus,

$$\psi(u) \underset{u\to+\infty}{\sim} \frac{\Gamma(\alpha,\lambda\theta_0)}{\Gamma(\alpha)} + \frac{\lambda^\alpha}{\Gamma(\alpha)}(1-q)^{\lambda-1}\theta_0^{\alpha-1}\frac{1}{u+2} + o\left(\frac{1}{u+2}\right).$$

with $\lambda > 1$ and $\theta_0 = -\log(1-q)$.

When Θ is Lévy distibuted, the function $\tilde{h}$ is

$$\tilde{h}(x) = \frac{\alpha}{2\sqrt{\pi}} \log\left(\frac{1}{1-xq}\right)^{-3/2} e^{-\frac{\alpha^2}{4\log\left(\frac{1}{1-xq}\right)}}.$$

Thus,

$$\psi(u) \underset{u\to+\infty}{\sim} \operatorname{erfc}\left(\frac{\alpha}{2\sqrt{\theta_0}}\right) + \frac{\alpha}{2\sqrt{\pi}}\theta_0^{-3/2} e^{-\frac{\alpha^2}{4\theta_0}} \frac{1}{u+2} + o\left(\frac{1}{u+2}\right),$$

with $\theta_0 = -\log(1-q)$.

4.3.4 Tail claim distributions

In this subsection, we analyze the tail of the claim distribution, i.e. $P(X > x)$ for large values of x. For the present model by mixing, the survival function is the following Stieltjes integral

$$P(X > x) = \int_0^{+\infty} P(X > x|\Theta = \theta) dF_\Theta(\theta).$$

In the continuous time framework, it leads to

$$P(X > x) = \int_0^{+\infty} e^{-\theta x} dF_\Theta(\theta),$$

which is the Laplace transform of the random variable Θ. Here too, one can see that a similar argument works only when Θ has a light-tailed distribution. In fact, we cannot obtain interesting results by applying the integration by part directly on this Stieltjes integral (as for the ruin probability). So, we assume that Θ has a continuous distribution and, similarly to the first subsection, we are going to derive the asymptotic survival function.

Proposition 4.3.5. *Let us consider the continuous time framework of Subsection 4.2.1 and assume Θ has a continuous distribution with density f_Θ.*

(i) If f_Θ is almost everywhere differentiable on $\mathbb{R}_+$ with f'_Θ being a Lebesgue-integrable, then for $x > 0$,

$$P(X > x) = \frac{f_\Theta(0)}{x} + o\left(\frac{1}{x}\right).$$

(ii) If f_Θ is C^∞ in the neighborhood of the origin, then for $x > 0$,

$$P(X > x) \underset{x\to+\infty}{\sim} \sum_{k=0}^{+\infty} \frac{f_\Theta^{(k)}(0)}{x^k}.$$

(iii) If f_Θ can be expanded in the neighborhood of the origin as

$$f_\Theta(t) \underset{t\to 0}{\sim} \sum_{k=0}^{+\infty} f_k t^{\frac{k+\eta}{\mu}-1},$$

for $\eta, \mu > 0$, then for $x > 0$,

$$P(X > x) \underset{x\to+\infty}{\sim} \sum_{k=0}^{+\infty} \Gamma\left(\frac{k+\eta}{\mu}\right) \frac{f_k}{x^{\frac{k+\eta}{\mu}}}.$$

Proof. (i) f_Θ satisfies the minimum requirement for an application of the integration by parts. We get

$$P(X > x) = \left[f_\Theta(t)\frac{e^{-\theta x}}{-x}\right]_0^{+\infty} + \frac{1}{x}\int_0^{+\infty} e^{-\theta x} f'_\Theta(\theta)d\theta = \frac{f_\Theta(0)}{x} + o\left(\frac{1}{x}\right).$$

(ii) It is a direct application of Property 2.3(i) of Olver et al. (2010).

(iii) It is a direct application of the Watson lemma, e.g. 2.3(ii) of Olver et al. (2010). □

Remark 4.3.6. *Parts (i) and (ii) of this proposition may be not applicable when the density is not defined or zero at the origin. This justifies the part (iii).*

Remark 4.3.7. *The reason, why the behavior of the integrand function f_Θ at the origin matters, is explained by the Laplace's method. The Laplace method studies the asymptotic of the following integral*

$$I(x) = \int_a^b e^{xp(t)} q(t)dt,$$

where p and q are continuous functions around the point a, assumed to be the minimum of p in $[a, b[$. In our case, $p(t) = t$, hence the minimum of the exponent on $\mathbb{R}_+$ is attained at the origin. See, e.g., 2.3(iii) of Olver et al. (2010).

Let us see if the two special cases studied in the previous section fall within the framework of the previous proposition. Firstly, assume that Θ follows a gamma distribution $\mathcal{G}a(\alpha, \lambda)$. Using the integral representation of the exponential function, the density function can be expanded as

$$f_\Theta(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \sum_{k=0}^{+\infty} \frac{(-\lambda)^k}{k!} t^{\alpha+k-1}.$$

Thus, we get

$$P(X > x) \underset{x\to+\infty}{\sim} \sum_{k=0}^{+\infty} (-1)^k \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)k!} \left(\frac{\lambda}{x}\right)^{k+\alpha}.$$

with $\eta = \alpha$ and $\mu = 1$. This (asymptotic) polynomial decrease of the survival function is consistent with the fact that X is Pareto distributed

$$P(X > x) = \frac{1}{\left(1 + \frac{x}{\lambda}\right)^\alpha}.$$

When Θ follows a Lévy distribution $\mathcal{L}e(\alpha)$,

$$f_\Theta(\theta) = \frac{\alpha}{2\sqrt{\pi\theta^3}} e^{-\alpha^2/4\theta}.$$

Although this function is not defined at zero, the density converges to zero, since we have

$$f_\Theta(1/t) = \frac{\alpha t^{3/2}}{2\sqrt{\pi}} e^{-\alpha^2 t/4} \underset{t\to+\infty}{\longrightarrow} 0.$$

However, we cannot find a valid series expansion of $f_\Theta(1/t)$ at $+\infty$, or equivalently of $f_\Theta(\theta)$ at 0, based on the series expansion of the exponential function. Therefore, the preceding proposition is of limited use in the Lévy case, where we already know that

$$P(X > x) = e^{-\alpha\sqrt{x}}, x \geq 0.$$

More generally, Proposition 4.3.5 is difficult to apply when the density function of Θ is not defined.

Now, we look at the tail of the claim distribution in the discrete time framework. We have

$$P(X > u) = \int_0^{+\infty} (1-q) \left(1 - e^{-\theta}\right)^u dF_\Theta(\theta).$$

One way to deal with such an integral is to use an integration by part directly on the integral. But, it does not lead to satifying results as for the ruin probability. Even if we assume Θ has a continuous distribution, we do not get a Laplace transform of a certain function as in the continuous time:

$$P(X > u) = \int_0^{+\infty} (1-q) f_\Theta(\theta) \left(1 - e^{-\theta}\right)^u d\theta.$$

To get a term easily integrable, one can try to make a change of variable, e.g. $x = 1 - e^{-\theta}$ or $x = -\log(1 - e^{-\theta})$. The latter is not possible on $\mathbb{R}_+$ because the derivative of the function $\theta \mapsto -\log(1 - e^{-\theta})$ is unbounded near 0. Let us try $x = 1 - e^{-\theta}$. We get

$$P(X > u) = \int_0^1 (1-q) \frac{f_\Theta(-\log(1-x))}{1-x} x^u dx.$$

Using an integration by part theorem requires that the following limit to exist

$$\lim_{x\to 1}\frac{f_\Theta(-\log(1-x))}{1-x} = \lim_{t\to+\infty} f_\Theta(t)e^t.$$

This requirement is strong and will be satisfied only for light-tailed distributions and a certain range of parameter values. For example, when Θ is exponentially distributed $\mathcal{E}(\lambda)$, the previous constraint imposes $\lambda > 1$.

Another way to deal with such integral asymptotic is to apply the Pascal formula, assuming u is an integer. We get

$$P(X > u) = \sum_{k=0}^{u} \binom{u}{k}(1-q)(-1)^k \int_0^{+\infty} e^{-k\theta} dF_\Theta(\theta).$$

The integral is (once again) the Laplace transform L_Θ of the random variable Θ at k. This binomial alternating sum requires a special treatment because finding an asymptotic for $L_\Theta(k)$ will not help us to derive an asymptotic of the sum. This issue is studied in the next subsection.

4.3.5 Binomial alternating sum for claim tails

In the discrete time framework, the survival function can be expressed as

$$P(X > u) = \sum_{k=0}^{u} \binom{u}{k}(1-q)(-1)^k L_\Theta(k),$$

where L_Θ denotes the Laplace transform of the random variable Θ.

This integral falls within the framework of the alternating binomial sum defined as

$$S_n(\phi) = \sum_{k=n_0}^{n} \binom{n}{k}(-1)^k \phi(k), \tag{4.7}$$

where $0 \le n_0 \le n$, ϕ is a real function and $n \in \mathbb{N}$ is large. n_0 can be used to exclude first few points of the sum that would not be defined.

Letting

$$\phi(k) = (1-q)\frac{\lambda^\alpha}{(\lambda+j)^\alpha},$$

in Equation (4.7), we get the distribution function of X when Θ is gamma distributed of Subsection 4.2.2. Note that, having started with the integral representation $\int_0^{+\infty} P(X = k|\Theta = \theta)dF_\Theta(\theta)$ of $P(X = k)$, we know that the alternating sum is valued on $[0, 1]$. This is not immediate without that integral representation.

Let us point out that the probability $P(X = k)$ is a decreasing function of k. This is not easy to see by using the alternating binomial sum representation (4.7). Here is a simle proof. To indicate the dependence on the parameter λ, denote by $P(X = k)_\lambda$. From algebraic manipulation and using the binomial recurrence equation $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$, we get

$$P(X = k+1)_\lambda = P(X = k)_\lambda - P(X = k)_{\lambda+1}\frac{\lambda^\alpha}{(\lambda+1)^\alpha}.$$

So, as announced, the probability mass function of X is decreasing. The binomial alternating sum representation of $P(X > k)$ is

$$P(X > k) = (1 - q) \sum_{j=0}^{k} \binom{k}{j} (-1)^j \frac{\lambda^\alpha}{(\lambda + j)^\alpha}.$$

There should be an exponential cancelling in the sum, since $\binom{k}{j}$ tends quickly to infinity for large values of k (we recall $\binom{n}{k} \sim e^{-k} n^k / k!$) and $P(X > k)$ is a decreasing function. A study of the alternating sum seems to be rather complex.

Going back to the alternating sum with $n_0 = 0$, the first few terms can be expressed as

$$S_0(\phi) = \phi(0), S_1(\phi) = \phi(0) - \phi(1), S_2(\phi) = \phi(0) - 2\phi(1) + \phi(2).$$

Let Δ be the forward difference operator. Then we have $S_0 = \Delta^0\phi(0)$, $S_1 = -\Delta\phi(0)$ and $S_2 = \Delta^2\phi(0)$. More generally, the binomial alternating sum can be rewritten as

$$S_n(\phi) = \sum_{k=n_0}^{n} \binom{n}{k} (-1)^k \phi(k) = (-1)^n \Delta^n(\phi)(0).$$

Some complex analysis

To deal with such sums, a standard method consists in using complex analysis and contour integrals. Flajolet and Sedgewick (1995) provide a complete overview of this topic. In this subsection, we consider that the complex extension of ϕ of the sum $S_n(\phi)$. Their Lemma 1 gives the so-called Rice integral representation of $S_n(\phi)$

$$\sum_{k=n_0}^{n} \binom{n}{k} (-1)^k \phi(k) = (-1)^n \oint_\gamma \phi(z) \frac{n!}{z(z-1)\dots(z-n)} dz, \tag{4.8}$$

where ϕ is assumed to be analytic on a domain Ω containing $[n_0, n[$ and γ is a closed curve in Ω encircling $[n_0, n[$ but not $[0, n_0 - 1]$. Let f be the integrand of the right-hand side of (4.8).

By the residue theorem (e.g. Chapter 10 of Bak and Newman (2010)), if the integrand is analytic except at a countable number of isolated singularities inside the domain γ, then the contour integral equals to the sum of the residues of the integrand taken at the singularities inside γ. The right-hand side of Equation (4.8) is still cumbersome to compute, since the integrand has at least $n + 1$ singularities at $0, 1, \dots, n$. So, in the present situation, the complex contour integration does not really simplify the problem.

As we want to derive some asymptotics when n tends to infinity, the domain γ of the contour integration has to be extended to the entire positive real half-plane $\{z, Re(z) > 0\}$. However, we do not want to compute the sum of the residuals at integers $\{n_0, n_0+1, \dots, +\infty\}$. Furthermore, we do not know if $f(z)$ does not explode as $Re(z) \to +\infty$. Nevertheless, the solution does come by extending the contour of integration. Let γ be the circle $C_{(0,R)}$ of radius R centered at 0 excluding poles in $\mathbb{N}$. Assuming f is of polynomial growth towards infinity, the integral

$$\oint_{C_{(0,R)}} f(z) dz$$

tends to 0 as $R \to +\infty$. By the residue theorem, the contour integral $\oint_{C_{(0,\infty)}} f(z)dz$ also equals to the sum of residuals of f at integers $\{0, \ldots, n_0 - 1\}$ and $\{n_0, n_0 + 1, \ldots, +\infty\}$. The first residual contribution is a finite sum, while the second contribution is the binomial alternating sum $S_n(\phi)$.

Thus, in the particular case of polynomial growth, the binomial sum $S_n(\phi)$ reduces to the computation of a limited number of residuals at $\{0, \ldots, n_0 - 1\}$, see the proof of Theorem 1 of Flajolet and Sedgewick (1995).

Theorem. *Let ϕ be a rational analytic function on $[n_0, \infty[$. Then we have*

$$\sum_{k=n_0}^{n} \binom{n}{k} (-1)^k \phi(k) = -(-1)^n \sum_{s} Res\left(\phi(s) \frac{n!}{s(s-1)\ldots(s-n)}\right),$$

where the summation of residues is done over poles not on $[n_0, \infty[$.

Theorem 2(i) of Flajolet and Sedgewick (1995) applies the same approach when f is meromorphic (i.e. complex differentiable everywhere except at a countable number of points) and not necessarily of polynomial growth.

The same argument applies when we replace the circle $C_{(0,R)}$ by a semicircle $S_{(0,R,d)} = \{z \in \mathbb{C}, Re(z) > d, |z| < R\}$, see part (ii) of Theorem 2 of Flajolet and Sedgewick (1995). But this time, we only get an asymptotic for the original problem. Furthermore, things are intrinsically more complicated when the function ϕ is not a rational function, because we consider the complex extension of ϕ. For instance, function $z \mapsto 1/z^2$ has a second-order pole at $z = 0$ but function $z \mapsto 1/z^{1.414}$ has algebraic singularity at $z = 0$.

To deal with algebraic singularities, the integration contour γ must exclude the singularities. Flajolet and Sedgewick (1995) examplify a keyhole structure approach (i.e. Hankel contour) when $\phi(z)$ has a (non polar) algebraic singularity $1/x^\lambda$, see proof of their Theorem 3. The keyhole structure captures the effect of the singularity at 0 by decreasing the radius of the hole in $1/\log(n)$ as n tends to infinity. Dealing with non-isolated singularities (i.e. branch points) need even more care than just skipping it as with a semicircle.

Let us consider, for example, the complex square root $\sqrt{z}$, the branch point is the negative real axis $]-\infty, 0]$. The branch point is of finite order, compared to the complex logarithm for instance. Indeed we have

$$\sqrt{\rho e^{i(\theta+2k\pi)}} = \sqrt{\rho} e^{i(\theta/2+k\pi)} = \begin{cases} -\sqrt{\rho} e^{i(\theta/2)} & \text{if } k = 1 \\ \sqrt{\rho} e^{i(\theta/2)} & \text{if } k = 2 \end{cases}$$

Flajolet and Sedgewick (1995) consider a local approximation of $z^{1/2}$ around the origin only at the contour part of the keyhole structure surrounding the origin. Otherwise, we keep the polynomial growth of the square root for the rest of the contour, see Example 7. Handling branch points of infinity order, say with the complex logarithm function $\log(z)$, is similar except that the resulting asymptotic as n tends to infinity is different, see Example 8.

We report below a table of correspondences between singularity and asymptotics.

Two simple illustrations

Let us consider for phi two particular functions of interest below. Firstly, we choose $f_1(z)$ be $1/(z + \beta)^\alpha$ where $z \in \mathbb{C}$. It has a singularity at $-\beta$, which is a multiple pole if $\alpha \in \mathbb{N}$.

Singular part $s_0 \notin \mathbb{N}$	Asymptotics
simple pole $(z-s_0)^{-1}$	$-\Gamma(-s_0)n^{s_0}$
multiple pole $(z-s_0)^{-m}$	$-\Gamma(-s_0)n^{s_0}\frac{(\log n)^{m-1}}{(m-1)!}$
algebraic singularity $(z-s_0)^{\lambda}$	$-\Gamma(-s_0)n^{s_0}\frac{(\log n)^{-\lambda-1}}{\Gamma(-\lambda)}$
+ logarithmic singularity $(z-s_0)^{\lambda}(\log(z-s_0))^r$	$-\Gamma(-s_0)n^{s_0}\frac{(\log n)^{-\lambda-1}}{\Gamma(-\lambda)}(\log\log n)^r$

Table 4.1: Correspondences between singularity and asymptotic

Using a keyhole structure centered at $-\beta$ and Table 4.1, we have an asymptotic of the form

$$S_n(f_1) \underset{n\to+\infty}{\sim} \frac{\Gamma(\beta)}{\Gamma(\alpha)}\frac{(\log n)^{\alpha-1}}{n^{\beta}}.$$

Secondly, we choose $f_2(z)$ be $e^{-\alpha\sqrt{z}}$. The function f_2 has a branch point at $z=0$, because of the complex square root. First, we use an inifinitesimal asymptotic of the exponential around 0. That is $f_2(z) = 1 - \alpha\sqrt{z} + o(z)$. Since the contour integral of a sum is the sum of contour integrals and that the contour integral of an analytic function is zero, we can drop the constant 1.

We use a right-oriented half-plane keyhole structure centered at 0 for $Re(z) > d$ (with $-\infty < d < 0$), similar to Theorem 3 of Flajolet and Sedgewick (1995), since the function f_2 has a exponential growth on the half-plane $Re(z) < d$ and cannot be integrated as the radius tends to infinity.

We cannot use the singularity correspondence table for the square root, because the singularity is zero. But, the square root can be approximated by Theorem 3 of Flajolet and Sedgewick (1995). And the term $o(z)$ is controlled by the small circle of the keyhole structure on which $|z| < 1/\log(n)$. Thus we get the following asymptotic of the alternating sum

$$S_n(f_2) \underset{n\to+\infty}{\sim} \frac{\alpha}{\sqrt{\pi}\log(n)} - \frac{\alpha\gamma_e}{2\sqrt{\pi\log^3(n)}},$$

where $\gamma_e = 0.5772156649$ is the Euler-Mascheroni constant.

Claim tail asymptotics

Based on the previous subsections, we are able to derive tail asymptotics of a claim X given a mixing distribution Θ in the discrete time framework presented in Subsection 4.2.2.

When Θ follows an exponential distribution $\mathcal{E}(\lambda)$, we use the asymptotic of the beta function $\beta(a,b)$ for large values of b, see Appendix 4.6.1. We get that the tail of the distribution is asymptotically

$$P(X > k) \underset{k\to+\infty}{\sim} (1-q)\frac{\Gamma(\lambda+1)}{(k+1)^{\lambda}},$$

which decreases like a discrete Pareto distribution (i.e. a Zipf distribution). This tail behavior of a Yule-Simon distribution was already reported in Simon (1955).

When Θ follows a gamma distribution $\mathcal{G}a(\alpha,\lambda)$, we use asymptotics of the alternating binomial sum with the function $f_1(z) = 1/(z+\lambda+1)^{\alpha}$. Therefore, the tail distribution is

asymptotically

$$P(X > k) \underset{k\to+\infty}{\sim} (1-q)\frac{\lambda^\alpha \Gamma(\lambda+1)}{\Gamma(\alpha)}\frac{(\log k)^{\alpha-1}}{k^{\lambda+1}},$$

which decreases slightly slower than a Zipf distribution due to the logarithm in the numerator.

When Θ follows a Lévy distribution $\mathcal{L}e(\alpha)$, again we use an asymptotic of alternating binomial sums with the function $f_2(z) = e^{-\alpha\sqrt{z}}$. Thus, the tail distribution is asymptotically

$$P(X > k) \underset{k\to+\infty}{\sim} (1-q)\alpha\left(\frac{1}{\sqrt{\pi\log(k)}} - \frac{\gamma_e}{2\sqrt{\pi\log^3(k)}}\right),$$

which decreases extremely slowly. Such a tail behaviour is heavier than for a Pareto distribution. With continuous distributions, a similar behaviour is obtained for the log-Cauchy distribution, for example.

Numerical illustrations

On Figure 4.1, we plot the tails of the distributions derived above. The exponential-geometric distribution has a very tractable survival function, since incomplete beta function is available most softwares, e.g., in R via the `pbeta` function. Therefore, we can benchmark the asymptotic with the true value. However, for the two other distributions, we have to compute two binomial alternating sums. These sums are particularly unstable because the central term $C_n^{n/2}$ reaches very quickly infinity, which drives the binomial alternating sum between $+\infty$ or $-\infty$.

In modern computers, a real number is stored in eight bytes (i.e. 64 bits), but only 53 bits are reserved for the precision (see, e.g., http://en.wikipedia.org/wiki/Double-precision_floating-point_format). In our numerical experiment, the alternating binomial sum $S_n(\phi)$ becomes unstable for $n \geq 48$ with the standard double precision. To compute the alternating sum $S_n(\phi)$ for large n, we have no other option than to use high precision floating-point arithmetic libraries such as the GMP library of Grandlund Torbjoern & the GMP Devel. Team (2011) and the MPFR library of Fousse et al. (2011). Using GMP and MPFR libraries allow us to a high number of bits, say 500 or 1000. Reliability of those libraries has been established in Fousse et al. (2007). Those libraries are interfaced in R via the **Rmpfr** package of Maechler (2012).

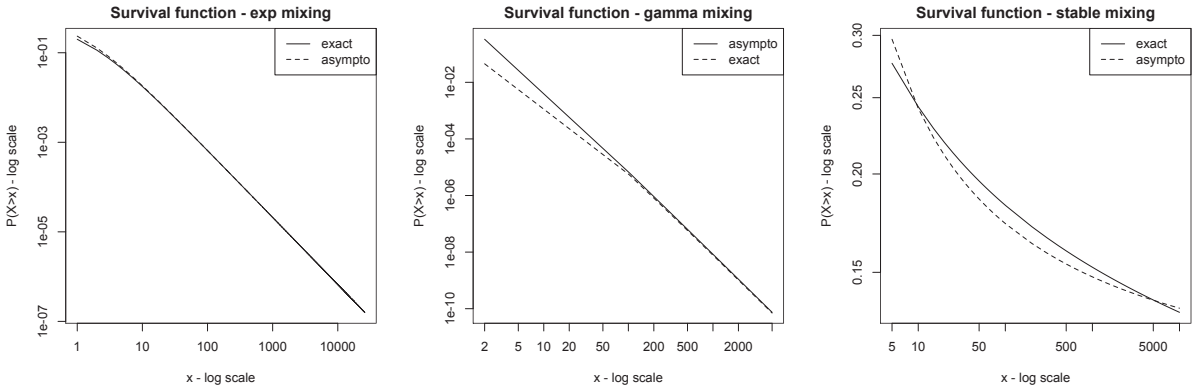
Figures 4.1a, 4.1b and 4.1c correspond to a mixing distribution when Θ is exponential, gamma and Lévy-stable distributed, respectively. Note that all plots have log scale for x and y axes. On Figures 4.1a, 4.1b, the distribution tail show a Pareto-type behavior, as we observe a straight line. The Lévy stable mixing on Figure 4.1c shows clearly a heavier tail.

4.4 Focus on the dependence structure

This section studies the dependence structure of the dependent risk models described in Subsections 4.2.1 and 4.2.2, respectively for discrete-time and continuous-time settings. Let us start by recalling Property 2.1 of Albrecher et al. (2011) for the continuous-time model.

Proposition. *When claim sizes fullfill for each $n \geq 1$,*

$$P(X_1 > x_1, \ldots, X_n > x_n | \Theta = \theta) = \prod_{i=1}^{n} e^{-\theta x_i},$$



(a) Exponential (b) Gamma (c) Lévy stable

Figure 4.1: Survival functions

then, they have a dependence structure due to an Archimedean survival copula with generator $\phi = L_\Theta^{-1}$*, the inverse Laplace transform of* Θ*.*

Therefore, in continuous time, the dependence structure is simply an Archimedean copula. Regarding the discrete-time setting, things are more complicated: the dependence among discrete random variables is a complex topic. Genest and Neslehova (2007) present issues linked to discrete copula. To better understand the problem, we recall the Sklar theorem, see, e.g., Joe (1997); Nelsen (2006).

Theorem (Sklar)**.** *Let* H *be the bivariate distribution function of a random pair* (X, Y)*. There exists a copula* C *such that for all* $x, y \in \mathbb{R}$*,*

$$H(x, y) = C(F_X(x), F_Y(y)). \tag{4.9}$$

Furthermore, C *is unique on the Cartesian product of the ranges of the marginal distributions.*

As a consequence, the copula C is not unique outside the support of the random variables X and Y. When X, Y are discrete variables in $\mathbb{N}$, C is only unique on $\mathbb{N}^2$ but not on $\mathbb{R}^2 \setminus \mathbb{N}^2$. The non-identifiability is a major source of issues. An example of discrete copulas is the empirical copula for observation sample $(X_i, Y_i)_{1 \leq i \leq n}$.

Some problems in the discrete case arise from the discontinuity of the distribution function of discrete random variables. Let B be a Bernoulli variable $\mathcal{B}(p)$. The distribution function F_B, given by $F_B(x) = (1-p)\mathbb{1}_{x \geq 0} + p\mathbb{1}_{x \geq 1}$, is discontinuous. Thus, the random variable $F_B(B)$ is not a uniform variable $\mathcal{U}(0, 1)$.

Let us introduce Genest and Neslehova (2007)'s notation. Let $\mathcal{A}$ be the class functions verifying Equation (4.9) for all $x, y \in \mathbb{R}$ for a given F_X and F_Y. Let us define the function B as for all $u, v \in [0, 1], B(u, v) = (F_X^{-1}(u), F_Y^{-1}(v))$. We also denote by D the distribution function of the pair $(F_X(X), F_Y(Y))$.

With a simple bivariate Bernoulli vector, Example 1 of Genest and Neslehova (2007) shows that (i) functions B and D are different, (ii) B is not a distribution function despite both B and D belong to the class $\mathcal{A}$. Even in that simple support $\{0, 1\}^2$, the identifiability issue of the copula C cannot be discarded. Proposition 1 of Genest and Neslehova (2007) extends to

any bivariate pair (X, Y): B is not a distribution, whereas D is a distribution function but not a copula.

Let us now consider two exponential random variables that are conditionally independent given a factor Θ. We choose to focus here only on the bivariate case. Suppose that these variables are discretized by taking their integer parts. We will see that the two resulting random variables have geometric distributions; these are, of course, correlated. Our purpose in Subsections 4.4.1 - 4.4.4 below is precisely to study the dependence involved.

4.4.1 Dependence induced by the mixing approach

We start by comparing the joint distributions of two mixed exponential random variables and of geometric approximations obtained by discretization.

Continous case

Now, consider $Y_i, i = 1, 2$ to be conditionnaly independent exponential random variables, i.e. $Y_i/\Theta = \theta \overset{i.i.d.}{\sim} \mathcal{E}(\theta)$. We have

$$F_{Y_1}(x) = P(Y_1 \leq x) = \int_0^\infty P(Y_1 \leq x/\Theta = \theta) dF_\Theta(\theta) = 1 - L_\Theta(x),$$

where L_Θ stands for the Laplace transform of the random variable Θ, assuming L_Θ exists. Using this formulation, we can check that the random variable $F_{Y_i}(Y_i)$ is uniformly distributed. Indeed we have

$$P(F_{Y_1}(Y_1) \leq u) = P(Y_1 \leq L_\Theta^{-1}(1-u)) = 1 - L_\Theta(L_\Theta^{-1}(1-u)) = u.$$

Furthermore, the (unique) copula of the couple (Y_1, Y_2) can be derived by

$$C_{Y_1,Y_2}(u, v) = P(F_{Y_1}(Y_1) \leq u, F_{Y_2}(Y_2) \leq v) = P(Y_1 \leq L_\Theta^{-1}(1-u), Y_2 \leq L_\Theta^{-1}(1-v))$$
$$= \int_0^\infty (1 - e^{-\theta L_\Theta^{-1}(1-u)})(1 - e^{-\theta L_\Theta^{-1}(1-v)}) dF_\Theta(\theta).$$

Hence the copula function is given by

$$C_{Y_1,Y_2}(u, v) = u + v - 1 + L_\Theta \left[L_\Theta^{-1}(1-u) + L_\Theta^{-1}(1-v) \right]. \tag{4.10}$$

We retrieve the fact that the couple (Y_1, Y_2) has an Archimedean survival copula with generator L_Θ^{-1}, see, e.g., Albrecher et al. (2011). In other words, the joint distribution of the tail is given by

$$P(\bar{F}_{Y_1}(Y_1) > u, \bar{F}_{Y_2}(Y_2) > v) = L_\Theta \left[L_\Theta^{-1}(u) + L_\Theta^{-1}(v) \right].$$

From Theorem 4.6.2 of Nelsen (2006), the above expression can be extended to any dimension. A n-dimension function $C(u_1, \ldots, u_n)$ defined by a generator ϕ and $C(u_1, \ldots, u_n) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_n))$ is a n-copula for all $n \geq 2$ if and only if the generator ϕ^{-1} is completely monotone, i.e. for all $k \in \mathbb{N}$,

$$(-1)^k \frac{d^k \phi^{-1}}{dt^k}(t) \geq 0.$$

As the generator is $\phi^{-1}(t) = L_\Theta(t)$ in our model, then the Laplace transform is completely monotone. In particular, L_Θ is a decreasing convex function. As ϕ is the generator of an Archimedean copula, ϕ is a convex decreasing function. Since $\phi(t) = L_\Theta^{-1}(t)$, L_Θ^{-1} is also a convex decreasing function.

Discrete case

If Y follows an exponential distribution $\mathcal{E}(\theta)$, then $W = \lfloor Y \rfloor$ follows a geometric distribution $\mathcal{G}(1-e^{-\theta})$, where $\lfloor x \rfloor$ denotes the floor function. Indeed, we have

$$P(W = k) = P(Y \in [k, k+1[) = e^{-\theta k}(1-e^{-\theta}).$$

Furthermore, $P(W > x) = \overline{F}_Y(\lfloor x \rfloor + 1) = e^{-\theta(\lfloor x \rfloor + 1)}$ and $P(W \leq x) = F_Y(\lfloor x \rfloor + 1) = 1 - e^{-\theta(\lfloor x \rfloor + 1)}$ for all $x \in \mathbb{R}$. Hence, the distribution function F_W and F_Y only coincide for integer values with a shift, i.e. $F_W(n) = F_Y(n+1)$ for all $n \in \mathbb{N}$.

Using the same assumption for Y_1 and Y_2 as in Subsection 4.4.1, we look at the dependence of W_1, W_2 with $W_i = \lfloor Y_i \rfloor, i = 1, 2$. The distribution function is given by

$$F_{W_1}(x) = P(W_1 \leq x) = P(\lfloor Y_1 \rfloor \leq x) = P(Y_1 < \lfloor x \rfloor + 1) = P(Y_1 \leq \lfloor x \rfloor + 1) = 1 - L_\Theta(\lfloor x \rfloor + 1),$$

for $x \in \mathbb{R}_+$. In this case, the random variable $F_{W_i}(W_i)$ is not uniform on [0,1] since

$$P(F_{W_1}(W_1) \leq u) = P(W_1 \leq L_\Theta^{-1}(1-u) - 1) = 1 - L_\Theta\left(\lfloor L_\Theta^{-1}(1-u) \rfloor\right) \neq u.$$

Their joint distribution function is given by

$$\begin{aligned} &D_{W_1,W_2}(u,v) = P(F_{W_1}(W_1) \leq u, F_{W_2}(W_2) \leq v) = P(W_1 \leq L_\Theta^{-1}(1-u) - 1, W_2 \leq L_\Theta^{-1}(1-v) - 1) \\ &= \int_0^\infty (1 - e^{\theta(\lfloor L_\Theta^{-1}(1-u) - 1 \rfloor + 1)})(1 - e^{\theta(\lfloor L_\Theta^{-1}(1-v) - 1 \rfloor + 1)}) dF_\Theta(\theta). \end{aligned}$$

The distribution function of $(F_{W_1}(W_1), F_{W_2}(W_2))$ can be rewritten as

$$D_{W_1,W_2}(u,v) = 1 - L_\Theta\left(\lfloor l_u \rfloor\right) - L_\Theta\left(\lfloor l_v \rfloor\right) + L_\Theta\left(\lfloor l_u \rfloor + \lfloor l_v \rfloor\right), \tag{4.11}$$

where $l_p = L_\Theta^{-1}(1-p)$. Equations (4.10) and (4.11) differ by the use of the floor function in the arguments of L_Θ.

Remark 4.4.1. *If one uses another parametrization for W_i than the geometric distribution $\mathcal{G}(1-e^{-\theta})$, then we have to replace all the Laplace transform by an appropriate expectation. For example, if we consider $\mathcal{G}(e^{-\theta})$, we use $E((1-e^{-\Theta})^x)$ instead of $L_\Theta(x)$.*

4.4.2 Differences between C_{Y_1,Y_2} and D_{W_1,W_2}

Before looking at the differences between C_{Y_1,Y_2} and D_{W_1,W_2}, we check that these distribution functions (defined in Equations (4.10) and (4.11)) are identical on the support of (F_{W_1}, F_{W_2}). Let $\mathrm{Im}(F_{W_i})$ be the inverse image by F_{W_i} of the integer set $\mathbb{N}$. Let $u \in \mathrm{Im}(F_{W_1})$ and $v \in \mathrm{Im}(F_{W_2})$, i.e. there exist $n, m \in \mathbb{N}$, such that $u = F_{W_1}(n)$ and $v = F_{W_2}(m)$. Firstly, we have

$$u = F_{W_1}(n) = F_{Y_1}(n+1) \quad \text{and} \quad v = F_{W_2}(m) = F_{Y_2}(m+1).$$

Secondly, functions C_{Y_1,Y_2} and D_{W_1,W_2} can be written as

$$C_{Y_1,Y_2}(u,v) = P(F_{Y_1}(Y_1) \leq F_{Y_1}(n+1), F_{Y_2}(Y_2) \leq F_{Y_2}(m+1)) = P(Y_1 \leq n+1, Y_2 \leq m+1).$$

and

$$D_{W_1,W_2}(u,v) = P(F_{W_1}(W_1) \leq F_{W_1}(n), F_{W_2}(W_2) \leq F_{W_2}(m)) = P(W_1 \leq n, W_2 \leq m).$$

Hence, both functions $C_{Y_1,Y_2}(u,v)$ and $D_{W_1,W_2}(u,v)$ are equal for $u, v \in \mathrm{Im}(F_{W_1}) \times \mathrm{Im}(F_{W_2})$.

Now, we are going to determine the maximal distance between $C_{Y_1,Y_2}(u,v)$ and $D_{W_1,W_2}(u,v)$. For that, we begin by deriving two elementary lemmas.

Lemma 4.4.2. *Let f be a continuous concave (resp. convex) function on a set S. Then the sequence $(f(x_i) - f(x_{i-1}))_i$ with $x_i = x_0 + i\delta$ is decreasing (resp. increasing).*

Proof. Let $\tilde{F}_i : [x_{i-1}, x_{i+1}] \mapsto \mathbb{R}$ be defined as

$$\tilde{F}_i(x) = \frac{x - x_{i-1}}{x_{i+1} - x_{i-1}} f(x_{i-1}) + \frac{x_{i+1} - x}{x_{i+1} - x_{i-1}} f(x_{i+1}).$$

Since f is concave, we have for all $x \in [x_{i-1}, x_{i+1}]$, $f(x) \geq \tilde{F}_i(x)$. In particular for $x = x_i$, we get

$$f(x_i) \geq (f(x_{i-1}) + f(x_{i+1}))/2 \Leftrightarrow f(x_i) - f(x_{i-1}) \geq f(x_{i+1}) - f(x_i).$$

□

Lemma 4.4.3. *Let f be a completely monotone function and $c > 0$. The function $f_c : x \mapsto f(x) - f(x + c)$ is also completely monotone.*

Proof. As f being completely monotone, $(-1)^n f^{(n)}$ is decreasing. Furthermore, we have

$$(-1)^n f_c^{(n)}(x) = (-1)^n f^{(n)}(x) - (-1)^n f^{(n)}(x + c) \geq 0$$

as $x < x + c$. Thus f_c is completely monotone. □

In particular, if f is convex decreasing, then $x \mapsto f(x) - f(x + c)$ is also convex decreasing, and $x \mapsto f(x + c) - f(x)$ is concave increasing.

Let $\Delta_{i,j}$ be the height of stairstep in the graphical representation of the distribution function D_{W_1,W_2}. We have $\Delta_{i,j} = D_{W_1,W_2}(x_i, y_j) - D_{W_1,W_2}(x_{i-1}, y_{j-1})$, where $x_i, y_j \in \mathrm{Im}(F_{W_1}) \times \mathrm{Im}(F_{W_2})$, i.e. $x_i = F_{W_1}(i)$ and $y_j = F_{W_2}(j)$.

Proposition 4.4.4. *The maximal stairstep, representing the highest difference between distribution functions D_{W_1,W_2} and C_{Y_1,Y_2}, is $\Delta_{\infty,0} = \Delta_{0,\infty}$.*

Proof. We are looking for the maximum of $\Delta_{i,j}$. By algebraic manipulations, we have

$$D_{W_1,W_2}(x_i, y_j) = P(W_1 \leq i, W_2 \leq j) = 1 - L_\Theta(i + 1) - L_\Theta(j + 1) + L_\Theta(i + j + 2).$$

Since L_Θ is completely monotone, the function $x \mapsto L_\Theta(x + c) - L_\Theta(x)$ is concave increasing, cf. Lemma 4.4.3. Similarly, for a constant $c > 0$, the function $g_c : x \mapsto 1 - L_\Theta(c+1) - L_\Theta(x + 1) + L_\Theta(x + 2 + c)$ is concave increasing.

Furthermore, we have

$$\Delta_{i,j} = D_{W_1,W_2}(x_i, y_j) - D_{W_1,W_2}(x_i, y_{j-1}) + D_{W_1,W_2}(x_i, y_{j-1}) - D_{W_1,W_2}(x_{i-1}, y_{j-1})$$

Using the function f defined above, we have $D_{W_1,W_2}(x_i, y_j) - D_{W_1,W_2}(x_i, y_{j-1}) = g_i(j) - g_i(j - 1)$. Hence, $(D_{W_1,W_2}(x_i, y_j) - D_{W_1,W_2}(x_i, y_{j-1}))_j$ is a decreasing sequence, using Lemma 4.4.2 and g_i is concave. $D_{W_1,W_2}(x_i, y_{j-1}) - D_{W_1,W_2}(x_{i-1}, y_{j-1}) = g_{i+1}(j + 1) - g_{i+1}(j) + L_\Theta(i) - L_\Theta(i + 1)$. By Lemma 4.4.3, the function g_{i+1} is concave and by Lemma 4.4.2 the sequence $(H(x_i, y_{j-1}) - H(x_{i-1}, y_{j-1}))_j$ is decreasing.

Therefore for a fixed i, $(\Delta_{i,j})_j$ is a sum of two decreasing sums of j. Since Archimedean copulas are symmetric, we deduce

$$\max_{i,j \geq 0} \Delta_{i,j} = \max_{i \geq j} \Delta_{i,j} = \max_{i \geq 0} \max_{i \geq j \geq 0} \Delta_{i,j} = \max_i \Delta_{i,0}.$$

Getting back the original definition, we have

$$\Delta_{i,0} = P(W_1 \leq i, W_2 \leq 0) = P(W_1 \leq i, W_2 \leq 0) = 1 - L_\Theta(i+1) - L_\Theta(1) + L_\Theta(i+2).$$

Since L_Θ is convex, the sequence $L_\Theta(i+2) - L_\Theta(i+1)$ is an increasing sequence. We conclude that $(\Delta_{i,1})_i$ is increasing, so the maximum is attained at $+\infty$. Therefore,

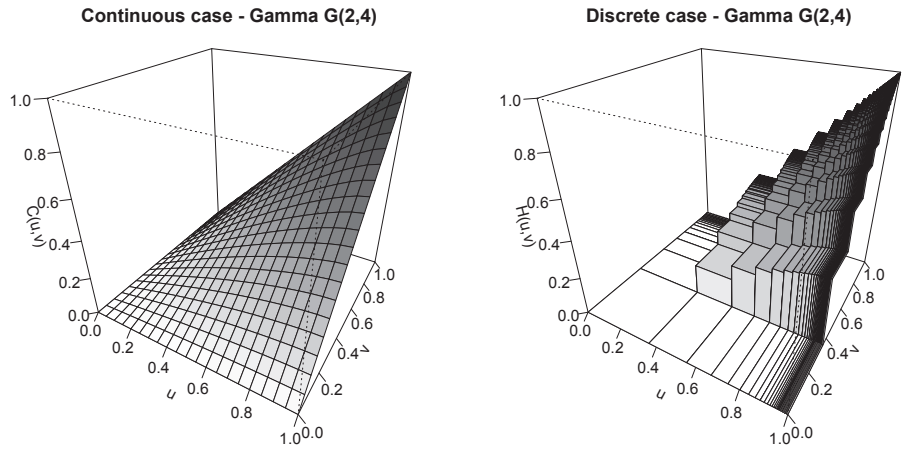
$$\max_{i,j} \Delta_{i,j} = \Delta_{+\infty,0} = P(W_1 \leq +\infty, W_2 \leq 0) = P(W_2 = 0).$$

□

On Figure 4.2, we investigate the numerical differences with a given distribution for the latent variable. We consider Θ follows a gamma distribution $\mathcal{G}a(2,4)$. In other words, we assume

$$L_\Theta(t) = \left(\frac{\lambda}{\lambda + t}\right)^\alpha \Leftrightarrow L_\Theta^{-1}(z) = \lambda(z^{-1/\alpha} - 1),$$

with $\alpha = 2, \lambda = 4$. We plot the unique continuous copula C_{Y_1,Y_2} and the distribution function D_{W_1,W_2}, left-hand and right-hand graphs, respectively.



(a) Continuous case (b) Discrete case

Figure 4.2: Comparison of copula C_{Y_1,Y_2} and distribution function D_{W_1,W_2}

C_{Y_1,Y_2} passes through all the top-left corners of the stairs described by H_{W_1,W_2}. The grid of points used for the graphs is $\{(F_{W_1}(n), F_{W_2}(m)), n, m \in \mathbb{N}\}$. As n increases, the distribution function $F_{W_1}(n)$ tends 1, so there infinitely many stairs towards the point $(1,1)$. Graphically, the maximal stairstep corresponds to $(u,v) = (1, F_{W_2}(0))$ or $(F_{W_1}(0), 1)$ and $\Delta_{\infty,0} = 0.36$. The maximal stairstep is also the maximum difference between copula C_{Y_1,Y_2} and distribution function D_{W_1,W_2}.

4.4.3 Non-identifiability issues

The function D_{W_1,W_2} is not a copula but only a distribution function. The maximal stairstep leads to the question of the maximal differences between two copulas satisfying

Equation (4.9)

$$H(x,y) = C(F_X(x), F_Y(y)),$$

when X, Y take discrete values. The answer is given by the Carley bounds C_H^-, C_H^+. For all copulas C_H verifying Equation (4.9), we have

$$C_H^- \leq C_H \leq C_H^+,$$

cf. Proposition 2 of Genest and Neslehova (2007).

The non-identifiability issue matters, since with discrete copulas the dependence measure, such as the tau of Kendall or the rho of Spearman, are no longer unique. Furthermore, if X and Y are independent, it does not imply that the copula of (X, Y) is the independent copula. In other words, the copula alone does not characterize the dependence: we need assumptions on margins.

In the copula literature, efforts have been done to tackle the non-identifiability issue of dependence measure. The current answer is the interpolated copula $C_{X,Y}^{\maltese}$, which is a bilinear interpolation of the distribution function D of the pair $(F(X), F(Y))$ on the discrete set $\mathrm{Im}(F_X) \times \mathrm{Im}(F_Y)$.

Let $(u_i, v_j) \in \mathrm{Im}(F_X) \times \mathrm{Im}(F_Y)$, i.e. $u_i = F_X(i)$ and $u_j = F_X(j)$. For all $(u, v) \in [u_i, u_{i+1}] \times [v_j, v_{j+1}]$, the interpolated copula is defined as

$$C_{X,Y}^{\maltese}(u,v) = \bar{\lambda}_u\bar{\lambda}_v D(u_i, v_j) + \lambda_u\bar{\lambda}_v D(u_{i+1}, v_j) + \bar{\lambda}_u\lambda_v D(u_i, v_{j+1}) + \lambda_u\lambda_v D(u_{i+1}, v_{j+1}), \quad (4.12)$$

where $\lambda_u = (u - u_i)/(u_{i+1} - u_i)$, $\lambda_v = (v - v_j)/(v_{j+1} - v_j)$. This copula was already mentioned in Lemma 2.3.5 of Nelsen (2006) to prove the Sklar theorem. The interpolated copula can also be interpreted as the copula of $(X + U, Y + V)$ where U, V are two independent uniform random variables, see, e.g., Section 4 of Denuit and Lambert (2005). This formulation is useful when doing random generation.

The properties of the interpolated copula of the pair (X, Y) are: (i) Kendall's tau $\tau(X, Y) = \tau(C_{X,Y}^{\maltese})$ and Spearman's rho $\rho(X, Y) = \rho(C_{X,Y}^{\maltese})$, (ii) $C_{X,Y}^{\maltese}$ is absolutely continuous and (iii) $X \perp Y \Leftrightarrow C_{X,Y}^{\maltese} = \Pi$, the independent copula. Unfortunately, $C_{X,Y}^{\maltese}$ depends on marginals and when $F_X = F_Y$ does not imply $C_{X,Y}^{\maltese}(u, v) = \min(u, v)$, as well as $F_X = 1 - F_Y \nRightarrow C_{X,Y}^{\maltese}(u, v) = (u + v - 1)_+$, see, e.g., Genest and Neslehova (2007).

As the interpolated copula is absolutely continuous, a density can be derived. Indeed, by differentiating Equation (4.12) with respect to u and v, we get

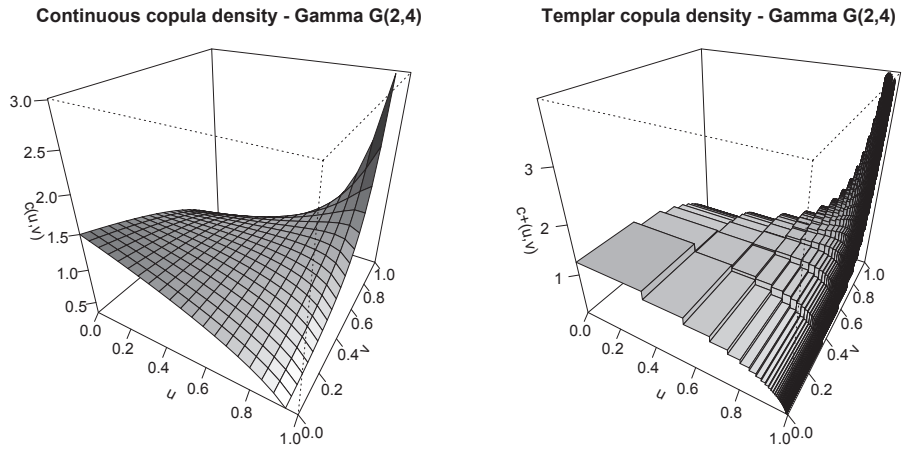
$$c_{X,Y}^{\maltese}(u,v) = \frac{D(u_i, v_j) - D(u_{i+1}, v_j) - D(u_i, v_{j+1}) + D(u_{i+1}, v_{j+1})}{(u_{i+1} - u_i)(v_{j+1} - v_j)}, \quad (4.13)$$

for $(u, v) \in [u_i, u_{i+1}] \times [v_j, v_{j+1}]$.

To better see the differences between the copula C_{Y_1,Y_2} and the interpolated copula $C_{W_1,W_2}^{\maltese}$, we choose to plot the densities on Figure 4.3 and not the distribution functions. On Figure 4.3, we plot densities of the continuous copula of the pair (Y_1, Y_2) and the interpolated copula of the pair (W_1, W_2), in the same setting as the previous subsection, i.e. Θ is gamma distributed $\mathcal{G}a(2, 4)$.

4.4.4 Zero-modified distributions

In the previous Subsections, we worked only with a simple geometric distribution. Following the model presented in Subsection 4.2.2, we use hereafter a zero-modified distribution



(a) Continuous case (b) Discrete case

Figure 4.3: Comparison of copula c_{Y_1,Y_2} and interpolated $c^{\Pi}_{W_1,W_2}$

$\mathcal{G}e(q,\rho)$ where the mixing is done over the parameter ρ. We want again to study the associated dependence structure.

Continuous case

Let $Y_i, i = 1, 2$ be conditionnaly independent exponential random variable $Y_i/\Theta = \theta \sim \mathcal{E}(\theta)$ and I_i be independent Bernoulli variable $\mathcal{B}(1-q)$. We define $Z_i = I_i Y_i$. By conditioning on the variable I_i, we get

$$P(Z_i \leq x) = 1 - (1-q)L_\Theta(x),$$

and

$$P(Z_1 \leq x, Z_2 \leq y) = 1 - (1-q)(L_\Theta(x) + L_\Theta(y)) + (1-q)^2 L_\Theta(x+y),$$

for $x, y \geq 0$. Let D_{Z_1,Z_2} be the distribution function of the pair $(F_{Z_1}(Z_1), F_{Z_2}(Z_2))$. We have

$$D_{Z_1,Z_2}(u,v) = \begin{cases} 0 & \text{if } u < q \text{ or } v < q, \\ P(Z_1 \leq l_u, Z_2 \leq l_v) & \text{otherwise.} \end{cases}$$

where for $p \geq q$, $l_p = L_\Theta^{-1}((1-p)/(1-q))$. Hence, we get for $u, v \geq q$,

$$D_{Z_1,Z_2}(u,v) = u + v - 1 + (1-q)^2 L_\Theta(l_u + l_v).$$

As q tends 0, we get back to the Archimedean continuous copula C_{Y_1,Y_2} of the previous subsection. Note that there is a jump when u or v equal q. Indeed, we have $D_{Z_1,Z_2}(q,v) = qv$ and $D_{Z_1,Z_2}(u,v) \geq q^2$ for $u, v \geq q$.

Discrete case

If Y follows an exponential distribution $\mathcal{E}(\theta)$ and I follows a Bernoulli distribution $\mathcal{B}er(1-q)$, then $W = I\lceil Y \rceil$ follows a 0-modified geometric distribution $\mathcal{G}(q, 1-e^{-\theta})$, where $\lceil x \rceil$ is the

ceiling function. Indeed, we have

$$P(W = k) = qP(0 = k) + (1 - q)P(\lceil Y \rceil = k) = q\delta_{k,0} + (1 - q)(1 - \delta_{k,0})e^{-\theta(k-1)}(1 - e^{-\theta}).$$

Furthermore, $P(W \leq x) = q + (1 - q)P(\lfloor Y \rfloor < \lfloor x \rfloor) = q + (1 - q)(1 - e^{-\nu(\lfloor x \rfloor)})$ for all $x \in \mathbb{R}_+$. In particular for $x = k \in \mathbb{N}$, we have $F_W(k) = q + (1 - q)(1 - e^{-\theta k})$.

Let $W_i = I_i\lceil Y_i \rceil$ with I_i's and Y_i be conditionnaly independent exponential random variables, i.e. $Y_i/\Theta = \theta \sim \mathcal{E}(\theta)$. We have

$$P(W_i \leq x) = 1 - (1 - q)L_\Theta(\lfloor x \rfloor),$$

and

$$P(W_1 \leq x, W_2 \leq y) = 1 - (1 - q)(L_\Theta(\lfloor x \rfloor) + L_\Theta(\lfloor y \rfloor)) + (1 - q)^2 L_\Theta(\lfloor x \rfloor + \lfloor y \rfloor).$$

Let D_{W_1,W_2} be the distribution function of the pair $(F_{W_1}(W_1), F_{W_2}(W_2))$. Similarly to the previous subsection, we get

$$D_{W_1,W_2}(u, v) = \begin{cases} 0 & \text{if } u < q \text{ or } v < q \\ P(W_1 \leq l_u, W_2 \leq l_v) & \text{otherwise} \end{cases}.$$

And we get for $u, v \geq q$,

$$D_{W_1,W_2}(u, v) = 1 - (1 - q)(L_\Theta(\lfloor l_u \rfloor) + L_\Theta(\lfloor l_v \rfloor)) + (1 - q)^2 L_\Theta(\lfloor l_u \rfloor + \lfloor l_v \rfloor).$$

There is also a jump when u or v equal q. Indeed, we have $D_{W_1,W_2}(q, v) = q - q(1-q)L_\Theta(\lfloor l_v \rfloor)$ and $D_{W_1,W_2}(u, v) \geq q^2$ for $u, v \geq q$.

Differences between D_{Z_1,Z_2} and D_{W_1,W_2}

Let $u \in \mathrm{Im}(F_{W_1})$ and $v \in \mathrm{Im}(F_{W_2})$, i.e. there exist $n, m \in \mathbb{N}$, such that $u = F_{W_1}(n)$ and $v = F_{W_2}(m)$. Firstly, we have

$$u = F_{W_1}(n) = F_{Z_1}(n + 1) \text{ and } v = F_{W_2}(m) = F_{Z_2}(m + 1).$$

Since we have $l_u = n$ and $l_v = m$, we have

$$D_{Z_1,Z_2}(u, v) = 2 - (1 - q)(L_\Theta(n) + L_\Theta(m)) - 1 + (1 - q)^2 L_\Theta(n + m) = D_{W_1,W_2}(u, v).$$

Now, we focus on the maximal difference between these two functions, after deriving an elementary lemma.

Lemma 4.4.5. *Let f be a completely monotone function, $c > 0$ and $p \in [0, 1]$. The function $f_{c,p} : x \mapsto f(x) - pf(x + c)$ is also completely monotone.*

Proof. As f being completely monotone, $(-1)^n f^{(n)}$ is decreasing. Thus, we have

$$(-1)^n f^{(n)}(x) \geq (-1)^n f^{(n)}(x + c) \geq p(-1)^n f^{(n)}(x + c),$$

as $x < x + c$. Hence, $(-1)^n f_{c,p}^{(n)}(x) \geq 0$, i.e. $f_{c,p}$ is completely monotone. □

Proposition 4.4.6. *Let $\Delta_{i,j}$ be the difference between functions D_{W_1,W_2} and D_{Z_1,Z_2} at the point $u = F_{W_1}(i)$ and $v = F_{W_1}(i)$. The maximal stairstep is $\Delta_{\infty,1} = \Delta_{1,\infty}$.*

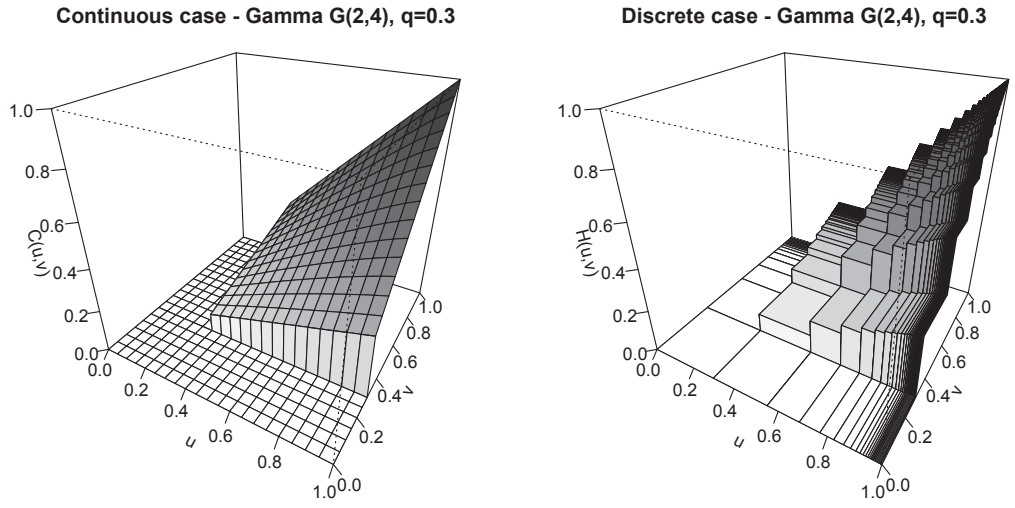
Proof. Since for $u, v \leq q = F_{W_1}(0)$, both functions D_{W_1,W_2} and D_{Z_1,Z_2} are zero, we have

$$\max_{i,j\geq 0} \Delta_{i,j} = \max_{i,j\geq 1} \Delta_{i,j}.$$

Using Lemma 4.4.5, the function $x \mapsto L_\Theta(x) - (1-q)L_\Theta(x+c)$ is convex decreasing. Thus, $x \mapsto L_\Theta(x) - (1-q)L_\Theta(x+c)$ is concave increasing. We get back to the proof of Proposition 4.4.4. So, it follows that the maximal difference is the stairstep $\Delta_{\infty,1}$. □

Graphical comparison

As already said, functions D_{Z_1,Z_2} and D_{W_1,W_2} are not copula but only distribution functions. We plot on Figure 4.4 these two functions when Θ is gamma distributed $\mathcal{G}a(2,4)$ with parameter $q = 0.3$. The appropriate way to deal with non-identifiability issues is again to use the interpolated copula C^{Π}.



(a) Continuous case

(b) Discrete case

Figure 4.4: Comparison of distribution functions D_{Z_1,Z_2} and D_{W_1,W_2}

4.5 Conclusion

This paper uses a new class of dependent risk models, where the dependence is based on a mixing approach. Emphasis has been put on infinite time ruin probability asymptotics. This paper validates the $A + B/u$ rule suggested in Example 2.3 of Albrecher et al. (2011). The asymptotic rule applies both when the claim amounts or the waiting times are correlated. In the ruin literature, even when some dependence is added in the claim arrival process, e.g., a Markovian setting or a specific dependence for the claim severity and waiting time, the decreasing shape of the ruin probability remains unchanged compared to the corresponding independent case, either exponential $e^{-\gamma u}$ or polynomial $u^{-\alpha}$. Hence, our particular mixing approach, leading to $A + B/u$ asymptotics, significantly worsens the situation for the insurer.

A large part of this paper focuses on the discrete time framework. We have seen that discrete time ruin probability asymptotics can be obtained in a similar way as in the continuous case. However, deriving asymptotics for the claim distribution is far more difficult: complex analysis is necessary to tackle the binomial alternating sum issue. Furthermore, the non-uniqueness of discrete copulas is also studied. We quantify the maximal difference between the continuous and the discrete settings. Despite the issues encountered with discrete copula, the latent variable approach is considered in many other articles, e.g., Joe (1997); Frees and Wang (2006); Channouf and L'Ecuyer (2009); Braeken et al. (2007); Leon and Wu (2010).

As mentioned in Albrecher et al. (2011), the approach proposed in this paper to derive new explicit formula can be used for more general risk models. It could be interesting to test whether the $A + B/u$ still applies for the ruin probability, say, with phase-type claim distributions. Beyond the study of ruin probability, the mixing approach and the asymptotic rule might probably be used for finite-time ruin probabilities and the Gerber-Shiu function.

4.6 Appendix

4.6.1 Usual special functions

List of common special functions

Let us recall the definition of the so-called special functions, see Olver et al. (2010) for a comprehensive and updated list. In most cases, the definition does not limit to $z \in \mathbb{R}$, but can be extended to $z \in \mathbb{C}$.

- the gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$,
- the lower incomplete gamma function $\gamma(\alpha, z) = \int_0^z x^{\alpha-1}e^{-x}dx$,
- the upper incomplete gamma function $\Gamma(\alpha, z) = \int_z^\infty x^{\alpha-1}e^{-x}dx$,
- the beta function $\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$,
- the incomplete beta function $\beta(a, b, z) = \int_0^z x^{a-1}(1-x)^{b-1}dx$,
- the complementary incomplete beta function $\bar{\beta}(a, b, z) = \int_z^1 x^{a-1}(1-x)^{b-1}dx = \beta(b, a, 1-z)$,
- the error function $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}dt$,
- the complementary error function $\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{+\infty} e^{-t^2}dt$,
- the binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Known asymptotics of usual special functions

In this subsection, we list asymptotics of the main special functions, see Olver et al. (2010).

Gamma function Known as the Stirling formula, the asymptotic expansion for $\Gamma(z)$ is

$$\Gamma(z) \underset{+\infty}{\sim} e^{-z}z^z\sqrt{\frac{2\pi}{z}}\left(1 + \frac{1}{12z} + \cdots + \frac{g_k}{z^k} + \dots\right),$$

for $z \in \mathbb{C}$ and $|\arg(z)| < \pi$ with $g_0 = 1$, $g_1 = 1/12$, $g_2 = 1/288$, $g_3 = -139/51480$. See page 141 of Olver et al. (2010).

We recall that the incomplete gamma functions are defined as

$$\gamma(a, z) = \int_0^z x^{\alpha-1}e^{-x}dx, \quad \text{and} \quad \Gamma(a, z) = \int_z^\infty x^{\alpha-1}e^{-x}dx.$$

for $z \in \mathbb{C}$.

For large value of z and fixed a, from page 179 of Olver et al. (2010), we have the following asymptotics

$$\Gamma(a,z) \underset{z>>a}{\sim} z^{a-1}e^{-z}\left(1+\frac{a-1}{z}+\cdots+\frac{u_n}{z^n}+\dots\right),$$

where $u_n=(a-1)(a-2)\dots(a-n)$ for $|\arg(z)|<3\pi/2$. For large value of z, no expansion is needed for $\gamma(a,z)$ since we have $\gamma(a,z)\sim\Gamma(a)$.

Beta function Using the Stirling formula, the beta function $\beta(x,y)$ can be approximated for large values x and y fixed. We recall $\beta(x,y)=\Gamma(x)\Gamma(y)/\Gamma(x+y)$. We have

$$\Gamma(x) \underset{x\to+\infty}{\sim} e^{-x}x^x\sqrt{\frac{2\pi}{x}}\sum_{k\geq 0}\frac{g_k}{x^k} \quad \text{and} \quad \Gamma(x+y) \underset{x\to+\infty}{\sim} e^{-x-y}(x+y)^{x+y}\sqrt{\frac{2\pi}{x+y}}\sum_{k\geq 0}\frac{g_k}{x^k},$$

where the term $e^{-x-y}(x+y)^{x+y}\sqrt{\frac{2\pi}{x+y}} \underset{+\infty}{\sim} e^{-x}x^{x+y}\sqrt{\frac{2\pi}{x}}$. Therefore we obtain

$$\beta(x,y) \underset{x\to+\infty}{\sim} \frac{\Gamma(y)}{x^y}\sum_{k\geq 0}\frac{d_k}{x^k},$$

where the coefficients d_k are given by

$$d_0=1 \quad \text{and} \quad d_k=g_k-\sum_{m=0}^{k-1}d_m g_{k-m}.$$

From Olver et al. (2010), page 184, we have asymptotics for the incomplete beta ratio function

$$I_\beta(a,b,x)=\frac{\beta(a,b,x)}{\beta(a,b)} \underset{a\to+\infty}{\sim} \Gamma(a+b)x^a(1-x)^{b-1}\sum_{k\geq 0}\frac{1}{\Gamma(a+k+1)\Gamma(b+k)}\left(\frac{x}{1-x}\right)^k,$$

for large values a and fixed values of $b>0$, $0<x<1$. Multplying by $\beta(a,b)$, we get

$$\beta(a,b,x) \underset{a\to+\infty}{\sim} x^a(1-x)^{b-1}\sum_{k\geq 0}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+k+1)\Gamma(b+k)}\left(\frac{x}{1-x}\right)^k,$$

with $x\leq 1$ and $a\to+\infty$.

Finally, to get the asymptotic of $\bar\beta(a,b,x)$ for large values of b, we use $\bar\beta(a,b,x)=\beta(b,a,1-x)$. We have

$$\bar\beta(a,b,x) \underset{b\to+\infty}{\sim} (1-x)^{b-1}x^a\sum_{k\geq 0}\frac{\Gamma(b)\Gamma(a)}{\Gamma(b+k+1)\Gamma(a+k)}\left(\frac{1-x}{x}\right)^k,$$

The Error function From page 164 of Olver et al. (2010), we have

$$\operatorname{erfc}(z) \sim \frac{e^{-z^2}}{z\sqrt{\pi}}\left(\sum_{k=0}^{N-1}(-1)^k\frac{1.3.5\ldots(2k-1)}{(2z^2)^k}+R_N(z)\right)$$

for $z \to +\infty$ and $|\arg(z)| < 3\pi/4$ We can deduce the asymptotic of the error function at $-\infty$ or $+\infty$ using $\operatorname{erf}(x) = 1 - \operatorname{erfc}(x)$, $\operatorname{erf}(-x) = -\operatorname{erf}(x)$ and $\operatorname{erfc}(-x) = 2 - \operatorname{erfc}(x)$. We get

$$\operatorname{erf}(x) \underset{+\infty}{\sim} 1-\frac{e^{-x^2}}{x\sqrt{\pi}}, \quad \operatorname{erfc}(x) \underset{+\infty}{\sim} \frac{e^{-x^2}}{x\sqrt{\pi}}, \quad \operatorname{erf}(x) \underset{-\infty}{\sim} -1+\frac{e^{-x^2}}{x\sqrt{\pi}}, \quad \operatorname{erfc}(x) \underset{-\infty}{\sim} 2-\frac{e^{-x^2}}{x\sqrt{\pi}}.$$

4.6.2 For the continuous time model

The special case of the Lévy distribution

As for the incomplete gamma function, the function $\Gamma(.,.;.)$ satisfies a recurrence on the a parameter,

$$\Gamma(a+1,x;b) = a\Gamma(a,x;b) + b\Gamma(a-1,x;b) + x^a e^{-x-b/x},$$

see Theorem 2.2 of Chaudry and Zubair (2002). Thus we deduce

$$\Gamma(-3/2,x;b) = \frac{1}{b}\left(\Gamma(1/2,x;b) + 1/2\Gamma(-1/2,x;b) - x^{-1/2}e^{-x-b/x}\right).$$

As reported in Theorem 2.6 and Corollary 2.7 of Chaudry and Zubair (2002), $\Gamma(a,x;b)$ has a simpler expresssion in terms of the error function when $a = 1/2, -1/2, \ldots,$

$$\Gamma(1/2,x;b) = \frac{\sqrt{\pi}}{2}\left[e^{2\sqrt{b}}\operatorname{erfc}\left(x+\frac{\sqrt{b}}{x}\right) + e^{-2\sqrt{b}}\operatorname{erfc}\left(x-\frac{\sqrt{b}}{x}\right)\right]$$

and

$$\Gamma(-1/2,x;b) = \frac{\sqrt{\pi}}{2\sqrt{b}}\left[-e^{2\sqrt{b}}\operatorname{erfc}\left(x+\frac{\sqrt{b}}{x}\right) + e^{-2\sqrt{b}}\operatorname{erfc}\left(x-\frac{\sqrt{b}}{x}\right)\right].$$

Therefore, we have

$$\Gamma(-3/2,x;b) = \frac{\sqrt{\pi}}{2b}\left[\left(1-\frac{1}{2\sqrt{b}}\right)e^{2\sqrt{b}}\operatorname{erfc}(d_+) + \left(1+\frac{1}{2\sqrt{b}}\right)e^{-2\sqrt{b}}\operatorname{erfc}(d_-) - \frac{2}{\sqrt{\pi x}}e^{-x-b/x}\right],$$

with

$$d_+ = \sqrt{x} + \frac{\sqrt{b}}{x} \quad \text{and} \quad d_- = \sqrt{x} - \frac{\sqrt{b}}{x}.$$

It yields

$$\Gamma(-3/2,\theta_0 u;\alpha^2 u/4) = \frac{2\sqrt{\pi}}{\alpha^2 u}\left[\left(1-\frac{1}{\alpha\sqrt{u}}\right)e^{\alpha\sqrt{u}}\operatorname{erfc}(d_+) + \left(1+\frac{1}{\alpha\sqrt{u}}\right)e^{-\alpha\sqrt{u}}\operatorname{erfc}(d_-) - \frac{2}{\sqrt{\pi u\theta_0}}e^{-u\theta_0-\alpha^2/(4\theta_0)}\right],$$

with $d_+ = \sqrt{u\theta_0} + \alpha/(2\sqrt{\theta_0})$ and $d_- = \sqrt{u\theta_0} - \alpha/(2\sqrt{\theta_0})$. We deduce that

$$I(u,\theta_0) = \frac{\theta_0\sqrt{u}}{\alpha} e^{u\theta_0} \left[\left(1 - \frac{1}{\alpha\sqrt{u}}\right) e^{\alpha\sqrt{u}} \text{erfc}\left(d_+\right) \right.$$
$$\left. + \left(1 + \frac{1}{\alpha\sqrt{u}}\right) e^{-\alpha\sqrt{u}} \text{erfc}\left(d_-\right) - \frac{2}{\sqrt{\pi u\theta_0}} e^{-u\theta_0 - \alpha^2/(4\theta_0)} \right].$$

By reordering the terms, we get the formula of Albrecher et al. (2011), which is

$$I(u,\theta_0) = e^{-\alpha c/4\lambda} \left[e^{(c\alpha + 2\lambda\sqrt{u})^2/4\lambda c} \left(-1 + \alpha\sqrt{u}\right) \text{erfc}\left(d_+\right) \right.$$
$$\left. + e^{(c\alpha - 2\lambda\sqrt{u})^2/4\lambda c} \left(1 + \alpha\sqrt{u}\right) \text{erfc}\left(d_-\right) - \frac{2\alpha}{\sqrt{\pi\lambda/c}} \right].$$

4.6.3 For the discrete time model

Geometric distribution

In this subsection, we study the geometric distribution, its properties and its minor extensions.

Classic geometric distribution The geometric distribution $\mathcal{G}(p)$ is characterized by the following probability (mass) function

$$P(X = k) = p(1-p)^k,$$

where $k \in \mathbb{N}$ and $0 \leq p \leq 1$. Note that it takes values in all integers $\mathbb{N}$. Another definition of the geometric distribution is $p(1-p)^{k-1}$ so that the random variable takes values in strictly positive integers. In this case, we can interpret this distribution as the distribution of the first moment we observe a specific event occuring with probability p in a serie of indepedent and identically distributed Bernoulli trials.

The probability generating function is given by

$$G_X(z) = \frac{p}{1-(1-p)z}.$$

With this characterizition, it is straightforward to see that summing n geometric random variates $\mathcal{G}(p)$ has a negative binomial distribution $\mathcal{NB}(n, 1-p)$, see, e.g., Chapter 5 of Johnson et al. (2005). The first two moments can be derived $E(X) = (1-p)/p$ and $Var(X) = (1-p)/p^2$ when $p > 0$. In the following subsection, a net profit condition will force $E(X) < 1$ which is quivalent to $p > 1/2$. Furthermore, we also have $P(X > k) = (1-p)^{k+1}$.

Modified geometric distributions The geometric distribution will be used to model the claim severity. It is very restrictive to model claims by a one-parameter distribution. Thus, we introduce two modified versions of the classic geometric distributions: 0-modified geometric distribution and 0,1-modified geometric distribution.

The principle is simple, we modify respectively the first and the first two probabilities of the probability function. Firstly, we introduce the 0-modified geometric distribution. That is $X \sim \mathcal{G}(q,\rho)$ when

$$P(X = k) = \begin{cases} q & \text{if } k = 0, \\ (1-q)\rho(1-\rho)^{k-1} & \text{otherwise.} \end{cases},$$

when $k \in \mathbb{N}$. Using the Kronecker product δ_{ij}, it can be rewritten as

$$P(X = k) = q\delta_{0,k} + (1-q)(1-\delta_{0,k})\rho(1-\rho)^{k-1}.$$

The expectation and variance are given by $E(X) = (1-q)/\rho$ and $Var(X) = (1-q)(1-\rho+q)/\rho^2$. We also have

$$P(X > k) = (1-q)(1-\rho)^k.$$

We get back to the classic geometric distribution with $\rho = q$.

Secondly, we present the 0,1-modified geometric distribution: $X \sim \mathcal{G}(q, \rho, 1-\alpha)$

$$P(X = k) = \begin{cases} q & \text{if } k = 0, \\ (1-q)\rho & \text{if } k = 1, \\ (1-q)(1-\rho)(1-\alpha)\alpha^{k-2} & \text{otherwise.} \end{cases},$$

which is stricly equivalent to

$$P(X = 0) = q = 1 - p \quad \text{and} \quad P(X = k/X > 0) = \rho\delta_{k,1} + (1-\rho)(1-\alpha)\alpha^{k-2}(1-\delta_{k,1}).$$

The mean and the variance are given by

$$E(X) = p\left(1 + \frac{1-\rho}{1-\alpha}\right) \quad \text{and} \quad Var(X) = q\rho + (1-q)(1-\rho)\frac{3-\alpha}{(1-\alpha)^2} - p^2\left(1 + \frac{1-\rho}{1-\alpha}\right)^2.$$

We get back to the 0-modified geometric distribution with $\alpha = 1 - \rho$.

4.6.4 Error function linked terms

We want to compute the following integral, linked to the error function

$$J(a, b, x) = \int_x^\infty e^{-ay^2 - b/y^2} dy,$$

where $a, b, x > 0$. The SAGE mathematical software (Stein et al. (2011)) suggests to do a change of variable in order to get $\int e^{-t^2} dt$. Since the equation $t^2 = ay^2 + b/y^2$ does not have a unique solution, we consider

$$t = \pm\sqrt{a}y + \sqrt{b}/y.$$

This leads to split the integral $J(a, b, x)$. With algebric manipulations, we get

$$2\sqrt{a}dy = \sqrt{a}dy + \frac{\sqrt{b}}{-y^2}dy + \sqrt{a}dy - \frac{\sqrt{b}}{-y^2}dy.$$

Therefore,

$$2\sqrt{a}J(a, b, x) = e^{2\sqrt{ab}} \int_{\tilde{x}_1}^\infty e^{-t^2} dt + e^{-2\sqrt{ab}} \int_{\tilde{x}_2}^\infty e^{-t^2} dt,$$

with $\tilde{x}_1 = \sqrt{a}x + \frac{\sqrt{b}}{x}$ and $\tilde{x}_2 = \sqrt{a}x - \frac{\sqrt{b}}{x}$. Hence

$$J(a, b, x) = \frac{\sqrt{\pi}}{4\sqrt{a}} \left[e^{2\sqrt{ab}} \text{erfc}\left(\sqrt{a}x + \frac{\sqrt{b}}{x}\right) + e^{-2\sqrt{ab}} \text{erfc}\left(\sqrt{a}x - \frac{\sqrt{b}}{x}\right) \right].$$

This result is in line with Theorem 7 of Chaudry and Zubair (1994) or Theorem 3.1 of Chaudry and Zubair (2002). It is closely related to the generalized error function

$$\text{erfc}(x;b) = \frac{\sqrt{\pi}}{4} e^{2\sqrt{b}} \left[e^{2\sqrt{b}} \text{erfc}\left(x + \frac{\sqrt{b}}{x} \right) + e^{-2\sqrt{b}} \text{erfc}\left(x - \frac{\sqrt{b}}{x} \right) \right].$$

If $x = 0$, we get

$$J(a,b,x) = \frac{\sqrt{\pi}}{2\sqrt{a}} e^{-2\sqrt{ab}} = J(a,b).$$

If $b = -1$, then it is equivalent as changing the occurence of $\sqrt{b}$ by the imaginary number i. We get

$$J(a,-1,x) = \frac{\sqrt{\pi}}{4\sqrt{a}} \left[e^{2i\sqrt{a}} \text{erfc}\left(\sqrt{a}x + \frac{i}{x} \right) + e^{-2i\sqrt{a}} \text{erfc}\left(\sqrt{a}x - \frac{i}{x} \right) \right].$$

This number is of type $z_1 z_2 + \bar{z}_1 \bar{z}_2$, where $z_1 = e^{2i\sqrt{a}}$ and $z_2 = \text{erfc}\left(\sqrt{a}x + \frac{i}{x}\right)$. It is easy to check that

$$z_1 z_2 + \bar{z}_1 \bar{z}_2 = 2|z_1 z_2| \cos(\arg(z_1) + \arg(z_2)) \in \mathbb{R}.$$

Thus, we extend the notation $J(a,b,x)$ for $b = -1$ by the above expression.

4.6.5 Analysis

Integration by parts

We give in this subsection the integration by part theorem for the Lebesgues and the Stieltjes integrals from Gordon (1994)'s graduate book. Other reference books include Hildebrandt (1971); Stroock (1994).

Lebesgues integral

Theorem (Theorem 12.5 of Gordon (1994))**.** *Let f be Lebesgues-integrable on $[a,b]$ and $F(x) = \int_a^x f$ for each $x \in [a,b]$. If G is absolutely continuous on $[a,b]$, then fG is Lebesgues-integrable and we have*

$$\int_a^b fG = F(b)G(b) - \int_a^b FG'.$$

If the integral limits as b tends to infinity exist, one can consider the indefinite integral version

$$\int_a^{+\infty} fG = \lim_{b \to +\infty} F(b)G(b) - \int_a^{+\infty} FG'.$$

Stieltjes integral

Theorem (Theorem 12.14 of Gordon (1994))**.** *Let f, g be bounded functions on a closed interval $[a,b]$. If f is Stieltjes-integrable with respect to g on $[a,b]$, then g is also Stieltjes-integrable with respect to g on $[a,b]$ and we have*

$$\int_a^b g(t)df(t) = [g(t)f(t)]_a^b - \int_a^b f(t)dg(t).$$

If the integral limits as b tends to infinity exist, one can consider the indefinite integral version

$$\int_a^{+\infty} g(t)df(t) = \lim_{b\to+\infty} g(b)f(b) - g(a)f(a) - \int_a^{+\infty} f(t)dg(t).$$

Rademacher theorem

For a proof of this theorem, see e.g. Clarke and Bessis (1999).

Theorem. *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a locally Lipschitz function. Then f is almost everywhere differentiable.*

Bibliography

Albrecher, H. and Asmussen, S. (2006), 'Ruin probabilities and aggregate claims distributions for shot noise Cox processes', *Scandinavian Actuarial Journal* **2006**(2), 86–110.

Albrecher, H. and Boxma, O. (2004), 'A ruin model with dependence between claim sizes and claim intervals', *Insurance: Mathematics and Economics* **35**(2), 245–254.

Albrecher, H., Constantinescu, C. and Loisel, S. (2011), 'Explicit ruin formulas for models with dependence among risks', *Insurance: Mathematics and Economics* **48**(2), 265–270.

Albrecher, H. and Teugels, J. L. (2006), 'Exponential behavior in the presence of dependence in risk theory', *Jounal of Applied Probability* **43**(1), 265–285.

Andersen, S. (1957), 'On the collective theory of risk in case of contagion between claims', *Bulletin of the Institute of Mathematics and its Applications* **12**, 2775–279.

Asmussen, S. and Albrecher, H. (2010), *Ruin Probabilities*, 2nd edition edn, World Scientific Publishing Co. Ltd. London.

Asmussen, S. and Rolski, T. (1991), 'Computational methods in risk theory: A matrix algorithmic approach', *Insurance: Mathematics and Economics* **10**(4), 259–274.

Bak, J. and Newman, D. J. (2010), *Complex Analysis*, 3rd edn, Springer.

Boudreault, M., Cossette, H., Landriault, D. and Marceau, E. (2006), 'On a risk model with dependence between interclaim arrivals and claim sizes', *Scandinavian Actuarial Journal* **2006**(5), 265–285.

Braeken, J., Tuerlinckx, F. and De Boeck, P. (2007), 'Copula functions for residual dependency', *Psychometrika* **72**(3), 393–411.

Cai, J. and Li, H. (2005), 'Multivariate risk model of phase type', *Insurance: Mathematics and Economics* **36**(2), 137–152.

Centeno, M. d. L. (2002*a*), 'Excess of loss reinsurance and Gerber's inequality in the Sparre Anderson model', *Insurance: Mathematics and Economics* **31**(3), 415–427.

Centeno, M. d. L. (2002*b*), 'Measuring the effects of reinsurance by the adjustment coefficient in the Sparre Anderson model', *Insurance: Mathematics and Economics* **30**(1), 37–49.

Channouf, N. and L'Ecuyer, P. (2009), Fitting a normal copula for a multivariate distribution with both discrete and continuous marginals, *in* 'Proceedings of the 2009 Winter Simulation Conference'.

Chaudry, M. A. and Zubair, S. M. (1994), 'Generalized incomplete gamma functions with applications', *Journal of Computational and Applied Mathematics* **55**(1), 99–123.

Chaudry, M. A. and Zubair, S. M. (2002), *On a Class of Incomplete Gamma Functions with Applications*, Chapman & Hall.

Clarke, F. H. and Bessis, D. N. (1999), 'Partial subdifferentials, derivates and Rademacher's theorem', *Transactions of the American Mathematical Society* **351**(7), 2899–2926.

Collamore, J. (1996), 'Hitting probabilities and large deviations', *The Annals of Probability* **24**(4), 2065–2078.

Constantinescu, C., Hashorva, E. and Ji, L. (2011), 'Archimedean copulas in finite and infinite dimensions—with application to ruin problems', *Insurance: Mathematics and Economics* **49**(3), 487–495.

Denuit, M. and Lambert, P. (2005), 'Constraints on concordance measures in bivariate discrete data', *Journal of Multivariate Analysis* **93**, 40–57.

Dutang, C., Lefèvre, C. and Loisel, S. (2012), A new asymptotic rule for the ultimate ruin probability. Working paper, ISFA.

Embrechts, P. and Veraverbeke, N. (1982), 'Estimates for the probability of ruin with special emphasis on the possibility of large claims', *Insurance: Mathematics and Economics* **1**(1), 55–72.

Flajolet, P. and Sedgewick, R. (1995), 'Mellin transforms and asymptotics: Finite differences and Rice's integrals', *Theoretical Computer Science* **144**(1-2), 101–124.

Fousse, L., Hanrot, G., Lefèvre, V., Pélissier, P. and Zimmermann, P. (2007), 'MPFR: A multiple-precision binary floating-point library with correct rounding', *ACM Trans. Math. Softw.* **33**(2).

Fousse, L., Hanrot, G., Lefèvre, V., Pélissier, P. and Zimmermann, P. (2011), *MPFR: A multiple-precision binary floating-point library with correct rounding.*
URL: *http://mpfr.org/*

Frees, E. W. and Wang, P. (2006), 'Copula credibility for aggregate loss models', *Insurance: Mathematics and Economics* **38**, 360–373.

Genest, C. and Neslehova, J. (2007), 'A primer on copulas for count data', *ASTIN Bulletin* .

Gerber, H. U. (1988), 'Mathematical fun with compound binomial process', *ASTIN Bulletin* **18**(2), 161–168.

Gerber, H. U. and Shiu, E. S. (1998), 'On the time value of ruin', *North American Actuarial Journal* **2**(1), 48–78.

Gerber, H. U. and Shiu, E. S. (2005), 'The time value of ruin in a Sparre Andersen model', *North American Actuarial Journal* **9**(2), 49–84.

Gordon, R. A. (1994), *The Integrals of Lebesgue, Denjoy, Perron and Henstock*, Vol. 4, American Mathematical Society.

Grandlund Torbjoern & the GMP Devel. Team (2011), *GNU MP - The GNU Multiple Precision Arithmetic Library.*
URL: *http://gmplib.org/*

Heil, C. (2007), 'Real analysis lecture notes: absolute continuous and singular functions', Lecture Notes, School of Mathematics, Georgia Institute of Technology.

Hildebrandt, T. (1971), *Introduction to the Theory of Integration*, Routledge.

Huang, H.-N., Marcantognini, S. and Young, N. (2006), 'Chain rules for higher derivatives', *The Mathematical Intelligencer* **28**(2), 61–69.

Jeffrey, A. and Dai, H.-H. (2008), *Handbook of Mathematical Formulas and Integrals*, Academic Press.

Joe, H. (1997), Multivariate dependence measure and data analysis, *in* 'Monographs on Statistics and Applied Probability', Vol. 73, Chapman & Hall.

Johnson, N. L., Kotz, S. and Kemp, A. W. (2005), *Univariate Discrete Distributions*, 3rd edn, Wiley Interscience.

Jones, D. S. (1997), *Introduction to Asymptotics: a Treatment using Nonstandard Analysis*, World Scientific.

Klueppelberg, C. and Stadtmueller, U. (1998), 'Ruin probabilities in the presence of heavy-tails and interest rates', *Scandinavian Actuarial Journal* **1998**(1), 49–58.

Leon, A. R. D. and Wu, B. (2010), 'Copula-based regression models for a bivariate mixed discrete and continuous outcome', *Statistics in Medicine* .

Li, S., Lu, Y. and Garrido, J. (2009), 'A review of discrete-time risk models', *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas* **103**(2), 321–337.

Lu, Y. and Garrido, J. (2005), 'Doubly periodic non-homogeneous Poisson models for hurricane data', *Statistical Methodology* **2**(1), 17–35.

Maechler, M. (2012), *Rmpfr: R MPFR - Multiple Precision Floating-Point Reliable*, ETH Zurich.

Merentes, N. (1991), 'On the Composition Operator in AC[a,b]', *Collect. Math.* **42**(3), 237–243.

Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F. and Clark, C. W., eds (2010), *NIST Handbook of Mathematical Functions*, Cambridge University Press.
URL: *http://dlmf.nist.gov/*

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org*

Shiu, E. S. W. (1989), 'The probability of eventual ruin in the compound binomial model', *ASTIN Bulletin* **19**(2), 179–190.

Silvia, E. (1999), 'Companion Notes for Advanced Calculus', Lecture Notes, University of California.

Simon, H. A. (1955), 'On a class of skew distribution functions', *Biometrika* **42**(3/4), 425–440.

Song, M., Meng, Q., Wu, R. and Ren, J. (2010), 'The Gerber-Shiu discounted penalty function in the risk process with phase-type interclaim times', *Applied Mathematics and Computation* **216**(2), 523–531.

Stein, W. A. et al. (2011), *Sage Mathematics Software (Version 4.6.2)*, The Sage Development Team.

Stroock, D. (1994), *A Concise Introduction to the Theory of Integration*, Birkhauser.

Sundt, B. and dos Reis, A. D. E. (2007), 'Cramér-Lundberg results for the infinite time ruin probability in the compound binomial model', *Bulletin of the Swiss Association of Actuaries* **2**.

Willmot, G. E. (1993), 'Ruin probabilities in the compound binomial model', *Insurance: Mathematics and Economics* **12**(2), 133–142.

Conclusion

Conclusion et perspectives

Dans cette thèse, nous nous sommes attachés au problème de modélisation des différentes composantes d'un marché d'assurance non-vie. Nous avons apporté de nouvelles contributions aux thèmes de la modélisation des résiliations, des primes et de la probabilité de ruine.

Tout d'abord, nous avons considéré des modèles de résiliation dans le chapitre 1. Les modèles statistiques de régression permettent de mesurer, à un instant donné, l'impact des évolutions de prix sur la résiliation des contrats d'assurance. Cependant, le chapitre souligne les déficiences possibles de ces modèles, si l'on ne possède pas les données appropriées. Le rôle capital dans la calibration d'une estimation de la prime marché (par police) est établi. Même dans le cas où les estimations des taux de résiliation sont considérés comme fiables, l'approche par régression ne permet pas d'expliquer complètement les interactions entre les assurés et les assureurs sur un marché.

Basé sur ce constat, un jeu non-coopératif est proposé dans le chapitre 2 pour modéliser le marché d'assurance dans sa globalité. Partant d'une vision sur une période, le modèle est composé de deux niveaux d'agents. D'un côté, les assurés réagissent aux fluctuations de prix par un modèle multinomial logit, et de l'autre les assureurs maximisent un critère de profitabilité approché sous contraintes de solvabilité. Nous démontrons l'existence et l'unicité de l'équilibre de Nash et sa sensibilité locale aux paramètres. Une version plus complexe du jeu où l'on prend mieux en compte la taille de portefeuille espérée par un assureur dans ses fonctions objective et contrainte est présentée. Mais l'unicité de l'équilibre de Nash est perdue et pose problème dans son application. La répétition du jeu simple sur plusieurs périodes met en évidence une corrélation sur la prime moyenne marché. Cette approche apporte un nouveau point de vue sur la modélisation des cycles de marché. L'évaluation des probabilités de ruine et de domination est rendue possible par une méthode de Monte-Carlo. Les perspectives pour ce chapitre sont assez nombreuses, nous en donnons deux des plus évidentes. En pratique, les assureurs différencient leur prime d'assurance en fonction du profil de risque de l'assuré. Dans notre jeu, nous supposons que tous les assurés ont le même profil de risque. Dans un premier temps, il serait intéressant de prendre en compte deux classes d'assurés dans le jeu d'assurance. Une seconde extension tout aussi pertinante serait de rajouter des sinistres catastrophes naturelles et des réassureurs. Ce troisième type d'agents permettrait de se rapprocher de la réalité des marchés d'assurance.

Le calcul de prime d'équilibre étant rendu nécessaire, le chapitre 3 présente en détails les méthodes d'optimisation les plus avancées permettant de résoudre les équilibres de Nash généralisé. Les méthodes d'optimisation étudiées réposent sur une reformulation des équations de Karush-Kuhn-Tucker (KKT) du problème d'équilibre de Nash. Elles permettent d'élargir le cadre scolaire des jeux simples à deux joueurs aux jeux généralisés à plusieurs joueurs. Un complément souhaitable serait de fournir un même panorama pour les jeux conjointement convexes pour lesquelles d'autres reformulations que la reformulation KKT peuvent être utilisés.

Enfin, le chapitre 4 s'intéresse à un tout autre point de vue du marché de l'assurance en étudiant la probabilité de ruine d'un assureur en temps infini. Dans un modèle de risque avec dépendance entre les montants de sinistre ou les temps d'attente, nous proposons une nouvelle formule asymptotique de la probabilité de ruine en temps continu et discret. La dépendance entre sinistres, introduite par une variable aléatoire mélange, permet des formules fermées de la probabilité de ruine ultime dans quelques cas particuliers. Mais surtout, une nouvelle forme d'asymptotique en $A + B/u$ est démontrée et est à comparer aux décroissances connues, $e^{-\gamma u}$ ou $1/u^{\alpha}$, pour les sinistres à queues de distribution légères ou lourdes, respectivement. En dernier lieu, ce chapitre étudie les problèmes liés à l'utilisation des copules pour les variables aléatoires discrètes. Une quantification de l'écart maximal entre les versions continue et discrète du modèle est réalisée. Comme souligné dans Albrecher *et al.* (2011), l'approche par mélange utilisé dans ce chapitre peut être utilisée pour des modèles de risque plus avancés que le modèle de Cramér-Lundberg. Il serait intéressant de voir si une formule asymptotique de la probabilité de ruine de ce type peut toujours être obtenue pour d'autres classes de modèles, par exemple, les modèles phase-type.

Bibliographie

Aalen, O., Borgan, O. et Gjessing, H. (2008). *Survival and Event History Analysis*. Springer.

Albrecher, H. et Asmussen, S. (2006). Ruin probabilities and aggregate claims distributions for shot noise Cox processes. *Scandinavian Actuarial Journal*, 2006(2):86–110.

Albrecher, H. et Boxma, O. (2004). A ruin model with dependence between claim sizes and claim intervals. *Insurance : Mathematics and Economics*, 35(2):245–254.

Albrecher, H., Constantinescu, C. et Loisel, S. (2011). Explicit ruin formulas for models with dependence among risks. *Insurance : Mathematics and Economics*, 48(2):265–270.

Albrecher, H. et Teugels, J. L. (2006). Exponential behavior in the presence of dependence in risk theory. *Jounal of Applied Probability*, 43(1):265–285.

Alesina, A. (1987). Macroeconomic Policy in a Two-Party System as a Repeated Game. *Quartely Journal of Economics*, 102(3):651–678.

Allgower, E. L. et Georg, K. (2003). *Introduction to Numerical Continuation Methods*. SIAM.

Andersen, P., Borgan, O., Gill, R. et Keiding, N. (1995). *Statistical Models Based on Counting Processes*. Springer, Corrected Edition.

Andersen, S. (1957). On the collective theory of risk in case of contagion between claims. *Bulletin of the Institute of Mathematics and its Applications*, 12:2775–279.

Anderson, S. P., Palma, A. D. et Thisse, J.-F. (1989). Demand for differentiated products, discrete choice models, and the characteristics approach. *The Review of Economic Studies*, 56(1):21–35.

Ania, A., Troeger, T. et Wambach, A. (2002). An evolutionary analysis of insurance markets with adverse selection. *Games and economic behavior*, 40(2):153–184.

Arrow, K. J. et Enthoven, A. C. (1961). Quasiconcave programming. *Econometrica*, 29(4):779–800.

Asmussen, S. (1989). Risk theory in a Markovian environment. *Scandinavian Actuarial Journal*, 1989(2):69–100.

Asmussen, S. (2000). *Ruin Probabilities*. World Scientific.

Asmussen, S. et Albrecher, H. (2010). *Ruin Probabilities*. World Scientific Publishing Co. Ltd. London, 2nd edition édition.

Asmussen, S., Henriksen, L. et Klueppelberg, C. (1994). Large claims approximations for risk processes in a Markovian environment. *Stochastic Processes and their Applications*, 54(1):29–43.

Asmussen, S. et Rolski, T. (1991). Computational methods in risk theory : A matrix algorithmic approach. *Insurance : Mathematics and Economics*, 10(4):259–274.

Atkins, D. C. et Gallop, R. J. (2007). Re-thinking how family researchers model infrequent outcomes : A tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, 21(4):726–735.

Aubin, J.-P. (1998). *Optima and Equilibria : An Introduction to Nonlinear Analysis*. Springer.

Aubin, J.-P. et Frankowska, H. (1990). *Set-valued Analysis*. Birkhauser, Boston.

Bak, J. et Newman, D. J. (2010). *Complex Analysis*. Springer, 3rd édition.

Barges, M., Loisel, S. et Venel, X. (2012). On finite-time ruin probabilities with reinsurance cycles influenced by large claims. *Scandinavian Actuarial Journal*, pages 1–23.

Basar, T. et Olsder, G. J. (1999). *Dynamic Noncooperative Game Theory*. SIAM.

Bazaraa, M. S., Sherali, H. D. et Shetty, C. M. (2006). *Nonlinear Programming : Theory and Algorithms*. Wiley interscience.

Bella, M. et Barone, G. (2004). Price-elasticity based on customer segmentation in the italian auto insurance market. *Journal of Targeting, Measurement and Analysis for Marketing*, 13(1):21–31.

Berge, C. (1963). *Topological Spaces, Including a Treatment of Multi-valued Functions Vector Spaces and Convexity*. Oliver and Boyd LTD.

Bertrand, J. (1883). Théorie mathématique de la richesse sociale. *Journal des Savants*.

Biard, R., Lefèvre, C. et Loisel, S. (2008). Impact of correlation crises in risk theory : Asymptotics of finite-time ruin probabilities for heavy-tailed claim amounts when some independence and stationarity assumptions are relaxed. *Insurance : Mathematics and Economics*, 43(3):412–421.

Biard, R., Loisel, S., Macci, C. et Veraverbeke, N. (2010). Asymptotic behavior of the finite-time expected time-integrated negative part of some risk processes and optimal reserve allocation. *Journal of Mathematical Analysis and Applications*, 367(2):535 – 549.

Bland, R., Carter, T., Coughlan, D., Kelsey, R., Anderson, D., Cooper, S. et Jones, S. (1997). Workshop - customer selection and retention. *In General Insurance Convention & ASTIN Colloquium*.

Blondeau, C. (2001). Les déterminants du cycle de l'assurance de dommages en france. *In Actes de XIXe Journées Internationales d'Economie*.

Bompard, E., Carpaneto, E., Ciwei, G., Napoli, R., Benini, M., Gallanti, M. et Migliavacca, G. (2008). A game theory simulator for assessing the performances of competitive electricity markets. *Electric Power Systems Research*, 78(2):217–227.

Bonnans, J. F., Gilbert, J. C., Lemaréchal, C. et Sagastizábal, C. A. (2006). *Numerical Optimization : Theoretical and Practical Aspects, Second edition*. Springer-Verlag.

Borch, K. (1960). Reciprocal reinsurance treaties seen as a two-person cooperative game. *Skandinavisk Aktuarietidskrift*, 1960(1-2):29–58.

Borch, K. (1975). Optimal insurance arrangements. *ASTIN Bulletin*, 8(3):284–290.

Boudreault, M., Cossette, H., Landriault, D. et Marceau, E. (2006). On a risk model with dependence between interclaim arrivals and claim sizes. *Scandinavian Actuarial Journal*, 2006(5):265–285.

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A. et Nesbitt, C. J. (1997). *Actuarial Mathematics*. The Society of Actuaries.

Braeken, J., Tuerlinckx, F. et De Boeck, P. (2007). Copula functions for residual dependency. *Psychometrika*, 72(3):393–411.

Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J. et Perez-Marin, A. M. (2008). Survival analysis of a household portfolio insurance policies : How much time do you have to stop total customer defection ? *Journal of Risk and Insurance*, 75(3):713–737.

Brouwer, L. E. J. (1912). Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen*, 71(1):97–115.

Broyden, C. G. (1965). A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19(92):577–593.

Bühlmann, H. (1984). The general economic premium principle. *ASTIN Bulletin*, 14(1):13–21.

Cai, J. et Li, H. (2005). Multivariate risk model of phase type. *Insurance : Mathematics and Economics*, 36(2):137–152.

Centeno, M. d. L. (2002a). Excess of loss reinsurance and Gerber's inequality in the Sparre Anderson model. *Insurance : Mathematics and Economics*, 31(3):415–427.

Centeno, M. d. L. (2002b). Measuring the effects of reinsurance by the adjustment coefficient in the Sparre Anderson model. *Insurance : Mathematics and Economics*, 30(1):37–49.

Channouf, N. et L'Ecuyer, P. (2009). Fitting a normal copula for a multivariate distribution with both discrete and continuous marginals. *In Proceedings of the 2009 Winter Simulation Conference*.

Chaudry, M. A. et Zubair, S. M. (1994). Generalized incomplete gamma functions with applications. *Journal of Computational and Applied Mathematics*, 55(1):99–123.

Chaudry, M. A. et Zubair, S. M. (2002). *On a Class of Incomplete Gamma Functions with Applications*. Chapman & Hall.

Chiappori, P.-A. et Salanié, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, 108(1):56–78.

Clark, D. R. et Thayer, C. A. (2004). A primer on the exponential family of distributions. *2004 call paper program on generalized linear models*.

Clarke, F. H. (1975). Generalized gradients and applications. *Transactions of the American Mathematical Society*, 205(1):247–262.

Clarke, F. H. (1990). *Optimization and Nonsmooth Analysis*. SIAM.

Clarke, F. H. et Bessis, D. N. (1999). Partial subdifferentials, derivates and Rademacher's theorem. *Transactions of the American Mathematical Society*, 351(7):2899–2926.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 368(829-836).

Collamore, J. (1996). Hitting probabilities and large deviations. *The Annals of Probability*, 24(4):2065–2078.

Committee, N. P. (2003). Time-series Econometrics : Cointegration and Autoregressive Conditional Heteroskedasticity. Nobel Prize in Economics documents, Nobel Prize Committee.

Constantinescu, C., Hashorva, E. et Ji, L. (2011). Archimedean copulas in finite and infinite dimensions—with application to ruin problems. *Insurance : Mathematics and Economics*, 49(3):487–495.

Cossette, H., Landriault, D. et Marceau, E. (2004). Compound binomial risk model in a Markovian environment. *Insurance : Mathematics and Economics*, 35(2):425–443.

Cossette, H. et Marceau, E. (2000). The discrete-time risk model with correlated classes of business. *Insurance : Mathematics and Economics*, 26(2-3):133–149.

Cournot, A. (1838). *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. Paris : Hachette.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society : Series B*, 34(2):187–200.

Cramér, H. (1930). *On the Mathematical Theory of Risk,*. Skandia Jubilee Volume, Stockholm.

Cummins, J. D. et Outreville, J. F. (1987). An international analysis of underwriting cycles in property-liability insurance. *The Journal of Risk Finance*, 54(2):246–262.

Cummins, J. D. et Venard, B. (2007). *Handbook of international insurance*. Springer.

Dardanoni, V. et Donni, P. L. (2008). Testing for asymmetric information in insurance markets with unobservable types. HEDG working paper.

Demgne, E. J. (2010). Etude des cycles de réassurance. Mémoire de D.E.A., ENSAE.

Dennis, J. E. et Morée, J. J. (1977). Quasi-newton methods, motivation and theory. *SIAM Review*, 19(1).

Dennis, J. E. et Schnabel, R. B. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM.

Denuit, M. et Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93:40–57.

Derien, A. (2010). *Solvabilité 2 : une réelle avancée ?* Thèse de doctorat, ISFA.

Dickson, D. (1992). On the distribution of surplus prior to ruin. *Insurance : Mathematics and Economics*, 11(3):191–207.

Diewert, W. E., Avriel, M. et Zang, I. (1981). Nine kinds of quasiconcavity and concavity. *Journal of Economic Theory*, 25(3).

Dionne, G., Gouriéroux, C. et Vanasse, C. (2001). Testing for evidence of adverse selection in the automobile insurance market : A comment. *Journal of Political Economy*, 109(2):444–453.

Doherty, N. et Garven, J. (1995). Insurance cycles : Interest rates and the capacty constraints model. *Journal of Business*, 68(3):383–404.

Dreves, A., Facchinei, F., Kanzow, C. et Sagratella, S. (2011). On the solutions of the KKT conditions of generalized Nash equilibrium problems. *SIAM Journal on Optimization*, 21(3):1082–1108.

Dreyer, V. (2000). Study the profitability of a customer. Mémoire de D.E.A., ULP - magistère d'actuariat. Confidential memoir - AXA Insurance U.K.

Dutang, C. (2011). Regression models of price elasticity in non-life insurance. Mémoire de D.E.A., ISFA.

Dutang, C. (2012a). A survey of GNE computation methods : theory and algorithms. Working paper, ISFA.

Dutang, C. (2012b). The customer, the insurer and the market. Working paper, ISFA.

Dutang, C. (2012c). Fixed-point-based theorems to show the existence of generalized Nash equilibria. Working paper, ISFA.

Dutang, C. (2012d). *GNE : computation of Generalized Nash Equilibria*. R package version 0.9.

Dutang, C., Albrecher, H. et Loisel, S. (2012a). A game to model non-life insurance market cycles. Working paper, ISFA.

Dutang, C., Lefèvre, C. et Loisel, S. (2012b). A new asymptotic rule for the ultimate ruin probability. Working paper, ISFA.

Edgeworth, F. (1881). *Mathematical Psychics : An Essay on the Application of Mathematics to the Moral Sciences*. London : Kegan Paul.

Embrechts, P. et Veraverbeke, N. (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance : Mathematics and Economics*, 1(1):55–72.

Emms, P., Haberman, S. et Savoulli, I. (2007). Optimal strategies for pricing general insurance. *Insurance : Mathematics and Economics*, 40(1):15–34.

Facchinei, F., Fischer, A. et Piccialli, V. (2007). On generalized Nash games and variational inequalities. *Operations Research Letters*, 35(2):159–164.

Facchinei, F., Fischer, A. et Piccialli, V. (2009). Generalized Nash equilibrium problems and Newton methods. *Math. Program., Ser. B*, 117(1-2):163–194.

Facchinei, F. et Kanzow, C. (1997). A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. *Mathematical Programming*, 76(3):493–512.

Facchinei, F. et Kanzow, C. (2009). Generalized Nash equilibrium problems. Updated version of the 'quaterly journal of operations research' version.

Facchinei, F. et Pang, J.-S. (2003). *Finite-Dimensional Variational Inequalities and Complementary Problems. Volume II.* Springer-Verlag New York, Inc.

Fahrmeir, L. (1994). Dynamic modelling and penalized likelihood estimation for discrete time survival data. *Biometrika*, 81(2):317–330.

Fan, J.-Y. (2003). A modified Levenberg-Marquardt algorithm for singular system of nonlinear equations. *Journal of Computational Mathematics*, 21(5):625–636.

Fan, J.-Y. et Yuan, Y.-X. (2005). On the quadratic convergence of the Levenberg-Marquardt Method without nonsingularity assumption. *Computing*, 74(1):23–39.

Faraway, J. J. (2006). *Extending the Linear Model with R : Generalized Linear, Mixed Effects and Parametric Regression Models.* CRC Taylor& Francis.

Feldblum, S. (2001). Underwriting cycles and business strategies. *In CAS proceedings.*

Fields, J. A. et Venezian, E. C. (1989). Interest rates and profit cycles : A disaggregated approach. *Journal of Risk and Insurance*, 56(2):312–319.

Fischer, A. (2002). Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Math. Program., Ser. A*, 94(1):91–124.

Flajolet, P. et Sedgewick, R. (1995). Mellin transforms and asymptotics : Finite differences and Rice's integrals. *Theoretical Computer Science*, 144(1-2):101–124.

Fousse, L., Hanrot, G., Lefèvre, V., Pélissier, P. et Zimmermann, P. (2007). MPFR : A multiple-precision binary floating-point library with correct rounding. *ACM Trans. Math. Softw.*, 33(2).

Fousse, L., Hanrot, G., Lefèvre, V., Pélissier, P. et Zimmermann, P. (2011). *MPFR : A multiple-precision binary floating-point library with correct rounding.*

Frees, E. W. (2004). *Longitudinal and Panel Data.* Cambridge University Press.

Frees, E. W. et Wang, P. (2006). Copula credibility for aggregate loss models. *Insurance : Mathematics and Economics*, 38:360–373.

Fudenberg, D. et Tirole, J. (1991). *Game Theory.* The MIT Press.

Fukushima, M. et Pang, J.-S. (2005). Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Comput. Manag. Sci.*, 2:21–56.

Fukushima, M. et Qi, L., éditeurs (1999). *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*. Kluwer Academic Publishers.

Ganji, A., Khalili, D. et Karamouz, M. (2007). Development of stochastic dynamic Nash game model for reservoir operation. I. The symmetric stochastic model with perfect information. *Advances in Water Resources*, 30(3):528–542.

Genc, T. S. et Sen, S. (2008). An analysis of capacity and price trajectories for the Ontario electricity market using dynamic Nash equilibrium under uncertainty. *Energy Economics*, 30(1):173–191.

Genest, C. et Neslehova, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*.

Geoffard, P. Y., Chiappori, P.-A. et Durand, F. (1998). Moral hazard and the demand for physician services : First lessons from a French natural experiment. *European Economic Review*, 42(3-5):499–511.

Gerber, H. U. (1988). Mathematical fun with compound binomial process. *ASTIN Bulletin*, 18(2):161–168.

Gerber, H. U. et Shiu, E. S. (1998). On the time value of ruin. *North American Actuarial Journal*, 2(1):48–78.

Gerber, H. U. et Shiu, E. S. (2005). The time value of ruin in a Sparre Andersen model. *North American Actuarial Journal*, 9(2):49–84.

Glicksberg, I. (1950). Minimax theorem for upper and lower semicontinuous payoffs. *Rand Corporation Research Memorandum RM-478, Santa Monica, CA*.

Golubin, A. Y. (2006). Pareto-Optimal Insurance Policies in the Models with a Premium Based on the Actuarial Value. *Journal of Risk and Insurance*, 73(3):469–487.

Gordon, R. A. (1994). *The Integrals of Lebesgue, Denjoy, Perron and Henstock*, volume 4. American Mathematical Society.

Gourieroux, C. et Monfort, A. (1997). *Time Series and Dynamic Models*. Cambridge University Press.

Grace, M. et Hotchkiss, J. (1995). External impacts on the Property-Liability insurance cycle. *Journal of Risk and Insurance*, 62(4):738–754.

Grandell, J. (1991). *Aspects of Risk Theory*. Springer Series in Statistics : Probability and its Applications.

Grandlund Torbjoern & the GMP Devel. Team (2011). *GNU MP - The GNU Multiple Precision Arithmetic Library*.

Granger, C. W. et Engle, R. F. (1987). Co-integration and error-correction : Representation, Estimation and Testing. *Econometrica*, 55(2):251–276.

Gron, A. (1994a). Capacity constraints and cycles in property-casualty insurance markets. *RAND Journal of Economics*, 25(1).

Gron, A. (1994b). Capacity constraints and cycles in property-casualty insurance markets. *RAND Journal of Economics*, 25(1).

Guillen, M., Parner, J., Densgsoe, C. et Perez-Marin, A. M. (2003). *Using Logistic Regression Models to Predict and Understand Why Customers Leave an Insurance Company*, chapitre 13. Volume 6 de Shapiro et Jain (2003).

Haley, J. (1993). A Cointegration Analysis of the Relationship Between Underwriting Margins and Interest Rates : 1930-1989. *Journal of Risk and Insurance*, 60(3):480–493.

Hamel, S. (2007). Prédiction de l'acte de résiliation de l'assuré et optimisation de la performance en assurance automobile particulier. Mémoire de D.E.A., ENSAE. Mémoire confidentiel - AXA France.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

Hardelin, J. et de Forge, S. L. (2009). Raising capital in an insurance oligopoly market. Working paper.

Harker, P. T. (1991). Generalized Nash games and quasi-variational inequalities. *European Journal of Operational Research*, 54(1):81–94.

Hastie, T. J. et Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall.

Hastie, T. J. et Tibshirani, R. J. (1995). Generalized additive models. *to appear in Encyclopedia of Statistical Sciences*.

Haurie, A. et Viguier, L. (2002). A Stochastic Dynamic Game of Carbon Emissions Trading. working paper, University of Geneva.

Heil, C. (2007). Real analysis lecture notes : absolute continuous and singular functions. Lecture Notes, School of Mathematics, Georgia Institute of Technology.

Hildebrandt, T. (1971). *Introduction to the Theory of Integration*. Routledge.

Hipp, C. (2005). Phasetype distributions and its application in insurance.

Hogan, W. W. (1973). Point-to-set maps in mathematical programming. *SIAM Review*, 15(3):591–603.

Huang, H.-N., Marcantognini, S. et Young, N. (2006). Chain rules for higher derivatives. *The Mathematical Intelligencer*, 28(2):61–69.

Ichiishi, T. (1983). *Game Theory for Economic Analysis*. Academic press.

Ip, C. et Kyparisis, J. (1992). Local convergence of quasi-Newton methods for B-differentiable equations. *Mathematical Programming*, 56(1-3):71–89.

Jablonowski, M. (1985). Earnings cycles in property/casualty insurance : A behavioral theory. *CPCU Journal*, 38(3):143–150.

Jackman, S. (2011). *pscl : Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University*. Department of Political Science, Stanford University. R package version 1.04.1.

Jeffrey, A. et Dai, H.-H. (2008). *Handbook of Mathematical Formulas and Integrals*. Academic Press.

Jeyakumar, V. (1998). Simple Characterizations of Superlinear Convergence for Semismooth Equations via Approximate Jacobians. Rapport technique, School of Mathematics, University of New South Wales.

Jiang, H. (1999). Global convergence analysis of the generalized Newton and Gauss-Newton methods for the Fischer-Burmeister equation for the complementarity problem. *Mathematics of Operations Research*, 24(3):529–543.

Jiang, H., Fukushima, M., Qi, L. et Sun, D. (1998). A Trust Region Method for Solving Generalized Complementarity Problem. *SIAM Journal on Optimization*, 8(1).

Jiang, H., Qi, L., Chen, X. et Sun, D. (1996). Semismoothness and superlinear convergence in nonsmooth optimization and nonsmooth equations. *In Nonlinear Optimization and Applications*. Plenum Press.

Jiang, H. et Ralph, D. (1998). Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares. *In* Fukushima, M. et Qi, L., éditeurs : *Reformulation - nonsmooth, piecewise smooth, semismooth and smoothing methods*. Boston MA : Kluwer Academic Publishers.

Joe, H. (1997). Multivariate dependence measure and data analysis. *In Monographs on Statistics and Applied Probability*, volume 73. Chapman & Hall.

Johnson, N. L., Kotz, S. et Kemp, A. W. (2005). *Univariate Discrete Distributions*. Wiley Interscience, 3rd édition.

Jones, D. S. (1997). *Introduction to Asymptotics : a Treatment using Nonstandard Analysis*. World Scientific.

Kagraoka, Y. (2005). Modeling insurance surrenders by the negative binomial model. Working Paper 2005.

Kakutani, S. (1941). A generalization of Brouwer's fixed point theorem. *Duke Math. J.*, 8(3):457–459.

Kanzow, C. et Kleinmichel, H. (1998). A new class of semismooth Newton-type methods for nonlinear complementarity problems. *Computational Optimization and Applications*, 11(3):227–251.

Kelsey, R., Anderson, D., Beauchamp, R., Black, S., Bland, R., Klauke, P. et Senator, I. (1998). Workshop - price/demand elasticity. *In General Insurance Convention & ASTIN Colloquium*.

Kim, C. (2005). Modeling surrender and lapse rates with economic variables. *North American Actuarial Journal*, 9(4):56–70.

Kliger, D. et Levikson, B. (1998). Pricing insurance contracts - an economic viewpoint. *Insurance : Mathematics and Economics*, 22(3):243–249.

Klueppelberg, C. et Stadtmueller, U. (1998). Ruin probabilities in the presence of heavy-tails and interest rates. *Scandinavian Actuarial Journal*, 1998(1):49–58.

Krawczyk, J. et Tidball, M. (2005). A Discrete-Time Dynamic Game of Seasonal Water Allocation. *Journal of Optimization Theory and Applications*, 128(2):441–429.

Krawczyk, J. et Uryasev, S. (2000). Relaxation algorithms to find Nash equilibria with economic applications. *Environmental Modeling and Assessment*, 5(1):63–73.

Kubota, K. et Fukushima, M. (2010). Gap function approach to the generalized Nash equilibrium problem. *Journal of optimization theory and applications*, 144(3):511–531.

Kulkarni, A. A. et Shanbhag, U. V. (2010). Revisiting generalized Nash games and variational inequalities. preprint.

Lemaire, J. et Quairière, J.-P. (1986). Chains of reinsurance revisited. *ASTIN Bulletin*, 16(2):77–88.

Leon, A. R. D. et Wu, B. (2010). Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine*.

Li, S., Lu, Y. et Garrido, J. (2009). A review of discrete-time risk models. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 103(2):321–337.

Lin, X. et Willmot, G. E. (2000). The moments of the time of ruin, the surplus before ruin, and the deficit at ruin. *Insurance : Mathematics and Economics*, 27(1):19–44.

Loisel, S. et Milhaud, X. (2011). From deterministic to stochastic surrender risk models : Impact of correlation crises on economic capital. *European Journal of Operational Research*, 214(2).

Lopes, V. L. R. et Martinez, J. M. (1999). On the convergence of quasi-Newton methods for nonsmooth problems. preprint.

Lu, Y. et Garrido, J. (2005). Doubly periodic non-homogeneous Poisson models for hurricane data. *Statistical Methodology*, 2(1):17–35.

Luce, R. (1959). *Individual Choice Behavior ; A Theoretical Analysis*. Wiley.

Lundberg, F. (1903). *1. Approximerad framställning af sannolikhetsfunktionen : 2. Återförsäkring af kollektivrisker*. Akademisk avhandling, Almqvist och Wiksell, Uppsala.

Maechler, M. (2012). *Rmpfr : R MPFR - Multiple Precision Floating-Point Reliable*. ETH Zurich.

Malinovskii, V. K. (2010). Competition-originated cycles and insurance companies. work presented at ASTIN 2009.

Manski, C. F. et McFadden, D. (1981). *Structural Analysis of Discrete Data with Econometric Applications*. The MIT Press.

Marceau, E. (2009). On the discrete-time compound renewal risk model with dependence. *Insurance : Mathematics and Economics*, 44(2):245–259.

Marceau, E. (2012). *Modélisation et évaluation des risques en actuariat*. Springer.

Markham, F. J. (2007). An investigation into underwriting cycles in the South African short-term insurance market for the period 1975 to 2006. Rapport technique, University of the Witwatersrand.

Marshall, A. W. (1996). Copulas, marginals and joint distributions. *In Distributions with fixed marginals and related topics*, volume 28. IMS Lecture Notes - Monograph Series.

Marshall, A. W. et Olkin, I. (1979). *Inequalities : theory of majorization and its applications*. Academic Press.

McCullagh, P. et Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, 2nd édition.

McFadden, D. (1981). Econometric Models of Probabilistic Choice. *In* Manski et McFadden (1981), chapitre 5.

Merentes, N. (1991). On the Composition Operator in AC[a,b]. *Collect. Math.*, 42(3):237–243.

Milhaud, X., Maume-Deschamps, V. et Loisel, S. (2011). Surrender triggers in Life Insurance : What main features affect the surrender behavior in a classical economic context ? *Bulletin Français d'Actuariat*, 22(11).

Mimra, W. et Wambach, A. (2010). A Game-Theoretic Foundation for the Wilson Equilibrium in Competitive Insurance Markets with Adverse Selection. CESifo Working Paper No. 3412.

Moler, C. et Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):300.

Monteiro, R. et Pang, J.-S. (1999). A Potential Reduction Newton Method for Constrained equations. *SIAM Journal on Optimization*, 9(3):729–754.

Moreno-Codina, J. et Gomez-Alvado, F. (2008). Price optimisation for profit and growth. *Towers Perrin Emphasis*, 4:18–21.

Nash, J. F. (1950a). Equilibrium points in N-Person Games. *Proc. Nat. Acad. Sci. U.S.A.*, 36(1):48–49.

Nash, J. F. (1950b). The Bargaining Problem. *Econometrica*, 18(2):155–162.

Nash, J. F. (1951). Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295.

Nash, J. F. (1953). Two Person Cooperative Games. *Econometrica*, 21(1):128–140.

Nelder, J. A. et Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer.

Neuts, M. (1975). *Probability Distributions of Phase-Type*. Liber Amicorum Prof. Emeritus H. Florin, Dept. Math., Univ. Louvain, Belgium.

Neuts, M. (1981). *Matrix-Geometric Solutions in Stochastic Models : An Algorithmic Approach*. The John Hopkins University Press.

Nikaido, H. et Isoda, K. (1955). Note on non-cooperative convex games. *Pacific J. Math.*, 5(1):807–815.

Nocedal, J. et Wright, S. J. (2006). *Numerical Optimization*. Springer Science+Business Media.

Ohlsson, E. et Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. Springer.

Ok, E. A. (2005). *Real Analysis with Economic Applications*. New York University.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F. et Clark, C. W., éditeurs (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press.

Osborne, M. et Rubinstein, A. (2006). *A Course in Game Theory*. Massachusetts Institute of Technology.

Picard, P. (2009). Participating insurance contracts and the Rothschild-Stiglitz equilibrium puzzle. working paper, Ecole Polytechnique.

Polborn, M. K. (1998). A model of an oligopoly in an insurance market. *The Geneva Paper on Risk and Insurance Theory*, 23(1):41–48.

Powell, M. (1970). A hybrid method for nonlinear algebraic equations. *In* Rabinowitz, P., éditeur : *Numerical Methods for Nonlinear Algebraic Equations*, chapitre 6. Gordon & Breach.

Powers, M. R. et Shubik, M. (1998). On the tradeoff between the law of large numbers and oligopoly in insurance. *Insurance : Mathematics and Economics*, 23(2):141–156.

Powers, M. R. et Shubik, M. (2006). A "square- root rule" for reinsurance. *Cowles Foundation Discussion Paper No. 1521.*

Qi, L. (1993). Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18(1):227–244.

Qi, L. (1997). On superlinear convergence of quasi-Newton methods for nonsmooth equations. *Operations Researchs Letters*, 20(5):223–228.

Qi, L. et Chen, X. (1995). A globally convergent successive approximation method for severely nonsmooth equations. *SIAM Journal on control and Optimization*, 33(2):402–418.

Qi, L. et Jiang, H. (1997). Semismooth KKT equations and convergence analysis of Newton and Quasi-Newton methods for solving these equations. *Mathematics of Operations Research*, 22(2):301–325.

Qi, L. et Sun, D. (1998). A Survey of Some Nonsmooth Equations and Smoothing Newton Methods. Rapport technique, Applied Mathematics Report AMR 98/10, School of Mathematics, the University of New South Wales.

Qi, L. et Sun, J. (1993). A nonsmooth version of Newton's method. *Mathematical Programming*, 58(1-3):353–367.

R Core Team (2012). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabehi, H. (2007). Study of multi-risk household. Mémoire de D.E.A., ISUP. Mémoire confidentiel - AXA France.

Rees, R., Gravelle, H. et Wambach, A. (1999). Regulation of insurance markets. *The Geneva Paper on Risk and Insurance Theory*, 24(1):55–68.

Rockafellar, R. T. et Wets, R. J.-V. (1997). *Variational Analysis*. Springer-Verlag.

Rolski, T., Schmidli, H., Schmidt, V. et Teugels, J. (1999). *Stochastic Processes for Insurance and Finance*. Wiley Series in Proability and Statistics.

Rosen, J. B. (1965). Existence and Uniqueness of Equilibrium Points for Concave N-person Games. *Econometrica*, 33(3):520–534.

Rothschild, M. et Stiglitz, J. E. (1976). Equilibrium in competitive insurance markets : An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):630–649.

Sergent, V. (2004). Etude de la sensibilité de l'assuré au prix en assurance auto de particuliers. Mémoire de D.E.A., ISUP. Mémoire confidentiel - AXA France.

Shaked, M. et Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer.

Shapiro, A. F. et Jain, L. C. (2003). *Intelligent and Other Computational Techniques in Insurance*. World Scientific Publishing.

Shapley, L. (1953). Stochastic games. *Proc. Nat. Acad. Sci. U.S.A.*, 39(10):1095–1100.

Shiu, E. S. W. (1989). The probability of eventual ruin in the compound binomial model. *ASTIN Bulletin*, 19(2):179–190.

Sigmund, K. et Hofbauer, J. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press.

Silvia, E. (1999). Companion Notes for Advanced Calculus. Lecture Notes, University of California.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440.

Simon, L. (2011). Mathematical methods. Rapport technique, Berkeley, Lecture notes.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'ISUP - Paris 8*, 8:229–231.

Sleet, C. (2001). On credible monetary policy and private government information. *Journal of Economic Theory*, 99(1-2):338–376.

Song, M., Meng, Q., Wu, R. et Ren, J. (2010). The Gerber-Shiu discounted penalty function in the risk process with phase-type interclaim times. *Applied Mathematics and Computation*, 216(2):523–531.

Steihaug, T. (2007). Splines and b-splines : an introduction. Rapport technique, University of Oslo.

Stein, W. A. *et al.* (2011). *Sage Mathematics Software (Version 4.6.2)*. The Sage Development Team.

Stroock, D. (1994). *A Concise Introduction to the Theory of Integration*. Birkhauser.

Su, L. et White, H. (2003). Testing conditional independence via empirical likelihood. UCSD Department of Economics Discussion Paper.

Sun, D. et Han, J. (1997). Newton and Quasi-Newton methods for a class of nonsmooth equations and related problems. *SIAM Journal on Optimization*, 7(2):463–480.

Sundt, B. et dos Reis, A. D. E. (2007). Cramér-Lundberg results for the infinite time ruin probability in the compound binomial model. *Bulletin of the Swiss Association of Actuaries*, 2.

Taksar, M. et Zeng, X. (2011). Optimal non-proportional reinsurance control and stochastic differential games. *Insurance : Mathematics and Economics*, 48(1):64–71.

Taylor, G. C. (1986). Underwriting strategy in a competitive insurance environment. *Insurance : Mathematics and Economics*, 5(1):59–77.

Taylor, G. C. (1987). Expenses and underwriting strategy in competition. *Insurance : Mathematics and Economics*, 6(4):275–287.

Teugels, J. et Veraverbeke, N. (1973). *Cramér-type Estimates for the Probability of Ruin*. Center for Operations Research and Econometrics Leuven : Discussion papers.

Thurstone, L. (1927). A Law of Comparative Judgment. *Psychological Review*, 34:273–286.

Tomala, T. et Gossner, O. (2009). Repeated games. *In Encyclopedia of Complexity and Systems Science*. Springer.

Trufin, J., Albrecher, H. et Denuit, M. (2009). Impact of underwriting cycles on the solvency of an insurance company. *North American Actuarial Journal*, 13(3):385–403.

Turner, H. (2008). Introduction to generalized linear models. Rapport technique, Vienna University of Economics and Business.

Venables, W. N. et Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, 4th édition.

Venezian, E. C. (1985). Ratemaking methods and profit cycles in property and liability insurance. *Journal of Risk and Insurance*, 52(3):477–500.

von Heusinger, A. et Kanzow, C. (2009). Optimization reformulations of the generalized Nash equilibrium problem using the Nikaido-Isoda type functions. *Computational Optimization and Applications*, 43(3).

von Heusinger, A., Kanzow, C. et Fukushima, M. (2010). Newton's method for computing a normalized equilibrium in the generalized Nash game through fixed point formulation. *Mathematical Programming*, 123(1):1–25.

von Neumann, J. et Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Wambach, A. (2000). Introducing heterogeneity in the Rothschild-Stiglitz model. *Journal of Risk and Insurance*, 67(4):579–591.

Wang, T., Monteiro, R. et Pang, J.-S. (1996). An interior point potential reduction method for constrained equations. *Mathematical Programming*, 74(2):159–195.

Willmot, G. E. (1993). Ruin probabilities in the compound binomial model. *Insurance : Mathematics and Economics*, 12(2):133–142.

Wood, S. N. (2001). mgcv : GAMs and Generalized Ridge Regression for R. *R News*, 1:20–25.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society : Series B*, 65(1):95–114.

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society : Series B*, 70(3).

Wood, S. N. (2010). Fast stable reml and ml estimation of semiparametric glms. *Journal of the Royal Statistical Society : Series B*, 73(1):3–36.

Yamashita, N. et Fukushima, M. (2000). On the rate of convergence of the Levenberg-Marquardt method. Rapport technique, Kyoto University.

Yeo, A. C. et Smith, K. A. (2003). *An integretated Data Mining Approach to Premium Pricing for the Automobile Insurance Industry*, chapitre 5. Volume 6 de Shapiro et Jain (2003).

Zeileis, A., Kleiber, C. et Jackman, S. (2008). Regression models for count data in r. *Journal of Statistical Software*, 27(8).

Zorich, V. (2000). *Mathematical Analysis I*, volume 1. Universitext, Springer.