



**HAL**  
open science

# Three Essays on the Design, Pricing, and Hedging of Insurance Contracts

Xiaoshan Su

► **To cite this version:**

Xiaoshan Su. Three Essays on the Design, Pricing, and Hedging of Insurance Contracts. Business administration. Université de Lyon, 2019. English. NNT : 2019LYSE2065 . tel-02475894

**HAL Id: tel-02475894**

**<https://theses.hal.science/tel-02475894>**

Submitted on 12 Feb 2020

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

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



N° d'ordre NNT : 2019LYSE2065

## THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

**École Doctorale : ED 486 Sciences Économique et de Gestion**

Discipline : Sciences de gestion

Soutenue publiquement le 11 décembre 2019, par :

Xiaoshan SU

---

# **Three Essays on the Design, Pricing, and Hedging of Insurance Contracts.**

---

Devant le jury composé de :

Laurent VILANOVA, Professeur des universités, Université Lumière Lyon 2, Président

André LAPIED, Professeur des universités, Aix-Marseille Université, Rapporteur

Jean-Luc PRIGENT, Professeur des universités, Université de Cergy-Pontoise, Rapporteur

Véronique MAUME-DESCHAMP, Professeure des universités, Université Claude Bernard Lyon 1,

Examinatrice

Donatien HAINAUT, Professeur Université Catholique de Louvain, Examineur

Olivier LE COURTOIS, Professeur, EM Lyon Business School, Directeur de thèse

## Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « [Paternité – pas d'utilisation commerciale – pas de modification](#) » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

**THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE  
LYON**

opérée au sein de  
**l'Université Lumière Lyon 2**

**École Doctorale ED486**  
**Sciences Économiques et de Gestion**

**Spécialité de doctorat: Sciences de Gestion**  
**Discipline: Quantitative Finance**

Soutenue publiquement le 11 septembre 2019, par :

**Xiaoshan SU**

---

**Three Essays on the Design, Pricing, and Hedging of  
Insurance Contracts**

---

Devant le jury composé de :

Mme. **Véronique MAUME-DESCHAMPS**, Suffragant  
Professeure des Universités, Université Claude Bernard Lyon 1

M. **Laurent VILANOVA**, Suffragant  
Professeur des Universités, Université Lumière Lyon 2

M. **Donatien HAINAUT**, Suffragant  
Professeur, Université Catholique de Louvain

M. **André LAPIED**, Rapporteur  
Professeur des Universités, Université d'Aix-Marseille

M. **Jean-Luc PRIGENT**, Rapporteur  
Professeur des Universités, Université de Cergy-Pontoise

M. **Olivier LE COURTOIS**, Directeur de thèse  
Professeur, emlyon business school

---

## Acknowledgments

My sincerest gratitude goes to my advisor, Professor Olivier Le Courtois. Many thanks for him to guide me into the fields of regime switching Lévy models, insurance and decision theory. His guidance, advices and encouragement support me a lot, not only in academics, but also in life, throughout my PhD study. His guidance leads us not only to become a good researcher, but also to be a better person. He is very nice and open-minded, encouraging us to be an independent scholar, introducing us to cooperate with others, and supporting us to attend conferences to know colleagues and broaden eyesight of research fields. I am impressed by his excellent expertise and insightful advices. Discussion with him helps me a lot in my research work. It is fortunate for me to be his student.

My thanks also go to whom I am lucky to cooperate with, François Quittard-Pinon, Eric André, Jirô Akahori, Manying Bai. I am very grateful for them to participate in our research projects, provide many useful comments and advance the progress of research work. I also send special thanks to Professor Jirô Akahori for his kind invitation and warm reception during my visit at Ritsumeikan University. I also thank Professor Manying Bai for his full support of my study in France.

Moreover, I would like to thank Professors Laurent Vilanova, Véronique Maume-Deschamps, Donatien Hainaut, André Lapied, Jean-Luc Prigent for being my committee members. My deep gratitude goes to Professors André Lapied and Jean-Luc Prigent for kindly accepting to be my thesis reviewers.

I also want to thank the financial risk analysis research centre of emlyon business school for the funding of my PhD study. I learnt a lot from high-quality seminars in this center. I am also grateful to Professor Isabelle Royer for her help in my PhD registration and Professor Pascal Langevin for his offer of a part-time job position. I thank my friends, professors, and colleagues, Yeming Gong, Lorenz Schneider, Bertrand Tavin, Guillaume Coqueret, Ryozi Miura, Jean Pascal Brivady, Li Shen, Xia Xu, Nicole Cardoso, Sandrine Brunet, El Mehdi Beqqali, Muriel Thiémé and others, for their constructive comments and kind assistance.

Finally, my special appreciation goes to my family: my parents Zhiqing Su, Shaozhen Lin, my brother Xiaoping Su, my niece Yuen Su. Thanks very much for their unconditional love, encouragement and full support.



# Contents

<b>1</b>	<b>Introduction (Version Française)</b>	<b>1</b>
1.1	Contexte . . . . .	1
1.1.1	Assurance vie et non vie . . . . .	2
1.1.2	Les limites des méthodes actuarielles traditionnelles . . . . .	5
1.1.3	Objectif . . . . .	8
1.2	Sujets spécifiques . . . . .	9
1.2.1	Contrats participatifs et sources de risque . . . . .	9
1.2.2	Conception du contrat . . . . .	12
1.2.3	Modèle fréquence-gravité . . . . .	13
1.2.4	Structure de la thèse . . . . .	14
<b>2</b>	<b>Introduction</b>	<b>17</b>
2.1	Background . . . . .	17
2.1.1	Life and non-life insurance . . . . .	18
2.1.2	The limitations of traditional actuarial methods . . . . .	21
2.1.3	Objective . . . . .	23
2.2	Specific topics . . . . .	24
2.2.1	Participating contracts and risk sources . . . . .	24
2.2.2	Contract design . . . . .	27
2.2.3	Frequency-Severity model . . . . .	28
2.2.4	Thesis structure . . . . .	28
<b>3</b>	<b>Background Knowledge</b>	<b>31</b>
3.1	Lévy process . . . . .	31
3.1.1	Definition and properties . . . . .	32
3.1.2	Examples of Lévy processes . . . . .	34
3.1.3	Itô formula and measure transform . . . . .	39
3.2	Regime Switching Lévy Process . . . . .	41
3.2.1	Definition and properties . . . . .	41
3.2.2	Itô formula and measure transform . . . . .	41
3.3	Stochastic orders . . . . .	42
3.3.1	The convex order . . . . .	42
3.3.2	The $m$ -convex order . . . . .	43
3.3.3	The monotone convex order . . . . .	45
3.3.4	The $m$ -increasing convex order . . . . .	46
3.4	Higher-order degree risk and stochastic dominance . . . . .	47
3.4.1	$n^{\text{th}}$ degree risk . . . . .	47
3.4.2	$n^{\text{th}}$ degree stochastic dominance . . . . .	48
3.5	The GLM and GAM model . . . . .	48
3.5.1	Generalized linear model . . . . .	48

3.5.2	Generalized additive model . . . . .	50
3.6	Machine learning algorithms . . . . .	51
3.6.1	Regression trees . . . . .	51
3.6.2	Gradient boosting . . . . .	53
<b>4</b>	<b>Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk</b>	<b>55</b>
4.1	Introduction . . . . .	56
4.2	The Basic Framework . . . . .	58
4.2.1	Contract Specifications . . . . .	58
4.2.2	Investment Funds and Interest Rate Dynamics . . . . .	60
4.2.3	Main Subcontract Terms . . . . .	61
4.3	The First Passage Time Results of Regime Switching Jump Diffusion Processes . . . . .	62
4.3.1	Matrix Wiener-Hopf Factorization . . . . .	62
4.3.2	The First Passage Time Results . . . . .	64
4.3.3	Numerical Algorithm . . . . .	65
4.4	Contract Valuation . . . . .	66
4.4.1	Computation of $GF$ and $LR$ . . . . .	67
4.4.2	Computation of $PO$ and $BO$ . . . . .	68
4.4.3	Numerical Analysis . . . . .	69
4.5	Constant Guaranteed Rates V.S. Floating Guaranteed Rates . . . . .	72
4.6	Two Hedging Strategies . . . . .	77
4.6.1	Hedging Strategy . . . . .	77
4.6.2	Hedging Effectiveness . . . . .	79
4.7	Conclusion . . . . .	81
4.8	Appendix . . . . .	82
4.8.1	Computations of Subcontract Terms for Contract with a Floating Guaranteed Rate . . . . .	82
<b>5</b>	<b>Optimal Insurance under Third Degree Risk</b>	<b>85</b>
5.1	Introduction . . . . .	85
5.2	The model . . . . .	88
5.3	Some motivations . . . . .	90
5.3.1	The cases of risk averters and risk lovers . . . . .	90
5.3.2	The advantage of third degree risk . . . . .	92
5.4	Optimal Contract . . . . .	92
5.4.1	General model . . . . .	93
5.4.2	Concave constraint . . . . .	95
5.5	Background risk . . . . .	98
5.5.1	The independent case . . . . .	99
5.5.2	The dependent case . . . . .	99
5.6	Conclusion . . . . .	101

## Contents

---

<b>6</b>	<b>Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims</b>	<b>103</b>
6.1	Introduction . . . . .	103
6.2	Stochastic Gradient Boosting Frequency-Severity Model . . . . .	106
6.2.1	The Dependent Frequency-Severity Model . . . . .	106
6.2.2	Stochastic Gradient Boosting . . . . .	107
6.2.3	The Dependent-FSBoost Model . . . . .	108
6.2.4	Tuning Parameters Choice and $\delta$ Estimation . . . . .	113
6.3	Simulation Study . . . . .	113
6.3.1	Simple Case . . . . .	114
6.3.2	Complex Case . . . . .	117
6.3.3	Examination of $\delta$ . . . . .	120
6.4	Application . . . . .	121
6.4.1	Data . . . . .	121
6.4.2	Model Comparison . . . . .	124
6.4.3	Model Interpretation . . . . .	125
6.5	Conclusion . . . . .	129
<b>7</b>	<b>General conclusion</b>	<b>131</b>
	<b>Bibliography</b>	<b>135</b>



# Introduction (Version Française)

---

## Contents

---

<b>1.1 Contexte</b> . . . . .	<b>1</b>
1.1.1 Assurance vie et non vie . . . . .	2
1.1.2 Les limites des méthodes actuarielles traditionnelles . . . . .	5
1.1.3 Objectif . . . . .	8
<b>1.2 Sujets spécifiques</b> . . . . .	<b>9</b>
1.2.1 Contrats participatifs et sources de risque . . . . .	9
1.2.2 Conception du contrat . . . . .	12
1.2.3 Modèle fréquence-gravité . . . . .	13
1.2.4 Structure de la thèse . . . . .	14

---

## 1.1 Contexte

Le risque existe dans presque toutes les activités humaines. Lors de la planification d'un pique-nique, il y a le risque de pleuvoir. Lorsque vous commandez un billet de théâtre, vous risquez d'être épuisé. En conduisant une voiture, il y a un risque d'accident de la circulation. Il existe une grande variété de risques qui nous influencent dans la vie. Dans la plupart des cas, le risque est peu susceptible d'entraîner une perte, alors que la taille de la perte est généralement importante. Pour l'individu, il est difficile de supporter le coût induit par des pertes sérieuses. Une bonne stratégie consiste à faire supporter les coûts à toutes les personnes confrontées au même type de risque. Chaque personne met une petite somme d'argent dans un pool de trésorerie, qui sert à couvrir les pertes. Une méthode de mise en œuvre consiste à transférer les risques à un tiers et ce dernier aide à compenser les pertes. Cette demande de partage des risques engendre un énorme secteur de l'assurance. Les compagnies d'assurance émettent des polices d'assurance pour mettre en œuvre le partage du risque entre les individus. Une variété de polices d'assurance est conçue pour satisfaire la demande de partage de divers risques, et les individus achètent celle qui leur est nécessaire pour transférer les risques. Par exemple, si votre voiture est assurée en cas d'accident de voiture, vous signalez le dommage à l'entreprise d'assurance et celle-ci peut indemniser votre perte si le montant dépend du contrat écrit. Selon que l'objectif de protection est une vie individuelle ou non, les contrats d'assurance sont classés dans les assurances vie et non-vie.

Dans un marché concurrentiel, il est très important que les compagnies d'assurance facturent un prix juste. Par exemple, en assurance automobile, si les compagnies d'assurance imposent trop peu aux jeunes conducteurs et aux anciens conducteurs, elles perdront leurs anciens conducteurs au profit de leurs concurrents tout en attirant de jeunes conducteurs. Ce problème de sélection adverse conduit les assureurs à perdre des contrats rentables et à obtenir des contrats sous-évalués, ce qui entraîne des pertes économiques. De nos jours, la conception de nombreux contrats d'assurance vie comprend de nombreuses clauses restrictives, telles que les garanties de taux d'intérêt, les contrats liés à des actions, les contrats de participation, etc. Ces contrats font en sorte que les sociétés d'assurance font face à davantage de risques, notamment économiques, de marché et de crédit. Ces risques doivent être pris en compte dans la tarification de ces contrats. L'ignorance de ces risques entraînera des difficultés. Par exemple, dans les années 90, de nombreuses sociétés d'assurances ont déclaré l'échec de leur entreprise. Une des raisons est la sous-évaluation du contrat en raison du risque ignoré de la garantie de taux d'intérêt. La compagnie d'assurance conçoit et établit les prix des contrats de manière à garantir sa compétence et ses profits. Entre temps, la compagnie d'assurance doit gérer les risques découlant de l'émission de ces contrats. Certains contrats liant des investissements sur des marchés financiers, la couverture devient un outil de gestion des risques utile pour la compagnie d'assurance. Ainsi, la conception, la tarification et la couverture des contrats d'assurance sont importantes pour la compagnie d'assurance et pour l'efficacité du marché de l'assurance.

### 1.1.1 Assurance vie et non vie

L'assurance vie traite de deux risques auxquels un individu est probablement confronté de son vivant. L'un meurt prématurément, laissant une famille à charge s'occuper de lui-même et un autre vit trop vieux pour pouvoir subvenir à ses besoins. Dans le premier cas, l'assurance vie offre une protection complète contre le risque de décès de l'assuré. En cas de décès, l'indemnité assurée sera intégralement versée. Dans ce dernier cas, l'assurance vie fonctionne également avec sa fonction d'épargne à long terme. La petite prime payée par versements faciles peut être accumulée beaucoup après une longue période. Ensuite, le contrat prévoit le paiement d'un montant à l'échéance du contrat ou périodiquement à certaines dates spécifiées. Assurément, l'assuré doit payer la prime à l'assureur périodiquement pendant la durée du contrat. Il convient de noter que l'assurance vie peut être souscrite après une période déterminée à partir de laquelle l'assuré obtiendra une partie du paiement de la prime. Il existe trois types d'assurance vie, à savoir l'assurance vie entière, l'assurance vie temporaire et la rente. Une assurance vie entière ne verse une prestation forfaitaire qu'au décès de l'assuré. Une assurance-vie temporaire fournit au bénéficiaire le montant du contrat à l'échéance ou au moment du décès de l'assuré si celui-ci décède avant l'échéance. Pour la rente, à l'expiration de sa durée, l'assureur verse périodiquement le montant de la police à l'assuré, tant que celui-ci est en vie.

## 1.1. Contexte

---

L'assurance non-vie, également connue sous le nom d'assurance de biens et risques divers, traite des risques encourus par des personnes et des biens, tels que les dommages causés par la maladie, la fraude, les accidents, les incendies, les tempêtes de vent, les tremblements de terre, le vol, etc. le particulier, son domicile, son automobile, son bateau, ses bagages, etc. Le contrat porte la garantie, pour laquelle l'assureur indemnise le sinistre subi par l'assuré lors de la survenance d'un événement incertain. En règle générale, les contrats sont à court terme, par exemple un an, et les assurés doivent les mettre à jour tous les ans s'ils nécessitent une protection à long terme. Le caractère des objets assurés peut être grossièrement classé en tant que personne, bien et intérêt. Pour personne, il existe une assurance accident et maladie. Pour les biens, il existe de nombreux exemples tels que l'assurance incendie, l'assurance dommages, l'assurance des coques marines, etc. Comme pour les intérêts, il existe une assurance responsabilité, une assurance dommages indirects, une assurance fraude, etc.

La comparaison détaillée des caractéristiques distinctes entre l'assurance-vie et l'assurance non-vie est présentée dans le tableau 1.

**Tableau 1** Comparaison entre assurance vie et assurance non vie

Fonctionnalité	Assurance-vie	Assurance non vie
couverture des risques	risque de la vie d'un individu	tout risque de personnes et de biens en dehors du risque de la vie, tel que la perte pour cause de maladie, fraude, accident, incendie, tempête de vent, tremblement de terre, vol, etc.
style de contrat	une forme d'investissement	un contrat d'indemnisation
terme du contrat	long terme	court terme
temps de paiement	soit au décès de l'assuré, soit à l'échéance	en cas d'événement incertain
prime	paiements périodiques pendant la durée du contrat	un paiement forfaitaire
intérêt assurable	présent au moment du contrat	présent à la fois au moment du contrat et au moment de la perte
valeur politique	basé sur le paiement de prime spécifique	limite au préjudice réel subi, quel que soit le montant de la police
des économies	épargne à long terme	pas d'économie

De nos jours, les activités d'assurance mondiales se sont développées dans une certaine mesure. Le tableau 2 répertorie les 10 principales régions du secteur de l'assurance en 2017 et nous pouvons constater que le marché de l'assurance a été très important dans les pays développés, qu'il s'agisse d'assurance vie ou non. Le quota de marché des trois principaux pays atteint près de 50% des activités d'assurance sur le marché mondial, qui présente un important potentiel de marché à développer, en particulier pour les marchés émergents.

Le secteur des assurances à grande échelle est très exigeant en matière de gestion des risques du secteur des assurances. Le risque est le fondement du secteur des assurances, pour lequel la police d'assurance transfère de nombreux risques de l'assuré aux sociétés d'assurance. Comme mentionné précédemment, la conception complexe

## Chapter 1. Introduction (Version Française)

**Tableau 2** **Top 10 des activités d'assurance en 2017 (US \$ millions)**

Rang	Région	Total des primes		Primes vie	
		Montant	Part du marché mondial	Montant	Part du marché mondial
1	états Unis	1,377,114	28.15%	546,800	20.58%
2	continentale (Chine)	541,466	11.07%	317,570	11.95%
3	Japon	422,050	8.63%	307,232	11.56%
4	Royaume-Uni	283,331	5.79%	189,833	7.14%
5	France	241,603	4.94%	153,520	5.78%
6	Allemagne	222,978	4.56%	96,973	3.65%
7	Corée du Sud	181,218	3.70%	102,839	3.87%
8	Italie	155,509	3.18%	113,947	4.29%
9	Canada	119,520	2.44%	51,592	1.94%
10	Taiwan (Chine)	117,474	2.40%	98,602	3.71%

Primes non-vie	
Montant	Part du marché mondial
830,315	37.16%
223,876	10.02%
114,818	5.14%
93,499	4.18%
88,083	3.94%
126,005	5.64%
78,378	3.51%
41,562	1.86%
67,927	3.04%
18,873	0.84%

Source: Swiss Re, Sigma, n° 3/2018.

## 1.1. Contexte

---

des contrats d'assurance oblige les sociétés d'assurance à faire face à de nombreux risques différents. Les compagnies d'assurance doivent donc prendre en charge ces risques. Les directives Solvabilité II et III, en tant que deux cadres réglementaires pour le secteur des assurances, obligent les sociétés d'assurance à se concentrer sur la gestion de tous les risques auxquels elles sont confrontées. Les normes de réglementation établissent un nouvel ensemble d'exigences de fonds propres, de techniques d'évaluation, ainsi que de normes de gouvernance et de reporting, dans lesquelles les actifs et les passifs sont mesurés de manière à garantir que les fonds propres soient suffisamment protégés contre les risques des assureurs. Actuellement, Solvency II est la norme de régulation largement utilisée. Elle est divisée en trois piliers. Le pilier 1 utilise des modèles cohérents sur le marché pour évaluer les actifs et les passifs, puis calculer les exigences de fonds propres réglementaires. Le pilier 2 définit les exigences de fonds propres (SCR) et le pilier 3 fournissent un rapport annuel privé aux autorités de contrôle, ainsi qu'un rapport d'information sur la solvabilité et les informations financières au public. Le pilier 1 encourage les sociétés d'assurance à développer leurs propres modèles internes d'évaluation. Afin d'améliorer leurs compétences, les principales sociétés d'assurance ont déployé de nombreux efforts de ce côté. Ainsi, pour relever les défis de la gestion des risques et d'une réglementation stricte, la précision des prix et la couverture des contrats d'assurance sont devenues extrêmement importantes.

### 1.1.2 Les limites des méthodes actuarielles traditionnelles

Les contrats d'assurance vie traditionnels se concentrent principalement sur la protection de la mortalité. Toutefois, du point de vue de l'investissement, les assurés sont conscients des opportunités d'investissement sur le marché financier et ont la demande de bénéficier des avantages d'un investissement financier en liaison avec la protection de la mortalité. Pour attirer les assurés, les compagnies d'assurance ont intégré cette demande d'investissement dans la conception des contrats d'assurance et ont mis au point de nombreux types de contrats modernes, tels que les contrats en unités de compte, les rentes variables et les contrats de fonds distincts, etc. paiements promis au décès ou à l'échéance. Une partie ou la totalité des primes est investie dans un fonds d'actions et les souscripteurs se partagent les bénéfices des investissements financiers. Ainsi, dans l'assurance moderne, le risque financier est une source de risque importante qui ne peut être ignorée. De nos jours, les contrats d'assurance vie modernes ont dominé le marché de l'assurance vie, en particulier dans les pays développés comme l'Amérique, le Canada, la France, l'Allemagne, l'Australie, etc.

Dans le passé, les techniques actuarielles sont l'outil le plus utilisé pour la tarification des contrats d'assurance vie. Cependant, ces méthodes ne conviennent pas à l'évaluation des contrats d'assurance modernes. Les raisons sont doubles. D'une part, seules les techniques actuarielles permettent d'évaluer et de gérer le risque d'assurance, mais ne peuvent rien faire contre le risque financier. Parce que leur gestion du risque repose fortement sur la diversification. Avec de nombreux

contrats d'assurance sur des durées de vie indépendantes, le théorème de la limite centrale garantit le faible degré d'incertitude du total des sinistres, ce qui permet aux techniques actuarielles d'utiliser une méthode déterministe pour la tarification. Par exemple, une compagnie d'assurance vend 10 000 contrats d'assurance à des personnes indépendantes, chacune ayant une probabilité de réclamation de 0.03. Si ces contrats sont des contrats traditionnels, la meilleure estimation du taux de mortalité est de 0.03 et la technique actuarielle ajuste la valeur de la meilleure estimation, par exemple 0.04. Ensuite, la probabilité que le taux de mortalité réel supérieur à 0.04 soit inférieur à  $10^{-8}$  et à la quasi-totalité du risque de mortalité est absorbée. Néanmoins, pour les contrats d'assurance modernes, le risque financier inhérent est un risque systématique ou non diversifiable. Lorsque l'investissement lié ne fonctionne pas correctement, tous les contrats sont affectés simultanément. Ainsi, le risque financier rend les contrats moins indépendants que les contrats traditionnels. Dans ce cas, le théorème central limite ne peut plus être utilisé et les techniques actuarielles ne peuvent donc pas non plus fonctionner. Par exemple, certains contrats effectuent le paiement promis. Supposons que la compagnie d'assurance indemnise le manque entre le paiement promis et la valeur du fonds si l'investissement a une mauvaise performance. Nous supposons que la probabilité que l'investissement ne soit pas performant soit de 0.03 et qu'il existe ensuite une probabilité de 0.03 que tous les contrats nécessitent le versement par la compagnie d'assurance de l'indemnisation. Ce risque systématique ne peut être géré par les techniques actuarielles traditionnelles.

Par ailleurs, comme nous l'avons déjà mentionné, les contrats d'assurance vie moderne comportent de nombreuses clauses restrictives telles que les options de bonus et de rachat, les garanties de taux d'intérêt, les contrats de participation, etc. Ces clauses sont des options incorporées dans les contrats et leur tarification doit être précise. Par exemple, la garantie de taux d'intérêt crédite chaque année le compte des assurés d'un taux minimal. Au moment de l'émission, si le taux d'intérêt garanti est très inférieur aux taux d'intérêt du marché, aucune prime pour ces garanties n'a de mauvaise conséquence. Toutefois, lorsque les taux d'intérêt du marché baissent et restent à un niveau bas pendant longtemps, le taux de garantie fixe entraînera une lourde charge de responsabilité pour la compagnie d'assurance. C'est la raison pour laquelle de nombreuses entreprises d'assurance ont fait faillite en Europe, au Japon et aux états-Unis dans les années 90. Ainsi, les options intégrées doivent vraiment être tarifées. Cependant, les techniques actuarielles traditionnelles ne peuvent rien faire de ce côté.

Les méthodes en ingénierie financière constituent des solutions naturelles pour résoudre le problème de tarification et de gestion des risques des contrats d'assurance modernes. En fait, les contrats d'assurance modernes peuvent être traités comme un type spécial de dérivé financier combinant une protection contre la mortalité. [Boyle and Schwartz \(1977\)](#); [Brennan and Schwartz \(1979\)](#); [Brennan, Schwartz, et al. \(1976\)](#) sont des travaux fondamentaux qui utilisent la théorie moderne de la tarification des options pour la tarification des contrats d'assurance vie. Il existe de nombreux ouvrages sur l'amélioration des méthodes de marché pour la tarification

## 1.1. Contexte

---

et la gestion des risques des contrats d'assurance modernes. Voir [Bernard, Le Courtois, and Quittard-Pinon \(2005\)](#); [Coleman, Kim, Li, and Patron \(2007\)](#); [Coleman, Li, and Patron \(2006\)](#); [Kélani and Quittard-Pinon \(2017\)](#); [Lin, Tan, and Yang \(2009\)](#); [Siu \(2005\)](#), pour n'en nommer que quelques-uns. Les méthodes de marché sont également encouragées par l'exigence réglementaire de Solvency II selon laquelle les actifs et les passifs des sociétés d'assurance doivent être tarifés à l'aide de modèles de marché, ainsi que par plusieurs normes internationales telles que l'International Accounting Standard Board (IASB), la Comptabilité financière. Standard Board (FASB) et la norme internationale d'information financière (IFRS), etc., qui imposent aux sociétés d'assurance-vie de divulguer leurs informations comptables à la juste valeur. L'étude des méthodes du marché pour la tarification des contrats d'assurance modernes a suscité un grand intérêt, tant dans la pratique que dans le monde universitaire.

Lors de la tarification des polices d'assurance non-vie, il existe de nombreux facteurs importants, tels que la perte de sinistre, la charge de sécurité, les coûts administratifs, etc. Apparemment, le facteur le plus crucial est la perte de sinistre. Ainsi, nous nous concentrons principalement sur la modélisation des pertes de sinistres. Le montant de la perte de sinistres attendue dépend fortement des caractéristiques d'un contrat individuel. Par exemple, dans une réclamation d'assurance automobile, les points d'âge, de sexe et d'enregistrement de véhicule du preneur d'assurance, par revenu ou densité de population du lieu de résidence du preneur d'assurance, l'âge et le modèle du véhicule, etc. Taille. Nous devons donc utiliser des modèles prédictifs.

Les technologies actuariales traditionnelles choisissent les modèles linéaires généralisés (GLM) pour modéliser la taille des revendications. Deux modèles standard basés sur GLM sont le modèle de fréquence-gravité GLM et le modèle de Poisson composé GLM Tweedie. Le premier modélise la taille des revendications en deux parties, la fréquence des revendications et la gravité des revendications, la fréquence des revendications examinant le nombre de revendications en utilisant une régression de Poisson ou une régression binomiale négative, et la gravité de la revendication prenant en charge le montant des revendications en fonction de en utilisant une régression gamma ou gaussienne inverse. Ce dernier modélise la taille de la revendication par une somme de Poisson de i.i.d. variables aléatoires gamma. La distribution Tweedie résultante appartient à la famille de dispersion exponentielle et le modèle GLM Tweedie peut donc être construit pour prédire directement la taille de la revendication. Bien que les modèles GLM soient largement utilisés, il existe une limite évidente. La forme linéaire des modèles GLM est trop stricte pour les applications réelles, ce qui affecte leur précision de prédiction. Par exemple, en assurance auto, la relation non linéaire entre âge et taille du sinistre est bien documentée. Les modèles additifs généralisés (GAM) surmontent la forme linéaire restrictive du modèle GLM en modélisant la variable continue avec des fonctions lisses. Cependant, le modèle GAM est également limité à la forme additive et les effets d'interaction complexes doivent être identifiés manuellement. En pratique, il est difficile de trouver tous les effets d'interaction, en particulier avec de nombreuses variables. Manquer des

interactions importantes affectera négativement la précision des prévisions. Nous avons donc besoin de modèles plus flexibles.

Les algorithmes d'apprentissage automatique sont les solutions naturelles dont nous avons besoin. Les algorithmes apprennent la structure du modèle à partir de données, ce qui est capable d'adapter une relation non linéaire flexible entre la variable de réponse et les prédicteurs et de capturer des interactions complexes et d'ordres plus élevés parmi les prédicteurs. De plus, les algorithmes peuvent automatiquement compléter la sélection des fonctionnalités. Ils se sont avérés très performants dans diverses applications, en raison de leur grande précision de prédiction. L'application des algorithmes d'apprentissage automatique en assurance non-vie est devenue un sujet de recherche brûlant. [Wuthrich and Buser \(2017\)](#) et [Noll, Salzmann, and Wuthrich \(2018\)](#) ont montré de manière empirique que les arbres de régression, les réseaux de stimulation et les réseaux de neurones surpassent le modèle GLM dans la prédiction de la fréquence des réclamations. [Yang, Qian, and Zou \(2018\)](#) développe un modèle de Poisson composé de Tweedie renforçant le gradient et montre que le modèle est supérieur aux autres modèles de pointe. L'ASTIN (études actuarielles en assurance non-vie) de l'Association actuarielle internationale (IAA) encourage l'adoption de technologies d'apprentissage automatique dans le domaine de l'assurance. Le concept de technologie d'assurance (InsurTech) est développé dans le livre blanc sur le développement de China InsurTech, dans lequel un point clé est lié à l'application des technologies d'apprentissage automatique. Ainsi, l'étude des algorithmes d'apprentissage automatique pour la tarification de l'assurance non-vie est devenue un sujet intéressant, tant dans la pratique que dans le monde universitaire.

### 1.1.3 Objectif

Cette thèse améliore la conception, la tarification et la couverture des contrats d'assurance, en utilisant les outils d'ingénierie financière, de théorie de la décision et d'apprentissage automatique. En particulier, nous développons un modèle avancé pour la tarification d'un type courant de contrat d'assurance vie, à savoir les contrats d'assurance vie participants, en présence de risques de crédit, de marché (saut), économiques (changement de régime). Nous introduisons également deux stratégies de couverture, à savoir la couverture semi-statique et la couverture dynamique, afin de couvrir les risques de défaut, de saut et de changement de régime dans les contrats avec participation. Outre la tarification des contrats d'assurance vie, nous étudions également la conception optimale de la police d'assurance en envisageant une vaste couverture d'assurés comprenant à la fois des précurseurs de risques et des passionnés des risques, puis nous montrons que les contrats optimaux peuvent prendre la forme de contrats à double limitation perte, changement-perte, double changement-perte et stop-loss dans différents contextes. En ce qui concerne la tarification de l'assurance non-vie, nous développons un modèle fréquence-sévérité de renforcement du gradient stochastique, dans lequel l'algorithme de renforcement du gradient stochastique est utilisé pour estimer les distributions de fréquence et de

## 1.2. Sujets spécifiques

---

gravité des sinistres, ainsi que la dépendance non linéaire entre fréquence et sévérité des sinistres. traiter la fréquence des réclamations comme facteur prédictif dans le modèle de régression de la gravité des réclamations. Nous montrons que le modèle proposé est supérieur aux autres modèles à la pointe de la technologie.

## 1.2 Sujets spécifiques

### 1.2.1 Contrats participatifs et sources de risque

Les contrats d'assurance-vie avec participation sont l'un des contrats d'assurance-vie les plus populaires. Dans ces contrats, le preneur d'assurance reçoit non seulement un rendement minimum garanti, mais partage également les bénéfices d'un investissement financier dépassant le rendement garanti. Le preneur d'assurance verse une somme forfaitaire à la compagnie d'assurance et l'assureur investit et gère les fonds dans un portefeuille de référence. L'assureur crédite chaque année un intérêt égal ou supérieur à un taux minimal garanti. La différence nette entre la valeur de marché du portefeuille de référence et la valeur comptable du compte du titulaire de la police constitue la réserve de bonus. Si la réserve de bonus terminal est positive, l'assuré reçoit un bonus terminal. Si l'assureur fait défaut pendant la durée du contrat, le preneur d'assurance recevra les actifs restants. Voir la discussion détaillée sur les différentes caractéristiques contractuelles des contrats participants dans [Grosen and Jørgensen \(2000\)](#).

Comme indiqué précédemment, le risque de crédit doit être chiffré. [Briys and De Varenne \(1994, 1997\)](#) fournit un cadre général pour l'évaluation des contrats de participation lorsque ceux-ci prennent explicitement en compte le risque de crédit de l'assureur. Cependant, le défaut ne peut se produire qu'à l'échéance. [Grosen and Jørgensen \(2002\)](#) corrige cette faiblesse en introduisant un contrôle permanent de la solvabilité de l'entreprise. Nous examinons les spécifications des contrats participants en utilisant le framework dans [Grosen and Jørgensen \(2002\)](#).

L'assuré souscrit un contrat d'assurance vie moyennant le versement d'une prime unique  $L_0$ . L'assureur gère un investissement dans un portefeuille de référence et les fonds sont en partie financés par la prime  $L_0$ . Indiquez par  $A_0$  la valeur initiale des fonds. Ensuite, les titulaires de police ont des actions  $\alpha = \frac{L_0}{A_0}$  des fonds et profitent des avantages d'un rendement du capital excédentaire. L'assureur promet aux souscripteurs un taux minimum de  $r_g$  pendant la durée du contrat. Ensuite, le paiement d'une échéance promise est égal à  $L_T = L_0 e^{r_g T}$ , où  $T$  correspond à l'échéance du contrat. Le paiement promis peut être honoré si la valeur finale du portefeuille est suffisamment grande, c'est-à-dire,  $A_T \geq L_T$ . Dans le cas contraire, le contrat est exécuté par défaut et les titulaires de police saisissent la valeur de portefeuille restante  $A_T$ . Outre le paiement à l'échéance promise, les titulaires de police ont le droit de recevoir un bonus si l'investissement des fonds est suffisamment performant. Les souscripteurs partagent une proportion  $\delta$  de la valeur des fonds supérieure au paiement promis, c'est-à-dire  $\delta(\alpha A_T - L_T)_+$ , où  $\delta$ , appelé taux

de participation, représente le niveau de participation. des assurés au potentiel de hausse des investissements financiers. Ensuite, les souscripteurs reçoivent un paiement à l'échéance, en dollars canadiens.

$$\Theta_L(T) = \begin{cases} A_T & \text{if } A_T < L_T \\ L_T & \text{if } L_T \leq A_T \leq \frac{L_T}{\alpha} \\ L_T + \delta(\alpha A_T - L_T) & \text{if } A_T > \frac{L_T}{\alpha} \end{cases}.$$

Ce gain peut être écrit sous une forme compacte

$$\Theta_L(T) = L_T + \delta(\alpha A_T - L_T)_+ - (L_T - A_T)_+.$$

où les trois termes du côté droit représentent le paiement à échéance promis, l'option de bonus et une option de vente à découvert liée au défaut survenant à l'échéance, respectivement.

La structure de paiement ci-dessus repose sur une hypothèse irréaliste selon laquelle le défaut de l'assureur ne survient qu'à la date d'échéance du contrat. En pratique, la défaillance de l'assureur peut survenir à tout moment  $t$ . Ainsi, supposons qu'à tout moment  $t$ , la frontière réglementaire soit proportionnelle au paiement promis,  $L_t = L_0 e^{r_g t}$ , et que la solvabilité de l'assureur soit contrôlée en permanence par les autorités de régulation. Soit  $\kappa$  le paramètre de niveau limite et la valeur par défaut de l'assureur se produit au moment opportun

$$\tau = \inf\{t \geq 0 : A_t \leq \kappa L_0 e^{r_g t}\}.$$

La valeur du portefeuille restant est de  $A_\tau$ . Ensuite, notons  $\{r_t, t \geq 0\}$  la dynamique des taux d'intérêt. La formule de tarification en vertu de la mesure neutre au risque  $Q$  peut être écrite sous forme de

$$V = E_Q \left( e^{-\int_0^T r_s ds} (L_T + \delta(\alpha A_T - L_T)_+ - (L_T - A_T)_+) \mathbb{1}_{\tau \geq T} + e^{-\int_0^\tau r_s ds} A_\tau \mathbb{1}_{\tau < T} \right).$$

Outre le risque de crédit, les investissements financiers rendent les contrats confrontés à des risques de marché et économiques. Le phénomène de sauts du prix des actifs, c'est-à-dire de rares et importants mouvements, est souvent observé sur le marché financier. Ce phénomène est induit par des chocs externes dus à certains événements extrêmes tels que les changements de politique. La caractéristique leptokurtique asymétrique du rendement des actifs financiers est bien documentée, voir, par exemple, [Cont \(2001\)](#). Ces caractéristiques peuvent être bien expliquées par le phénomène de saut. [Merton \(1976\)](#) est le travail fondateur qui modélise le risque de saut dans la dynamique des prix des actifs. [Ball and Torous \(1983\)](#); [Jarrow and Rosenfeld \(1984\)](#); [Jorion \(1988\)](#) fournissent des preuves convaincantes de la présence de hausses du prix des actifs. Dans la littérature sur les prix des options,

## 1.2. Sujets spécifiques

---

Bakshi, Cao, and Chen (1997); Duffie, Pan, and Singleton (2000); Kou (2002); Kou and Wang (2004); Merton (1976); Naik and Lee (1990); Pan (2002) démontrent que l'intégration de sauts est essentielle pour expliquer le prix d'option observé, tel qu'un phénomène empirique appelé "volatilité volatile", observé dans des scénarios optionnels. En outre, une classe importante de modèles stochastiques, appelés modèles de Lévy, est développée pour prendre en compte le risque de saut sur le marché financier, tels que le Variance Gamma (VG), le Normal Inverse Gaussian (NIG), le Tempered Stable (TS), les processus hyperboliques généralisés (GH), les processus de Meixner et CGMY, etc. Il existe de nombreux ouvrages sur la tarification des produits dérivés avec les modèles de Lévy. Voir Asmussen, Avram, and Pistorius (2004); Cariboni and Schoutens (2007); Carr and Wu (2003); Cont and Voltchkova (2005a,b); Fusai and Meucci (2008); Kou (2002); Kou and Wang (2004); Tankov (2003), pour n'en nommer que quelques-uns. La preuve du risque de saut est très suffisante.

Le risque de changement de régime est un risque économique important. Ce risque économique est constitué par les modifications structurelles de l'environnement macroéconomique ou des cycles économiques, qui induisent des modifications de la structure dynamique du prix des actifs ou de la structure par terme des taux d'intérêt. Le risque a été examiné dans de nombreuses études empiriques, telles que Bollen, Gray, and Whaley (2000); Engel and Hamilton (1990); Guidolin and Timmermann (2005, 2008). Ils apportent un soutien convaincant à la présence d'un changement de régime du prix des actifs. Une classe importante de modèles stochastiques, appelée modèle de changement de régime, est développée pour prendre en charge le risque de changement de régime. Hamilton (1989) vulgarise le modèle de série chronologique à changement de régime dans la littérature économique et économétrique. Hardy (2001) montre le succès empirique du modèle de changement de régime en ajustant les données mensuelles à long terme des indices Standard and Poor's 500 et Toronto Stock Exchange 300. Les modèles de changement de régime ont été largement utilisés dans différents domaines de la finance, tels que l'évaluation des options, la répartition de l'actif et la gestion des risques, etc. Voir Buffington and Elliott (2002); Chollete, Heinen, and Valdesogo (2009); Elliott, Chan, and Siu (2005); Guidolin and Timmermann (2007); Guo (2001); Zhang and Guo (2004); Zhou and Yin (2003), to name just a few.

Les contrats d'assurance vie sont relativement anciens. Il peut y avoir des changements substantiels dans les conditions économiques sur une longue période. Il est donc essentiel d'intégrer le risque de changement de régime à l'évaluation des produits d'assurance vie. En général, nous considérons que le saut et le changement de régime sont des risques à court terme et à long terme, respectivement. Ainsi, nous finalisons la tarification des contrats d'assurance vie avec participation en présence de risques de crédit, de saut et de changement de régime.

## 1.2.2 Conception du contrat

La conception optimale de l'assurance a attiré une grande attention de la part des praticiens et des universitaires. Depuis l'ouvrage sémantique de Arrow (1963), il existe une littérature abondante sur ce sujet, par exemple, voir Chi (2017); Chi and Tan (2011); Chi and Zhou (2017); Cummins and Mahul (2004); Raviv (1979); Young (1999).

Nous examinons le modèle d'assurance optimal avec un exemple simple. Soit  $X$  le montant de la perte subie par un assuré. La conception de l'assurance fonctionne sur la répartition de la perte en  $f(X)$  et  $R_f(X)$ , où  $f(X)$  représente une partie de la perte cédée à l'assureur et  $R_f(X)$  enregistre perte restante conservée par l'assuré. La conception vise à maximiser l'utilité attendue de la richesse finale d'un assuré. Supposons que l'assuré a une aversion pour le risque avec une fonction d'utilité concave croissante,  $u(x)$ . Nous supposons que l'assuré avec un patrimoine initial  $w_0$  souscrit un contrat d'assurance moyennant le paiement d'une prime  $P$ . Ensuite, la richesse finale de l'assuré est  $w(X) = w_0 - R_f(X) - P$ . Le contrat d'assurance optimal maximise l'utilité escomptée de la richesse finale de l'assuré, c'est-à-dire

$$\arg \max_f E(u(w(X))).$$

En termes de prime  $P$ , il existe de nombreux principes de prime. Young (2014) énumère onze principes de prime communs. Si nous utilisons le principe de la prime de valeur attendue, où la prime d'assurance ne dépend que de la valeur attendue de l'indemnité, nous avons la solution du problème d'optimisation ci-dessus comme la fonction  $f(x) = (x - b)_+$ , où  $b$  est une franchise à montant fixe.

Ceci est un exemple de base de la conception d'assurance optimale. Les travaux susmentionnés enrichissent le contenu de ce champ. Le critère d'optimisation utilise parfois la minimisation de l'exposition au risque de l'assuré avec certaines mesures de risque spécifiques, telles que la valeur à risque et la valeur à risque conditionnelle. Voir, par exemple, Cai and Tan (2007); Cai, Tan, Weng, and Zhang (2008); Chi and Tan (2011); Tan, Weng, and Zhang (2009). Chi (2017) et Chi and Wei (2018) prennent également en compte les attitudes de risque plus élevées de l'assuré. Cependant, ces travaux supposent tous que l'assuré a une aversion pour le risque. Bien qu'une majorité de personnes s'aperçoive qu'elles sont peu enclines à prendre des risques, il existe encore une minorité significative d'entre elles qui aiment les risques. Deck and Schlesinger (2014) examine les préférences de risque plus élevées des avertisseurs de risque et des amateurs de risque et découvre les deux faits suivants. Premièrement, les avertisseurs de risques n'apprécient pas une augmentation du risque à tous les degrés. Deuxièmement, les amoureux du risque aiment les risques qui augmentent de degré pair, mais n'aiment pas les augmentations de degré impair. Ainsi, les avertisseurs de risques et les passionnés de risques ont les mêmes attitudes de risque à des degrés étranges, telles que la prudence de troisième ordre. Crainich, Eeckhoudt, and Trannoy (2013) confirme que les amoureux du risque sont prudents. Par conséquent, nous supposons seulement que l'assuré est prudent afin de couvrir à la fois

## 1.2. Sujets spécifiques

---

les avertisseurs de risques et les amoureux des risques. Ensuite, avec une large couverture des assurés, nous montrons les formes spécifiques du contrat optimal dans différents contextes.

### 1.2.3 Modèle fréquence-gravité

Le modèle fréquence-gravité est un modèle de réclamation d'assurance largement utilisé. Nous illustrons le modèle fréquence-gravité dans le cadre GLM. Soit  $x = (x_1, \dots, x_p)'$  et  $y = (y_1, \dots, y_q)'$  les variables prédictives de la fréquence et de la gravité des réclamations, respectivement. Indiquez par  $N$  et  $X_i$  les variables aléatoires respectives de la fréquence et de la gravité des réclamations. Avec deux fonctions de liaison  $g$  et  $f$ , nous avons les modèles de régression comme

$$g(\nu) = \alpha x \quad \text{and} \quad f(\mu) = \beta y,$$

où  $\nu = E(N|x)$  et  $\mu = E(X_i|y)$ , et  $\alpha$  et  $\beta$  sont deux vecteurs de coefficients de régression de taille  $1 \times p$  et  $1 \times q$ , respectivement. L'inférence des modèles GLM donne les valeurs des coefficients de régression  $\alpha$  et  $\beta$ , ainsi que d'autres paramètres des distributions de  $N$  et  $X_i$ . Avec les distributions connues de fréquence de réclamation  $N$  et de gravité  $X_i$ , nous avons la distribution du montant global de la réclamation comme suit:

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

Ensuite, en fonction du montant total de la demande d'indemnisation, nous pouvons calculer les primes de la police d'assurance non-vie.

Comme nous l'avons mentionné précédemment, les formes linéaires ou additives des modèles de fréquence et de gravité GML et GAM sont trop rigides pour des applications réelles. En outre, ces modèles supposent souvent que la fréquence et la gravité des réclamations sont indépendantes. En réalité, la fréquence et la gravité des réclamations sont souvent liées. Par exemple, les réclamations d'assurance habitation dues aux inondations tendent à être à la fois importantes et fréquentes dans la zone touchée. Pour l'assurance automobile, la fréquence et la gravité des réclamations sont souvent négativement corrélées, car les conducteurs qui font plusieurs réclamations par an ne provoquent généralement que quelques accidents mineurs. Il faut donc modéliser la dépendance entre fréquence des sinistres et gravité. [Erhardt and Czado \(2012\)](#); [Frees, Gao, and Rosenberg \(2011\)](#); [Gschlößl and Czado \(2007\)](#) capturer la dépendance en traitant la fréquence des revendications en tant que variable prédictive dans le modèle de régression de la gravité moyenne des revendications et [Czado, Kastenmeier, Brechmann, and Min \(2012\)](#); [Krämer, Brechmann, Silvestrini, and Czado \(2013\)](#); [Shi, Feng, and Ivantsova \(2015\)](#) employez les copules paramétriques pour modéliser la distribution conjointe de la fréquence et de la gravité moyenne des réclamations.

Ainsi, nous développons un modèle stochastique amplificateur fréquence-gravité pour surmonter les formes restreintes des modèles GLM et GAM, où nous traitons la

fréquence des réclamations comme une variable prédictive dans le modèle de régression amplifiant le gradient de la sévérité moyenne des réclamations pour capturer de manière flexible la dépendance non linéaire. entre la fréquence et la gravité des réclamations.

### 1.2.4 Structure de la thèse

Cette thèse porte sur la conception, la tarification et la couverture des contrats d'assurance afin d'améliorer la gestion des risques des sociétés d'assurance et l'efficacité du marché de l'assurance. Le chapitre 2 passe en revue certaines connaissances de base du processus de Lévy et du changement de régime du processus de Lévy pour le chapitre 3, de quelques notions d'ordre stochastique, du risque de degré supérieur et de la dominance stochastique pour le chapitre 4, ainsi que du modèle GLM et GAM, arbre de régression et algorithme de renforcement du gradient pour le chapitre 5.

Le chapitre 3 développe une approche basée sur la transformation pour la tarification des contrats d'assurance-vie avec participation à taux garanti constant et à taux garanti flottant, dans laquelle nous incorporons les risques de crédit, de marché (bond) et économiques (changement de régime), et dont du portefeuille de référence est décrit par un modèle de diffusion à sauts double exponentiel à changement de régime. Nous fournissons des formules de forme fermée pour la valeur du contrat en utilisant une transformation de Laplace ou de Laplace-Fourier, dans laquelle seuls certains facteurs matriciels de Wiener-Hopf sont impliqués. Ensuite, le prix est obtenu en effectuant une inversion numérique de Laplace et Fourier et en mettant en oeuvre la factorisation matricielle de Wiener-Hopf. En comparant les résultats avec les simulations de Monte-Carlo, nous montrons que notre méthode de tarification est facile à mettre en œuvre et exacte. Nous montrons également que le contrat à taux garanti flottant est un produit plus risqué mais plus rentable que le contrat à taux garanti constant. Deux stratégies de couverture sont introduites pour couvrir les risques de saut et de changement de régime dans les contrats avec participation. Ce chapitre est basé sur le document de [Le Courtois, Quittard Pinon, and Su \(2018\)](#), "Tarification et couverture des contrats de participation défaillants avec basculement de régime et risque de renversement".

Le chapitre 4 étudie un contrat d'assurance optimal consistant à envisager une couverture étendue des assurés comprenant à la fois des précurseurs de risques et des amoureux des risques, en supposant que les assurés ont une tolérance au troisième degré. Nous indiquons la différence de forme de contrat optimale entre les avertisseurs de risques et les amateurs de risques. En les considérant comme une cible, nous montrons que la forme d'assurance optimale est un double contrat de limitation des pertes, ce qui montre que, y compris les passionnés du risque, le contrat ne change que de manière modeste. Nous limitons les contrats d'intérêt aux types convexes et les modifications optimales des contrats se transforment en une assurance contre les pertes de change ou en une assurance contre les risques de changement de change, qui dépend du coefficient de variation de la perte conservée. Enfin, nous montrons que le contrat optimal peut réduire à la forme privilégiée par les avertis-

## 1.2. Sujets spécifiques

---

seurs de risque en présence de risque de fond. Ce chapitre est basé sur le document de [André, Le Courtois, and Su \(2018\)](#), "Assurance optimale sous risque de troisième degré".

Le chapitre 5 développe un modèle stochastique d'accentuation du gradient fréquence-gravité, dans lequel nous appliquons l'algorithme d'accentuation du gradient stochastique pour estimer les modèles de régression marginaux des composantes de fréquence et de gravité des réclamations et introduisons la dépendance en traitant la fréquence des réclamations comme un prédicteur du modèle de régression pour la réclamation moyenne. gravité. Le modèle est capable d'adapter une relation non linéaire flexible entre la fréquence des réclamations (gravité) et les prédicteurs et de capturer les interactions complexes entre les prédicteurs et la dépendance non linéaire entre la fréquence des réclamations et la gravité. Une étude de simulation montre l'excellente performance de prévision de notre modèle. Ensuite, nous démontrons l'application de notre modèle avec une donnée de réclamation d'assurance automobile française. Les résultats montrent que le modèle proposé est supérieur aux autres modèles de pointe. Ce chapitre est basé sur le document de [Su and Bai \(2018\)](#), "Modèle de réclamations d'assurance fréquence-sévérité de gradient stochastique".



# Introduction

## Contents

<b>2.1</b>	<b>Background</b>	<b>17</b>
2.1.1	Life and non-life insurance	18
2.1.2	The limitations of traditional actuarial methods	21
2.1.3	Objective	23
<b>2.2</b>	<b>Specific topics</b>	<b>24</b>
2.2.1	Participating contracts and risk sources	24
2.2.2	Contract design	27
2.2.3	Frequency-Severity model	28
2.2.4	Thesis structure	28

## 2.1 Background

Risk exists in almost every human activity. When planning a picnic, there is a risk that it will rain. When ordering a theatre ticket, there is a risk of selling out. When driving a car, there is a risk of a traffic accident. There is a large variety of risks that influence us in the life. In most cases, the risk has a small probability to incur loss, whereas the loss size is usually large. For the individual, it is difficult to bear the cost induced by serious losses. One good strategy is to make all the persons who face the same kind of risk share the costs. Each person puts a small amount of money into a cash pool and the cash pool serves to cover the loss. One implementation method is to transfer the risks to a third party and the third party helps to compensate for the losses. This risk-sharing demand spawns a huge insurance industry. The insurance firms issue insurance policies to implement the risk sharing among individuals. A variety of insurance policies are designed to satisfy the sharing demand of various risks, and the individuals purchase the necessary one to make shift of the risks. For instance, if your car is insured when a car accident occurs, you report the damage to the insurance firm and the firm can make compensation to your loss, where the amount depends on the written contract. Based on whether the protection aim is an individual life, the insurance contracts are categorized into life insurance and non-life insurance.

In a competitive market, it is of great importance for the insurance firms to charge a fair price. For instance, in auto insurance, if insurance companies charge

too little for young drivers and too much for old drivers, they will lose old drivers to competitors while attracting young drivers. This adverse selection issue leads the insurers to lose profitable and gain underpriced policies, both resulting in economic losses. Nowadays, the design of many life insurance contracts incorporates many covenants, such as interest rate guarantees, equity-linked policies, participating policies, etc. These covenants make the insurance companies face more risks, such as economic, market and credit risks. These risks should be considered in the pricing of these contracts. The ignorance of these risks will cause difficulties. For example, in the 1990s, many insurance companies declare the business failure. One reason is the underpricing of the contract due to the ignored risk from interest rate guarantee. The insurance company makes the appropriate design and pricing of contracts to ensure its competence and profits. Meanwhile, the insurance company need manage the risks arising from the issuance of these contracts. As some contracts have linked investments in financial markets, the hedging becomes a useful risk management tool to the insurance company. Thus, the design, pricing and hedging of insurance contracts are significant to the insurance company and the efficiency of insurance market.

### 2.1.1 Life and non-life insurance

Life insurance deals with two hazards that is probably faced by an individual in his lifetime. One is dying prematurely leaving a dependent family to take care of itself and another one is living too old to have any means of support. In the former case, life insurance provides full protection against death risk of the insured. In case of death, the assured compensation will be fully paid. In the latter case, life insurance also works with its long-term saving function. The small premium paid by easy installments can be accumulated much after a long period. Then, the contract provides the payment of an amount at contract maturity or periodically at some specified dates. Surely, the insured has to pay the premium to the insurer periodically during the lifetime of the contract. It is worth noting that life insurance can be surrendered after a specified period, from which the insured will obtain a proportion of premium payment. There are three types of life insurance, including whole life insurance, term life insurance, and annuity. A whole life insurance pays out a lump sum benefit only on the death of the insured. A term life insurance provides the beneficiary with the policy amount at maturity or at death time of the insured if the insured dies before the maturity. For the annuity, when the term of the annuity expires, the insurer pays the policy amount to the insured periodically, as long as the insured is alive.

Non-life insurance, also known as property and casualty insurance, deals with the exposure of risk of individuals and property, such as the loss caused by sickness, fraud, accident, fire, windstorm, earthquake, theft, etc. The insured objects cover the individual, home, car, ship, luggage, etc. The contract bears the indemnity feature, for which the insurer compensates the loss suffered by the insured on the occurrence of an uncertain event. Generally, the contracts are short-term, for in-

## 2.1. Background

---

stance, one year, and the insured have to update the contracts every year if requiring a long-term protection. The character of the insured objects can be roughly classified as person, property and interest. For person, there are accident and sickness insurance. For property, there are numerous examples such as fire insurance, motor vehicle damage insurance, marine hull insurance, etc. As with interest, there are liability insurance, consequential loss insurance, fraud insurance, etc.

The detail comparison of distinct features between life insurance and non-life insurance is listed in Table 1.

**Table 1** Comparison between life insurance and non-life insurance

Feature	Life insurance	Non-life insurance
risk coverage	life risk of an individual	any risk of individuals and property apart from the risk of life, such as the loss caused by sickness, fraud, accident, fire, windstorm, earthquake, theft, etc
contract style	a form of investment	a contract of indemnity
contract term	long term	short term
payment time	either on the death of the insured or on maturity	on the occurrence of uncertain event
premium	periodical payments during the lifetime of the contract	a lump sum payment
insurable interest	present at the time of contract	present both at the time of contract and at the time of loss
policy value	based on the specific premium payment	limit to the actual loss suffered, irrespective of the policy amount
savings	long-term saving	no saving

Nowadays, global insurance business has developed to a certain degree. Table 2 lists top 10 regions of insurance business in 2017 and we can see that the insurance market has been very large in developed countries, no matter life insurance or non-life insurance business. The market quota of top 3 countries reaches almost 50% of insurance business in global market, which exhibits a large market potential yet to be developed, especially for the emerging markets.

The large-scale insurance business makes high demands on risk management of insurance industry. Risk is the foundation of insurance industry, for which insurance policy transfers lots of risks from the insured to the insurance firms. As mentioned before, the complex design of insurance contracts make insurance firms confront many different risks. So the insurance firms have to take care of these risks. The Solvency II and III Directive, as two regulatory frameworks for the insurance industry, have required insurance firms to focus on managing all of the risks they are facing. The regulation standards establish a new set of capital requirements, valuation techniques and governance and reporting standards, where assets and liabilities are measured to ensure enough capital to be hold against insurers' risks. Currently, the Solvency II is the widely used standard of regulation, which is split into three pillars, where pillar 1 employs market-consistent models to value the assets and liabilities and then calculate the regulatory capital requirements, pillar 2 sets the

**Table 2** Top 10 regions of insurance business in 2017 (US\$ millions)

Rank	Region	Total premiums		Life premiums	
		Amount	Share of global market	Amount	Share of global market
1	United States	1,377,114	28.15%	546,800	20.58%
2	Mainland (China)	541,466	11.07%	317,570	11.95%
3	Japan	422,050	8.63%	307,232	11.56%
4	United Kingdom	283,331	5.79%	189,833	7.14%
5	France	241,603	4.94%	153,520	5.78%
6	Germany	222,978	4.56%	96,973	3.65%
7	South Korea	181,218	3.70%	102,839	3.87%
8	Italy	155,509	3.18%	113,947	4.29%
9	Canada	119,520	2.44%	51,592	1.94%
10	Taiwan (China)	117,474	2.40%	98,602	3.71%

Non-life premiums	
Amount	Share of global market
830,315	37.16%
223,876	10.02%
114,818	5.14%
93,499	4.18%
88,083	3.94%
126,005	5.64%
78,378	3.51%
41,562	1.86%
67,927	3.04%
18,873	0.84%

Source: Swiss Re, sigma, No. 3/2018.

## 2.1. Background

---

Solvency Capital Requirement (SCR) and pillar 3 provides a private annual report to supervisors, and a solvency and financial information report to public. The pillar 1 encourages insurance firms to develop their own internal models for valuation. In order to improve competency, the major insurance firms have made lots of efforts on this side. Thus, to meet the challenges of risk management and strict regulation, the accurate pricing and hedging of insurance contracts have become extremely important.

### 2.1.2 The limitations of traditional actuarial methods

Traditional life insurance contracts mainly focus on mortality protection. However, from investment sides, the insureds are aware of investment opportunities in financial market and have the demand to enjoy the benefits of financial investment in conjunction with mortality protection. In order to attract the insureds, the insurance companies have incorporated this investment demand into the design of insurance contracts and have developed many modern contract types, such as unit-linked insurance, variable annuities and segregated fund contracts, etc. These new contracts often make promised payments on death or maturity. Some or all of the premiums are invested in an equity fund and the policyholders share the profits of the financial investments. Thus, in modern insurance, the financial risk is an important risk source that can't be ignored. Nowadays, modern life insurance contracts have dominated life insurance market, especially in developed countries, such as America, Canada, France, Germany, Australia, etc.

In the past, the actuarial techniques are the mostly used tool for the pricing of life insurance contracts. However, these methods are not fit for the valuation of modern insurance contracts. The reasons are twofold. On the one hand, the actuarial techniques only can assess and manage insurance risk, but can do nothing for financial risk. Because their management of risk relies heavily on diversification. With lots of insurance contracts on independent lives, the central limit theorem ensures the little uncertainty of the total claims, which enables the actuarial techniques to utilize deterministic method for pricing. For instance, an insurance company sells 10,000 insurance contracts to independent lives, each having a probability of claim of 0.03. If these contracts are traditional ones, the best estimate of mortality rate is 0.03 and the actuarial technique makes an adjustment to the best estimate value, for instance, 0.04. Then, the probability that the actual mortality rate over 0.04 is less than  $10^{-8}$  and almost all the mortality risk is absorbed. Nonetheless, for the modern insurance contracts, the inherent financial risk is a systematic or nondiversifiable risk. When the linked investment doesn't perform well, all the contracts are affected simultaneously. Thus, the financial risk makes the contracts not as independent as traditional ones. In this case, the central limit theorem can no longer be used, and hence the actuarial techniques also can't work. For example, some contracts make the promised payment. Suppose that the insurance company makes compensation for the shortage between the promised payment and the funds value if the investment has bad performance. We assume the probability that the invest-

ment doesn't perform well is 0.03 and then there is a probability of 0.03 that all the contracts need the insurance company to make the compensation. This systematic risk can't be handled by the traditional actuarial techniques.

On the other hand, as we mentioned before, modern life insurance contracts bear many covenants such as bonus and surrender options, interest rate guarantees, and participating policies, etc. These covenants are embedded options in the contracts and they should be accurately priced. For example, the interest rate guarantee makes the account of the policyholders credited with a minimum rate each year. At the time of issuance, if the guaranteed interest rate is much lower than market interest rates, no premium for these guarantees has no bad consequence. However, when market interest rate declines and remains in a low level for long, the fixed guarantee rate will result in a large liability burden to insurance company. This is the reason why lots of insurance businesses default in Europe, Japan, and the United States in the 1990s. Thus, the embeded options really need to be priced. However, the traditional actuarial techniques can do nothing on this side.

The methods in financial engineering field are natural solutions to solve the pricing and risk management problem of modern insurance contracts. In fact, modern insurance contracts can be treated as a special kind of financial derivative that combines mortality protection. [Boyle and Schwartz \(1977\)](#); [Brennan and Schwartz \(1979\)](#); [Brennan, Schwartz, et al. \(1976\)](#) are seminal works to employ modern option pricing theory for pricing life insurance contracts. There has been a large body of literature on the improvement of market methods for the pricing and risk management of modern insurance contracts. See [Bernard, Le Courtois, and Quittard-Pinon \(2005\)](#); [Coleman, Kim, Li, and Patron \(2007\)](#); [Coleman, Li, and Patron \(2006\)](#); [Kélani and Quittard-Pinon \(2017\)](#); [Lin, Tan, and Yang \(2009\)](#); [Siu \(2005\)](#), to name just a few. The market methods are also promoted by the regulatory requirement of the Solvency II that the assets and liabilities of insurance company need to be priced using market models, and also by serveral international standards such as the International Accounting Standard Board (IASB), the Financial Accounting Standard Board (FASB) and the International Financial Reporting Standard (IFRS), etc, which require life insurance companies to disclose their accounting information at fair value. The study of the market methods for pricing modern insurance contracts has been of great interests in both practice and academics.

When pricing non-life insurance policies, there are many important factors, such as the claim loss, safety loading, administrative costs, etc. Apparently, among them the most crucial factor is the claim loss. Thus, we mainly focus on the modeling of claim loss. The amount of expected claim loss highly depends on the characteristics of an individual policy. For example, in an auto insurance claim, the age, gender and motor vehicle record points of the policyholder, per captial income or population density of the policyholder's residential area, age and model of the vehicle, etc, all have significant effects on the claim size. Thus, we need make use of predictive models.

Traditional acturial technologies choose the generalized linear models (GLM) to model the claim size. Two standard GLM-based models are the GLM frequency-

## 2.1. Background

---

severity model and the GLM Tweedie compound Poisson model. The former models the claim size by two parts, claim frequency and claim severity, where the claim frequency examines the number of claims by employing a Poisson or negative binomial regression, and the claim severity takes care of the amount of claims conditional on occurrence, by using a gamma or inverse Gaussian regression. The latter models the claim size by a Poisson sum of i.i.d. gamma random variables. The resulting Tweedie distribution belongs to the exponential dispersion family and hence the GLM Tweedie model can be built to predict claim size directly. Though the GLM models are widely used, there is an obvious limitation. The linear form of the GLM models is too strict for real applications, which affects its prediction accuracy. For example, in auto insurance, the nonlinear relation between age and claim size is well documented. The generalized additive models (GAM) overcome the restrictive linear form of the GLM model, by modeling the continuous variable with smooth functions. However, the GAM model is also restricted to the additive form and the complex interaction effects have to be manually identified. In practice, it is difficult to find all the interaction effects, especially with many variables. Missing some important interactions will adversely affect prediction accuracy. Thus, we need more flexible models.

Machine learning algorithms are the natural solutions that we need. The algorithms learn the model structure from data, which is capable of fitting a flexible non-linear relation between response variable and predictors and capturing complex and higher order interactions among predictors. Further, the algorithms can automatically complete feature selection. They have proven to be very successful in a variety of applications, due to the high prediction accuracy. The application of machine learning algorithms in non-life insurance have become a hot research topic. [Wuthrich and Buser \(2017\)](#) and [Noll, Salzmann, and Wuthrich \(2018\)](#) have empirically shown that regression trees, boosting and neural networks outperform the GLM model in the prediction of claim frequency. [Yang, Qian, and Zou \(2018\)](#) develops a gradient boosting Tweedie compound Poisson model and show that the model is superior to other state-of-the-art models. The ASTIN (Actuarial Studies in Non-life Insurance) of the International Actuarial Association (IAA) encourages the adoption of machine learning technologies in insurance domain. The concept of Insurance Technology (InsurTech) is developed in the China InsurTech Development White Paper, in which one key point is related to the application of machine learning technologies. Thus, the study of machine learning algorithms for pricing non-life insurance has become an interesting topic in both practice and academics.

### 2.1.3 Objective

This thesis improves the design, pricing and hedging of insurance contracts, by making use of the tools in financial engineering, decision theory and machine learning. Specifically, we develop an advanced model for the pricing of a popular type of life insurance contracts, i.e., participating life insurance contracts, in the presence of credit, market (jump), economic (regime switching) risks. We also introduce two

hedging strategies, i.e., semi-static hedging and dynamic hedging, to hedge default, jump and regime switching risks in the participating contracts. Besides the pricing of life insurance contract, we also investigate the optimal design of insurance policy by considering a wide coverage of insureds that include both risk averters and risk lovers and then we show that the optimal contracts can be in the forms of dual limited stop-loss, change-loss, dual change-loss, and stop-loss in different settings. As for the pricing of non-life insurance, we develop a stochastic gradient boosting frequency-severity model, in which the stochastic gradient boosting algorithm is employed to estimate distributions of claim frequency and severity and the nonlinear dependence between claim frequency and severity is accommodated by treating the claim frequency as a predictor in the regression model of claim severity. We show the proposed model is superior to other state-of-the-art models.

## 2.2 Specific topics

In this section, we introduce the participating contracts and explain the important risk sources that we incorporate into the pricing. Then, we illustrate the optimal contract design with a simple example and indicate the difference between our work and previous works. We also exhibit the frequency-severity model within the GLM framework and point out the restrictions of the currently popular frequency-severity models. Finally, we show the basic contents of our works.

### 2.2.1 Participating contracts and risk sources

Participating life insurance contracts are one of the most popular life insurance contracts. In these contracts, the policyholder not only receives a minimum guaranteed return, but also shares the profits of financial investment exceeding the guaranteed return. The policyholder pays a lump sum to the insurance company and the insurer invests and manages the funds in a reference portfolio. The insurer credits interest at or above a minimum guaranteed rate to the insured every year. The net difference between the market value of the reference portfolio and the book value of the policyholder's account is the bonus reserve. If the terminal bonus reserve is positive, the policyholder receives a terminal bonus. If the insurer defaults during the lifetime of the contract, the policyholder will receive the remaining assets. See comprehensive discussion on different contractual features of participating contracts in [Grosen and Jørgensen \(2000\)](#).

As stated before, the credit risk need be priced. [Briys and De Varenne \(1994, 1997\)](#) provide a general framework for the valuation of the participating contracts where they take the credit risk of the insurer explicitly into account. However, the default is only allowed to happen at maturity. [Grosen and Jørgensen \(2002\)](#) correct this weakness, through introducing continuous monitoring of the solvency of the firm. We take a look at the specification of the participating contracts using the framework in [Grosen and Jørgensen \(2002\)](#).

## 2.2. Specific topics

---

The insured purchases a life insurance contract by a single sum premium payment  $L_0$ . The insurer manages an investment in a reference portfolio and the funds are partly financed with the premium  $L_0$ . Denote by  $A_0$  the initial value of the funds. Then, the policyholders have  $\alpha = \frac{L_0}{A_0}$  shares of the funds and enjoy the benefits of excess investment return. The insurer promises the policyholders a minimum rate  $r_g$  during the life of the contract. Then, a promised maturity payment is  $L_T = L_0 e^{r_g T}$ , where  $T$  is the maturity of the contract. The promised payment can be honored if the terminal value of the portfolio is enough large, i.e.,  $A_T \geq L_T$ . Otherwise, the contract defaults and the policyholders seize the remaining portfolio value  $A_T$ . Besides the promised maturity payment, policyholders are entitled to receive a bonus if the investment of the funds performs enough well. The policyholders share a proportion  $\delta$  of the funds value exceeding the promised payment, i.e.,  $\delta(\alpha A_T - L_T)_+$ , where  $\delta$ , named as the participation rate, represents the participation level of the policyholders in the upside potential of financial investments. Then, the policyholders receive a payoff at maturity  $T$ ,

$$\Theta_L(T) = \begin{cases} A_T & \text{if } A_T < L_T \\ L_T & \text{if } L_T \leq A_T \leq \frac{L_T}{\alpha} \\ L_T + \delta(\alpha A_T - L_T) & \text{if } A_T > \frac{L_T}{\alpha} \end{cases}.$$

This payoff can be written in a compact form as

$$\Theta_L(T) = L_T + \delta(\alpha A_T - L_T)_+ - (L_T - A_T)_+.$$

where the three terms on the right-hand side represent the promised maturity payment, the bonus option and a short put option related to the default happening at maturity, respectively.

The above payoff structure is given under an unrealistic assumption that the default of the insurer only happens at the maturity of the contract. In practice, the default of the insurer can happen at any time  $t$ . Thus, suppose that at any time  $t$  the regulatory boundary is proportional to the promised payment,  $L_t = L_0 e^{r_g t}$ , and the solvency of the insurer is continuously monitored by regulatory authorities. Let  $\kappa$  be the boundary level parameter and the default of the insurer happens at time

$$\tau = \inf\{t \geq 0 : A_t \leq \kappa L_0 e^{r_g t}\}.$$

The remaining portfolio value is  $A_\tau$ . Then, denote by  $\{r_t, t \geq 0\}$  the dynamics of interest rates. The pricing formula under the risk-neutral measure  $Q$  can be written in the form of

$$V = E_Q \left( e^{-\int_0^T r_s ds} (L_T + \delta(\alpha A_T - L_T)_+ - (L_T - A_T)_+) \mathbb{1}_{\tau \geq T} + e^{-\int_0^\tau r_s ds} A_\tau \mathbb{1}_{\tau < T} \right).$$

Besides the credit risk, the financial investments make the contracts face market and economic risks. The jumps phenomenon of asset price, i.e., rare large movements, is often observed in the financial market. This phenomenon is induced by external shocks from some extreme events such as policy changes. The asymmetric leptokurtic feature of financial asset return is well documented, see, for instance, [Cont \(2001\)](#). These features can be well explained by jump phenomenon. [Merton \(1976\)](#) is the seminal work that models jump risk in the dynamics of asset prices. [Ball and Torous \(1983\)](#); [Jarrow and Rosenfeld \(1984\)](#); [Jorion \(1988\)](#) provide convincing evidences to support the presence of jumps in asset price. In option pricing literature, [Bakshi, Cao, and Chen \(1997\)](#); [Duffie, Pan, and Singleton \(2000\)](#); [Kou \(2002\)](#); [Kou and Wang \(2004\)](#); [Merton \(1976\)](#); [Naik and Lee \(1990\)](#); [Pan \(2002\)](#) demonstrate that incorporating jumps is essential to explain observed option price, such as an empirical phenomena called "volatility smile" observed in option markets. Further, an important class of stochastic models, named Lévy models, is developed to take account of jump risk in the financial market, such as the Variance Gamma (VG), the Normal Inverse Gaussian (NIG), the Tempered Stable (TS), the Generalized Hyperbolic (GH), the Meixner and the CGMY processes, etc. There has been a large stream of literature on the derivatives pricing with Lévy models. See [Asmussen, Avram, and Pistorius \(2004\)](#); [Cariboni and Schoutens \(2007\)](#); [Carr and Wu \(2003\)](#); [Cont and Voltchkova \(2005a,b\)](#); [Fusai and Meucci \(2008\)](#); [Kou \(2002\)](#); [Kou and Wang \(2004\)](#); [Tankov \(2003\)](#), to name only a few. The evidence of jump risk is very sufficient.

Regime switching risk is an important economic risk. This economic risk is the structural changes in the macroeconomic environment or in the business cycles, which induces changes in the dynamic patterns of asset price or in the term structure of interest rates. The risk has been examined in many empirical studies, such as [Bollen, Gray, and Whaley \(2000\)](#); [Engel and Hamilton \(1990\)](#); [Guidolin and Timmermann \(2005, 2008\)](#). They provide convincing supports for the presence of regime switching in asset price. One prominent class of stochastic models, named regime switching model, is developed to take care of regime switching risk. [Hamilton \(1989\)](#) popularizes the regime switching time series model in the economic and econometric literature. [Hardy \(2001\)](#) shows the empirical success of the regime switching model, by fitting the long-term monthly data from the Standard and Poor's 500 and the Toronto Stock Exchange 300 indices. The regime switching models have been widely used in different areas in finance, such as option valuation, asset allocation, and risk management, etc. See [Buffington and Elliott \(2002\)](#); [Chollete, Heinen, and Valdesogo \(2009\)](#); [Elliott, Chan, and Siu \(2005\)](#); [Guidolin and Timmermann \(2007\)](#); [Guo \(2001\)](#); [Zhang and Guo \(2004\)](#); [Zhou and Yin \(2003\)](#), to name just a few.

The life insurance contracts are relatively long dated. There can be substantial changes in economic conditions over a long period of time. Thus, it is essential to incorporate regime switching risk into the valuation of life insurance products. In general, we understand jump and regime switching as short-term and long-term risks, respectively. Thus, we complete the pricing of participating life insurance contracts in the presence of credit, jump and regime switching risks.

## 2.2. Specific topics

---

### 2.2.2 Contract design

The optimal insurance design has attracted great attention from both practitioners and academics. Since the seminal work of Arrow (1963), there has been an extensive literature on this topic, for instance, see Bernard, He, Yan, and Zhou (2015); Bernard and Tian (2009); Chi (2017); Chi and Tan (2011); Chi and Zhou (2017); Cummins and Mahul (2004); Gollier and Schlesinger (1996); Kaluszka (2001, 2005); Kaluszka and Okolewski (2008); Raviv (1979); Young (1999); Zhou, Wu, and Wu (2010).

We take a look at the optimal insurance design with a simple example. Let  $X$  be an amount of loss that is faced by an insured. The insurance design works on the partition of the loss into  $f(X)$  and  $R_f(X)$ , where  $f(X)$  represents part of the loss ceded to the insurer and  $R_f(X)$  captures the remaining loss retained by the insured. The design seeks a goal of maximizing the expected utility of final wealth of an insured. Suppose that the insured is risk averse with an increasing concave utility function  $u(x)$ . We assume that the insured with an initial wealth  $w_0$  purchases an insurance contract by a premium payment  $P$ . Then, the final wealth of the insured is  $w(X) = w_0 - R_f(X) - P$ . The optimal insurance contract maximizes the expected utility of the final wealth of the insured, i.e.,

$$\arg \max_f E(u(w(X))).$$

In terms of the premium  $P$ , there are many premium principles. Young (2014) lists eleven common premium principles. If we use the expected value premium principle, where the insurance premium depends only on the expected value of the indemnity, we have the the solution of the above optimization problem as the function  $f(x) = (x - b)_+$ , where  $b$  is a fixed amount deductible.

This is a basic example of optimal insurance design. The afore-mentioned works enrich the contents of this field. The optimization criterion sometimes uses the minimization of the insured's risk exposure with some specific risk measures, such as value at risk and conditional value at risk. See, for instance, Cai and Tan (2007); Cai, Tan, Weng, and Zhang (2008); Chi and Tan (2011); Tan, Weng, and Zhang (2009). Chi (2017) and Chi and Wei (2018) also consider the higher-order risk attitudes of the insured. However, these works all assume the insured to be risk averse. Though a majority of people are found to be risk averse, there is still a significant minority of them who are risk lovers. Deck and Schlesinger (2014) examines the higher order risk preferences of risk averters and risk lovers and finds the following two facts. First, risk averters dislike an increase in risk at every degree. Second, risk lovers like risk increases of even degrees, but dislike increases of odd degrees. Thus, both risk averters and risk lovers have the same risk attitudes at odd degrees, such as third-order prudence. Crainich, Eeckhoudt, and Trannoy (2013) confirms that the risk lovers are prudent. Therefore, we only assume that the insured is prudent so as to cover both of risk averters and risk lovers. Then, with a wide coverage of insureds, we show the specific forms of the optimal contract under different settings.

### 2.2.3 Frequency-Severity model

The frequency-severity model is a widely used model of insurance claims. We illustrate the frequency-severity model within the GLM framework. Let  $x = (x_1, \dots, x_p)'$  and  $y = (y_1, \dots, y_q)'$  be the predictor variables of claim frequency and severity, respectively. Denote by  $N$  and  $X_i$  the respective random variables of claim frequency and severity. With two link functions  $g$  and  $f$ , we have the regression models as

$$g(\nu) = \alpha x \quad \text{and} \quad f(\mu) = \beta y,$$

where  $\nu = E(N|x)$  and  $\mu = E(X_i|y)$ , and  $\alpha$  and  $\beta$  are two vectors of regression coefficients with size  $1 \times p$  and  $1 \times q$ , respectively. The inference of GLM models gives the values of regression coefficients  $\alpha$  and  $\beta$ , and other parameters of the distributions of  $N$  and  $X_i$ . With the known distributions of claim frequency  $N$  and severity  $X_i$ , we have the distribution of aggregate claim amount as

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

Then, based on the aggregate claim amount, we can calculate the premiums for the non-life insurance policy.

As we mentioned before, the linear or additive forms of the GLM and GAM frequency-severity models are too rigid for real applications. Further, these models often assume the claim frequency and severity to be independent. In reality, the claim frequency and severity are often dependent. For example, home insurance claims due to flood tend to be both large and frequent in the affected area. For automobile insurance, the claim frequency and severity are often negatively correlated because drivers who claims several times per year generally only involves in some minor accidents. Thus, the dependence between claim frequency and severity need to be modeled. Erhardt and Czado (2012); Frees, Gao, and Rosenberg (2011); Gschlößl and Czado (2007) capture the dependence by treating claim frequency as a predictor variable in the regression model of average claim severity and Czado, Kastenmeier, Brechmann, and Min (2012); Krämer, Brechmann, Silvestrini, and Czado (2013); Shi, Feng, and Ivantsova (2015) employ the parametric copulas to model the joint distribution of claim frequency and average claim severity. Thus, we develop a stochastic gradient boosting frequency-severity model to overcome the restricted forms of the GLM and GAM models, where we treat claim frequency as a predictor variable in the gradient boosting regression model of the average claim severity to flexibly capture the nonlinear dependence between claim frequency and severity.

### 2.2.4 Thesis structure

This thesis focuses on the design, pricing and hedging of insurance contracts to improve the risk management of insurance company and the efficiency of insurance market. Chapter 2 reviews some basic knowledge of Lévy process and regime

## 2.2. Specific topics

---

switching Lévy process for chapter 3, of some stochastic order notions, higher-order degree risk and stochastic dominance for chapter 4, and of the GLM and GAM model, regression tree and gradient boosting algorithm for chapter 5.

Chapter 3 develops a transform-based approach for the pricing of participating life insurance contracts with a constant guaranteed rate and with a floating guaranteed rate, in which we incorporate credit, market (jump), economic (regime switching) risks, and where the evolution of the reference portfolio is described by a regime switching double exponential jump diffusion model. We provide closed-form formulas for the contract value by using a Laplace or Laplace-Fourier transform, where only some matrix Wiener-Hopf factors are involved. Then, the price is obtained by performing numerical Laplace and Fourier inversion and by implementing the matrix Wiener-Hopf factorization. By comparing the results with Monte-Carlo simulations, we show that our pricing method is easy to implement and is accurate. We also show that the contract with a floating guaranteed rate is a riskier but more worthy product when comparing to the contract with a constant guaranteed rate. Two hedging strategies are introduced to hedge jump and regime switching risks in the participating contracts. This chapter is based on the paper of [Le Courtois, Quittard Pinon, and Su \(2018\)](#), "Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk".

Chapter 4 studies an optimal insurance contract of considering a wide coverage of insureds that include both risk averters and risk lovers, by assuming that the insureds are third degree risk averse. We indicate the difference of optimal contract form between risk averters and risk lovers. Treating them as one target, we show that the optimal insurance form is a dual limited stop-loss contract, which manifests that including risk lovers only change the contract in the small loss part. We narrow down contracts of interest to convex types and the optimal contract changes into a change-loss insurance or a dual change-loss insurance, which depends on the coefficient of variation of the retained loss. Finally, we show that the optimal contract can reduce to the form favored by risk averters in the presence of background risk. This chapter is based on the paper of [André, Le Courtois, and Su \(2018\)](#), "Optimal Insurance under Third Degree Risk".

Chapter 5 develops a stochastic gradient boosting frequency-severity model, where we apply stochastic gradient boosting algorithm to estimate the marginal regression models of both claim frequency and severity components and introduce dependence by treating claim frequency as a predictor in the regression model for the average claim severity. The model is capable of fitting a flexible nonlinear relation between claim frequency (severity) and predictors and capturing complex interactions among predictors and nonlinear dependence between claim frequency and severity. A simulation study shows the excellent prediction performance of our model. Then, we demonstrate the application of our model with a French auto insurance claim data. The results show the proposed model is superior to other state-of-the-art models. This chapter is based on the paper of [Su and Bai \(2018\)](#), "Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims".



# Background Knowledge

---

## Contents

---

<b>3.1 Lévy process</b> . . . . .	<b>31</b>
3.1.1 Definition and properties . . . . .	32
3.1.2 Examples of Lévy processes . . . . .	34
3.1.3 Itô formula and measure transform . . . . .	39
<b>3.2 Regime Switching Lévy Process</b> . . . . .	<b>41</b>
3.2.1 Definition and properties . . . . .	41
3.2.2 Itô formula and measure transform . . . . .	41
<b>3.3 Stochastic orders</b> . . . . .	<b>42</b>
3.3.1 The convex order . . . . .	42
3.3.2 The $m$ -convex order . . . . .	43
3.3.3 The monotone convex order . . . . .	45
3.3.4 The $m$ -increasing convex order . . . . .	46
<b>3.4 Higher-order degree risk and stochastic dominance</b> . . . . .	<b>47</b>
3.4.1 $n^{\text{th}}$ degree risk . . . . .	47
3.4.2 $n^{\text{th}}$ degree stochastic dominance . . . . .	48
<b>3.5 The GLM and GAM model</b> . . . . .	<b>48</b>
3.5.1 Generalized linear model . . . . .	48
3.5.2 Generalized additive model . . . . .	50
<b>3.6 Machine learning algorithms</b> . . . . .	<b>51</b>
3.6.1 Regression trees . . . . .	51
3.6.2 Gradient boosting . . . . .	53

---

## 3.1 Lévy process

As we use a special regime switching Lévy model for pricing the participating life insurance contract, we briefly review some basic knowledge of Lévy process and regime switching Lévy process. We begin with the introduction of Lévy process.

### 3.1.1 Definition and properties

**Definition 1. (Lévy Process)** A stochastic process  $X = \{X_t, t \geq 0\}$ , defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , is a Lévy process if the following conditions hold:

- (i) The paths of  $X$  are  $\mathbb{P}$ -almost surely right-continuous with left limits.
- (ii)  $P(X_0 = 0) = 1$ .
- (iii) For  $0 \leq s \leq t$ ,  $X_t - X_s$  is equal in distribution to  $X_{t-s}$ .
- (iv) For  $0 \leq s \leq t$ ,  $X_t - X_s$  is independent of  $\{X_u, u \leq s\}$ .

The distribution of a Lévy process at any time  $t$  is completely determined by the distribution of  $X_1$ . If we define a distribution of a Lévy process at a single time  $t$ , we have specified the whole Lévy process. For example, if the distribution of  $X_1$  has a characteristic function  $\phi(u)$ , the distribution of an increment of  $X$  over  $[s, s + t]$ ,  $s, t \geq 0$ , i.e.,  $X_{t+s} - X_s$ , has a characteristic function  $(\phi(u))^t$ . Thus, for any time  $t > 0$ , the random variable  $X_t$  belongs to the class of infinitely divisible distributions.

Let  $X$  be a random variable taking values in  $\mathbb{R}$  with law  $\mu_X$ . We say that  $X$  has an infinitely divisible distribution if, for each  $n = 1, 2, \dots$ , there exists a group of i.i.d. random variables  $Y_{1,n}, \dots, Y_{n,n}$  such that

$$X \stackrel{d}{=} Y_{1,n} + \dots + Y_{n,n},$$

where  $\stackrel{d}{=}$  denotes equality in distribution. Let  $\phi_X(u) = E(e^{iuX})$  be the characteristic function of  $X$ , where  $u \in \mathbb{R}$ . Denote by  $\mu^{*n}$  the  $n$ -fold convolution of a law  $\mu$ . We have the following proposition:

**Proposition 1.** *The following are equivalent:*

- (i)  $X$  is infinitely divisible;
- (ii) For each  $n = 1, 2, \dots$ , there exists another law  $\mu_n$  of a real-valued random variable such that  $\mu_X = \mu_n^{*n}$ ;
- (iii) For each  $n = 1, 2, \dots$ , there exists a characteristic exponent of a probability distribution  $\phi_n$  such that  $\phi_X(u) = n\phi_n(u)$ , for all  $u \in \mathbb{R}$ .

Some common examples of infinitely divisible distributions are the Gaussian distribution, the Poisson distribution, the gamma distribution and the  $\alpha$ -stable distributions. For example, if  $X \sim N(\mu, \sigma^2)$ , then we have i.i.d. random variables  $Y_i, i = 1, \dots, n$  with distribution  $N(\frac{\mu}{n}, \frac{\sigma^2}{n})$  such that  $X \stackrel{d}{=} \sum_{i=1}^n Y_i$ . If  $X \sim P(\lambda)$ , then we have  $X \stackrel{d}{=} \sum_{i=1}^n Y_i$ , where  $Y_i, i = 1, \dots, n$  are i.i.d. with  $Y_i \sim P(\frac{\lambda}{n})$ .

There is a one-to-one mapping between Lévy process and infinitely divisible distributions. This relation is characterized by the following proposition:

### 3.1. Lévy process

**Proposition 2. (Infinite Divisibility of Lévy Processes)** (*Schoutens and Cariboni (2010)*) Let  $X = \{X_t, t \geq 0\}$  be a Lévy process. Then,  $X_t$  has an infinitely divisible distribution  $F$ . Conversely, if  $F$  is an infinitely divisible distribution, there exists a Lévy process  $X = \{X_t, t \geq 0\}$  such that the distribution of  $X_1$  is  $F$ .

Next, a general form of Lévy process can be given. Based on the Lévy-Itô decomposition, the Lévy process  $X = \{X_t, t \geq 0\}$  is in the form of:

$$X_t = \mu t + \sigma W_t + \int_0^t \int_{|x| \geq 1} x N(ds, dx) + \int_0^t \int_{|x| < 1} x (N(ds, dx) - v(dx) ds),$$

where  $\mu \in \mathbb{R}$ ,  $\sigma \in \mathbb{R}$ ,  $W$  is a standard Brownian motion,  $N$  is a Poisson random measure on  $([0, \infty) \times \mathbb{R}, \mathcal{B}([0, \infty)) \times \mathcal{B}(\mathbb{R}), dt \times v(dx))$ , and  $v$  is a measure on  $\mathbb{R} \setminus \{0\}$  with  $\int_{-\infty}^{\infty} (1 \wedge x^2) v(dx) < \infty$ . The  $(\mu, \sigma, v(dx))$ , termed as Lévy triplet, fully determines the Lévy process  $X$ . The measure  $v$  is called the Lévy measure of  $X$ , which indicates that jumps with sizes in the set  $A$  occur according to a Poisson process with parameter  $v(A) = \int_A v(dx)$ . One can see that the Lévy process consists of a linear Brownian motion, a compound Poisson process and a square-integrable martingale with an almost surely countable number of jumps with sizes less than 1 on each finite time interval. The Lévy-Khintchine formula gives the characteristic exponent  $\phi(u)$  of the Lévy process  $X$ :

$$\phi(u) = i\mu u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{\infty} (e^{iux} - 1 - iux \mathbf{1}_{\{|x| < 1\}}) v(dx),$$

where  $\mathbf{1}_A$  is the indicator function of  $A$ .

The Lévy process  $X$  is of finite variation if and only if  $\sigma = 0$  and  $\int_{-\infty}^{\infty} (1 \wedge |x|) v(dx) < \infty$ . In this case,  $\mu$  and  $-\int_{|x| < 1} xv(dx)$  can be combined as  $d = \mu - \int_{|x| < 1} xv(dx)$  and the Lévy process can be rewritten in the form of

$$X_t = dt + \int_0^t \int_{\mathbb{R}} x N(ds, dx).$$

If  $v(\mathbb{R}) < \infty$ , there is finite jumps in any finite interval and the process is said to be of finite activity. Otherwise, the process is said to be of infinite activity. The Lévy process  $X$  is a compound Poisson process with drift if and only if  $\sigma = 0$  and  $v(\mathbb{R}) < \infty$ . As the Brownian motion is of infinite variation, a Lévy process with a Brownian component is of infinite variation. A pure jump Lévy process, i.e., one with no Brownian component  $\sigma = 0$ , is of infinite variation if and only if

$\int_{|x|<1} |x|v(dx) = \infty$ . In this case, the sum of all jumps smaller than  $\varepsilon > 0$  does not converge, whereas the sum of the jumps compensated by their mean converges. The Lévy process  $X$  is a subordinator if and only if  $v(-\infty, 0) = 0$ ,  $\int_{(0,\infty)} (1 \wedge x)v(dx) < \infty$ ,  $\sigma = 0$  and  $d \geq 0$ . If  $v(-\infty, 0) = 0$  and  $X$  does not have monotone paths, such kind of Lévy process is referred to as a spectrally positive Lévy process. Conversely, if  $v(0, \infty) = 0$  and  $X$  does not have monotone paths, the Lévy process  $X$  is named as a spectrally negative Lévy process. These two classes of processes are called as spectrally one-side Lévy process.

### 3.1.2 Examples of Lévy processes

The Lévy processes often used in financial modeling can be separated into two classes. The first one is jump diffusion models, in which the diffusion part describes the normal evolution and the jumps characterize rare events. The Lévy process  $X$  of jump diffusion type has the following form:

$$X_t = \gamma t + \sigma W_t + \sum_{i=1}^{N_t} Y_i,$$

where  $\{N_t, t \geq 0\}$  is the Poisson process with parameter  $\lambda$  counting jump times of  $X$  and  $Y_i$  are i.i.d. random variables denoting jump sizes. Two popular models are the Merton jump diffusion model and the Kou model. We can look at some basic properties of these two models.

In the Merton jump diffusion model, the jump size  $Y_i$  follows a Normal distribution:

$$Y_i \sim N(\mu, \delta^2).$$

As the sum of independent normally distributed variables is normal, the

$$P(X_t \in A) = \sum_{k=0}^{\infty} P(X_t \in A | N_t = k) P(N_t = k)$$

entails the probability density of  $X_t$  as

$$p_t(x) = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\frac{(x - \gamma t - k\mu)^2}{2(\sigma^2 t + k\delta^2)}}}{k! \sqrt{2\pi(\sigma^2 t + k\delta^2)}}.$$

The similar way makes the European option in the Merton model can also be expressed in a series expansion, where each term is a Black-Scholes formula. The Lévy measure of the Merton model is

$$v(dx) = \frac{\lambda}{\delta\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\delta^2}} dx.$$

### 3.1. Lévy process

---

and the characteristic exponent is

$$\phi(u) = i\mu u - \frac{1}{2}\sigma^2 u^2 + \lambda \left( e^{i\mu u - \frac{1}{2}\delta^2 u^2} - 1 \right).$$

In the Kou model, the jump size  $Y_i$  follows an asymmetric double exponential distribution with the density

$$f_Y(y) = p\eta_1 e^{-\eta_1 y} \mathbf{1}_{\{y \geq 0\}} + q\eta_2 e^{\eta_2 y} \mathbf{1}_{\{y < 0\}}, \eta_1 > 1, \eta_2 > 0,$$

where  $p, q \geq 0$  and  $p + q = 1$ . The  $p$  and  $q$  denote the probability of an upward and downward jump. The means of positive and negative jump sizes are  $\frac{1}{\eta_1}$  and  $\frac{1}{\eta_2}$ , respectively. The probability distribution of  $X_t$  has semi-heavy tails. The Lévy measure of the Kou model is

$$v(dx) = (p\lambda\eta_1 e^{-\eta_1 x} \mathbf{1}_{\{x \geq 0\}} + q\lambda\eta_2 e^{\eta_2 x} \mathbf{1}_{\{x < 0\}}) dx$$

and the characteristic exponent is

$$\phi(u) = i\mu u - \frac{1}{2}\sigma^2 u^2 + \lambda \left( \frac{p\eta_1}{\eta_1 - iu} + \frac{q\eta_2}{\eta_2 + iu} - 1 \right).$$

In contrast to the Merton model, the memoryless property of exponential distribution enables the Kou model to obtain the explicit solution of Laplace transform of the first passage time problem, which is important in the barrier style derivatives pricing and the credit risk.

The second one consists of infinite activity models. In these models, one don't need to introduce a Brownian component. Because the Brownian component is always used to capture the frequent small moves while the jumps in these models are able to capture both rare large moves and frequent small moves. The empirical evidence shows that these models are typically not improved by adding a Brownian component. We list some popular models of this type.

There are two ways of defining a Variance Gamma process. One way is to subordinate Brownian motion with drift  $\theta$  by a Gamma process  $G = \{G_t, t \geq 0\}$ , whose parameters  $a = \frac{1}{\delta} > 0$  and  $b = \frac{1}{\delta} > 0$ . Denote by  $W = \{W_t, t \geq 0\}$  the standard Brownian motion independent from  $G$ . Let  $\sigma > 0$  and the Variance Gamma process  $X = \{X_t, t \geq 0\}$ , with parameters  $\sigma > 0, \delta > 0$  and  $\theta$ , can be defined as

$$X_t = \theta G_t + \sigma W_{G_t}.$$

The characteristic exponent is

$$\phi(u; \sigma, \delta, \theta) = -\frac{1}{\delta} \ln(1 - iu\theta\delta + \frac{1}{2}\sigma^2\delta u^2).$$

Another way is to define the Variance Gamma process as the difference of two independent Gamma processes as

$$X_t = G_t^1 - G_t^2,$$

where  $G^1 = \{G_t^1, t \geq 0\}$  is a Gamma process with parameters  $a = C$  and  $b = M$  and  $G^2 = \{G_t^2, t \geq 0\}$  is another independent Gamma process with parameters  $a = C$  and  $b = G$ . The parameters of two definitions have the following relation

$$C = \frac{1}{\delta} > 0,$$

$$G = \left( \sqrt{\frac{1}{4}\theta^2\delta^2 + \frac{1}{2}\sigma^2\delta} - \frac{1}{2}\theta\delta \right)^{-1} > 0,$$

$$M = \left( \sqrt{\frac{1}{4}\theta^2\delta^2 + \frac{1}{2}\sigma^2\delta} + \frac{1}{2}\theta\delta \right)^{-1} > 0.$$

In terms of parameters  $C, G, M$ , the characteristic exponent is rewritten in the form of

$$\phi(u; C, G, M) = C \ln \left( \frac{GM}{GM + (M - G)iu + u^2} \right).$$

The Lévy measure is

$$v(dx) = \begin{cases} Ce^{Gx}|x|^{-1}dx, & x < 0, \\ Ce^{-Mx}x^{-1}dx, & x > 0. \end{cases}$$

As  $v(\mathbb{R}) = \infty$ , the Variance Gamma process is of infinite activity. As  $\int_{|x|<1} |x|v(dx) < \infty$  and  $\sigma = 0$ , the Variance Gamma process is of finite variation. The Lévy triplet is given by  $(\mu, 0, v(dx))$ , where

$$\mu = \frac{-C(G(e^{-M} - 1) - M(e^{-G} - 1))}{MG}.$$

The Normal Inverse Gaussian process can be built by subordinating drifted Brownian motion with an Inverse Gaussian process  $I = \{I_t, t \geq 0\}$ , whose parameters  $a = 1$  and  $b = \delta\sqrt{\alpha^2 - \beta^2}$ , with  $\alpha > 0$ ,  $-\alpha < \beta < \alpha$  and  $\delta > 0$ . Denote by  $W = \{W_t, t \geq 0\}$  the standard Brownian motion independent from  $I$ . The Variance Gamma process  $X = \{X_t, t \geq 0\}$ , with parameters  $\alpha, \beta$  and  $\delta$ , can be defined as

$$X_t = \beta\delta^2 I_t + \delta W_{I_t}.$$

The characteristic exponent is

$$\phi(u; \alpha, \beta, \delta) = -\delta(\sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2}).$$

### 3.1. Lévy process

---

The  $X_1$  has the density function

$$f(x; \alpha, \beta, \delta) = \frac{\alpha\delta}{\pi} e^{\delta\sqrt{\alpha^2 - \beta^2} + \beta x} \frac{K_1(\alpha\sqrt{\delta^2 + x^2})}{\sqrt{\delta^2 + x^2}},$$

where  $K_v(x)$  denotes the modified Bessel function of the third kind as

$$K_v(x) = \frac{\pi}{2} \frac{I_v(x) - I_{-v}(x)}{\sin(v\pi)},$$

where the right side of this equation is its limit when  $v$  is an integer, and

$$I_v(x) = \left(\frac{x}{2}\right)^v \sum_{k=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^k}{k! \Gamma(v + k + 1)}.$$

The Lévy measure is

$$v(dx) = \frac{\delta\alpha}{\pi} \frac{e^{\beta x} K_1(\alpha|x|)}{|x|} dx,$$

and the Lévy triplet is  $(\mu, 0, v(dx))$ , where

$$\mu = \frac{2\delta\alpha}{\pi} \int_0^1 \sinh(\beta x) K_1(\alpha x) dx.$$

The characteristic exponent of the CGMY process is

$$\phi(u; C, G, M, Y) = C\Gamma(-Y)((M - iu)^Y - M^Y + (G + iu)^Y - G^Y),$$

where  $C, G, M > 0$  and  $Y < 2$ . The Lévy measure is

$$v(dx) = \begin{cases} Ce^{Gx}(-x)^{-1-Y} dx, & x < 0, \\ Ce^{-Mx}x^{-1-Y} dx, & x > 0, \end{cases}$$

and the Lévy triplet is  $(\mu, 0, v(dx))$ , where

$$\mu = C \left( \int_0^1 e^{-Mx} x^{-Y} dx - \int_{-1}^0 e^{Gx} |x|^{-Y} dx \right).$$

If  $Y < 0$ , the process has finite activity. Otherwise, the process is of infinite activity. If the  $Y$  falls into the range  $[1, 2)$ , the process is of infinite variation. If  $Y = 0$ , the CGMY process reduces to the Variance Gamma process, i.e.,  $CGMY(C, G, M, 0) = VG(C, G, M)$ .

The characteristic exponent of the Meixner process  $\{X_t, t \geq 0\}$  is

$$\phi(u; \alpha, \beta, \delta) = 2\delta \ln \left( \frac{\cos\left(\frac{\beta}{2}\right)}{\cosh\left(\frac{\alpha u - i\beta}{2}\right)} \right).$$

The  $X_1$  has the density function

$$f(x; \alpha, \beta, \delta) = \frac{\left(2 \cos\left(\frac{\beta}{2}\right)\right)^{2\delta}}{2\alpha\pi\Gamma(2\delta)} e^{\frac{bx}{a}} \left| \Gamma\left(\delta + \frac{ix}{\alpha}\right) \right|^2,$$

where  $\alpha > 0, -\pi < \beta < \pi, \delta > 0$ . The Lévy measure is

$$v(dx) = \delta \frac{e^{\frac{\beta x}{\alpha}}}{x \sinh\left(\frac{\pi x}{\alpha}\right)} dx,$$

and the Lévy triplet is  $(\mu, 0, v(dx))$ , where

$$\mu = \alpha\delta \tan\left(\frac{\beta}{2}\right) - 2\delta \int_1^\infty \frac{\sinh\left(\frac{\beta x}{\alpha}\right)}{\sinh\left(\frac{\pi x}{\alpha}\right)} dx.$$

As  $\int_{|x|<1} |x|v(dx) = \infty$ , the process is of infinite variation.

The characteristic exponent of the Generalized Hyperbolic process  $\{X_t, t \geq 0\}$  is

$$\phi(u; \alpha, \beta, \delta, \kappa) = \ln \left( \left( \frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + iu)^2} \right)^{\frac{\kappa}{2}} \frac{K_\kappa(\delta\sqrt{\alpha^2 - (\beta + iu)^2})}{K_\kappa(\delta\sqrt{\alpha^2 - \beta^2})} \right).$$

The  $X_1$  has the density function

$$f(x; \alpha, \beta, \delta, \kappa) = a(\alpha, \beta, \delta, \kappa) (\delta^2 + x^2)^{\frac{\kappa}{2} - \frac{1}{4}} K_{\kappa - \frac{1}{2}}(\alpha\sqrt{\delta^2 + x^2}) e^{\beta x},$$

$$a(\alpha, \beta, \delta, \kappa) = \frac{(\alpha^2 - \beta^2)^{\frac{\kappa}{2}}}{\sqrt{2\pi\alpha}^{\kappa - \frac{1}{2}} \delta^\kappa K_\kappa(\delta\sqrt{\alpha^2 - \beta^2})},$$

### 3.1. Lévy process

---

where

$$\begin{cases} \delta \geq 0, & |\beta| < \alpha & \text{if } \kappa > 0, \\ \delta > 0, & |\beta| < \alpha & \text{if } \kappa = 0, \\ \delta > 0, & |\beta| \leq \alpha & \text{if } \kappa < 0. \end{cases}$$

The Lévy measure is

$$v(dx) = \begin{cases} \frac{e^{\beta x}}{|x|} \left( \int_0^\infty \frac{e^{-|x|\sqrt{2y+\alpha^2}}}{\pi^2 y (J_\kappa^2(\delta\sqrt{2y}) + N_\kappa^2(\delta\sqrt{2y}))} dy + \kappa e^{-\alpha|x|} \right), & \kappa \geq 0, \\ \frac{e^{\beta x}}{|x|} \int_0^\infty \frac{e^{-|x|\sqrt{2y+\alpha^2}}}{\pi^2 y (J_{-\kappa}^2(\delta\sqrt{2y}) + N_{-\kappa}^2(\delta\sqrt{2y}))} dy, & \kappa < 0, \end{cases}$$

where the functions  $J_\kappa$  and  $N_\kappa$  are the Bessel functions of the first kind and the second kind as

$$J_v(x) = \left(\frac{x}{2}\right)^v \sum_{k=0}^{\infty} \frac{\left(-\frac{x^2}{4}\right)^k}{k! \Gamma(v+k+1)},$$

and

$$N_v(x) = \frac{J_v(x) \cos(v\pi) - J_{-v}(x)}{\sin(v\pi)},$$

where the right side of this equation is its limit value when  $v$  is an integer. The Generalized Hyperbolic process has the Variance Gamma process and the Normal Inverse Gaussian process as subcases, where the Variance Gamma process with parameters  $\hat{\sigma}, \hat{\delta}, \hat{\theta}$  is obtained by taking  $\kappa = \frac{\hat{\sigma}^2}{\hat{\delta}}, \alpha = \sqrt{\frac{2}{\hat{\delta}} + \frac{\hat{\theta}^2}{\hat{\sigma}^4}}, \beta = \frac{\hat{\theta}}{\hat{\sigma}^2}, \delta \rightarrow 0$ , and the Normal Inverse Gaussian process with parameters  $\hat{\alpha}, \hat{\beta}, \hat{\delta}$  is obtained by making  $\alpha = \hat{\alpha}, \beta = \hat{\beta}, \delta = \hat{\delta}, \kappa = -\frac{1}{2}$ .

#### 3.1.3 Itô formula and measure transform

We conclude the introduction of Lévy process with two most useful tools: Itô formula and measure transform.

**Proposition 3. (Itô formula for Lévy process) (Tankov (2003))**

Let  $\{X_t, t \geq 0\}$  be a Lévy process with Lévy triplet  $(\mu, \sigma, v(dx))$ . For any  $C^{1,2}$  function  $f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\begin{aligned} f(t, X_t) - f(0, X_0) &= \int_0^t \frac{\partial f}{\partial s}(s, X_s) ds + \int_0^t \frac{\partial f}{\partial x}(s, X_{s-}) dX_s + \frac{\sigma^2}{2} \int_0^t \frac{\partial^2 f}{\partial x^2}(s, X_{s-}) ds \\ &+ \sum_{0 \leq s \leq t, \Delta X_s \neq 0} \left( f(s, X_s) - f(s, X_{s-}) - \Delta X_s \frac{\partial f}{\partial x}(s, X_{s-}) \right). \end{aligned}$$

**Proposition 4. (Measure transform for Lévy process) (Tankov (2003))**

Let  $(X_t, \mathbb{P})$  and  $(X_t, \mathbb{Q})$  be two Lévy processes on  $\mathbb{R}$  with Lévy triplets  $(\mu, \sigma, v)$  and  $(\mu_Q, \sigma_Q, v_Q)$ . Then  $P|_{\mathcal{F}_t}$  and  $Q|_{\mathcal{F}_t}$  are equivalent for all  $t$  if and only if the following three conditions are satisfied:

1.  $\sigma = \sigma_Q$
2. The Lévy measures are equivalent with

$$\int_{-\infty}^{\infty} \left( e^{\frac{\phi(x)}{2}} - 1 \right)^2 v(dx) < \infty,$$

where  $\phi(x) = \ln \left( \frac{dv_Q}{dv} \right)$ .

3. If  $\sigma = 0$  then we must in addition have

$$\mu_Q - \mu = \int_{-1}^1 x(v_Q - v)(dx).$$

When  $\mathbb{P}$  and  $\mathbb{Q}$  are equivalent, the Radon-Nikodym derivative is

$$\frac{dQ|_{\mathcal{F}_t}}{dP|_{\mathcal{F}_t}} = e^{U_t}$$

with

$$U_t = \eta X_t^c - \frac{\eta^2 \sigma^2 t}{2} - \eta \mu t + \lim_{\varepsilon \downarrow 0} \left( \sum_{s \leq t, |\Delta X_s| > \varepsilon} \phi(\Delta X_s) - t \int_{|x| > \varepsilon} (e^{\phi(x)} - 1) v(dx) \right).$$

Here  $X_t^c$  is the continuous part of  $X_t$  and  $\eta$  is such that

$$\mu_Q - \mu - \int_{-1}^1 x(v_Q - v)(dx) = \sigma^2 \eta$$

if  $\sigma > 0$  and zero if  $\sigma = 0$ .

$U_t$  is a Lévy process with Lévy triplet  $(\mu_U, \sigma_U, v_U)$  given by:

$$\begin{aligned} \mu_U &= -\frac{1}{2} \sigma^2 \eta^2 - \int_{-\infty}^{\infty} (e^y - 1 - y \mathbf{1}_{|y| \leq 1}) (v \phi^{-1})(dy), \\ \sigma_U^2 &= \sigma^2 \eta^2, \\ v_U &= v \phi^{-1}. \end{aligned}$$

### 3.2. Regime Switching Lévy Process

---

## 3.2 Regime Switching Lévy Process

In this section, we introduce the definition, properties and useful tools for regime switching Lévy process.

### 3.2.1 Definition and properties

Let  $\{J_t, t \geq 0\}$  be an irreducible Markov process with finite state space  $E = \{e_1, e_2, \dots, e_n\}$  and infinitesimal generator matrix  $Q = (q_{i,j})_{e_i, e_j \in E}$ . Let the process  $\{X_t, t \geq 0\}$  evolve as a Lévy process with Lévy triplets  $(\hat{\mu}_i, \hat{\sigma}_i, \hat{\nu}_i)$  when the state of  $J$  equals  $e_i \in E$ . The process of such kind is called regime switching Lévy process. The regime switching Lévy process  $X$  can be written in the form of:

$$X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_{|x| \geq 1} x N(ds, dx) + \int_0^t \int_{|x| < 1} x (N(ds, dx) - \nu(dx) ds),$$

where  $W$  is a standard Brownian motion,  $\mu_s = \langle \hat{\mu}, J_s \rangle$ ,  $\sigma_s = \langle \hat{\sigma}, J_s \rangle$ ,  $N(s, dx) = \langle \hat{N}(s, dx), J_s \rangle$ , and where  $\langle \cdot, \cdot \rangle$  denotes the inner product,  $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_n)$ ,  $\hat{\sigma} = (\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_n)$  and  $\hat{N}(s, dx) = (\hat{N}_1(s, dx), \hat{N}_2(s, dx), \dots, \hat{N}_n(s, dx))$ . For each state  $e_i \in E$ ,  $\hat{\mu}_i \in \mathbb{R}$ ,  $\hat{\sigma}_i \in \mathbb{R}$ ,  $\hat{N}_i$  is a Poisson random measure on  $([0, \infty) \times \mathbb{R}, \mathcal{B}([0, \infty)) \times \mathcal{B}(\mathbb{R}), dt \times \hat{\nu}_i(dx))$ , and  $\hat{\nu}_i$  is a measure on  $\mathbb{R} \setminus \{0\}$  with  $\int_{-\infty}^{\infty} (1 \wedge x^2) \hat{\nu}_i(dx) < \infty$ .

Let  $J_0 = e_i$  and  $J_t = e_j$ . Then, the characteristic exponent of  $X$  is given by

$$\phi(u) = e_i e^{Q + \text{diag}(\varphi_k(u))} e_j,$$

where  $\varphi_k(u)$  is the characteristic exponent of the Lévy process with Lévy triplets  $(\hat{\mu}_k, \hat{\sigma}_k, \hat{\nu}_k)$ , defined as follows:

$$\varphi_k(u) = i\hat{\mu}_k u - \frac{1}{2} \hat{\sigma}_k^2 u^2 + \int_{-\infty}^{\infty} (e^{iux} - 1 - iux \mathbf{1}_{\{|x| < 1\}}) \hat{\nu}_k(dx).$$

### 3.2.2 Itô formula and measure transform

As regime switching Lévy process is a semimartingale, we directly show Itô formula and measure transform for semimartingales.

**Proposition 5.** (*Itô formula for semimartingales*) (*Tankov (2003)*)

Let  $\{X_t, t \geq 0\}$  be a semimartingale. For any  $C^{1,2}$  function  $f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\begin{aligned} f(t, X_t) - f(0, X_0) &= \int_0^t \frac{\partial f}{\partial s}(s, X_s) ds + \int_0^t \frac{\partial f}{\partial x}(s, X_{s-}) dX_s \\ &\quad + \frac{1}{2} \int_0^t \frac{\partial^2 f}{\partial x^2}(s, X_{s-}) d[X, X]_s^c \\ &\quad + \sum_{0 \leq s \leq t, \Delta X_s \neq 0} \left( f(s, X_s) - f(s, X_{s-}) - \Delta X_s \frac{\partial f}{\partial x}(s, X_{s-}) \right). \end{aligned}$$

**Proposition 6. (Girsanov-Meyer Theorem) (Protter (2005))**

Let  $\mathbb{Q}$  and  $\mathbb{P}$  be equivalent and  $Z_t = \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t}$ . Let  $\{X_t, t \geq 0\}$  be a semimartingale under  $\mathbb{P}$  with decomposition  $X = M + A$ . Then,  $X$  is also a semimartingale under  $\mathbb{Q}$  and has a decomposition  $X = L + C$ , where

$$L_t = M_t - \int_0^t \frac{1}{Z_s} d[Z, M]_s$$

is a  $\mathbb{Q}$  local martingale, and  $C = X - L$  is a  $\mathbb{Q}$  finite variation process.

### 3.3 Stochastic orders

As we use stochastic orders in the contract design, we simply introduce some related concepts and properties.

#### 3.3.1 The convex order

The convex order is an important order that compares the dispersion of random variables. Let  $X$  and  $Y$  be two random variables satisfy

$$E(\phi(X)) \leq E(\phi(Y)) \quad \text{for all convex functions } \phi : \mathbb{R} \rightarrow \mathbb{R}, \quad (3.1)$$

when the expectations exist. Then,  $X$  is said to be smaller than  $Y$  in the convex order, denoted as  $X \leq_{\text{cx}} Y$ . The inequality (2.1) says that  $Y$  is "more variable" than  $X$ . Note that it is sufficient to consider only convex functions  $\phi$  defined on the union of the supports of  $X$  and  $Y$ . If the inequality (2.1) holds for all concave functions  $\phi$ ,  $X$  is said to be smaller than  $Y$  in the concave order, denoted as  $X \leq_{\text{cv}} Y$ . The  $X \leq_{\text{cv}} Y$  if and only if  $Y \leq_{\text{cx}} X$ .

Denote by  $F(\bar{F})$  and  $G(\bar{G})$  the distribution (survival) functions of  $X$  and  $Y$ , respectively. The  $X \leq_{\text{cx}} Y$  can give some useful necessary conditions, such as

- a)  $E(X) = E(Y)$
- b)  $\text{Var}(X) \leq \text{Var}(Y)$

### 3.3. Stochastic orders

---

c)  $E((X - a)_+) \leq E((Y - a)_+)$  for all  $a$  (or equivalently  $\int_x^\infty \bar{F}(u)du \leq \int_x^\infty \bar{G}(u)du$  for all  $x$ )

d)  $E((a - X)_+) \leq E((a - Y)_+)$  for all  $a$  (or equivalently  $\int_{-\infty}^x F(u)du \leq \int_{-\infty}^x G(u)du$  for all  $x$ ).

If  $E(X) = E(Y)$ , the  $X \leq_{cx} Y$  if and only if the conditions (c) or (d) holds. If  $E(X^2) = E(Y^2)$  and  $X \leq_{cx} Y$  ( $X \leq_{cv} Y$ ), the  $X$  and  $Y$  have the same distribution. If  $E(\phi(X)) = E(\phi(Y))$  holds for some strictly convex function  $\phi$  and  $X \leq_{cx} Y$  ( $X \leq_{cv} Y$ ), the  $X$  and  $Y$  also have the same distribution.

Denote by  $S^-$  the sign change operator. Given any function  $\phi : [a, b] \rightarrow \mathbb{R}$ ,  $S^-(\phi)$  denotes the number of sign changes of  $\phi$  in  $[a, b]$ , defined by

$$S^-(\phi) = \sup S^-(\phi(x_1), \phi(x_2), \dots, \phi(x_n)),$$

where the supremum is taken over all sets  $x_1 < x_2 < \dots < x_n$  such that  $x_i \in [a, b]$  and  $n < \infty$ , and  $S^-(y_1, y_2, \dots, y_n)$  denotes the number of sign changes of the sequence  $\{y_1, y_2, \dots, y_n\}$ , zero terms being discarded. Then, we show two useful conditions that imply the convex order.

**Proposition 7. (Ohlin (1969))**

Let  $X$  be a random variable and  $f(x)$  and  $g(x)$  two non-decreasing functions of  $x$ , such that  $E(f(X)) = E(g(X))$ . If there exists a  $x_0$ , such that

$$g(x) \geq f(x), \quad \text{for } x < x_0,$$

$$g(x) \leq f(x), \quad \text{for } x > x_0,$$

then  $g(X) \leq_{cx} f(X)$ .

**Proposition 8. (Shaked and Shanthikumar (2007))**

Let  $X$  and  $Y$  be two random variables with equal means, density functions  $f$  and  $g$ , distribution functions  $F$  and  $G$ , and survival functions  $\bar{F}$  and  $\bar{G}$ , respectively. Then  $X \leq_{cx} Y$  if any of the following conditions hold:

- $S^-(g - f) = 2$  and the sign sequence is  $+, -, +$ .
- $S^-(\bar{F} - \bar{G}) = 1$  and the sign sequence is  $+, -$ .
- $S^-(G - F) = 1$  and the sign sequence is  $+, -$ .

#### 3.3.2 The $m$ -convex order

For a positive integer  $m$ , denote  $M_{m-cx}$  the class of all functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  whose  $m^{\text{th}}$  derivative  $\phi^{(m)}$  exists and satisfies  $\phi^{(m)}(x) \geq 0$ , for all  $x \in \mathbb{R}$ , or which are limits

of sequences of functions whose  $m^{\text{th}}$  derivative  $\phi^{(m)}$  is continuous and nonnegative. Let  $X$  and  $Y$  be two random variables such that

$$E(\phi(X)) \leq E(\phi(Y)), \quad \text{for all functions } \phi \in M_{m\text{-cx}}, \quad (3.2)$$

Then,  $X$  is said to be smaller than  $Y$  in the  $m$ -convex order, denoted as  $X \leq_{m\text{-cx}} Y$ . If the inequality (2.2) holds for all functions whose  $m^{\text{th}}$  derivative  $\phi^{(m)}$  exists and satisfies  $(-1)^{m-1}\phi^{(m)} \geq 0$ , for all  $x \in \mathbb{R}$ , or which are limits of sequences of functions whose  $(-1)^{m-1}\phi^{(m)}$  is continuous and nonnegative,  $X$  is said to be smaller than  $Y$  in the  $m$ -concave order, denoted as  $X \leq_{m\text{-cv}} Y$ . They have relation that

$$X \leq_{m\text{-cx}} Y \iff \begin{cases} X \leq_{m\text{-cv}} Y, & \text{when } m \text{ is odd,} \\ Y \leq_{m\text{-cv}} X, & \text{when } m \text{ is even.} \end{cases}$$

If  $X \leq_{m\text{-cx}} Y$ , the  $E(X^k) \leq E(Y^k)$  for  $k \geq m$  such that  $k - m$  is even. If  $X \leq_{m\text{-cx}} Y$  and  $X$  and  $Y$  are nonnegative, the  $E(X^k) \leq E(Y^k)$  for  $k \geq m$ . If  $E(X^m) = E(Y^m)$  and  $X \leq_{m\text{-cx}} Y$ , the  $X$  and  $Y$  have the same distribution. If  $X \leq_{m\text{-cx}} Y$  and  $E(\phi(X)) = E(\phi(Y))$  holds for some functions  $\phi \in M_{m\text{-cx}}$  such that  $\phi^{(m)}(x) > 0$  for all  $x \in \mathbb{R}$ , the  $X$  and  $Y$  also have the same distribution.

Let  $F(\bar{F})$  and  $G(\bar{G})$  be distribution (survival) functions of  $X$  and  $Y$ , respectively. Denote by  $F^0(x) = F(x)$  and  $F^k(x) = \int_{-\infty}^x F^{k-1}(z)dz$ , for  $k \geq 1$ . Let  $\bar{F}^0(x) = \bar{F}(x)$  and  $\bar{F}^k(x) = \int_x^{\infty} \bar{F}^{k-1}(z)dz$ , for  $k \geq 1$ . Define  $G^k$  and  $\bar{G}^k$  in the same manner. The  $X \leq_{m\text{-cx}} Y$  is equivalent to any of the following conditions:

a)

$$\begin{cases} E(X^k) = E(Y^k), & k = 1, 2, \dots, m-1, \\ (-1)^m(G^{m-1}(x) - F^{m-1}(x)) \geq 0, & \text{for all } x \in \mathbb{R}, \end{cases}$$

b)

$$\begin{cases} E(X^k) = E(Y^k), & k = 1, 2, \dots, m-1, \\ \bar{G}^{m-1}(x) - \bar{F}^{m-1}(x) \geq 0, & \text{for all } x \in \mathbb{R}. \end{cases}$$

Let  $\mathcal{B}_m(S; \mu_1, \mu_2, \dots, \mu_{m-1})$  be the class of all the random variables  $X$  whose distribution functions are defined in  $S$  and which have the first  $m - 1$  moments  $E(X^k) = \mu_k, k = 1, 2, \dots, m - 1$ . Then, we show some useful sufficient conditions that lead to  $X \leq_{m\text{-cx}} Y$ .

**Proposition 9. (*Shaked and Shanthikumar (2007)*)**

Let  $X$  and  $Y$  be two random variables in  $\mathcal{B}_m(S; \mu_1, \mu_2, \dots, \mu_{m-1})$  with distribution functions  $F$  and  $G$ , respectively, and with density functions  $f$  and  $g$ , respectively.

a) If  $S^-(F - G) = m - 1$  and the last sign of  $F - G$  is a  $+$ , then  $X \leq_{m\text{-cx}} Y$ .

b) If  $S^-(f - g) = m$  and the last sign of  $g - f$  is a  $+$ , then  $X \leq_{m\text{-cx}} Y$ .

### 3.3. Stochastic orders

---

#### 3.3.3 The monotone convex order

Let  $X$  and  $Y$  be two random variables such that

$$E(\phi(X)) \leq E(\phi(Y)) \quad \text{for all increasing convex functions } \phi: \mathbb{R} \rightarrow \mathbb{R}, \quad (3.3)$$

when the expectations exist. Then,  $X$  is said to be smaller than  $Y$  in the increasing convex order, denoted as  $X \leq_{icx} Y$ . The inequality (2.3) says that  $Y$  is "larger" and "more variable" than  $X$ . If the inequality (2.3) holds for all increasing concave functions  $\phi$ ,  $X$  is said to be smaller than  $Y$  in the increasing concave order, denoted as  $X \leq_{icv} Y$ . They have relation that  $X \leq_{icx} (\leq_{icv}) Y$  if and only if  $-X \geq_{icv} (\geq_{icx}) -Y$ .

Denote by  $F(\bar{F})$  and  $G(\bar{G})$  the distribution (survival) functions of  $X$  and  $Y$ , respectively. The  $X \leq_{icx} Y$  ( $X \leq_{icv} Y$ ) have some useful necessary conditions, such as

- a)  $E(X) \leq E(Y)$
- b)  $E((X - a)_+) \leq E((Y - a)_+)$  for all  $a$  ( $E((X - a)_-) \leq E((Y - a)_-)$  for all  $a$ )
- c)  $\int_x^\infty \bar{F}(u) du \leq \int_x^\infty \bar{G}(u) du$  for all  $x$  ( $\int_{-\infty}^x F(u) du \geq \int_{-\infty}^x G(u) du$  for all  $x$ ).

The  $X \leq_{icx} Y$  ( $X \leq_{icv} Y$ ) if and only if the conditions (b) or (c) holds.

We have the following relation between the increasing convex (concave) order and the convex order.

**Proposition 10. (*Shaked and Shanthikumar (2007)*)**

Let  $X$  and  $Y$  be two random variables.

- a) If  $X \leq_{cx} Y$ , then  $X \leq_{icx} Y$  and  $Y \leq_{icv} X$ .
- b) If  $X \leq_{icx} Y$  and  $E(X) = E(Y)$ , then  $X \leq_{cx} Y$ .
- c) If  $X \leq_{icv} Y$  and  $E(X) = E(Y)$ , then  $Y \leq_{cx} X$ .

There are some useful conditions that imply the increasing convex and the increasing concave orders.

**Proposition 11. (*Shaked and Shanthikumar (2007)*)**

Let  $X$  and  $Y$  be two random variables with distribution functions  $F$  and  $G$ , and survival functions  $\bar{F}$  and  $\bar{G}$ , respectively, and with finite means such that  $E(X) \leq E(Y)$ . Then  $X \leq_{icx} Y$  ( $X \leq_{icv} Y$ ) if any of the following conditions hold:

- a)  $S^-(\bar{F} - \bar{G}) \leq 1$  and the sign sequence is  $+, - (-, +)$ ,
- b)  $S^-(G - F) \leq 1$  and the sign sequence is  $+, - (-, +)$ .

### 3.3.4 The $m$ -increasing convex order

For a positive integer  $m$ , denote  $M_{m-icx}$  the class of all functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  whose first  $m$  derivatives  $\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(m)}$  exist and satisfy  $\phi^{(k)}(x) \geq 0, k = 1, \dots, m$ , for all  $x \in \mathbb{R}$ , or which are limits of sequences of functions whose  $\phi^{(k)}, k = 1, \dots, m$  is continuous and nonnegative. Let  $X$  and  $Y$  be two random variables such that

$$E(\phi(X)) \leq E(\phi(Y)), \quad \text{for all functions } \phi \in M_{m-icx}, \quad (3.4)$$

Then,  $X$  is said to be smaller than  $Y$  in the  $m$ -icx order, denoted as  $X \leq_{m-icx} Y$ . If the inequality (2.2) holds for all functions whose first  $m$  derivatives  $\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(m)}$  exist and satisfy  $(-1)^{k-1}\phi^{(k)}(x) \geq 0, k = 1, \dots, m$ , for all  $x \in \mathbb{R}$ , or which are limits of sequences of functions whose  $(-1)^{k-1}\phi^{(k)}, k = 1, \dots, m$  is continuous and nonnegative,  $X$  is said to be smaller than  $Y$  in the  $m$ -icv order, denoted as  $X \leq_{m-icv} Y$ . The orders  $\leq_{m-icx}$  and  $\leq_{m-icv}$  have the following relation:

$$X \leq_{m-icx} Y (X \leq_{m-icv} Y) \iff -X \geq_{m-icv} -Y (-X \geq_{m-icx} -Y).$$

The order  $\leq_{2-icx}$  ( $\leq_{2-icv}$ ) is equivalent to the order  $\leq_{icx}$  ( $\leq_{icv}$ ). The  $m$ -convex (concave) order and  $m$ -increasing convex (concave) order have the following relation:

$$X \leq_{m-cx} Y (X \leq_{m-cv} Y) \implies X \leq_{m-icx} Y (X \leq_{m-icv} Y).$$

Let  $F(\bar{F})$  and  $G(\bar{G})$  be distribution (survival) functions of  $X$  and  $Y$ , respectively. Denote by  $F^0(x) = F(x)$  and  $F^k(x) = \int_{-\infty}^x F^{k-1}(z)dz$ , for  $k \geq 1$ . Let  $\bar{F}^0(x) = \bar{F}(x)$  and  $\bar{F}^k(x) = \int_x^\infty \bar{F}^{k-1}(z)dz$ , for  $k \geq 1$ . Define  $G^k$  and  $\bar{G}^k$  in the same manner. The  $X \leq_{m-icx} Y$  ( $X \leq_{m-icv} Y$ ) is equivalent to any of the following conditions:

a)

$$\bar{F}^{m-1}(x) \leq \bar{G}^{m-1}(x) \text{ for all } x \text{ (} F^{m-1}(x) \geq G^{m-1}(x) \text{ for all } x),$$

b)

$$E((X-a)_+^{m-1}) \leq E((Y-a)_+^{m-1}) \text{ for all } a \text{ (} E((X-a)_-^{m-1}) \leq E((Y-a)_-^{m-1}) \text{ for all } a).$$

The  $X \leq_{m-icx} Y$  ( $X \leq_{m-icv} Y$ ) gives the following useful moment relation.

**Proposition 12. (Shaked and Shanthikumar (2007))**

Let  $X$  and  $Y$  be two random variables with finite first  $m-1$  moments. If  $X \leq_{m-icx} Y$  ( $X \leq_{m-icv} Y$ ), then  $E(X^k) < E(Y^k)$  ( $(-1)^{k+1}E(X^k) < (-1)^{k+1}E(Y^k)$ ) for the smallest  $k$  for which  $E(X^k) \neq E(Y^k)$ .

The  $X \leq_{m_1-icx} Y$  ( $X \leq_{m_1-icv} Y$ ) and  $X \leq_{m_2-icx} Y$  ( $X \leq_{m_2-icv} Y$ ),  $m_2 \geq m_1$  have the following relation:

### 3.4. Higher-order degree risk and stochastic dominance

---

**Proposition 13. (*Shaked and Shanthikumar (2007)*)**

Let  $X$  and  $Y$  be two random variables. If  $X \leq_{m_1-icx} Y$  ( $X \leq_{m_1-icv} Y$ ), then  $X \leq_{m_2-icx} Y$  ( $X \leq_{m_2-icv} Y$ ) for all  $m_2 \geq m_1$ .

There are some useful sufficient conditions that lead to  $X \leq_{m-icx} Y$ .

**Proposition 14. (*Shaked and Shanthikumar (2007)*)**

Let  $X$  and  $Y$  be two nonnegative random variables with distribution functions  $F$  and  $G$ , respectively, and with density functions  $f$  and  $g$ , respectively, such that  $E(X^i) = E(Y^i)$ ,  $i = 1, 2, \dots, m - 2$ , and  $E(X^{m-1}) \leq E(Y^{m-1})$ .

- a) If  $S^-(F - G) \leq m - 1$  and the last sign of  $F - G$  is a +, then  $X \leq_{m-icx} Y$ ,
- b) If  $S^-(f - g) \leq m$  and the last sign of  $g - f$  is a +, then  $X \leq_{m-icx} Y$ .

### 3.4 Higher-order degree risk and stochastic dominance

As we discuss the optimal contract design under third degree risk, we introduce the definition of higher-order degree risk and stochastic dominance.

#### 3.4.1 $n^{\text{th}}$ degree risk

Let  $F$  and  $G$  be distribution functions of  $X$  and  $Y$ , respectively, defined on  $[a, b]$ . Denote by  $F^0(x) = F(x)$  and  $F^k(x) = \int_{-\infty}^x F^{k-1}(z)dz$ , for  $k \geq 1$ . Define  $G^k$  in the same manner. The definition of  $n^{\text{th}}$  degree risk can be given:

**Definition 2.  $n^{\text{th}}$  degree risk (*Ekern (1980)*)**

$Y$  has more  $n^{\text{th}}$  degree risk than  $X$ , if and only if

$$G^k(b) = F^k(b), \quad k = 1, 2, \dots, n - 1, \tag{3.5}$$

$$G^{n-1}(x) \geq F^{n-1}(x) \text{ for all } x \in [a, b]. \tag{3.6}$$

The conditions (2.5) and (2.6) imply the first  $n - 1$  moments of  $X$  and  $Y$  are equal, i.e.,  $E(X^k) = E(Y^k)$ ,  $k = 1, \dots, n - 1$ , and the  $n^{\text{th}}$  moment of  $Y$  adjusted by  $(-1)^n$  is greater than the  $n^{\text{th}}$  moment of  $X$  adjusted by  $(-1)^n$ , i.e.,  $(-1)^n E(Y^n) \geq (-1)^n E(X^n)$ . For example, if  $Y$  has more first degree risk than  $X$ , the  $E(X) \geq E(Y)$ . If  $Y$  has more second degree risk, the  $E(X) = E(Y)$  and  $\text{Var}(Y) \geq \text{Var}(X)$ . If  $Y$  exhibits more third degree risk, the  $E(X) = E(Y)$ ,  $\text{Var}(Y) = \text{Var}(X)$  and the  $Y$  has less skewness than  $X$ . The  $Y$  having more  $n^{\text{th}}$  degree risk than  $X$  and  $Y \leq_{n-cv} X$  are the same concept.

Let  $u$  be a utility function whose  $n^{\text{th}}$  derivative  $u^{(n)}$  exists. Then, we have the following result:

**Proposition 15. (Ekern (1980))**

A person is  $n^{\text{th}}$  degree risk averse, if and only if  $(-1)^{n-1}u^{(n)}(x) \geq 0$  for all  $x \in [a, b]$ .

**3.4.2  $n^{\text{th}}$  degree stochastic dominance**

The definition of  $n^{\text{th}}$  degree stochastic dominance can be given:

**Definition 3.  $n^{\text{th}}$  degree stochastic dominance (Jean (1980))**

$Y$  is dominated by  $X$  in  $n^{\text{th}}$  degree stochastic dominance, denoted as  $Y \leq_{n-SD} X$ , if and only if

$$G^k(b) \geq F^k(b), \quad k = 1, 2, \dots, n - 2, \tag{3.7}$$

$$G^{n-1}(x) \geq F^{n-1}(x), \quad \text{for all } x \in [a, b]. \tag{3.8}$$

The  $Y \leq_{n-SD} X$  and  $Y \leq_{n-icv} X$  are the same concept. Then, we have the following result:

**Proposition 16. (Jean (1980))**

All the persons with utility function  $u(x)$  such that  $(-1)^{k-1}u^{(k)}(x) \geq 0, k = 1, \dots, n$ , prefer  $X$  to  $Y$  if and only if  $Y \leq_{n-SD} X$ .

**3.5 The GLM and GAM model**

As we compare our model with the GLM and GAM ones, we simply review the GLM and GAM models.

**3.5.1 Generalized linear model**

The GLM model generalizes ordinary linear regression to allow for response distribution other than normal. The GLM models an  $n$ -vector of independent response variables  $\mathbf{Y} = (Y_1, \dots, Y_n)$  as

$$g(\mu_i) = \mathbf{X}_i\boldsymbol{\beta},$$

where  $\mu_i = E(Y_i)$ ,  $g$  is a monotonic differentiable link function,  $\mathbf{X}_i$  is the  $i^{\text{th}}$  row of predictor matrix  $\mathbf{X}$ ,  $\boldsymbol{\beta}$  is the regression coefficients, and where  $Y_i \sim f_{\theta_i}(y_i)$ ,  $f_{\theta_i}(y_i)$  is an exponential dispersion family distribution, with canonical parameter  $\theta_i$ . The common choices of link function include the identity, log, reciprocal, logit, and probit.

The density function of exponential dispersion family distribution is in the form of

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

where  $a(\cdot), b(\cdot), c(\cdot)$  are arbitrary functions,  $\phi$  is the dispersion parameter,  $\theta$  is the canonical parameter. Denote  $Y$  the random variable with density function  $f_{\theta}(y)$ .

### 3.5. The GLM and GAM model

---

The mean and variance of  $Y$  can be given:

$$E(Y) = b'(\theta) \quad \text{and} \quad \text{Var}(Y) = b''(\theta)a(\phi).$$

The table 1 lists some exponential dispersion family distributions.

**Table 1** **Some exponential dispersion family distributions**

	Normal	Poisson	Binomial	Gamma	Inverse Gaussian
$f_{\theta}(y)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$	$\frac{\mu^y e^{-\mu}}{y!}$	$\binom{n}{y} \left(\frac{\mu}{n}\right)^y \left(1 - \frac{\mu}{n}\right)^{n-y}$	$\frac{1}{\Gamma(v)} \left(\frac{v}{\mu}\right)^v y^{v-1} e^{-\frac{vy}{\mu}}$	$\sqrt{\frac{\gamma}{2\pi y^3}} e^{-\frac{\gamma(y-\mu)^2}{2\mu^2 y}}$
Range	$(-\infty, \infty)$	$0, 1, 2, \dots$	$0, 1, \dots, n$	$y > 0$	$y > 0$
$\theta$	$\mu$	$\log(\mu)$	$\log\left(\frac{\mu}{n-\mu}\right)$	$-\frac{1}{\mu}$	$-\frac{1}{2\mu^2}$
$\phi$	$\sigma^2$	1	1	$\frac{1}{v}$	$\frac{1}{\gamma}$
$a(\phi)$	$\phi$	$\phi$	$\phi$	$\phi$	$\phi$
$b(\theta)$	$\frac{\theta}{2}$	$e^{\theta}$	$n \log(1 + e^{\theta})$	$-\log(-\theta)$	$-\sqrt{-2\theta}$
$c(y, \phi)$	$-\frac{1}{2} \left( \frac{y^2}{\phi} + \log(2\pi\phi) \right)$	$-\log(y!)$	$\log \binom{n}{y}$	$v \log(vy) - \log(\Gamma(v))$	$-\frac{1}{2} \left( \log(2\pi y^3 \phi) + \frac{1}{\phi y} \right)$

The GLM model can be estimated by the iteratively re-weighted least square (IRLS) algorithm. Denote

$$\alpha(\mu_i) = 1 + (y_i - \mu_i) \left( \frac{V'(\mu_i)}{V(\mu_i)} + \frac{g''(\mu_i)}{g'(\mu_i)} \right),$$

where  $V(\mu) = \frac{b''(\theta)}{\omega}$  and  $\omega$  is a known constant. The algorithm can be summarized as follows:

---

#### The IRLS algorithm for GLM model

---

1. Initialize  $\hat{\mu}_i = y_i + \delta_i$  and  $\hat{\eta}_i = g(\hat{\mu}_i)$ , where  $\delta_i$  is zero or a small constant ensuring that  $\hat{\eta}_i$  is finite.

2. While not reaching stopping rule do

1. Compute pseudodata  $z_i = \frac{g'(\hat{\mu}_i)(y_i - \hat{\mu}_i)}{\alpha(\hat{\mu}_i)} + \hat{\eta}_i$  and iterative weights

$$w_i = \frac{\alpha(\hat{\mu}_i)}{g'(\hat{\mu}_i)^2 V(\hat{\mu}_i)}.$$

2. Compute  $\hat{\beta}$  by minimizing the weighted least square sum

$$\sum_{i=1}^n \omega_i (z_i - \mathbf{X}_i \boldsymbol{\beta})^2.$$

3. Update  $\hat{\boldsymbol{\eta}} = \mathbf{X} \hat{\boldsymbol{\beta}}$  and  $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$ .

end

3. Return  $\hat{\boldsymbol{\beta}}$ .

---

The deviance of the GLM model measures the goodness of fit, defined as

$$D = \sum_{i=1}^n 2\omega_i (y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)),$$

where  $\omega_i$  is sample weight and the  $\tilde{\theta}$  and  $\hat{\theta}$  denote the maximum likelihood estimates of the canonical parameters for the saturated model and model of interest, respectively. The deviance, depending on the scale parameter, is the scaled deviance, defined as

$$D^* = \frac{D}{\phi}.$$

Another important measure of the goodness of fit is the Pearson statistic, which takes the form

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}.$$

Table 2 lists the deviance forms of some exponential dispersion family distributions.

**Table 2** Deviances of some exponential dispersion family distributions

Distribution	Deviance
Normal	$(y - \hat{\mu})^2$
Poisson	$2y \log \left( \frac{y}{\hat{\mu}} \right) - 2(y - \hat{\mu})$
Binomial	$2 \left( y \log \left( \frac{y}{\hat{\mu}} \right) + (n - y) \log \left( \frac{n - y}{n - \hat{\mu}} \right) \right)$
Gamma	$2 \left( \frac{y - \hat{\mu}}{\hat{\mu}} - \log \left( \frac{y}{\hat{\mu}} \right) \right)$
Inverse Gaussian	$\frac{(y - \hat{\mu})^2}{\hat{\mu}^2 y}$

### 3.5.2 Generalized additive model

The GAM model generalizes the GLM model to a more flexible additive form. The GAM models an  $n$ -vector of independent response variables  $\mathbf{Y} = (Y_1, \dots, Y_n)$  as

$$g(\mu_i) = \alpha + f_1(x_{i1}) + f_2(x_{i2}) + \dots + f_p(x_{ip}),$$

where  $\mu_i = E(Y_i)$ ,  $Y_i \sim f_{\theta_i}(y_i)$ ,  $f_{\theta_i}(y_i)$  is an exponential dispersion family distribution, with canonical parameter  $\theta_i$ ,  $g$  is a monotonic differentiable link function,  $x_{ij}$  is the  $i^{\text{th}}$  observation of predictor variable  $x_j$ , and the  $f_j(\cdot)$  is a smooth function of predictor variable  $x_j$ , in the linear basis expansion form

$$f_j(x) = \sum_{k=1}^K \beta_k^{(j)} b_k^{(j)}(x),$$

where  $\beta_k^{(j)}$  is the regression coefficient, and  $b_k^{(j)}(x)$  is the  $k^{\text{th}}$  basis function.

The GAM model can be estimated by the penalized iterative least square (PIRLS)

### 3.6. Machine learning algorithms

---

algorithm. Denote  $\beta_j = (\beta_1^{(j)}, \dots, \beta_K^{(j)})$  the coefficients of basis expansion for function  $f_j$ . Define the vector

$$\mathbf{L}_j = (f_j(x_1) - 2f_j(x_2) + f_j(x_3), f_j(x_2) - 2f_j(x_3) + f_j(x_4), \dots, f_j(x_{n-2}) - 2f_j(x_{n-1}) + f_j(x_n)).$$

Let  $\mathbf{D}_j$  be the matrix such that

$$\mathbf{L}_j = \mathbf{D}_j \beta_j$$

and  $\mathbf{S}_j = \mathbf{D}_j^T \mathbf{D}_j$ . Denote  $\mathbf{X} = (\mathbf{1}_n, \mathbf{X}_1, \dots, \mathbf{X}_p)$  and  $\beta = (\alpha, \beta_1, \dots, \beta_p)$ , where  $b_k^{(j)}(x_{ij})$  is the  $(i, k)$ <sup>th</sup> element of  $\mathbf{X}_j$ . Given the smooth parameters  $(\lambda_1, \dots, \lambda_p)$ , the algorithm can be summarized as follows:

---

#### The PIRLS algorithm for GAM model

---

1. Initialize  $\hat{\mu}_i = y_i + \delta_i$  and  $\hat{\eta}_i = g(\hat{\mu}_i)$ , where  $\delta_i$  is zero or a small constant ensuring that  $\hat{\eta}_i$  is finite.

2. While not reaching stopping rule do

1. Compute

$$w_i = \frac{1}{g'(\hat{\mu}_i)^2 V(\hat{\mu}_i)} \quad \text{and} \quad z_i = g'(\hat{\mu}_i)(y_i - \hat{\mu}_i) + \hat{\eta}_i.$$

2. Compute  $\hat{\beta}$  by minimizing the penalized least square sum

$$\|\sqrt{\mathbf{W}}\mathbf{z} - \sqrt{\mathbf{W}}\mathbf{X}\beta\|^2 + \sum_{j=1}^p \lambda_j \beta_j^T \mathbf{S}_j \beta_j,$$

where  $\mathbf{W}$  is the diagonal matrix such that  $W_{ii} = w_i$ .

3. Update  $\hat{\eta} = \mathbf{X}\hat{\beta}$  and  $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$ .

end

3. Return  $\hat{\beta}$ .

---

## 3.6 Machine learning algorithms

As we employ gradient boosting algorithm to develop a new dependent frequency-severity model, for which we use regression trees as weaker learners, we simply introduce the regression trees and the gradient boosting algorithm.

### 3.6.1 Regression trees

The algorithm produces a tree structure, where inputting covariates  $X = (X_1, \dots, X_p)$  can output the value of response variable  $Y$ . The regression tree has an additive form:

$$f(x) = \sum_{m=1}^M c_m \mathbf{1}_{\{X \in R_m\}},$$

where  $R_1, \dots, R_M$  are disjoint regions that collectively cover the space of all joint values of the covariates  $X$  and  $c_m$  is the response value in region  $R_m$ .

Denote  $\Phi(y, f(x))$  the loss function. The algorithm makes use of a greedy search to find the best binary partition in terms of minimum empirical loss. Starting from root node (all of the data), consider a splitting variable  $X_j$  and split point  $s$ , and define the pair of half-planes

$$R_1(X_j, s) = \{X|X_j \leq s\} \quad \text{and} \quad R_2(X_j, s) = \{X|X_j > s\}.$$

Denote  $(x_i, y_i), i = 1, \dots, N$  the observations and  $d_k$  the response value in region  $R_k(X_j, s)$ . The algorithm seeks the best splitting variable  $X_j$  and split point  $s$  by solving

$$\arg \min_{j,s} \left( \min_{d_1} \sum_{x_i \in R_1(X_j,s)} \Phi(y_i, d_1) + \min_{d_2} \sum_{x_i \in R_2(X_j,s)} \Phi(y_i, d_2) \right).$$

For each splitting variable, the best split point  $s$  can be quickly found and hence by scanning through all of the covariates, the best pair  $(X_j, s)$  can be determined. Then, the splitting process is repeated on each of the two regions. This process is repeated on all of the resulting regions until a stopping rule is reached.

A very large tree might overfit the data, while a small tree can not capture the important data structure. The preferred strategy is to grow a large tree  $T_0$ , stopping the splitting process until the minimum node size is reached, and then to prune this large tree using cost-complexity pruning.

Define a subtree  $T \subset T_0$  to be any tree that can be obtained by pruning  $T_0$ , through collapsing any number of its internal nodes. Denote  $m$  the terminal node that represents region  $R_m$ ,  $|T|$  the number of terminal nodes in  $T$ ,

$$\hat{c}_m = \arg \min_{c_m} \sum_{x_i \in R_m} \Phi(y_i, c_m),$$

$$Q_m(T) = \sum_{x_i \in R_m} \Phi(y_i, \hat{c}_m)$$

Define the cost complexity criterion

$$C_\alpha(T) = \sum_{n=1}^{|T|} Q_n(T) + \alpha|T|,$$

where  $\alpha$  is a penalty parameter to govern the tradeoff between tree size and its goodness of fit to the data.

For each  $\alpha$ , one can find the unique smallest subtree  $T_\alpha$  that minimizes  $C_\alpha(T)$ . The  $T_\alpha$  can be found by making use of weakest link pruning: it successively collapses the internal node that produces the smallest per-node increase in  $\sum_n Q_n(T)$  and continues until the root tree is produced. This finite sequence of subtrees contains

### 3.6. Machine learning algorithms

---

$T_{\hat{\alpha}}$ . The  $\hat{\alpha}$  is chosen by cross-validation method to minimize the cross-validated empirical loss. Then, the final tree is  $T_{\hat{\alpha}}$ .

#### 3.6.2 Gradient boosting

Denote  $\mathbf{x} = (x_1, \dots, x_p)$  the set of predictor variables and  $y$  the response variable. Given the training sample  $\{y_i, \mathbf{x}_i\}_1^n$  and the loss function  $\Psi(y, f(\mathbf{x}))$ , the algorithm is to estimate the function  $\hat{f}(\mathbf{x})$  that minimizes the loss over the training sample,

$$\hat{f}(\mathbf{x}) = \arg \min_{f(\mathbf{x})} \frac{1}{n} \sum_{i=1}^n \Psi(y_i, f(\mathbf{x}_i)),$$

where  $f(\mathbf{x})$  is constrained to the form of a sum of weak learners as

$$f(\mathbf{x}) = h(\mathbf{x}; \mathbf{a}_0) + \sum_{m=1}^M \beta_m h(\mathbf{x}; \mathbf{a}_m),$$

where  $h(\mathbf{x}; \mathbf{a}_m)$  is a weak learner with parameters  $\mathbf{a}_m$ ,  $\beta_m$  is the expansion coefficient,  $M$  is the number of weak learners.

The algorithm works in a forward stagewise manner. Let  $f_0(\mathbf{x})$  be an initial estimate of  $\hat{f}(\mathbf{x})$  as

$$f_0(\mathbf{x}) = h(\mathbf{x}; \mathbf{a}_0) = \arg \min_{\rho} \sum_{i=1}^n \Psi(y_i, \rho).$$

Denote  $f_{m-1}(\mathbf{x})$  the estimate at the  $(m-1)^{\text{th}}$  step. Then, at the  $m^{\text{th}}$  step, the algorithm computes the negative gradient

$$z_i = - \left. \frac{\partial \Psi(y_i, f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)} \right|_{f(\mathbf{x}_i) = f_{m-1}(\mathbf{x}_i)},$$

and fits the weak learner  $h(\mathbf{x}; \mathbf{a}_m)$  by minimizing the following least square sum

$$\mathbf{a}_m = \arg \min_{\mathbf{a}} \sum_{i=1}^n (z_i - h(\mathbf{x}_i; \mathbf{a}))^2.$$

The optimal  $\beta_m$  is the solution to

$$\beta_m = \arg \min_{\beta} \sum_{i=1}^n \Psi(y_i, f_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i; \mathbf{a}_m)).$$

Then, the estimate is updated as

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \nu \beta_m h(\mathbf{x}; \mathbf{a}_m),$$

where  $0 < \nu \leq 1$  is the shrinkage factor that controls the learning rate. **Friedman**

(2001) points out that small  $\nu$  reduces overfitting and increases predictive performance. The algorithm can be summarized as follows:

---

**Gradient Boosting Algorithm**

---

1. Initialize  $f_0(\mathbf{x}) = \arg \min_{\rho} \sum_{i=1}^n \Psi(y_i, \rho)$ .

2. For  $m = 1$  to  $M$ :

1. Compute

$$z_i = -\frac{\partial \Psi(y_i, f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)} \Big|_{f(\mathbf{x}_i)=f_{m-1}(\mathbf{x}_i)}, i = 1, \dots, n.$$

2. Fits the weak learner  $h(\mathbf{x}; \mathbf{a}_m)$  by minimizing the least square sum

$$\mathbf{a}_m = \arg \min_{\mathbf{a}} \sum_{i=1}^n (z_i - h(\mathbf{x}_i; \mathbf{a}))^2.$$

3. Compute

$$\beta_m = \arg \min_{\beta} \sum_{i=1}^n \Psi(y_i, f_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i; \mathbf{a}_m)).$$

4. Update

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \nu \beta_m h(\mathbf{x}; \mathbf{a}_m).$$

end

3. Return  $\hat{f}(\mathbf{x}) = f_M(\mathbf{x})$ .

---

# Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

---

## Contents

---

<b>4.1</b>	<b>Introduction</b>	<b>56</b>
<b>4.2</b>	<b>The Basic Framework</b>	<b>58</b>
4.2.1	Contract Specifications	58
4.2.2	Investment Funds and Interest Rate Dynamics	60
4.2.3	Main Subcontract Terms	61
<b>4.3</b>	<b>The First Passage Time Results of Regime Switching Jump Diffusion Processes</b>	<b>62</b>
4.3.1	Matrix Wiener-Hopf Factorization	62
4.3.2	The First Passage Time Results	64
4.3.3	Numerical Algorithm	65
<b>4.4</b>	<b>Contract Valuation</b>	<b>66</b>
4.4.1	Computation of $GF$ and $LR$	67
4.4.2	Computation of $PO$ and $BO$	68
4.4.3	Numerical Analysis	69
<b>4.5</b>	<b>Constant Guaranteed Rates V.S. Floating Guaranteed Rates</b>	<b>72</b>
<b>4.6</b>	<b>Two Hedging Strategies</b>	<b>77</b>
4.6.1	Hedging Strategy	77
4.6.2	Hedging Effectiveness	79
<b>4.7</b>	<b>Conclusion</b>	<b>81</b>
<b>4.8</b>	<b>Appendix</b>	<b>82</b>
4.8.1	Computations of Subcontract Terms for Contract with a Float- ing Guaranteed Rate	82

---

## 4.1 Introduction

Modern life insurance contracts bear various covenants (e.g., interest rate guarantees, bonus and surrender options, participating policies). These covenants make the contracts subject to more than mortality risk (e.g., market, credit, economic risks). The ignorance of these covenants will cause difficulties. For example, in the 1990s, the traditional actuarial method ignores the pricing of interest rate guarantee and thus induces a wide range of default insurance businesses. They are embedded options in the contracts, for which the valuation has to rely on option pricing method. Boyle and Schwartz (1977); Brennan and Schwartz (1979); Brennan, Schwartz, et al. (1976) are the seminal works to price life insurance contracts with market pricing method. The popularity of this method is also promoted by the regulatory requirement of the Solvency II that the assets and liabilities of insurance company need to be priced using market models, and also by the fair value concept in several international standards such as the International Accounting Standard Board (IASB), the Financial Accounting Standard Board (FASB) and the International Financial Reporting Standard (IFRS), etc.

Participating contracts is a popular class of insurance contracts. The premiums of these contracts are invested in a reference portfolio. The policyholder not only receives the guaranteed minimum return, but also participates in yields of the reference investment exceeding the minimum guarantee. Bacinello (2001) incorporate mortality risk into the pricing of an active participating contract in the Italian market. Bacinello (2003a,b); Grosen and Jørgensen (2000) price the participating contract with a surrender option via a binomial tree method whereas Jensen, Jørgensen, and Grosen (2001) makes use of a finite difference approach. Miltersen and Persson (2003) provide a closed-form solution for the participating contract with a cliquet-style guarantee and Tanskanen and Lukkarinen (2003) value the participating contract with a Bermudan-style surrender option and with a switch option of changing the reference investment and bonus policy. Both of the latter two papers work with constant interest rates. Briys and De Varenne (1994, 1997) take care of the credit risk and interest rate risk in the valuation of participating contracts. However, they only allow default to happen at maturity. Grosen and Jørgensen (2002) introduce early default and obtain closed-form pricing formulas, whereas the interest rate risk is ignored. Bernard, Le Courtois, and Quittard-Pinon (2005) consider early default risk and model the term structure of interest rates using an extended Vasicek model, where they compute the contract price by using an extended Fortet method. Bernard, Le Courtois, and Quittard-Pinon (2006) exploit the same model and obtain closed-form formulas for the price of a new type of participating contract, whose minimum guarantee is a proportion of zero-coupon bonds. Note that default risk is non-neglectable if we recall the unfortunate events happening in the 1990s. Recently, Cheng and Li (2018) and Le Courtois and Nakagawa (2013) also examine the mutual effects between early default risk and surrender risk when pricing participating contracts. However, all the above models are restricted to the Black-Schole model. As pointed out by Bacinello (2001), the life insurance contracts are

## 4.1. Introduction

---

usually long-term contracts and a Black-Schole framework doesn't fit the long-term contracts.

There are two main strands of the literature that is dedicated to an improvement of the Black-Schole model. The first one is the Lévy models where the market jump risk, i.e., sudden movements of large magnitude in asset prices, is incorporated. [Ballotta \(2005\)](#) employs a Merton jump diffusion model to price the participating contracts and also analyzes mispricing impacts of ignoring jump risk. [Kassberger, Kiesel, and Liebmann \(2008\)](#) prices several popular types of participating contracts with Meixner and NIG models. [Le Courtois and Quittard-Pinon \(2008\)](#) considers both of default risk and market jump risk by using a double exponential jump diffusion model. The second strand is the regime switching model where the economic risk (also named regime switching risk), i.e., structural changes of macro-economic environment or business cycles inducing changes in the dynamics of asset price or in the term structure of interest rates, is considered. [Siu \(2005\)](#) provides the pricing results of a participating contract with a surrender option, when the dynamics of the reference portfolio is modeled by a regime switching Brownian motion. [Siu, Lau, and Yang \(2008\)](#) and [Fard and Siu \(2013\)](#) price participating contracts under a regime switching jump diffusion model, where [Siu, Lau, and Yang \(2008\)](#) adopts a simulation method and [Fard and Siu \(2013\)](#) uses a numerical PIDE approach, but no early default is allowed in both of the two works. It is obviously important to incorporate the economic risk into the pricing of life insurance contracts if we recall that the life insurance contracts are long-term contracts. Many works such as [Cui, Kirkby, and Nguyen \(2017\)](#); [Fan, Shen, Siu, and Wang \(2015\)](#); [Hieber \(2017\)](#); [Lin, Tan, and Yang \(2009\)](#); [Siu, Yam, and Yang \(2015\)](#) have contributed to the pricing of insurance contracts using regime switching models and the American Academy of Actuaries and the Canadian Institute of Actuaries have also recommended the use of regime switching models.

In this paper, we incorporate market, economic, credit risks into the pricing of participating contracts, where the dynamics of the reference portfolio value is modeled by a regime switching double exponential jump diffusion model (regime switching Kou model). Following the works of [Bernard, Le Courtois, and Quittard-Pinon \(2005, 2006\)](#); [Grosen and Jørgensen \(2002\)](#); [Le Courtois and Quittard-Pinon \(2008\)](#), we assume the default of insurer happens when the value of the reference portfolio falls below a proportion of a minimum guarantee, which involves the first passage time problem of the regime switching Kou process. We introduce from [Le Courtois and Su \(2018\)](#) a complete solution, which includes a practical first passage time result that involves some matrix Wiener-Hopf factors and also a numerical algorithm of computing matrix Wiener-Hopf factorization. Wiener-Hopf factorization is a powerful tool to study the first passage time problem of stochastic process, specifically through decomposing the stochastic process into its ascending and descending ladder processes. The related theory has been widely used in the pricing work, for instance, see [Asmussen, Avram, and Pistorius \(2004\)](#); [Fusai, Germano, and Marazzina \(2016\)](#); [Green, Fusai, and Abrahams \(2010\)](#); [Jiang and Pistorius \(2008\)](#); [Jobert and Rogers \(2006\)](#). The Wiener-Hopf factorization built on Markov chains is also

## Chapter 4. Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

---

named as the matrix Wiener-Hopf factorization, the concept of which is introduced by Barlow, Rogers, and Williams (1980); London, McKean, Rogers, and Williams (1982a,b). Rogers et al. (1994) studies the matrix Wiener-Hopf factorization of a regime switching Brownian motion in the context of noisy fluid models. Rogers and Shi (1994) develops a numerical method to compute the corresponding matrix Wiener-Hopf factorization. Jiang and Pistorius (2008) and Mijatović and Pistorius (2011) develop matrix Wiener-Hopf factorization for regime switching double phase-type jump diffusion models. Le Courtois and Su (2018) provide a numerical method of computing matrix Wiener-Hopf factorization for the regime switching jump diffusion model. To the best of our knowledge, the matrix Wiener-Hopf factorization has not yet been used in the insurance domain. Based on matrix Wiener-Hopf factors, we deduce closed-form formulas for the price of participating contracts with a floating guarantee rate and with a constant guarantee rate up to Laplace or Laplace-Fourier transform, where only some matrix Wiener-Hopf factors are involved. Then, we obtain the price by performing numerical Laplace and Fourier inversion and by implementing matrix Wiener-Hopf factorization. Numerical illustrations show that our pricing methods are accurate and efficient and also indicate that the contract with a floating guarantee rate is a riskier but more worthy product when comparing to the contract with a constant guarantee rate. We also introduce both the dynamic and semi-static hedging strategies to hedge the participating contracts.

The rest of the paper is organized as follows. Section 2 presents the participating contracts and the regime switching Kou model. Section 3 introduces the first passage time result of the regime switching Kou process in terms of matrix Wiener-Hopf factors and also the numerical algorithm of computing matrix Wiener-Hopf factorization. Section 4 derives closed-form formulas for the price of participating contracts with a constant guarantee rate and demonstrates the accuracy and efficiency of the method. Section 5 makes comparison between participating contracts with a constant guarantee rate and participating contracts with a floating guarantee rate. Section 6 presents the dynamic and semi-static hedging strategies for the participating contracts.

### 4.2 The Basic Framework

This section begins with a specification of a participating contract. Then, we conclude with the modeling of investment funds and interest rate dynamics.

#### 4.2.1 Contract Specifications

We study a type of participating contract with a constant minimum guaranteed rate  $\tilde{r}_g$ . The interest rate guarantee promises the policyholders that the premium payment  $L_0$  will accumulate by the rate  $\tilde{r}_g$  during the life of the contract. Then, a guaranteed maturity payment is  $L_T^g = L_0 e^{\tilde{r}_g T}$ , where  $L_0$  is a premium payment from policyholders and  $T$  is the maturity of the contract.

## 4.2. The Basic Framework

---

The life insurance company is supposed to invest  $A_0$  in a reference portfolio. The initial capital of investment funds  $A_0$  is financed by the premium payment of policyholders  $L_0 = \alpha A_0$  and the initial investment of equityholders  $E_0 = (1 - \alpha)A_0$  at time zero. The parameter  $\alpha$ , termed as the wealth distribution coefficient, captures the initial proportion of funds provided by policyholders.

If the maturity value of investment funds is sufficient, i.e.,  $A_T \geq L_T^g$ , the promised maturity payment will be attained. Otherwise, the policyholders seize the total funds value. Once the funds run enough well, the policyholders obtain the bonus payment  $\delta(\alpha A_T - L_T^g)_+$ , i.e., a fraction  $\delta$  of the surplus of the funds maturity value contributed by policyholders  $\alpha A_T$  minus the promised maturity payment  $L_T^g$ . The parameter  $\delta$  is referred to as the participation coefficient. The bankruptcy of insurance company will force closure of the policy prior to the maturity date. The insurance company holds a large amount of financial investments and the performance of financial investments is tightly related to the solvency of insurance company. We assume that the insurance company is continuously monitored and the default happens when the funds value  $A_t$  falls below a default boundary  $B_t = \lambda L_t^g$  where  $0 < \lambda < 1$  is a boundary level parameter and  $L_t^g = L_0 e^{\bar{r}g t}$  is a nominal promise payment at time  $t \in [0, T]$ . Let

$$\tau = \inf\{t \geq 0 : A_t \leq \lambda L_t^g\}$$

be the bankruptcy time. Denote by  $r_s$  the market interest rate at time  $s$ . Then, the discounted payoffs of the components of the participating policy are as follows:

- $P_1$ : a promised maturity payment,

$$e^{-\int_0^T r_s ds} L_T^g \mathbf{1}_{\tau \geq T};$$

- $P_2$ : short position of a put option to default at maturity,

$$-e^{-\int_0^T r_s ds} (L_T^g - A_T)_+ \mathbf{1}_{\tau \geq T};$$

- $P_3$ : bonus option,

$$e^{-\int_0^T r_s ds} \delta(\alpha A_T - L_T^g)_+ \mathbf{1}_{\tau \geq T};$$

- $P_4$ : a rebate at early default case,

$$e^{-\int_0^\tau r_s ds} A_\tau \mathbf{1}_{\tau < T}.$$

## Chapter 4. Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

---

Then, the pricing formula under the risk-neutral measure  $\mathbb{Q}$  is as follows:

$$V = E_{\mathbb{Q}} \left( e^{-\int_0^T r_s ds} (L_T^g - (L_T^g - A_T)_+ + \delta(\alpha A_T - L_T^g)_+) \mathbf{1}_{\tau \geq T} + e^{-\int_0^{\tau} r_s ds} A_{\tau} \mathbf{1}_{\tau < T} \right). \quad (4.1)$$

To do the valuation, we specify the dynamics of the funds value and the interest rates in the following subsection.

### 4.2.2 Investment Funds and Interest Rate Dynamics

Let  $(\Omega, \mathcal{F}, \mathbb{Q})$  be a complete probability space. The states of the economy are modeled by a continuous time Markov chain process  $J = \{J_t; t \geq 0\}$  on  $(\Omega, \mathcal{F}, \mathbb{Q})$  with a finite state space  $E_0 = \{e_1, e_2, \dots, e_n\}$ , where  $e_i = (0, \dots, 1, \dots, 0) \in \mathbb{R}^n$ . Let  $Q$  be the generator matrix of  $J$ , where

$$Q = \{q_{ij}\}_{1 \leq i, j \leq n},$$

$q_{ij} \geq 0, \forall i \neq j$  and  $\sum_{j=1}^n q_{ij} = 0, i = 1, \dots, n$ . Then, the transition probabilities matrix is

$$P(s, t) = e^{Q(t-s)} \quad \forall s \leq t$$

and the  $(i, j)^{\text{th}}$  element  $p_{i,j}(s, t)$  is the probability of switching from state  $e_i$  at time  $s$  to state  $e_j$  at time  $t$ .

The dynamics of the funds value is supposed to follow changes of an exponential regime switching jump diffusion process under the risk-neutral measure  $\mathbb{Q}$ :

$$A_t = A_0 e^{X_t},$$

where  $A_0$  is the initial funds value and  $X$  is a regime switching jump diffusion process:

$$X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \int_0^t dN_s,$$

where  $W$  is a standard Brownian motion,  $\sigma_s = \langle \hat{\sigma}, J_s \rangle$ ,  $N_s = \langle \hat{N}, J_s \rangle$ , and where  $\langle \cdot, \cdot \rangle$  denotes the inner product in  $\mathbb{R}^n$ ,  $\hat{\sigma} = (\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_n)$  and  $\hat{N} = (\hat{N}_1, \hat{N}_2, \dots, \hat{N}_n)$ . For each state  $e_i \in E_0$ ,  $\hat{\sigma}_i \geq 0$ ,  $\hat{N}_i = \{\hat{N}_{it}; t \geq 0\}$  is a compound Poisson process with jump intensity  $\hat{\lambda}_i > 0$  and with jump size that follows a double exponential distribution whose density function is given by:

$$f_i(y) = \hat{p}_i \hat{\eta}_i e^{-\hat{\eta}_i y} I_{\{y \geq 0\}} + \hat{q}_i \hat{\theta}_i e^{\hat{\theta}_i y} I_{\{y < 0\}},$$

where  $\hat{p}_i \geq 0, \hat{q}_i \geq 0, \hat{\eta}_i > 1, \hat{\theta}_i > 0$ , and  $\hat{p}_i + \hat{q}_i = 1$ . The stochastic processes  $W, \hat{N}_i, i = 1, \dots, n$  and  $J$  are all independent.

The riskless rate also changes with the state of the economy. Let the riskless rate be  $r_t = \langle \hat{r}, J_t \rangle$ , where  $\hat{r} = (\hat{r}_1, \hat{r}_2, \dots, \hat{r}_n)$  and  $\hat{r}_i > 0$  is the riskless rate at state

## 4.2. The Basic Framework

---

$e_i$ . Then, the martingale condition  $E_Q(e^{-\int_0^t r_s ds} A_t) = A_0$  gives

$$\hat{\mu}_i = \hat{r}_i - \frac{1}{2}\hat{\sigma}_i^2 - \hat{\lambda}_i \left( \frac{\hat{p}_i \hat{\eta}_i}{\hat{\eta}_i - 1} + \frac{\hat{q}_i \hat{\theta}_i}{\hat{\theta}_i + 1} - 1 \right).$$

Denote  $X$  at state  $e_i$  as  $X^i$ . Let the  $n \times n$  matrix  $M_t(z)$  be the moment generating function of  $X_t$ , where the  $(i, j)$ <sup>th</sup> element equals to  $E(e^{zX_t}; J_t = e_j | J_0 = e_i)$ . Then,  $M_t(z) = e^{tG(z)}$ , where

$$G(z) = Q + \text{diag}\{\varphi_k(z)\},$$

and  $\varphi_k(z)$  is the Laplace exponent of  $X^k$  defined as follows:

$$\varphi_k(z) = \hat{\mu}_k z + \frac{1}{2}\hat{\sigma}_k^2 z^2 + \hat{\lambda}_k \left( \frac{\hat{p}_k \hat{\eta}_k}{\hat{\eta}_k - z} + \frac{\hat{q}_k \hat{\theta}_k}{\hat{\theta}_k + z} - 1 \right).$$

Let  $\mathcal{G}$  and  $\mathcal{H}$  be the filtrations generated by  $J$  and  $\{X^i; i = 1, 2, \dots, n\}$ , respectively. Then,  $\mathcal{F} = \mathcal{G} \vee \mathcal{H}$ .

### 4.2.3 Main Subcontract Terms

We rewrite the pricing formula (3.1) into four terms as follows:

$$V = GF + PO + BO + LR,$$

where

$$\left\{ \begin{array}{l} GF = E_{\mathbb{Q}} \left( e^{-\int_0^T r_s ds} L_T^g \mathbf{1}_{\tau \geq T} \right) \\ PO = E_{\mathbb{Q}} \left( -e^{-\int_0^T r_s ds} (L_T^g - A_T)_+ \mathbf{1}_{\tau \geq T} \right) \\ BO = E_{\mathbb{Q}} \left( e^{-\int_0^T r_s ds} \delta(\alpha A_T - L_T^g)_+ \mathbf{1}_{\tau \geq T} \right) \\ LR = E_{\mathbb{Q}} \left( e^{-\int_0^{\tau} r_s ds} A_{\tau} \mathbf{1}_{\tau < T} \right) \end{array} \right. \quad (4.2)$$

and  $GF, PO, BO, LR$  are expectations of the four components  $P_1, P_2, P_3, P_4$ , respectively.

Let  $Z_t = X_t - \tilde{r}_g t$  and  $Z_t$  is still a regime switching Kou process. Let  $Z^i$  be  $Z$  at state  $e_i$  and the parameters of  $Z^i$  are the same as  $X^i$  except drift term  $\hat{\mu}_i^Z$ , where

$$\hat{\mu}_i^Z = \hat{r}_i - \tilde{r}_g - \frac{1}{2}\hat{\sigma}_i^2 - \hat{\lambda}_i \left( \frac{\hat{p}_i \hat{\eta}_i}{\hat{\eta}_i - 1} + \frac{\hat{q}_i \hat{\theta}_i}{\hat{\theta}_i + 1} - 1 \right).$$

## Chapter 4. Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

---

We rewrite the default time  $\tau$  in a constant barrier form as

$$\tau = \inf\{t \geq 0 : Z_t \leq \ln \frac{\lambda L_0}{A_0}\},$$

and the default characterization becomes related to a first passage time problem of the regime switching Kou process  $Z$ .

### 4.3 The First Passage Time Results of Regime Switching Jump Diffusion Processes

We now shall introduce the first passage time result of regime switching Kou processes, which is built on the matrix Wiener-Hopf factorization, and based on which we can deduce closed-form solutions of four subcontract terms under Laplace or Laplace-Fourier transforms. We start with the definition of matrix Wiener-Hopf factorization of a regime switching jump diffusion process from [Jiang and Pistorius \(2008\)](#). Then, from [Le Courtois and Su \(2018\)](#), we introduce a practical first passage time result and also a numerical algorithm to compute the Wiener-Hopf factors. To save notations, the result is illustrated with  $X$  but in fact used with  $Z$ .

#### 4.3.1 Matrix Wiener-Hopf Factorization

The matrix Wiener-Hopf factorization of  $X$  is built on its fluid embedding. Let  $M = \{M_t; t \geq 0\}$  be the fluid embedding of  $X$ , where  $M$  is a continuous stochastic process whose sample paths are constructed from the sample paths of  $X$  by replacing positive jumps with linear segments of slope +1 and negative jumps with linear segments of slope -1. Let  $Y = \{Y_t; t \geq 0\}$  be an irreducible continuous time Markov chain with a finite state space:

$$E = E^+ \cup E^0 \cup E^- = \{s_1, s_2, \dots, s_n\} \cup \{s_{n+1}, s_{n+2}, \dots, s_{2n}\} \cup \{s_{2n+1}, s_{2n+2}, \dots, s_{3n}\},$$

where the unit vector  $s_i = (0, \dots, 1, \dots, 0) \in \mathbb{R}^{3n}$ . The spaces  $E^0$ ,  $E^+$  and  $E^-$  correspond to the states where  $X$  moves as a pure diffusion, makes a positive jump and makes a negative jump, respectively.

Conditional on the enlarged Markov chain  $Y$ , the process  $M$  can be expressed as follows:

$$M_t = M_0 + \int_0^t u(Y_s) ds + \int_0^t v(Y_s) dW_s,$$

where

$$u(s_j) = \begin{cases} 1 & \text{if } s_j \in E^+ \\ \hat{\mu}_{j-n} & \text{if } s_j \in E^0 \\ -1 & \text{if } s_j \in E^- \end{cases} \quad \text{and} \quad v(s_j) = \begin{cases} \hat{\sigma}_{j-n} & \text{if } s_j \in E^0 \\ 0 & \text{otherwise.} \end{cases}$$

### 4.3. The First Passage Time Results of Regime Switching Jump Diffusion Processes

---

Denote by the  $3n \times 3n$  matrix:

$$Q_a = \begin{pmatrix} T^+ & -T^+ & O_{n \times n} \\ B^+ & Q - D_a & B^- \\ O_{n \times n} & -T^- & T^- \end{pmatrix},$$

where the  $Q$  is the generator matrix of  $J$ ,  $a = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n)$ ,  $D_a = \text{diag}(\hat{\lambda}_i + \hat{a}_i)$ , and  $O_{n \times n}$  is a zero matrix of size  $n \times n$ . The

$$B^+ = \begin{pmatrix} \hat{\lambda}_1 \hat{p}_1 & & \\ & \ddots & \\ & & \hat{\lambda}_n \hat{p}_n \end{pmatrix}, \quad B^- = \begin{pmatrix} \hat{\lambda}_1 \hat{q}_1 & & \\ & \ddots & \\ & & \hat{\lambda}_n \hat{q}_n \end{pmatrix}$$

and

$$T^+ = \begin{pmatrix} -\hat{\eta}_1 & & \\ & \ddots & \\ & & -\hat{\eta}_n \end{pmatrix}, \quad T^- = \begin{pmatrix} -\hat{\theta}_1 & & \\ & \ddots & \\ & & -\hat{\theta}_n \end{pmatrix}.$$

The generator matrix of  $Y$  is  $Q_0$ . The regime switching discounting rate  $a$  with  $\hat{a}_i$  at state  $e_i$  can be incorporated into the exponential of the process  $X$  by changing the generator matrix of  $Y$  from  $Q_0$  to  $Q_a$ . For a general statement, we assume that the generator matrix of  $Y$  is  $Q_a$ .

Then, the matrix Wiener-Hopf factorization of  $(M, Y)$  is defined as follows:

**Definition 1** Let  $(G^{(a,+)}, G^{(a,-)})$  be a pair of irreducible  $2n \times 2n$  matrices, that is, matrices with non-negative off-diagonal elements and non-positive row sums, and  $(\Pi^{(a,+)}, \Pi^{(a,-)})$  be a pair of  $n \times 2n$  matrices with rows made of sub-probability vectors. The quadruple

$$\left( \Pi^{(a,+)}, G^{(a,+)}, \Pi^{(a,-)}, G^{(a,-)} \right)$$

is the Wiener-Hopf factorization of  $(M, Y)$  that is associated with  $a > 0$  if

$$\Xi(-G^{(a,+)}, W^{(a,+)}) = \Xi(G^{(a,-)}, W^{(a,-)}) = O_{3n \times 2n},$$

where

$$\Xi(S, W) = \frac{1}{2} \Sigma^2 W S^2 + V W S + Q_a W,$$

**Chapter 4. Pricing and Hedging Defaultable Participating Contracts  
with Regime Switching and Jump Risk**

---

with the  $3n \times 3n$  diagonal matrices:

$$\Sigma = \begin{pmatrix} O_{n \times n} & & & \\ & \begin{pmatrix} \hat{\sigma}_1 & & \\ & \ddots & \\ & & \hat{\sigma}_n \end{pmatrix} & \\ & & & O_{n \times n} \end{pmatrix}, \quad V = \begin{pmatrix} I_n & & & \\ & \begin{pmatrix} \hat{\mu}_1 & & \\ & \ddots & \\ & & \hat{\mu}_n \end{pmatrix} & \\ & & & -I_n \end{pmatrix},$$

and the  $3n \times 2n$  matrices:

$$W^{(a,+)} = \begin{pmatrix} I_{2n} \\ \Pi^{(a,+)} \end{pmatrix}, \quad W^{(a,-)} = \begin{pmatrix} \Pi^{(a,-)} \\ I_{2n} \end{pmatrix},$$

where  $I_n$  and  $I_{2n}$  are identity matrices of size  $n \times n$  and  $2n \times 2n$ , respectively, and  $O_{3n \times 2n}$  denotes a zero matrix of size  $3n \times 2n$ .

### 4.3.2 The First Passage Time Results

The up-crossing and down-crossing ladder processes  $\tilde{Y}^+$ ,  $\tilde{Y}^-$  of  $(M, Y)$  are defined as follows:

$$\tilde{Y}_z^+ = Y_{\tau_z^+} \quad \text{and} \quad \tilde{Y}_z^- = Y_{\tau_z^-},$$

where

$$\tau_z^+ = \inf\{s \geq 0 : M_s \geq z\} \quad \text{and} \quad \tau_z^- = \inf\{s \geq 0 : M_s \leq z\}.$$

The  $\tilde{Y}^+$  and  $\tilde{Y}^-$  are Markov processes with state spaces  $E^0 \cup E^+$  and  $E^0 \cup E^-$ , respectively. Let  $Q^{(a,+)}$  and  $Q^{(a,-)}$  be the generator matrices of  $\tilde{Y}^+$  and  $\tilde{Y}^-$ , respectively. Denote by  $\zeta^{(a,+)}$  and  $\zeta^{(a,-)}$  the initial distributions of  $\tilde{Y}^+$  and  $\tilde{Y}^-$  where the  $(i, j)^{\text{th}}$  element is defined by:

$$\left(\zeta^{(a,+)}\right)_{i,j} = P_{0,s_{2n+i}}(\tilde{Y}_0^+ = s_j, \tau_0^+ < \infty), \quad \text{for } s_{2n+i} \in E^-, s_j \in E^0 \cup E^+ \quad (4.3)$$

and

$$\left(\zeta^{(a,-)}\right)_{i,j} = P_{0,s_i}(\tilde{Y}_0^- = s_{n+j}, \tau_0^- < \infty), \quad \text{for } s_i \in E^+, s_{n+j} \in E^0 \cup E^-, \quad (4.4)$$

where  $P_{0,s_k}(\cdot) = P(\cdot | M_0 = 0, Y_0 = s_k)$ . Jiang and Pistorius (2008) have proved that the quadruple  $(\zeta^{(a,+)}, Q^{(a,+)}, \zeta^{(a,-)}, Q^{(a,-)})$  is a unique Wiener-Hopf factorization of  $(M, Y)$  when  $a > 0$ .

Let  $\tilde{\theta} = \min\{\hat{\theta}_i; i = 1, \dots, n\}$ . From Le Courtois and Su (2018), we introduce the first passage time result of  $X$  as follows:

**Proposition 1** Denote the first passage time of  $X$  below a constant level  $b$  as  $\tau$ , so

### 4.3. The First Passage Time Results of Regime Switching Jump Diffusion Processes

---

that

$$\tau = \inf\{t \geq 0 : X_t \leq b\}.$$

Let  $a_t = \langle a, J_t \rangle$  and the contingent payoff be  $h_\tau = \langle J_\tau, \hat{h} \rangle$  where  $\hat{h} = (\hat{h}_1, \dots, \hat{h}_n)$ . For any  $w > -\hat{\theta}$ , we have

$$E \left( e^{-\int_0^\tau a_s ds + w X_\tau} h_\tau \right) = Y_0 W^{(a,-)} e^{Q^{(a,-)}(x-b) + w b I_{2n}} \tilde{h},$$

where  $x$  is the initial point of  $X$ ,  $Y_0$  is the initial state vector of  $Y$ ,

$$\tilde{h} = \left( \left( \hat{h}_1, \dots, \hat{h}_n \right), \left( \frac{\hat{\theta}_1}{w + \hat{\theta}_1} \hat{h}_1, \dots, \frac{\hat{\theta}_n}{w + \hat{\theta}_n} \hat{h}_n \right) \right)',$$

and  $W^{(a,-)}$  is given in the definition 1 with  $\Pi^{(a,-)} = \zeta^{(a,-)}$ .

#### 4.3.3 Numerical Algorithm

From [Le Courtois and Su \(2018\)](#), we introduce a numerical method to compute matrix Wiener-Hopf factorization. Once we have  $(\zeta^{(a,+)}, Q^{(a,+)}, \zeta^{(a,-)}, Q^{(a,-)})$ , we can compute the first passage time results and also the value of four subcontract terms, whose closed-form expressions will be built on these matrix Wiener-Hopf factors in next section.

**Numerical Algorithm for the Computation of  $(\zeta^{(a,+)}, Q^{(a,+)}, \zeta^{(a,-)}, Q^{(a,-)})$**

- Step 1: Compute  $4n$  roots

$$\Re(\nu_1) \leq \Re(\nu_2) \leq \dots \leq \Re(\nu_{2n}) \leq 0 \leq \Re(\nu_{2n+1}) \leq \Re(\nu_{2n+2}) \leq \dots \leq \Re(\nu_{4n})$$

from the equation  $f(\nu) = 0$  where  $f(\nu) = \det(K(\nu)) = 0$  and  $K(\nu) = \frac{1}{2} \Sigma^2 \nu^2 - V\nu + Q_a$ . Let

$$\begin{aligned} \tilde{\beta}_i &= \nu_i, & i &= 1, \dots, 2n, \\ \bar{\beta}_j &= -\nu_{2n+j}, & j &= 1, \dots, 2n. \end{aligned}$$

- Step 2: For  $i = 1, \dots, 4n$ , compute the  $3n \times 1$  vector  $\gamma_i$  by solving a system of linear equations  $K(\nu_i) \gamma_i = 0$ .

- Step 3: Let

$$\begin{aligned} \tilde{\vartheta}_i &= (\gamma_{i,1}, \dots, \gamma_{i,2n})', & i &= 1, \dots, 2n, \\ \bar{\vartheta}_j &= (\gamma_{2n+j,n+1}, \dots, \gamma_{2n+j,3n})', & j &= 1, \dots, 2n, \end{aligned}$$

and

$$\tilde{Z} = [\tilde{\vartheta}_1, \tilde{\vartheta}_2, \dots, \tilde{\vartheta}_{2n}], \quad \bar{Z} = [\bar{\vartheta}_1, \bar{\vartheta}_2, \dots, \bar{\vartheta}_{2n}].$$

Then, we obtain

$$Q^{(a,+)} = \tilde{Z} \operatorname{diag}\{\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_{2n}\} \tilde{Z}^{-1},$$

$$Q^{(a,-)} = \bar{Z} \operatorname{diag}\{\bar{\beta}_1, \bar{\beta}_2, \dots, \bar{\beta}_{2n}\} \bar{Z}^{-1}.$$

The matrix exponential is computed as:

$$e^{Q^{(a,+)}x} = \tilde{Z} \operatorname{diag}\{e^{\tilde{\beta}_1 x}, e^{\tilde{\beta}_2 x}, \dots, e^{\tilde{\beta}_{2n} x}\} \tilde{Z}^{-1},$$

$$e^{Q^{(a,-)}x} = \bar{Z} \operatorname{diag}\{e^{\bar{\beta}_1 x}, e^{\bar{\beta}_2 x}, \dots, e^{\bar{\beta}_{2n} x}\} \bar{Z}^{-1}.$$

- Step 4: For  $k = 1, \dots, n$ , compute  $2n \times 1$  vector  $\tilde{\xi}_k$  by solving a system of linear equations  $\tilde{Z}'\tilde{\xi}_k = (\gamma_{1,2n+k}, \dots, \gamma_{2n,2n+k})'$  and compute  $2n \times 1$  vector  $\bar{\xi}_k$  by solving a system of linear equations  $\bar{Z}'\bar{\xi}_k = (\gamma_{2n+1,k}, \dots, \gamma_{4n,k})'$ . Then,

$$\zeta^{(a,+)} = [\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_n]'$$

and

$$\zeta^{(a,-)} = [\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_n]'$$

## 4.4 Contract Valuation

In this section, we deduce closed-form formulas for Laplace or Laplace-Fourier transforms of the subcontract terms. Before we proceed, we introduce a useful lemma and some notations that will be used in the sequel.

Let  $\mathbf{1}_n$  be the vector of ones with size  $n \times 1$ . Recall that  $Q$  is the generator matrix of  $J$ , we introduce the following lemma:

**Lemma 1** (Mijatović and Pistorius (2011))

Let  $\hat{d} \in \mathbb{C}^n$  and  $V = \operatorname{diag}\{\hat{d}\}$ , then

$$E \left( e^{-\int_0^t \langle \hat{d}, J_s \rangle ds} \right) = J_0 e^{(Q-V)t} \mathbf{1}_n.$$

If  $\Re(\hat{d}) > 0$ , the matrix  $Q - V$  is invertible and the following formula holds:

$$\int_0^\infty e^{(Q-V)t} dt = (V - Q)^{-1}.$$

Let  $W_Z^{(a,-)}, Q_Z^{(a,-)}$  be the  $W^{(a,-)}, Q^{(a,-)}$  given in definition 1 with  $X = Z$ . De-

#### 4.4. Contract Valuation

---

note by  $H$  the following  $2n \times n$  matrix:

$$H(w) = \begin{pmatrix} & I_n & \\ \begin{pmatrix} \hat{\theta}_1 \\ w + \hat{\theta}_1 \end{pmatrix} & & \\ & \ddots & \\ & & \begin{pmatrix} \hat{\theta}_n \\ w + \hat{\theta}_n \end{pmatrix} \end{pmatrix}.$$

##### 4.4.1 Computation of $GF$ and $LR$

Let  $\widehat{GF}$  be the Laplace transform of  $GF$ . For any  $u > 0$ , we obtain with the Proposition 1

$$\begin{aligned} \widehat{GF}(u) &= L_0 \int_0^\infty e^{-uT} E_{\mathbb{Q}} \left( e^{-\int_0^T (r_s - \tilde{r}_g) ds} \mathbf{1}_{\tau \geq T} \right) dT \\ &= L_0 E_{\mathbb{Q}} \left( \int_0^\tau e^{-\int_0^T (r_s + u - \tilde{r}_g) ds} dT \right) \\ &= L_0 E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_0^T (r_s + u - \tilde{r}_g) ds} dT - e^{-\int_0^\tau (r_s + u - \tilde{r}_g) ds} \int_0^\infty E_{\mathbb{Q}} \left( e^{-\int_\tau^{\tau+T} (r_s + u - \tilde{r}_g) ds} | \mathcal{F}_\tau \right) dT \right) \\ &= L_0 \left( Y_0 W_{\tilde{Z}}^{(\hat{r} + u - \tilde{r}_g, -)} e^{Q_{\tilde{Z}}^{(\hat{r} + u - \tilde{r}_g, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} H(0) - J_0 \right) (Q - \text{diag}\{\hat{r} + u - \tilde{r}_g\})^{-1} \mathbf{1}_n. \end{aligned}$$

Let  $\widehat{LR}$  be the Laplace transform of  $LR$ . In the same manner, we have

$$\begin{aligned} \widehat{LR}(u) &= \int_0^\infty e^{-uT} E_{\mathbb{Q}} \left( e^{-\int_0^\tau r_s ds} A_\tau \mathbf{1}_{\tau < T} \right) dT \\ &= A_0 E_{\mathbb{Q}} \left( \int_\tau^\infty e^{-uT} e^{-\int_0^\tau (r_s - \tilde{r}_g) ds} e^{Z_\tau} dT \right) \\ &= \frac{A_0}{u} E_{\mathbb{Q}} \left( e^{-\int_0^\tau (r_s + u - \tilde{r}_g) ds + Z_\tau} \right) \\ &= \frac{A_0}{u} Y_0 W_{\tilde{Z}}^{(\hat{r} + u - \tilde{r}_g, -)} e^{Q_{\tilde{Z}}^{(\hat{r} + u - \tilde{r}_g, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right) + \ln \frac{\lambda L_0}{A_0}} I_{2n} H(1) \mathbf{1}_n. \end{aligned}$$

Now we have closed-form formulas for  $\widehat{GF}$  and  $\widehat{LR}$ . The numerical Laplace inversion and numerical method of computing matrix Wiener-Hopf factors enable us to compute the  $GF$  and  $LR$ .

#### 4.4.2 Computation of $PO$ and $BO$

The two subcontract terms  $PO$  and  $BO$  are two single barrier options. For simple notations, we use the auxiliary  $z_1 = \log(\alpha)$  and  $z_2 = 0$  and rewrite  $PO$  and  $BO$  as

$$\begin{cases} PO = -A_0 E_{\mathbb{Q}} \left( e^{-\int_0^T (r_s - \tilde{r}_g) ds} (e^{z_1} - e^{Z_T})_+ \mathbf{1}_{\tau \geq T} \right), \\ BO = \delta \alpha A_0 E_{\mathbb{Q}} \left( e^{-\int_0^T (r_s - \tilde{r}_g) ds} (e^{Z_T} - e^{-z_2})_+ \mathbf{1}_{\tau \geq T} \right). \end{cases}$$

The exponential decaying term  $e^{-\alpha_1 z_1}$ ,  $\alpha_1 > 0$  is introduced into the  $PO$  to guarantee that  $e^{-\alpha_1 z_1} PO$  is square integrable in  $z_1$  over the positive line. Let  $\widetilde{PO}$  be the Laplace-Fourier transform of  $PO$ . Denote by  $\phi_k^Z(z)$  the Laplace exponent of  $Z^k$ . For any  $1 < \alpha_1 < \tilde{\theta} + 1$  and  $u > \max(\Re(\phi_1^Z(iv + 1 - \alpha_1)) + \tilde{r}_g - \hat{r}_1, \dots, \Re(\phi_n^Z(iv + 1 - \alpha_1)) + \tilde{r}_g - \hat{r}_n, 0)$ , we obtain

$$\begin{aligned} \widetilde{PO}(u, v) &= -A_0 \int_0^\infty e^{-uT} \int_{-\infty}^\infty e^{ivz_1} e^{-\alpha_1 z_1} E_{\mathbb{Q}} \left( e^{-\int_0^T (r_s - \tilde{r}_g) ds} (e^{z_1} - e^{Z_T})_+ \mathbf{1}_{\tau \geq T} \right) dz_1 dT \\ &= -A_0 E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_0^T (r_s + u - \tilde{r}_g) ds} \mathbf{1}_{\tau \geq T} \int_{Z_T}^\infty e^{-(\alpha_1 - iv)z_1} (e^{z_1} - e^{Z_T}) dz_1 dT \right) \\ &= -\frac{A_0}{(\alpha_1 - iv)(\alpha_1 - iv - 1)} E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_0^T (r_s + u - \tilde{r}_g) ds} \mathbf{1}_{\tau \geq T} e^{-(\alpha_1 - iv - 1)Z_T} dT \right) \\ &= -\frac{A_0}{(\alpha_1 - iv)(\alpha_1 - iv - 1)} \left( E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_0^T (r_s + u - \tilde{r}_g) ds} e^{-(\alpha_1 - iv - 1)Z_T} dT \right) \right. \\ &\quad \left. - \sum_{k=1}^n E_{\mathbb{Q}} \left( e^{-\int_0^{\tau} (r_s + u - \tilde{r}_g) ds} e^{-(\alpha_1 - iv - 1)Z_\tau} \mathbf{1}_{J_\tau = e_k} \right) \right. \\ &\quad \left. E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_\tau^{\tau+T} (r_s + u - \tilde{r}_g) ds} e^{-(\alpha_1 - iv - 1)(Z_{T+\tau} - Z_\tau)} dT | J_\tau = e_k \right) \right) \\ &= -\frac{A_0}{(\alpha_1 - iv)(\alpha_1 - iv - 1)} \left( Y_0 W_{\tilde{Z}}^{(\hat{r} + u - \tilde{r}_g, -)} e^{Q_{\tilde{Z}}^{(\hat{r} + u - \tilde{r}_g, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} \right. \\ &\quad \left. e^{-(\alpha_1 - iv - 1) \ln \frac{\lambda L_0}{A_0}} H(iv + 1 - \alpha_1) - J_0 \right) \\ &\quad (Q - \text{diag}\{\hat{r} + u - \tilde{r}_g - \phi_k^Z(iv + 1 - \alpha_1)\})^{-1} \mathbf{1}_n, \end{aligned}$$

where we make use of the strong Markov property of  $Z$  in the fourth equality and

#### 4.4. Contract Valuation

obtain the fifth equality by using the Proposition 1 and the Lemma 1.

Let  $\widetilde{BO}$  be the Laplace-Fourier transform of  $BO$ . For any  $\alpha_2 > 0$  and  $u > \max(\Re(\phi_1^Z(\alpha_2 - iv + 1)) + \tilde{r}_g - \hat{r}_1, \dots, \Re(\phi_n^Z(\alpha_2 - iv + 1)) + \tilde{r}_g - \hat{r}_n, 0)$ , the same derivation gives the  $\widetilde{BO}$  as follows:

$$\begin{aligned}
\widetilde{BO}(u, v) &= \delta\alpha A_0 \int_0^\infty e^{-uT} \int_{-\infty}^\infty e^{ivz_2} e^{-\alpha_2 z_2} E_{\mathbb{Q}} \left( e^{-\int_0^T (r_s - \tilde{r}_g) ds} (e^{Z_T} - e^{-z_2})_+ \mathbf{1}_{\tau \geq T} \right) dz_2 dT \\
&= \delta\alpha A_0 E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_0^T (r_s + u - \tilde{r}_g) ds} \mathbf{1}_{\tau \geq T} \int_{-Z_T}^\infty e^{-(\alpha_2 - iv)z_2} (e^{Z_T} - e^{-z_2}) dz_2 dT \right) \\
&= \frac{\delta\alpha A_0}{(\alpha_2 - iv)(\alpha_2 - iv + 1)} E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_0^T (r_s + u - \tilde{r}_g) ds} \mathbf{1}_{\tau \geq T} e^{(\alpha_2 - iv + 1)Z_T} dT \right) \\
&= \frac{\delta\alpha A_0}{(\alpha_2 - iv)(\alpha_2 - iv + 1)} \left( E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_0^T (r_s + u - \tilde{r}_g) ds} e^{(\alpha_2 - iv + 1)Z_T} dT \right) \right. \\
&\quad \left. - \sum_{k=1}^n E_{\mathbb{Q}} \left( e^{-\int_0^\tau (r_s + u - \tilde{r}_g) ds} e^{(\alpha_2 - iv + 1)Z_\tau} \mathbf{1}_{J_\tau = e_k} \right. \right. \\
&\quad \left. \left. E_{\mathbb{Q}} \left( \int_0^\infty e^{-\int_\tau^{\tau+T} (r_s + u - \tilde{r}_g) ds} e^{(\alpha_2 - iv + 1)(Z_{T+\tau} - Z_\tau)} dT | J_\tau = e_k \right) \right) \right) \\
&= \frac{\delta\alpha A_0}{(\alpha_2 - iv)(\alpha_2 - iv + 1)} \left( Y_0 W_{\tilde{Z}}^{(\hat{r} + u - \tilde{r}_g, -)} e^{Q_{\tilde{Z}}^{(\hat{r} + u - \tilde{r}_g, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} \right. \\
&\quad \left. e^{(\alpha_2 - iv + 1) \ln \frac{\lambda L_0}{A_0}} H(\alpha_2 - iv + 1) - J_0 \right) \\
&\quad (Q - \text{diag}\{\hat{r} + u - \tilde{r}_g - \phi_k^Z(\alpha_2 - iv + 1)\})^{-1} \mathbf{1}_n.
\end{aligned}$$

We have the explicit formulas of  $\widetilde{PO}$  and  $\widetilde{BO}$ . Then, the  $PO$  and  $BO$  can be computed by performing the Laplace-Fourier inversion of  $\widetilde{PO}$  and  $\widetilde{BO}$  and then by multiplying the term  $e^{\alpha_1 z_1}$  or  $e^{\alpha_2 z_2}$ , where the involved matrix Wiener-Hopf factors are also calculated by using the numerical method in subsection 3.3.

#### 4.4.3 Numerical Analysis

This subsection demonstrates our method by comparing with Monte Carlo simulation. Then, we examine the sensitivity of the fair participating rate  $\delta$  to the guaranteed rate  $r_g$  and to the wealth distribution coefficient  $\alpha$ .

We illustrate the numerical results with two economic states for simplicity. We preset the parameter values of contracts and funds value dynamics in Table 1 and Table 2. The two states  $e_1$  and  $e_2$  represent a "good" and a "bad" macroeconomic

## Chapter 4. Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

environment, respectively, where the "good" one bears higher interest rates and makes the funds value dynamics exhibit favorable features, such as less fluctuations, smaller average size of negative jumps and larger average size of positive jumps, etc. Note that some parameters are not listed yet: the initial state  $J_0$ ,  $Y_0$  and the state transition intensities  $q_{12}$  and  $q_{21}$ . They will be specified when used. Some of the parameters will be changed during the numerical study. We shall state the change when needed.

<b>Table 1</b>					<b>Contract Parameters</b>				
$\alpha$	$\tilde{r}_g$	$\delta$	$T$	$\lambda$					
0.85	0.02	0.9	10	0.8					

<b>Table 2</b>		<b>Funds Dynamics Parameters</b>					
$A_0$	State	$\hat{r}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{p}$	$\hat{\eta}$	$\hat{\theta}$
100	$e_1$	0.03	0.2	1.5	0.5	35	45
	$e_2$	0.02	0.4	3	0.5	45	30

<b>Table 3</b>		<b>Contract and Subcontract Values</b>						
		$(q_{12}, q_{21})$	GF	PO	BO	LR	Contract	Time
		(1, 0.5)	13.6173	-0.0345	21.7168	55.8348	91.1343	0.2555 min
Laplace-Fourier	(0.75, 0.75)	16.0942	-0.0457	21.4250	53.3456	90.8190	90.8190	0.3301 min
	(0.5, 1)	19.3834	-0.0628	21.0055	50.1079	90.4340	90.4340	0.2810 min
Monte-Carlo	(1, 0.5)	13.8766	-0.0379	21.9741	55.2439	91.0567	(90.9310, 91.1823)	29.8889 min
	(0.75, 0.75)	16.3868	-0.0497	21.6857	52.7593	90.7821	(90.6696, 90.8946)	29.8177 min
	(0.5, 1)	19.7011	-0.0674	21.1984	49.5445	90.3765	(90.2834, 90.4696)	30.5151 min

Then, we obtain our pricing results by performing numerical Laplace inversion of  $\widehat{GF}$  and  $\widehat{LR}$  and by making numerical Laplace-Fourier inversion of  $\widehat{PO}$  and  $\widehat{BO}$ . The numerical Laplace inversion we use is the Gaver-Stehfest algorithm and parameter choices of the algorithm are  $n = 6$  and  $B = 2$  (see the algorithm in [Kou and Wang \(2003\)](#)). The numerical Fourier inversion we use is the recursive adaptive Simpson quadrature method, where we implement the method by using the "quadv" function in matlab programming. In our numerical example, we set  $J_0 = (1, 0)$  and  $Y_0 = (0, 0, 1, 0, 0, 0)$  and make  $\alpha_1 = 15$  and  $\alpha_2 = 0.75$ . The pricing results are robust to a wide range of parameter choices  $\alpha_1$  and  $\alpha_2$ , for instance,  $\alpha_1 \in [10, 50]$  and  $\alpha_2 \in (0, 1.25]$ .

The pricing results are listed in Table 3, where we also show the Monte Carlo simulation results as a benchmark, reported in the form of a point estimate of contract value together with its associated 95% confidence interval and also of point estimates of the components  $GF, PO, BO, LR$ . The simulation results are based on 10,000 time steps and 100,000 sample paths. We can find that the price computed from our transform approach agrees with the simulation results, where our contract value stays within the 95% confidence interval and our contract value as well as the component value are all close to the point estimates. We finish the computation

#### 4.4. Contract Valuation

---

using matlab programming based on a PC with Intel Core i5 2 CPU 2.9 GHz and 8 GB of RAM. The computation time of obtaining one price by using our method is only around 0.3 minute while the simulation method takes about 30 minutes to generate one value. We can also see that the values of  $GF$  and  $BO$  from simulation are greater than the values from our method and the result is opposite for the components  $PO$  and  $LR$ . This observation is induced by the overestimation of Monte Carlo simulation method in default probabilities due to the systematic discretization bias. Thus, we can draw the conclusion that our pricing model is accurate and efficient.

When  $q_{12}$  decreases and  $q_{21}$  increases, the contract value depreciates. This is because the falling departing intensity of "good" state and rising departing intensity of "bad" state increases the time length of the funds dynamics at "good" state, which reduces the possibility of seizing more remaining asset value once default. Though the decreasing default risk has the opposite effect, the change trend is precisely dominated by the former effect.

The fair contract follows the equilibrium condition: the policyholders' premium investment  $L_0$  is equal to the initial market value of the contract  $V$ . This condition enables us to compute the fair participating rate in terms of the other parameters:

$$\delta = \frac{L_0 - GF - PO - LR}{BO^*},$$

where  $BO^* = E_{\mathbb{Q}} \left( (\alpha A_0 e^{Z_T} - L_0 e^{r^f T})_+ \mathbf{1}_{\tau \geq T} \right)$  is the stochastic component of  $BO$ , which can be solved by using the same method as in computation of  $BO$ .

We graph in Figure 1 the change of fair participation coefficient  $\delta$  with respect to  $\tilde{r}_g$  under three different market environments. We observe that the fair value of  $\delta$  decreases with  $\tilde{r}_g$ . This change trend follows our intuition that a high level of participation in benefit share should be provided to compensate for a low guaranteed rate. The change trend is susceptible to the market environment in the low guaranteed rates. However, the sensitivity weakens when the guaranteed rate raises, which is induced by the upper limit of the default risk. For the same  $\tilde{r}_g$ , the  $\delta$  decreases when  $q_{12}$  increases and  $q_{21}$  decreases. The change of departing intensity increases the time length of the funds dynamics at "bad" state, which leads to a more volatile financial market. The highly fluctuated market implies a better return but also brings a higher default risk. Undoubtedly, the positive effect plays a leading role here, thus adding the contract value.

## Chapter 4. Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

Figure 1 Participation Coefficient  $\delta$  (w.r.t.  $\tilde{r}_g$ ).

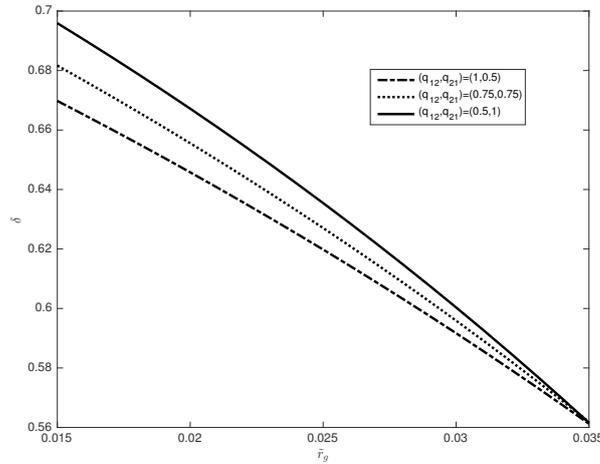


Figure 2 Participation Coefficient  $\delta$  (w.r.t.  $\alpha$ ).

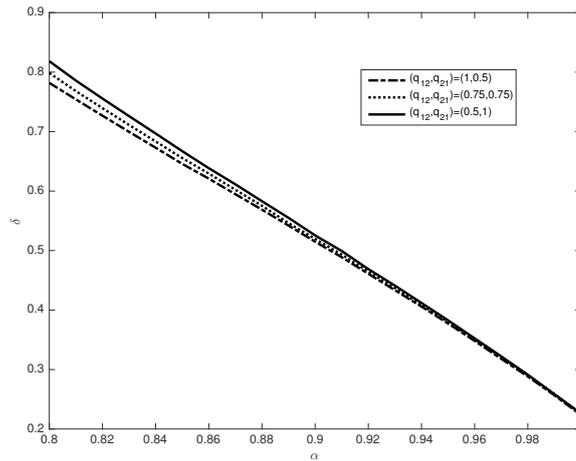


Figure 2 shows that the fair participation coefficient  $\delta$  decreases with  $\alpha$ . Increasing  $\alpha$  exposes the insurance company to higher default risk but also makes itself enjoy more remaining assets upon default. The former has an adverse effect on contract value, whereas the effect from the latter is favorable. Here we can see the latter dominates the change trend of the contract value and thus guides the decreasing trend in Figure 2.

### 4.5 Constant Guaranteed Rates V.S. Floating Guaranteed Rates

In this section, we compare our contracts (constant contracts) with the contracts having a floating guaranteed rate (floating contracts), whose minimum guaranteed

#### 4.5. Constant Guaranteed Rates V.S. Floating Guaranteed Rates

---

rate is linked to market interest rates. Let the guaranteed interest rates of the floating contracts be  $r_s^g = r^f + r_s$  at time  $s$  where  $r_s$  is the market interest rate and  $r^f$  is a constant used to control the difference between the guaranteed rates and market interest rates. Then, the guaranteed maturity payment becomes  $L_T^g = L_0 e^{\int_0^T r_s^g ds}$  at time  $T$ . The floating guaranteed rates reduce risk exposure to the fluctuations of interest rate, which circumvents an issue of a dramatic narrowing in the safety margin when low interest rates persist for long. The use of floating guaranteed rates see also in [Bernard, Le Courtois, and Quittard-Pinon \(2006\)](#) where [Bernard, Le Courtois, and Quittard-Pinon \(2006\)](#) use an implicit floating rate implied by zero-coupon bonds as guaranteed ones.

The payoff structure of the floating contracts is still in the form of the  $P_1, P_2, P_3, P_4$  and the pricing formulas of the four subcontract terms are also written as the equation (3.2), except now where the minimum guarantee at time  $t$  is  $L_t^g = L_0 e^{\int_0^t r_s^g ds}$ . Then, the equation (3.2) can be further simplified as follows:

$$\begin{cases} GF = L_0 e^{r^f T} \mathbb{Q}(\tau \geq T) \\ PO = E_{\mathbb{Q}} \left( -e^{r^f T} (L_0 - A_0 e^{Z_T})_+ \mathbb{1}_{\tau \geq T} \right) \\ BO = E_{\mathbb{Q}} \left( \delta e^{r^f T} (\alpha A_0 e^{Z_T} - L_0)_+ \mathbb{1}_{\tau \geq T} \right) \\ LR = E_{\mathbb{Q}} \left( A_0 e^{r^f \tau} e^{Z_\tau} \mathbb{1}_{\tau < T} \right) \end{cases}$$

where  $Z_t = X_t - \int_0^t r_s^g ds$  and the drift term  $\hat{\mu}_i^Z$  becomes

$$\hat{\mu}_i^Z = -r^f - \frac{1}{2} \hat{\sigma}_i^2 - \hat{\lambda}_i \left( \frac{\hat{p}_i \hat{\eta}_i}{\hat{\eta}_i - 1} + \frac{\hat{q}_i \hat{\theta}_i}{\hat{\theta}_i + 1} - 1 \right).$$

We can find that the value of the contract is not related to interest rate risk and the insurer saves the hedging cost to interest rate risk by issuing floating contracts. The Laplace or Laplace-Fourier transform results of the four subcontracts terms of

**Chapter 4. Pricing and Hedging Defaultable Participating Contracts  
with Regime Switching and Jump Risk**

---

the floating contracts are

$$\left\{ \begin{array}{l}
 \widehat{GF}(u) = \frac{L_0}{u - r^f} \left( 1 - Y_0 W_Z^{(\tilde{u}-r^f, -)} e^{Q_Z^{(\tilde{u}-r^f, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} H(0) \mathbf{1}_n \right) \\
 \widehat{LR}(u) = \frac{A_0}{u} Y_0 W_Z^{(\tilde{u}-r^f, -)} e^{Q_Z^{(\tilde{u}-r^f, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right) + \ln \frac{\lambda L_0}{A_0}} I_{2n} H(1) \mathbf{1}_n \\
 \widetilde{PO}(u, v) = -\frac{A_0}{(\alpha - iv)(\alpha - iv - 1)} \left( Y_0 W_Z^{(\tilde{u}-r^f, -)} e^{Q_Z^{(\tilde{u}-r^f, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} \right. \\
 \left. e^{-\frac{(\alpha - iv - 1) \ln \frac{\lambda L_0}{A_0}}{A_0}} H(iv + 1 - \alpha) - J_0 \right) \\
 (Q - \text{diag}\{u - r^f - \phi_k^Z(iv + 1 - \alpha)\})^{-1} \mathbf{1}_n \\
 \widetilde{BO}(u, v) = \frac{\delta \alpha A_0}{(\alpha - iv)(\alpha - iv + 1)} \left( Y_0 W_Z^{(\tilde{u}-r^f, -)} e^{Q_Z^{(\tilde{u}-r^f, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} \right. \\
 \left. e^{\frac{(\alpha - iv + 1) \ln \frac{\lambda L_0}{A_0}}{A_0}} H(\alpha - iv + 1) - J_0 \right) \\
 (Q - \text{diag}\{u - r^f - \phi_k^Z(\alpha - iv + 1)\})^{-1} \mathbf{1}_n
 \end{array} \right. \quad (4.5)$$

where  $W_Z^{(a, -)}$ ,  $Q_Z^{(a, -)}$  are the  $W^{(a, -)}$ ,  $Q^{(a, -)}$  given in definition 1 with  $X = Z$  (see detail proofs in Appendix A.1).

Let  $y_0$  be a discount rate that makes the discounted promised maturity payment equal to the premium, i.e.,

$$L_0 e^{y_0 T} = L_0 E_Q \left( e^{\int_0^T r_s^g ds} \right).$$

This indication is analogous to the concept of "yield to maturity" in bonds and is an internal rate of return earned by a policyholder from the minimum guarantee, assuming that the guaranteed maturity payment is made on schedule. We obtain with the Lemma 1 the explicit form of  $y_0$  as

$$y_0 = \frac{\log(J_0 e^{(Q + \text{diag}\{r^f + \hat{r}\})T} \mathbf{1}_n)}{T}.$$

When we compare the constant contracts and floating contracts, we make them bear the same promised maturity payment, i.e.,  $y_0 = \tilde{r}_g$ , and keep the same in other settings.

#### 4.5. Constant Guaranteed Rates V.S. Floating Guaranteed Rates

---

Then, we obtain the price of the floating contracts by making numerical Laplace or Laplace-Fourier inversion of the formulas (3.5), where we use the same parameters of numerical inversion method as in Section 4.3. The results are shown in Table 4, which are obtained with  $y_0$  equal to  $\tilde{r}_g$  in Table 3. We also compute the price by the Monte-Carlo simulation method and we can also find that the results of our transform approach agree with Monte-Carlo simulation results, where the contract value and its component values are all very close to the sample point estimates and the contract value also falls into the 95% confidence interval. The computation time of obtaining one price is almost 0.3 minute for our approach, whereas the simulation with 10,000 time steps and 100,000 sample paths takes around 30 minutes. The systematic discretization bias in the simulation also leads to the higher values of GF and BO and lower values of PO and LR for the overestimation of default probability. Therefore, we can have the conclusion that our method can also price floating contracts, both accurately and efficiently. By comparing the contract values in Table 3 and Table 4, we can also see that the floating contracts are more worthy than the constant ones.

	$(q_{12}, q_{21})$	GF	PO	BO	LR	Contract	Time
	(1, 0.5)	13.5018	-0.0344	21.5684	56.1638	91.1997	0.3707 min
Laplace-Fourier	(0.75, 0.75)	15.9877	-0.0458	21.3005	53.6339	90.8763	0.3458 min
	(0.5, 1)	19.2957	-0.0632	20.9117	50.3350	90.4793	0.3503 min
	(1, 0.5)	13.7843	-0.0375	21.8485	55.5522	91.1475 (91.0225, 91.2726)	29.4393 min
Monte-Carlo	(0.75, 0.75)	16.3144	-0.0500	21.5504	53.0190	90.8338 (90.7216, 90.9460)	29.8192 min
	(0.5, 1)	19.6178	-0.0678	21.1404	49.7671	90.4574 (90.3633, 90.5516)	30.5365 min

Now we compare the floating contracts and constant contracts on the change of default risk with respect to a guaranteed rate, where the two compared contracts keep the same promised maturity payments, i.e.,  $\tilde{r}_g = y_0$ . We fix the  $(q_{12}, q_{21}) = (1, 0.5)$  and make  $\tilde{r}_g$  range from 0.015 to 0.035. We find in Figure 3 that the floating contracts bear higher default risks. So the floating contracts are riskier products accompanied with higher returns. When the insurance company has the plan of issuing some contracts to attract the risk-seeking investors, the floating contract is a better choice.

## Chapter 4. Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

Figure 3 Default Probability (w.r.t.  $\tilde{r}_g$ ).

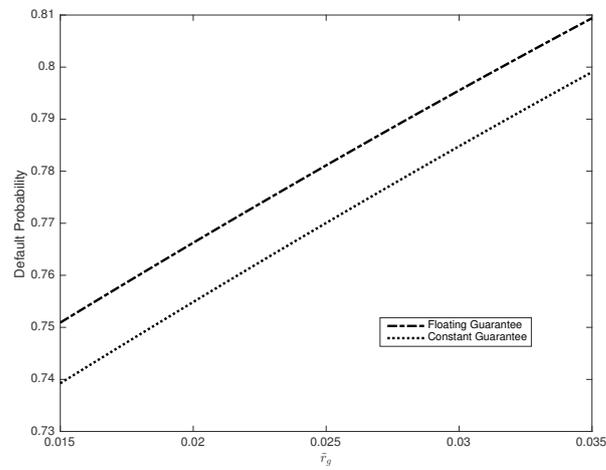
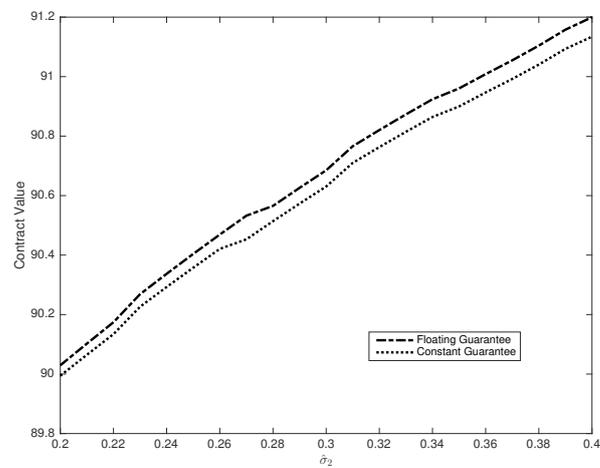


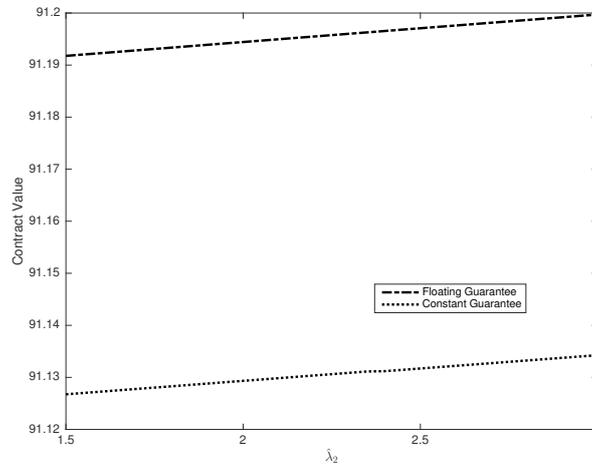
Figure 4 Contract Value (w.r.t.  $\hat{\sigma}_2$ ).



## 4.6. Two Hedging Strategies

---

Figure 5 Contract Value (w.r.t.  $\hat{\lambda}_2$ ).



Next we observe the impact of the fluctuation of financial market on the contract market value. We keep  $(q_{12}, q_{21}) = (1, 0.5)$  and have  $\hat{\sigma}_2$  range between 0.2 and 0.4 and make  $\hat{\lambda}_2$  change from 1.5 to 3. Increasing volatility or jump risk leads to a more fluctuated financial market. The fluctuated market implies a better return, but also induces a higher default risk. They are two opposite effects on the contract value. Obviously, in our numerical example, the favorable effect dominates the change trend of the contract value and we can find in Figure 4 and Figure 5 that the contract value increases if we increase  $\hat{\sigma}_2$  or  $\hat{\lambda}_2$ . The market value of the floating contracts are always higher than the constant contracts. At this stage, we can confirm that the floating contract is a more worthy participating contract when compared with a standard one, regardless of market conditions.

## 4.6 Two Hedging Strategies

In this section, we develop a dynamic hedging strategy where we choose the optimal delta to minimize the quadratic risk of the change differences between the value of the hedging portfolio and participating contracts at each rebalance time. We also introduce a semi-static hedging strategy that is developed in [He, Kennedy, Coleman, Forsyth, Li, and Vetzal \(2006\)](#), where we make some adjustments to fit our regime switching case. Then, we demonstrate the effectiveness of both the dynamic and semi-static methods in the hedging of participating contracts.

### 4.6.1 Hedging Strategy

For the ease of exposition, we neglect the basis risk that is induced by the mismatch between the inactively traded funds and the available hedging instruments and assume that the funds are fully invested in a market index, such as S&P 500. Then, besides the market index, we can also use the liquid standard options written on

## Chapter 4. Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

---

the market index as hedging instruments. Let  $\Delta$  be the time step size of the  $m$  rebalance times and then the hedge portfolio is rebalanced at times

$$0 = t_0 < t_1 < \dots < t_{m-1} < t_m = T,$$

where  $t_i = i\Delta$ . At each rebalance time  $t_i$ , the hedge portfolio contains an amount  $B_i$  in a bank account and a long position of  $e_i$  units of the underlying asset  $S_i$ . In the case of semi-static hedge, the hedge portfolio also possesses  $\tilde{n}$  additional options  $\hat{I}_i = (\hat{I}_{i,1}, \dots, \hat{I}_{i,\tilde{n}})$  with weights  $\hat{w}_i = (\hat{w}_{i,1}, \dots, \hat{w}_{i,\tilde{n}})$ . Let  $C_i$  be the time- $t_i$  value of the participating contract. Denote by  $\Delta C_i, \Delta S_i, \Delta \hat{I}_i$  the value change from  $t_i$  to  $t_{i+1}$  in the participating contract, in the underlying asset and in the options, respectively. Then, we can exhibit the specific contents of the two hedging strategies.

Within our dynamic hedging strategy, the optimal delta is computed to minimize the exposure of diffusion risk, jump risk and regime switching risk, simultaneously, where the regime switching risk is eliminated by making the optimal delta decision based on the specific economic state at each rebalance time. We compute the delta by solving the following optimization problem:

$$\arg \min_{\delta_k} E((\Delta C_k - \delta_k \Delta S_k)^2 | \mathcal{F}_k).$$

Then, at  $t_0$ , the bank account  $B_0 = C_0 - \delta_0 S_0$ . At each rebalance time  $t_k$ , the current underlying asset price  $S_k$  and economic state  $J_{t_k}$  determine the  $\delta_k$  that minimizes the risk exposure between  $t_k$  and  $t_{k+1}$ . The long position in the underlying asset is updated by purchasing  $\delta_k - \delta_{k-1}$  units of underlying assets. The transactions are financed by the bank account, which after rebalancing changes into

$$B_k = e^{\int_{t_{k-1}}^{t_k} r_s ds} B_{k-1} - (\delta_k - \delta_{k-1}) S_k.$$

Then, at time  $t_k$  after rebalancing the overall hedged position has value

$$\Pi(t_k) = -C_k + \delta_k S_k + B_k.$$

The semi-static hedging strategy in [He, Kennedy, Coleman, Forsyth, Li, and Vetzal \(2006\)](#) compute the optimal holding  $\{e_k, \hat{w}_k\}$  at each rebalance time  $t_k$  by solving the following optimization problem

$$\arg \min_{e_k, \hat{w}_k} E \left( \left( \Delta C_k - e_k \Delta S_k - \sum_{j=1}^{\tilde{n}} \hat{w}_{k,j} \Delta \hat{I}_{k,j} \right)^2 \middle| \mathcal{F}_k \right),$$

## 4.6. Two Hedging Strategies

---

where

$$\begin{cases} \Delta C_k = C_{k+1} - e^{\int_{k\Delta}^{(k+1)\Delta} r_s ds} C_k \\ \Delta S_k = S_{k+1} - e^{\int_{k\Delta}^{(k+1)\Delta} r_s ds} S_k \\ \Delta \hat{I}_{k,j} = \hat{I}_{k+1,j} - e^{\int_{k\Delta}^{(k+1)\Delta} r_s ds} \hat{I}_{k,j} \end{cases},$$

which takes time value into consideration because of the large rebalance time interval that is induced by infrequent rebalances in the semi-static hedge. Then, at  $t_0$ , the bank account  $B_0 = C_0 - e_0 S_0 - \sum_{j=1}^{\tilde{n}} \hat{w}_{0,j} \hat{I}_{0,j}$ . The self-financing constraint makes the bank account at time  $t_k$  after rebalancing become

$$B_k = e^{\int_0^{t_k} r_s ds} B_{k-1} - (e_k - e_{k-1}) S_k - \sum_{j=1}^{\tilde{n}} (\hat{w}_{k,j} - \hat{w}_{k-1,j}) \hat{I}_{k,j}.$$

Then, at time  $t_k$  after rebalancing the entire hedged position has value

$$\Pi(t_k) = -C_k + e_k S_k + \sum_{j=1}^{\tilde{n}} \hat{w}_{k,j} \hat{I}_{k,j} + B_k.$$

### 4.6.2 Hedging Effectiveness

We now illustrate the hedging effectiveness of the two hedging strategies. Before we proceed, we show the implementation of the hedging strategy under the regime switching jump diffusion case. We illustrate with the semi-static hedge. The dynamic hedge can be calculated in a similar manner. The derivative conditions of the optimality for the quadratic objective function give

$$\begin{cases} E \left( \left( \Delta C_k - e_k \Delta S_k - \sum_{j=1}^{\tilde{n}} \hat{w}_{k,j} \Delta \hat{I}_{k,j} \right) (-\Delta S_k) | \mathcal{F}_k \right) = 0 \\ E \left( \left( \Delta C_k - e_k \Delta S_k - \sum_{j=1}^{\tilde{n}} \hat{w}_{k,j} \Delta \hat{I}_{k,j} \right) (-\Delta \hat{I}_{k,l}) | \mathcal{F}_k \right) = 0, l = 1, \dots, \tilde{n} \end{cases}. \quad (4.6)$$

We discretize the value space of the underlying asset price into a series of grid points  $\{\bar{S}^1, \dots, \bar{S}^{N_1}\}$ . Note that the grid space can be different at each rebalance time but for simple statement we assume that the grid space keeps the same. At each rebalance time  $t_k$ , we compute the hedging positions  $(e_k, \hat{w}_k)$  at each grid point  $\bar{S}^i, i = 1, \dots, N_1$  for each economic state  $e_j, j = 1, \dots, n$ . Then, we form a discrete

## Chapter 4. Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk

approximation to the above linear system as follows:

$$\begin{cases} \sum_{s=1}^{N_1} \left( e_k (\Delta S_k^s)^2 + \sum_{j=1}^{\tilde{n}} \hat{w}_{k,j} \Delta \hat{I}_{k,j}^s \Delta S_k^s - \Delta C_k^s \Delta S_k^s \right) f(S_{t_{k+1}} = \bar{S}^s | J_{t_k} = e_j, S_{t_k} = \bar{S}^i) = 0 \\ \sum_{s=1}^{N_1} \left( e_k \Delta S_k^s \Delta \hat{I}_{k,l}^s + \sum_{j=1}^{\tilde{n}} \hat{w}_{k,j} \Delta \hat{I}_{k,j}^s \Delta \hat{I}_{k,l}^s - \Delta C_k^s \Delta \hat{I}_{k,l}^s \right) f(S_{t_{k+1}} = \bar{S}^s | J_{t_k} = e_j, S_{t_k} = \bar{S}^i) = 0, l = 1, \dots, \tilde{n} \end{cases}, \quad (4.7)$$

where  $\Delta C_k^s, \Delta S_k^s, \Delta \hat{I}_{k,j}^s$  denote the value change in the participating contract, in the underlying asset and in the options, respectively, when the underlying asset price changes from  $\bar{S}^i$  at time  $t_k$  to the  $\bar{S}^s$  at time  $t_{k+1}$  assuming the economic state  $J_{t_k} = e_j$ . The  $f(S_{t_{k+1}} = \bar{S}^s | J_{t_k} = e_j, S_{t_k} = \bar{S}^i)$  denotes the corresponding transitional probability of the underlying asset price, which can be computed by inverting the characteristic function of the regime switching Kou process. The linear system (3.7) is solved by using a truncated singular value decomposition, where small singular values are set to zero. Then, at each grid point we form  $n$  optimal hedging positions for  $n$  different economic states and the decision of using which hedging position is determined by the specific economic state of the hedging time. Note that the hedging position for the underlying asset price that is not exactly a grid point can be computed by the interpolation method.

In common practice, the insurance company is discretely monitored. In order to hedge default risk, we assume that the rebalancing times of the hedging strategy exactly coincide with the discretely monitored times. In the semi-static hedge, at each rebalance time  $t_k$  we employ five call options with strike prices  $0.8S_k, 0.9S_k, S_k, 1.1S_k, 1.2S_k$  and maturity  $t_{k+1}$ . We make  $m = 10$  and other parameters are set as in Section 4.3 except making  $\tilde{r}_g = 0.015$ . We generate 100000 sample

paths and for each path, we calculate the discounted profit and loss  $e^{-\int_0^{t^*} r_s ds} \Pi(t^*)$  (APL) and its relative value  $e^{-\int_0^{t^*} r_s ds} \frac{\Pi(t^*)}{V_0}$  (RPL) at default or maturity date  $t^*$  of the participating contract where  $V_0$  is the contract value. We use no hedge case as a benchmark and we can observe from Table 5 that both the dynamic and semi-static hedging strategies can significantly reduce risk and the dynamic hedge perform better especially in the hedging of extreme risk. The Figure 6 also indicates that the hedging position in the dynamic hedge bears smaller risk.

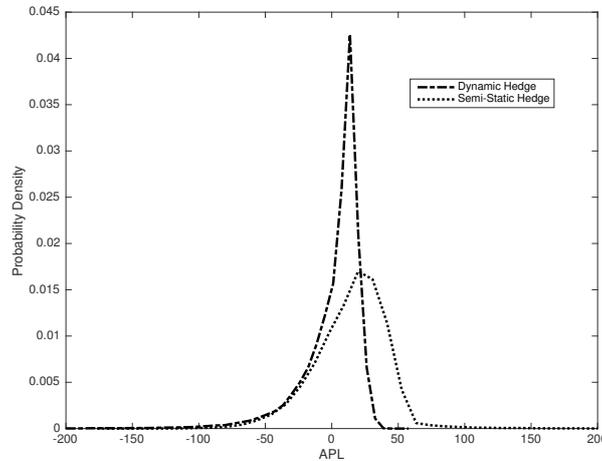
**Table 5** Comparison of hedging effectiveness

Measure	No hedge	Dynamic hdege	Semi-static hedge
Mean(APL)	89.3036	0.9949	14.7473
Std(APL)	108.0681	24.2676	29.3182
VaR <sub>95%</sub> (APL)	246.8931	23.1145	49.4263
CVaR <sub>95%</sub> (APL)	434.2527	26.5566	71.9829
Mean(RPL)	1.0000	0.1157	0.2133
Std(RPL)	1.2101	0.2580	0.3919

## 4.7. Conclusion

---

Figure 6 Comparison of probability densities of APL



## 4.7 Conclusion

This paper proposes a Laplace-Fourier transform approach to value participating life insurance contracts with a floating guaranteed rate and with a constant guaranteed rate, in the presence of credit, market (jump) and economic (regime switching) risks, based on a matrix Wiener-Hopf factorization of a regime switching Kou process. We demonstrate the accuracy and efficiency of this method. We also find that some typical contractual features are only understood from a short-term viewpoint. When we take the economic risk into account, we can obtain some characteristics of participating contracts, which is not consistent with the common knowledge that is obtained from the Black-Schole or Lévy models, such as the usual relation between the participating coefficient  $\delta$  and the wealth distribution coefficient  $\alpha$ .

In this paper, the floating contracts remove the impact of interest rate risk. In terms of the constant contracts, the effective modeling of interest rate term structure is vital to the valuation and hedging. The regime switching HJM model in [Valchev \(2004\)](#) and [Elliott and Siu \(2016\)](#) can be an attractive alternative, where using an exponential volatility structure can lead to an extended Vasicek model of short rates in each economic state and all the three parameters of the extended Vasicek model switch between different economic states. The correlation structure between the dynamics of the investment funds and interest rates also changes with the economic state and such a regime switching correlation structure can be incorporated by correlating the instantaneous covariance of the two Brownian motions. This modeling is more flexible and practical but also greatly enhances the complexity of the pricing, especially the involved first passage time problem. The problem of pricing and hedging standard participating contracts under such a practical model will be left for future research.

## 4.8 Appendix

### 4.8.1 Computations of Subcontract Terms for Contract with a Floating Guaranteed Rate

For any  $u > 0$ , we have

$$\begin{aligned}
 \widehat{GF}(u) &= \int_0^{\infty} e^{-uT} L_0 e^{r^f T} \mathbb{Q}(\tau \geq T) dT \\
 &= L_0 E_{\mathbb{Q}} \left( \int_0^{\tau} e^{-(u-r^f)T} dT \right) \\
 &= \frac{L_0}{u-r^f} \left( 1 - E_{\mathbb{Q}} \left( e^{-(u-r^f)\tau} \right) \right) \\
 &= \frac{L_0}{u-r^f} \left( 1 - Y_0 W_Z^{(\tilde{u}-r^f, -)} e^{Q_Z^{(\tilde{u}-r^f, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} H(0) \mathbf{1}_n \right),
 \end{aligned}$$

and

$$\begin{aligned}
 \widehat{LR}(u) &= \int_0^{\infty} e^{-uT} E_{\mathbb{Q}} \left( A_0 e^{r^f \tau} e^{Z_{\tau}} \mathbf{1}_{\tau < T} \right) dT \\
 &= A_0 E_{\mathbb{Q}} \left( \int_{\tau}^{\infty} e^{-uT} e^{r^f \tau} e^{Z_{\tau}} dT \right) \\
 &= \frac{A_0}{u} E_{\mathbb{Q}} \left( e^{-(u-r^f)\tau + Z_{\tau}} \right) \\
 &= \frac{A_0}{u} Y_0 W_Z^{(\tilde{u}-r^f, -)} e^{Q_Z^{(\tilde{u}-r^f, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right) + \ln \frac{\lambda L_0}{A_0}} I_{2n} H(1) \mathbf{1}_n,
 \end{aligned}$$

where  $\tilde{u} = (u, \dots, u)$  is a vector of size  $1 \times n$  with each element equal to  $u$ .

Let  $z_1 = \log(\alpha)$ . For any  $1 < \alpha < \tilde{\theta} + 1$  and  $u > \max(\Re(\phi_1^Z(iv + 1 - \alpha)) +$

## 4.8. Appendix

---

$r^f, \dots, \Re(\phi_n^Z(iv + 1 - \alpha)) + r^f, 0)$ , we obtain

$$\begin{aligned}
\widetilde{\widetilde{PO}}(u, v) &= -A_0 \int_0^\infty e^{-uT} \int_{-\infty}^\infty e^{ivz_1} e^{-\alpha z_1} e^{r^f T} E_{\mathbb{Q}} \left( (e^{z_1} - e^{Z_T})_+ \mathbf{1}_{\tau \geq T} \right) dz_1 dT \\
&= -A_0 E_{\mathbb{Q}} \left( \int_0^\infty e^{-(u-r^f)T} \mathbf{1}_{\tau \geq T} \int_{Z_T}^\infty e^{-(\alpha-iv)z_1} (e^{z_1} - e^{Z_T}) dz_1 dT \right) \\
&= -\frac{A_0}{(\alpha-iv)(\alpha-iv-1)} E_{\mathbb{Q}} \left( \int_0^\infty e^{-(u-r^f)T} \mathbf{1}_{\tau \geq T} e^{-(\alpha-iv-1)Z_T} dT \right) \\
&= -\frac{A_0}{(\alpha-iv)(\alpha-iv-1)} \left( E_{\mathbb{Q}} \left( \int_0^\infty e^{-(u-r^f)T} e^{-(\alpha-iv-1)Z_T} dT \right) \right. \\
&\quad \left. - \sum_{k=1}^n E_{\mathbb{Q}} \left( e^{-(u-r^f)\tau} e^{-(\alpha-iv-1)Z_\tau} \mathbf{1}_{J_\tau = e_k} \right. \right. \\
&\quad \left. \left. E_{\mathbb{Q}} \left( \int_0^\infty e^{-(u-r^f)T} e^{-(\alpha-iv-1)(Z_{T+\tau} - Z_\tau)} dT \mid J_0 = e_k \right) \right) \right) \\
&= -\frac{A_0}{(\alpha-iv)(\alpha-iv-1)} \left( Y_0 W_Z^{(\bar{u}-r^f, -)} e^{Q_Z^{(\bar{u}-r^f, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} \right. \\
&\quad \left. e^{-(\alpha-iv-1) \ln \frac{\lambda L_0}{A_0}} H(iv + 1 - \alpha) - J_0 \right) \\
&\quad (Q - \text{diag}\{u - r^f - \phi_k^Z(iv + 1 - \alpha)\})^{-1} \mathbf{1}_n.
\end{aligned}$$

Let  $z_2 = 0$ . For any  $\alpha > 0$  and  $u > \max(\Re(\phi_1^Z(\alpha - iv + 1)) + r^f, \dots, \Re(\phi_n^Z(\alpha -$

**Chapter 4. Pricing and Hedging Defaultable Participating Contracts  
with Regime Switching and Jump Risk**

---

$iv + 1)) + r^f, 0)$ , we also obtain

$$\begin{aligned}
\widetilde{BO}(u, v) &= \delta\alpha A_0 \int_0^\infty e^{-uT} \int_{-\infty}^\infty e^{ivz_2} e^{-\alpha z_2} e^{r^f T} E_{\mathbb{Q}} \left( (e^{Z_T} - e^{-z_2})_+ \mathbf{1}_{\tau \geq T} \right) dz_2 dT \\
&= \delta\alpha A_0 E_{\mathbb{Q}} \left( \int_0^\infty e^{-(u-r^f)T} \mathbf{1}_{\tau \geq T} \int_{-Z_T}^\infty e^{-(\alpha-iv)z_2} (e^{Z_T} - e^{-z_2}) dz_2 dT \right) \\
&= \frac{\delta\alpha A_0}{(\alpha-iv)(\alpha-iv+1)} E_{\mathbb{Q}} \left( \int_0^\infty e^{-(u-r^f)T} \mathbf{1}_{\tau \geq T} e^{(\alpha-iv+1)Z_T} dT \right) \\
&= \frac{\delta\alpha A_0}{(\alpha-iv)(\alpha-iv+1)} \left( E_{\mathbb{Q}} \left( \int_0^\infty e^{-(u-r^f)T} e^{(\alpha-iv+1)Z_T} dT \right) \right. \\
&\quad \left. - \sum_{k=1}^n E_{\mathbb{Q}} \left( e^{-(u-r^f)\tau} e^{(\alpha-iv+1)Z_\tau} \mathbf{1}_{J_\tau = e_k} \right) \right. \\
&\quad \left. E_{\mathbb{Q}} \left( \int_0^\infty e^{-(u-r^f)T} e^{(\alpha-iv+1)(Z_{T+\tau} - Z_\tau)} dT \mid J_0 = e_k \right) \right) \\
&= \frac{\delta\alpha A_0}{(\alpha-iv)(\alpha-iv+1)} \left( Y_0 W_Z^{(\bar{u}-r^f, -)} e^{Q_Z^{(\bar{u}-r^f, -)} \left( x - \ln \frac{\lambda L_0}{A_0} \right)} \right. \\
&\quad \left. e^{(\alpha-iv+1) \ln \frac{\lambda L_0}{A_0}} H(\alpha-iv+1) - J_0 \right) \\
&\quad (Q - \text{diag}\{u - r^f - \phi_k^Z(\alpha-iv+1)\})^{-1} \mathbf{1}_n.
\end{aligned}$$

# Optimal Insurance under Third Degree Risk

---

## Contents

<b>5.1</b>	<b>Introduction</b>	<b>85</b>
<b>5.2</b>	<b>The model</b>	<b>88</b>
<b>5.3</b>	<b>Some motivations</b>	<b>90</b>
5.3.1	The cases of risk averters and risk lovers	90
5.3.2	The advantage of third degree risk	92
<b>5.4</b>	<b>Optimal Contract</b>	<b>92</b>
5.4.1	General model	93
5.4.2	Concave constraint	95
<b>5.5</b>	<b>Background risk</b>	<b>98</b>
5.5.1	The independent case	99
5.5.2	The dependent case	99
<b>5.6</b>	<b>Conclusion</b>	<b>101</b>

---

## 5.1 Introduction

The optimal insurance design is of great interest for both the insurance industry and academics. There has been an extensive literature on this topic. The early works date back to [Arrow \(1963\)](#); [Borck \(1960\)](#). They investigate the criterion of maximizing the expected utility of a risk-averse insurer and obtains the optimal contract in the form of stop-loss. The results have been extended by many works. For instance, [Cummins and Mahul \(2004\)](#) shows that the optimal contract is a limited stop-loss contract when the coverage has an upper limit. [Kaluszka \(2001, 2005\)](#); [Kaluszka and Okolewski \(2008\)](#) reveal that the change-loss, limited stop-loss and truncated stop-loss insurance can be optimal contracts under different premium principles. [Chi and Zhou \(2017\)](#) shows change-loss or dual change-loss insurance are optimal ones in a general mean-variance model. Based on minimizing specific risk measures, such as value at risk and conditional value at risk, [Cai and Tan \(2007\)](#); [Cai, Tan, Weng, and Zhang \(2008\)](#) reveal the stop-loss insurance to be optimal among the class of increasing convex ceded loss functions and derive the explicit

optimal retention level for the insurance policy. Chi and Tan (2011) shows that the limited stop-loss and truncated stop-loss can also be optimal ones under VaR-based model while stop-loss contract is always optimal to the CVaR-based model. The extension of optimal contract primarily focuses on optimization criteria, generalizing premium principles, different constraints on the ceded loss functions. See Bernard, He, Yan, and Zhou (2015); Bernard and Tian (2009); Chi and Lin (2014); Gollier and Schlesinger (1996); Raviv (1979); Young (1999); Zhou, Wu, and Wu (2010), to name just a few.

The optimization criterion of a majority of aforementioned works is either to maximize the expected utility of risk averse insured or to minimize the insured's risk exposure. For the latter, it also implicitly assumes the insureds to be risk averse. Lots of works have given the facts that one's financial decision depends on more than just risk aversion. Higher order risk attitudes have been shown to be important, especially the prudence (third-order) and temperance (fourth-order), which have implications in many economic behaviours. For instance, Kimball et al. (1990) find that the prudence implies precautionary saving. Kimball (1993) manifests the negative correlation between temperance and the riskiness of portfolio choices. Noussair, Trautmann, and Van de Kuilen (2013) make use of an experimental method to measure risk attitudes and a large demographically representative sample supports the fact that the majority of individuals' decisions are consistent with risk aversion, prudence and temperance.

In the literature, many works place different constraints on the ceded loss functions and seek the optimal contract among a specific range of insurance contracts. Arrow (1963); Kaluszka (2005); Raviv (1979) only assume the indemnity of contracts to be nonnegative and less than the loss. Cai, Tan, Weng, and Zhang (2008) confine the ceded loss functions to the class of increasing convex functions. The constraint is too strict such that a very practical type of contracts, the limited full insurance, is excluded from the analysis. To take account of such type of contracts, Lu, Liu, and Meng (2013) assume the ceded loss functions to be increasing concave functions, whereas this set of contracts also removes popular ones, such as stop-loss and change-loss insurance policies. Chi and Zhou (2017) introduce a new set of admissible insurance policies in which the indemnity above a deductible is increasing and concave. This set encompasses many popular contracts, such as limited full insurance, limited stop-loss, stop-loss and change-loss insurance, etc.

In practice, an insured often confronts multiple risks. In addition to the insurable risk, the insured is also exposed to many other sources of risks, such as investment risk, human capital risk, etc. The literature refers to these uninsurable risk as the background risk. Doherty and Schlesinger (1983) studies optimal retention for stop-loss insurance with initial random wealth. Gollier (1996) finds that if background risk increases with the size of insurable loss, the optimal contract contains a "disappearing deductible" when the insured is risk averse and prudent. Dana and Scarsini (2007) shows that the optimal contract form crucially depends on the dependence between the insurable and noninsurable risks, where stochastic dependence is characterized by making use of the stochastic increasingness order.

## 5.1. Introduction

---

Chi and Wei (2018) shows that the stop-loss insurance is always favored by the higher order risk averse insured when background risk and insurable risk satisfies a general positive dependence structure, based on the higher degree right tail increasingness and stochastic increasingness orders.

Though a majority of people are found to exhibit risk aversion, there is still a significant minority of them who are risk lovers. Deck and Schlesinger (2014) examine the higher order risk preferences of risk averters and risk lovers and find the following two facts. First, Risk averters dislike an increase in risk at every degree. Second, risk lovers like risk increases of even degrees, but dislike increases of odd degrees. Thus, both risk averters and risk lovers have the same risk attitudes at odd orders, such as third-order prudence. Crainich, Eeckhoudt, and Trannoy (2013) have also confirmed that risk lovers are prudent.

Therefore, this paper for the first time considers the risk lovers in the contract design. We present optimal insurance contract form when both risk averters and risk lovers are covered. We first introduce a general set of insurance policies in which we assume that the indemnity is increasing in the loss. The general set contains most practical insurance contracts, such as stop-loss, limited stop-loss, change-loss insurance, etc. Further, the set is larger than the set used in Chi (2017). We will point out that an optimal contract for both of risk averters and risk lovers is a new type of contract, i.e., a dual limited stop-loss contract. This type of insurance policy is excluded out of the set in Chi (2017). Our set includes such type of contracts and indicates a wider range of admissible insurance policies. Based on such a set, we reveal the differences of optimal contracts between risk averters and risk lovers. Then, we assume that the insureds are prudent (third degree risk averse in Ekern (1980)) such that both of risk averters and risk lovers are covered and show that the optimal contract is a dual limited stop-loss contract. The work most closely related to ours is Chi (2017), where they study the insurance choice under third degree stochastic dominance (TSD), equivalently maximizing the expected utility of risk averse and prudent insured. The concept of stochastic dominance, see, for instance, Hadar and Russell (1969); Hanoch and Levy (1969); Jean (1980); Rothschild and Stiglitz (1970); Whitmore (1970). In contrast to the TSD-consistent optimization criterion, there are two obvious advantages of making use of third degree risk. First, the third degree risk averters cover a larger group of insureds than risk averse and prudent insureds since the former includes both of risk averters and risk lovers. Second, we will illustrate in section 2 that the third degree risk circumvents some ambiguous comparison in stochastic dominance.

Next, we narrow the interests of contracts to the concave types. The concave contracts enable the insurer and insured to pay more for a larger realization of the loss. As pointed out by Chi and Tan (2011); Huberman, Mayers, and Smith Jr (1983), this condition reduces ex post moral hazard. We find the optimal contract changes into a change-loss insurance or a dual change-loss insurance, which depends on the coefficient of variation of the retained loss. This conclusion is similar to the results in Chi (2017), but obtained under different optimization criteria and giving different economic implications. Finally, we take account of the background

risk, in which we assume the insured still to be third degree risk averse on the final wealth. We find that the optimal contract is a stop-loss insurance policy when the background risk and insurable risk is comonotonic.

The rest of this paper is organized as follows. In section 2, we introduce an optimal insurance model under third degree risk. Section 3 introduces some notions of stochastic order. Meanwhile, we also show the difference of optimal contract form between risk averters and risk lovers and illustrate another one advantage of third degree risk over third degree stochastic dominance. Section 4 shows that the dual limited stop-loss insurance is the optimal choice of third degree risk averters. Further, we narrow the admissible insurance policies to concave types and manifest the optimal concave insurance policies. Section 5 introduces the background risk and shows the form of optimal contract changes into the stop-loss insurance policy.

## 5.2 The model

Let  $X$  be the amount of loss faced by an insured over a fixed time period.  $X$  is a non-negative random variable defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . To reduce the risk exposure, the insured purchases an insurance policy to cover in part or in full the loss. The insurance policy covers a portion  $f(X)$  of the loss, where  $0 \leq f(X) \leq X$ . Therefore, the insured has to bear the uncovered part of the loss  $R_f(X) \equiv X - f(X)$ . The functions  $f(x)$  and  $R_f(x)$  are called the insured's ceded and retained loss functions, respectively.

In order to cover more types of insurance policies, we assume that the admissible insurance contract follows the principle of indemnity and the indemnity is increasing. The corresponding set of admissible ceded loss functions is given by:

$$\mathcal{S} = \{f(x) \mid 0 \leq f(x) \leq x, f'(x) \geq 0\},$$

where  $(x)_+ = \max(0, x)$ . The proposed set contains most insurance contracts, such as the stop-loss, limited stop-loss, chang-loss, dual chang-loss, quota-share, limited full insurance, etc, which have been shown to be optimal under different constraints.

We introduce ceded loss functions of a new contract and of some other insurance policies that we will use behind:

- stop-loss insurance (Arrow (1963); Borck (1960)):

$$f(x; d) = (x - d)_+, \quad d \geq 0,$$

- limited full insurance (Lu, Liu, and Meng (2013)):

$$f(x; p) = \min(x, p), \quad p > 0,$$

- change-loss insurance (Cai, Tan, Weng, and Zhang (2008)):

$$f(x; \theta, d) = \theta(x - d)_+, \quad \theta \in [0, 1], \quad d \geq 0,$$

## 5.2. The model

---

- dual change-loss insurance (Chi and Zhou (2017)):

$$f(x; \theta, d) = x - \theta(x - d)_+, \quad \theta \in [0, 1], \quad d \geq 0,$$

- dual limited stop-loss insurance:

$$f(x; d, p) = x - \min((x - d)_+, p), \quad d \geq 0, \quad p > 0.$$

We rewrite the ceded loss function of dual limited stop-loss insurance in the form of

$$f(x; d, p) = \begin{cases} x & \text{if } x \leq d \\ d & \text{if } d < x \leq d + p \\ x - p & \text{if } x > d + p \end{cases}.$$

We can find that the limited stop-loss insurance fully covers small and large losses and make constant compensation for medium loss.

The insured need make premium payment to initiate an insurance contract. We assume the amount of insurance premium paid by the insured is a function of the net premium. Denote by  $\pi$  the insurance premium principle, which we cast in the form of

$$\pi(Y) = \mathcal{G}(\mathbb{E}(Y)),$$

for any non-negative random variable  $Y$ , where  $\mathcal{G} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a strictly increasing function with  $\mathcal{G}(0) = 0$  and  $\mathcal{G}(y) \geq y$ . If  $\mathcal{G}(y) = y$ , we recover the net premium principle. If  $\mathcal{G}(y) = (1 + \rho)y$ , where  $\rho > 0$  is a safety loading coefficient, we obtain the expected value premium principle.

Denote  $w_0$  the initial wealth of the insured. With an insurance contract, the final wealth of the insured  $w_f(X)$  becomes the initial wealth minus the retained loss and the insurance premium:

$$w_f(X) = w_0 - R_f(X) - \pi(f(X)).$$

We use the usual objective that maximizes the expected utility of the insured's final wealth. Denote  $U(x)$  the utility function of the insured. The optimization problem is in the form of

$$f^* = \arg \max_{f \in \mathcal{F}} \mathbb{E}(U(w_f(X))). \quad (5.1)$$

Different choices of utility function  $U(x)$  leads to the optimal contract favored by distinct group of insureds, each choice corresponding to the group who has such kind of utility function. For example, Arrow (1963) shows that the stop-loss insurance policy is favored by the risk averse insured, who has utility function  $U(x)$  such that  $U'(x) \geq 0$  and  $U''(x) \leq 0$ . Chi (2017) manifests that the change-loss or dual change-loss contract is favorable for the risk averse and prudent insured, whose utility function  $U(x)$  satisfies  $U'(x) \geq 0, U''(x) \leq 0$  and  $U'''(x) \geq 0$ . In this paper,

we model the utility function  $U(x)$  as

$$U'''(x) \geq 0.$$

The person endowed with this utility function is said to be prudent or third degree risk averse (Ekern (1980)). The prudent persons constitute a very large population group. As pointed out by Deck and Schlesinger (2014) and Crainich, Eeckhoudt, and Trannoy (2013), both risk averters and risk lovers are prudent. Thus, we solve an optimal insurance problem when both of risk averters and risk lovers are covered.

Next, we show the optimization problem can reduce to the ordering problem of distributions of final wealth. Given two random wealths  $W_1$  and  $W_2$ , Menezes, Geiss, and Tressler (1980) and Ekern (1980) demonstrate that all the third risk averters prefer  $W_1$  to  $W_2$ , that is,

$$\mathbb{E}(U(W_1)) \geq \mathbb{E}(U(W_2))$$

if and only if  $W_2$  has more third degree risk than  $W_1$ , where third degree risk is defined as follows:

**Definition 4.** (Ekern (1980); Menezes, Geiss, and Tressler (1980)) *Random variable  $Z$  has more third degree risk than  $Y$ , if and only if  $\mathbb{E}(Y) = \mathbb{E}(Z)$ ,  $\text{Var}(Y) = \text{Var}(Z)$  and, for all  $x$ ,*

$$\int_{-\infty}^x \int_{-\infty}^v (F_Z(u) - F_Y(u)) \, du \, dv \geq 0,$$

where the latter inequality should hold strictly for at least one value of  $x$ .

Thus, we transform the optimization problem into seeking the final wealth  $w_f^*(X)$  that has the least third degree risk.

### 5.3 Some motivations

In this section, we shall introduce the motivations. We show the difference of optimal contract form between risk averter and risk lovers. Then, we illustrate another one advantage of third degree risk over third degree stochastic dominance.

#### 5.3.1 The cases of risk averters and risk lovers

Before we proceed to the results, we first review the relevant definitions of stochastic orders, i.e., the increasing convex (concave) order.

**Definition 5.** (Shaked and Shanthikumar (2007)) *Random variable  $Z$  is said to be smaller than  $Y$  in the increasing convex (concave) order, denoted as  $Z \leq_{icx} Y$  ( $Z \leq_{icv} Y$ ) if and only if*

$$\mathbb{E}(\phi(Z)) \leq \mathbb{E}(\phi(Y)), \tag{5.2}$$

### 5.3. Some motivations

---

for all increasing convex (concave) functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ .

If we only consider the second order risk attitude, the utility function of risk averter satisfies  $U'(x) \geq 0, U''(x) \leq 0$ , and for the risk lovers,  $U'(x) \geq 0, U''(x) \geq 0$ . Given two random wealths  $W_1$  and  $W_2$ , the risk averters (risk lovers) prefer  $W_1$  over  $W_2$  if and only if  $W_2 \leq_{\text{icv}} W_1$  ( $W_2 \leq_{\text{icx}} W_1$ ). Then, we can obtain optimal contracts for risk averters and risk lovers by making use of some properties of the orders  $\leq_{\text{icx}}, \leq_{\text{icv}}$ . Before we proceed to the results, we show the definition of an important stochastic order, the concave order, which is often used in actuarial context. Random variable  $Z$  is said to be *smaller than*  $Y$  in the convex order, denoted as  $Z \leq_{\text{cx}} Y$ , when the inequality 5.2 holds for all convex functions  $\phi$ . Now, we can give the optimal contract for risk averters and risk lovers, respectively.

**Proposition 17.** *For any admissible insurance policy  $f \in \mathcal{S}$ , there exists a stop-loss insurance  $f_d$  and a limited full insurance  $f_p$  such that*

$$\begin{cases} w_f(X) \leq_{\text{icx}} w_{f_p}(X) \\ w_f(X) \leq_{\text{icv}} w_{f_d}(X) \end{cases}.$$

*Proof.* From  $0 \leq f(x) \leq x$ , we get  $\mathbb{E}(f(X)) \leq \mathbb{E}(X)$ . Therefore, there exists  $p_0 \geq 0$  and  $m_0 \geq 0$  such that  $\mathbb{E}((X - p_0)_+) = \mathbb{E}(f(X))$  and  $\mathbb{E}(\min(X, m_0)) = \mathbb{E}(f(X))$ . Then, we have a stop-loss insurance with a ceded loss function:

$$f_d(x; p_0) = (x - p_0)_+,$$

and a limited full insurance with a ceded loss function:

$$f_p(x; m_0) = \min(x, m_0),$$

such that  $\mathbb{E}(f_d(X; p_0)) = \mathbb{E}(f_p(X; m_0)) = \mathbb{E}(f(X))$ . The function  $R_{f_p}(x)$  up-crosses  $R_f(x)$  and the function  $R_f(x)$  up-crosses  $R_{f_d}(x)$ . Combining these facts with  $\mathbb{E}(R_{f_d}(X; p_0)) = \mathbb{E}(R_{f_p}(X; m_0)) = \mathbb{E}(R_f(X))$ , we obtain from Lemma 3 in [Ohlin \(1969\)](#) that

$$R_{f_d}(X) \leq_{\text{cx}} R_f(X) \leq_{\text{cx}} R_{f_p}(X),$$

and then

$$-R_{f_d}(X) \leq_{\text{cx}} -R_f(X) \leq_{\text{cx}} -R_{f_p}(X).$$

Based on relations between  $\leq_{\text{cx}}$  and  $\leq_{\text{icx}}, \leq_{\text{icv}}$ , we obtain

$$\begin{cases} -R_f(X) \leq_{\text{icx}} -R_{f_p}(X) \\ -R_f(X) \leq_{\text{icv}} -R_{f_d}(X) \end{cases}.$$

As  $\pi(f_d(X; p_0)) = \pi(f_p(X; m_0)) = \pi(f(X))$ , we obtain the results.  $\square$

**Remark 1.** *Note that the optimal contract for risk averters has been obtained early in [Borck \(1960\)](#) and [Arrow \(1963\)](#). We list it here to ease comparison of optimal*

contract between risk averters and risk lovers. The difference of optimal contract between risk averters and risk lovers motives us to seek one type of contract favored by both of them.

### 5.3.2 The advantage of third degree risk

Besides one advantage that covers a wider range of insureds, we show another advantage of third degree risk over third degree stochastic dominance. Before we proceed, we review the relevant stochastic orders, i.e., the  $n$ -concave order and the  $n$ -increasing concave order. Denote  $\bar{\mathcal{U}}_{n-cv}$  the class of all functions  $g(x)$  such that  $(-1)^{n-1}g^{(n)}(x) \geq 0$  and  $\bar{\mathcal{V}}_{n-icv}$  the class of all functions  $g(x)$  such that  $(-1)^{k-1}g^{(k)}(x) \geq 0, k = 1, \dots, n$ , where  $g^{(k)}(x)$  denotes the  $k^{\text{th}}$  derivative of  $g(x)$ .

**Definition 6.** (*Shaked and Shanthikumar (2007)*) Random variable  $Z$  is said to be smaller than  $Y$  in the  $n$ -concave order ( $n$ -increasing concave order), denoted as  $Z \leq_{n-cv} Y$  ( $Z \leq_{n-icv} Y$ ), if and only if

$$\mathbb{E}(\phi(Z)) \leq \mathbb{E}(\phi(Y)), \quad (5.3)$$

for all functions  $\phi \in \bar{\mathcal{U}}_{n-cv}$  ( $\phi \in \bar{\mathcal{V}}_{n-icv}$ ).

Given two random wealths  $W_1$  and  $W_2$ , the  $W_1$  having more third degree risk than  $W_2$  is equivalent to  $W_1 \leq_{3-cv} W_2$  and the  $W_1$  dominated by  $W_2$  in third degree stochastic dominance is equivalent to  $W_1 \leq_{3-icv} W_2$ . Note that the 3-concave order (3-increasing concave order) and the third degree risk (third degree stochastic dominance) are the same concepts, just different names used in respective fields. The equivalent relation is kept for all  $n^{\text{th}}$  degree cases. In this paper, we only focus on the third degree case.

Now, we can illustrate another advantage of third degree risk over third degree stochastic dominance. In comparison of third degree risk, the first two moments of two random wealth are required to be equal, which produces an unambiguous comparison in terms of third order moment. However, for the third degree stochastic dominance, there exist ambiguous situations of, e.g., both a higher variance (undesirable) and a higher skewness (desirable).

For instance, let  $X \sim \Gamma(\frac{1}{16}, 64)$ ,  $Y \sim \text{IG}(4, 1)$ ,  $Z \sim \text{LN}(\frac{3}{2} \ln 2, \ln 2)$ . The table 1 lists the descriptive statistics. We have  $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z)$  and  $\text{Var}(Y) < \text{Var}(X), \text{Var}(Z) < \text{Var}(X)$ . Then, we obtain  $X \leq_{3-icv} Y$  and  $X \leq_{3-icv} Z$  (Klar (2002)). Thus, the risk averse and prudent person will prefer  $Y$  and  $Z$  due to less variance though with less skewness.

## 5.4 Optimal Contract

In this section, we solve the optimal insurance model established in section 2. Then, we confine the interests of insurance policy to a concave set and give the optimal concave contract.

## 5.4. Optimal Contract

Table 5.1: The descriptive statistics

	Mean	Variance	Skewness
Gamma	4	256	8
Inverse Gaussian	4	64	6
Lognormal	4	16	4

### 5.4.1 General model

Before we proceed, we introduce the  $n$ -convex order. Denote  $\overline{\mathcal{W}}_{n-cx}$  the class of all functions  $g(x)$  such that  $g^{(n)}(x) \geq 0$ , where  $g^{(n)}(x)$  denotes the  $n^{\text{th}}$  derivative of  $g(x)$ . Random variable  $Z$  is said to be smaller than  $Y$  in the  $n$ -convex order, denoted as  $Z \leq_{n-cx} Y$ , if and only if the inequality 5.3 holds for all functions  $\phi \in \overline{\mathcal{W}}_{n-cx}$ . Then, in the general set  $\mathcal{S}$ , we have the following optimal contract.

**Proposition 18.** *For any admissible insurance policy  $f \in \mathcal{S}$ , there exists a dual limited stop-loss insurance  $f_d$  such that*

$$w_f(X) \leq_{3-cv} w_{f_d}(X).$$

*Proof.* From  $0 \leq f(x) \leq x$ , we get  $\mathbb{E}(R_f(X)) \leq \mathbb{E}(X)$ . Therefore there exists  $p_0 \geq 0$  and  $m_0 \geq 0$  such that  $\mathbb{E}((X - p_0)_+) = \mathbb{E}(R_f(X))$  and  $\mathbb{E}(\min(X, m_0)) = \mathbb{E}(R_f(X))$ . For any  $p \in [0, p_0]$ , denote by  $m(p)$  the unique solution to the equation:

$$\mathbb{E}(\min((X - p)_+, m(p))) = \mathbb{E}(R_f(X)).$$

The function  $m(p)$  is increasing and continuous, with  $m(0) = m_0$  and  $m(p_0) = \infty$ .

The function  $R_f(x)$  up-crosses  $\min(x, m_0)$  and the function  $(x - p_0)_+$  up-crosses  $R_f(x)$ . Combining these facts with  $\mathbb{E}((X - p_0)_+) = \mathbb{E}(R_f(X)) = \mathbb{E}(\min(X, m_0))$ , we obtain from Lemma 3 in [Ohlin \(1969\)](#) that

$$\min(X, m_0) \leq_{cx} R_f(X) \leq_{cx} (X - p_0)_+.$$

Note that the function  $\phi(x) = x^2$  is convex. From the definition of convex order and  $\mathbb{E}(\min(X, m_0)) = \mathbb{E}(R_f(X)) = \mathbb{E}((X - p_0)_+)$ , we have:

$$\sigma(\min(X, m_0)) \leq \sigma(R_f(X)) \leq \sigma((X - p_0)_+).$$

Therefore, we can conclude that

$$cv_{\min(X, m_0)} \leq cv_{R_f(X)} \leq cv_{(X - p_0)_+}. \quad (5.4)$$

For  $p \in [0, p_0]$ , we define  $g_p(x) \equiv \min((x - p)_+, m(p))$ , where we observe that  $g_{p'}(x)$  up-crosses  $g_p(x)$  for  $0 \leq p < p' \leq p_0$ . By Lemma 3 in [Ohlin \(1969\)](#), we have:  $g_p(X) \leq_{cx} g_{p'}(X)$ , and hence  $\sigma(g_p(X)) \leq \sigma(g_{p'}(X))$  and  $cv_{g_p(X)}$  is increasing in  $p$ .

Note that  $g_0(X) = \min(X, m_0)$  and  $g_{p_0}(X) = (X - p_0)_+$ . Therefore, the inequality 5.4 turns into:

$$\text{cv}_{g_0(X)} \leq \text{cv}_{R_f(X)} \leq \text{cv}_{g_{p_0}(X)}.$$

Because  $\text{cv}_{g_p(X)}$  is continuous in  $p$ , it follows from the intermediate value theorem that there exists a  $p_1 \in [0, p_0]$  such that

$$\text{cv}_{g_{p_1}(X)} = \text{cv}_{R_f(X)}.$$

Then, we have a dual limited stop-loss insurance, where the ceded loss function is:

$$f_d(x; p_1, m(p_1)) = x - \min((x - p_1)_+, m(p_1)),$$

with  $p_1 \in [0, p_0]$  and  $m(p_1) \in [m_0, \infty)$ , such that  $E(f_d(X; p_1, m(p_1))) = E(f(X))$  and  $\sigma(R_{f_d}(X; p_1, m(p_1))) = \sigma(R_f(X))$ .

Now we need to prove that the above dual limited stop-loss insurance is the optimal solution to the problem 5.1 for third degree risk averse agents. Note that the functions  $R_f(x)$  and  $R_{f_d}(x; p_1, m(p_1))$  cross at most twice.

If they only cross once, Lemma 3 in [Ohlin \(1969\)](#) shows that these two non-negative stochastic variables are ranked in convex order. Because  $E(R_f(X)^2) = E(R_{f_d}(X; p_1, m(p_1))^2)$ , by Theorem 3.A.42 in [Shaked and Shanthikumar \(2007\)](#),  $R_f(X)$  and  $R_{f_d}(X; p_1, m(p_1))$  have the same distribution. As  $\pi(f_d(X; p_1, m(p_1))) = \pi(f(X))$ , we obtain the results.

If  $R_f(x)$  and  $R_{f_d}(x; p_1, m(p_1))$  cross twice, we are in the case depicted in Fig 1. Let  $0 < x_1 < x_2 < \infty$  be the two intersection points of  $R_f(x)$  and  $R_{f_d}(x; p_1, m(p_1))$ . We obtain:

$$\begin{cases} w_f(x) \leq w_{f_d}(x) & x \leq x_1 \\ w_f(x) \geq w_{f_d}(x) & x_1 < x \leq x_2 \\ w_f(x) < w_{f_d}(x) & x > x_2. \end{cases}$$

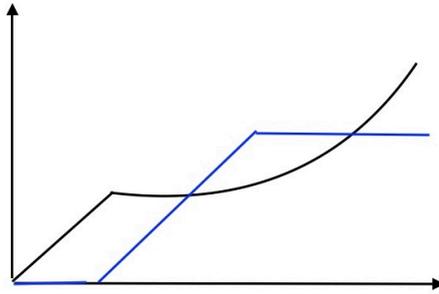


Figure 1: The intersection between  $R_f(X)$  and  $R_{f_d}(X; p_1, m(p_1))$

Let  $F_f$  and  $F_{f_d}$  be the cumulative distribution functions of  $w_f(X)$  and  $w_{f_d}(X)$ ,

## 5.4. Optimal Contract

---

respectively. Denote  $z_i = w_f(x_i)$ ,  $i = 1, 2$ . Note that  $z_1 > z_2$  and that

$$\begin{cases} F_f(z) \geq F_{f_d}(z) & z \leq z_2 \\ F_f(z) \leq F_{f_d}(z) & z_2 < z \leq z_1 \\ F_f(z) > F_{f_d}(z) & z > z_1. \end{cases}$$

Then  $F_f$  and  $F_{f_d}$  cross twice and the last sign of  $F_f(z) - F_{f_d}(z)$  is a  $+$ . In conjunction with

$$\begin{cases} \pi(f_d(X; p_1, m(p_1))) = \pi(f(X)) \\ \mathbb{E}(R_{f_d}(X; p_1, m(p_1))) = \mathbb{E}(R_f(X)) \\ \sigma(R_{f_d}(X; p_1, m(p_1))) = \sigma(R_f(X)) \end{cases} \implies \begin{cases} \mathbb{E}(w_f(X)) = \mathbb{E}(w_{f_d}(X)) \\ \sigma(w_f(X)) = \sigma(w_{f_d}(X)) \end{cases},$$

by Theorem 3.A.66 in [Shaked and Shanthikumar \(2007\)](#), we have  $w_f(X) \leq_{3-cx} w_{f_d}(X)$ . The equivalence of  $w_f(X) \leq_{3-cx} w_{f_d}(X)$  and  $w_f(X) \leq_{3-cv} w_{f_d}(X)$  gives our results.  $\square$

**Remark 2.** *As the stop-loss insurance and limited full insurance insure large loss and small loss, respectively, the optimal contract favored by both of risk averters and risk lovers is a compromise solution, which protects the insureds against both of large loss and small loss.*

### 5.4.2 Concave constraint

We confine admissible insurance policies to concave types and the set of admissible ceded loss functions changes into

$$\mathcal{Q} = \{f((x-d)_+) \mid d \geq 0, 0 \leq f(x) \leq x, f'(x) \geq 0, f''(x) \leq 0\},$$

which is the same as the set of admissible insurance contracts in [Chi \(2017\)](#). As pointed out by [Chi and Tan \(2011\)](#), the concave constraint can make both the insured and the insurer pay more for a larger amount of loss, so as to reduce ex post moral hazard. We now present the optimal concave contract when both of risk averters and risk lovers are covered.

**Proposition 19.** *For any admissible insurance policy  $f \in \mathcal{Q}$ , there exists a change-loss insurance  $f_c$  or a dual change-loss insurance  $f_d$  such that*

$$w_f(X) \leq_{3-cv} \begin{cases} w_{f_c}(X) & \text{if } cv_{R_f(X)} \leq cv_X \\ w_{f_d}(X) & \text{if } cv_{R_f(X)} > cv_X, \end{cases}$$

where  $cv_Y$  denotes the coefficient of variation of a random variable  $Y$ .

*Proof.* From  $0 \leq f(x) \leq x$ , we get  $\mathbb{E}(f(X)) \leq \mathbb{E}(X)$ . Therefore, there exists  $p_0 \geq 0$  and  $m_0 \geq 0$  such that  $\mathbb{E}((X-p_0)_+) = \mathbb{E}(f(X))$  and  $\mathbb{E}(\min(X, m_0)) = \mathbb{E}(f(X))$ .

For any  $p \in [0, p_0]$ , denote by  $\theta_1(p)$  the unique solution to the equation:

$$\mathbb{E}(\theta_1(p) (X - p)_+) = \mathbb{E}(f(X)).$$

The function  $\theta_1(p)$  is increasing and continuous, with  $\theta_1(0) = \mathbb{E}(f(X))/\mathbb{E}(X) \leq 1$  and  $\theta_1(p_0) = 1$ .

For any  $m \in [0, m_0]$ , denote by  $\theta_2(m)$  the unique solution to the equation:

$$\mathbb{E}(\theta_2(m) (X - \min(X, m))) = \mathbb{E}(\theta_2(m) (X - m)_+) = \mathbb{E}(R_f(X)).$$

The function  $\theta_2(m)$  is also increasing and continuous. Furthermore, we have:  $\theta_2(0) = \mathbb{E}(R_f(X))/\mathbb{E}(X) \leq 1$  and  $\theta_2(m_0) = 1$ .

The function  $R_f(x)$  up-crosses  $x - (x - p_0)_+$  and the function  $x - \min(x, m_0)$  up-crosses  $R_f(x)$ . Combining these facts with  $\mathbb{E}(X - (X - p_0)_+) = \mathbb{E}(R_f(X)) = \mathbb{E}(X - \min(X, m_0))$ , we obtain from Lemma 3 in [Ohlin \(1969\)](#) that

$$X - (X - p_0)_+ \leq_{cx} R_f(X) \leq_{cx} X - \min(X, m_0).$$

Note that the function  $\phi(x) = x^2$  is convex. From the definition of convex order, we have:

$$\sigma(X - (X - p_0)_+) \leq \sigma(R_f(X)) \leq \sigma(X - \min(X, m_0)).$$

Therefore, we can conclude that

$$cv_{X - (X - p_0)_+} \leq cv_{R_f(X)} \leq cv_{X - \min(X, m_0)}. \quad (5.5)$$

At this stage, we divide the discussion into two cases.

First case  $[cv_{R_f(X)} \leq cv_X]$ : For  $p \in [0, p_0]$ , we define  $g_p(x) \equiv x - \theta_1(p)(x - p)_+$ , where we observe that  $g_p(x)$  up-crosses  $g_{p'}(x)$  for  $0 \leq p < p' \leq p_0$ . By Lemma 3 in [Ohlin \(1969\)](#), we have:  $g_{p'}(X) \leq_{cx} g_p(X)$ , and hence  $\sigma(g_{p'}(X)) \leq \sigma(g_p(X))$  and  $cv_{g_p(X)}$  is decreasing in  $p$ . Note that  $g_{p_0}(X) = X - (X - p_0)_+$  and  $g_0(X) = (1 - \theta_1(0))X$ , so that  $cv_{g_0(X)} = cv_X$ . Therefore, by the inequality 5.5 and the first case assumption, we obtain:

$$cv_{g_{p_0}(X)} \leq cv_{R_f(X)} \leq cv_{g_0(X)}.$$

Because  $cv_{g_p(X)}$  is continuous in  $p$ , it follows from the intermediate value theorem that there exists a  $p_1 \in [0, p_0]$  such that

$$cv_{g_{p_1}(X)} = cv_{R_f(X)}.$$

Then, we have a change-loss insurance with ceded loss function:

$$f_c(x; \theta_1(p_1), p_1) = \theta_1(p_1)(X - p_1)_+,$$

with  $p_1 \in [0, p_0]$  and  $\theta_1(p_1) \in [\mathbb{E}(f(X))/\mathbb{E}(X), 1]$ , such that  $E(f_c(X; \theta_1(p_1), p_1)) = \mathbb{E}(f(X))$  and  $\sigma(R_{f_c}(X; \theta_1(p_1), p_1)) = \sigma(R_f(X))$ .

## 5.4. Optimal Contract

---

Second case [ $\text{cv}_{R_f(X)} > \text{cv}_X$ ]: For  $m \in [0, m_0]$ , we define  $h_m(x) \equiv \theta_2(m)(x-m)_+$  and see that  $h_{m'}(x)$  up-crosses  $h_m(x)$  for  $0 \leq m < m' \leq m_0$ . By Lemma 3 in [Ohlin \(1969\)](#), we have:  $h_m(X) \leq_{cx} h_{m'}(X)$ , and hence  $\sigma(h_m(X)) \leq \sigma(h_{m'}(X))$  and  $\text{cv}_{h_m(X)}$  is increasing in  $m$ . Note that  $h_{m_0}(X) = X - \min(X, m_0)$  and  $h_0(X) = \theta_2(0)X$ , so that  $\text{cv}_{h_0(X)} = \text{cv}_X$ . Therefore, by the inequality 5.5 and the second case assumption, we obtain:

$$\text{cv}_{h_0(X)} \leq \text{cv}_{R_f(X)} \leq \text{cv}_{h_{m_0}(X)}.$$

Because  $\text{cv}_{h_m(X)}$  is continuous in  $m$ , it follows from the intermediate value theorem that there exists a  $m_1 \in [0, m_0]$  such that

$$\text{cv}_{\theta_2(m_1)(X-m_1)_+} = \text{cv}_{R_f(X)}.$$

Then, we have a dual change loss insurance with a ceded loss function:

$$f_d(x; \theta_2(m_1), m_1) = x - \theta_2(m_1)(x - m_1)_+,$$

with  $m_1 \in [0, m_0]$  and  $\theta_2(m_1) \in [\mathbb{E}(R_f(X))/\mathbb{E}(X), 1]$ , such that  $\mathbb{E}(f_d(X; \theta_2(m_1), m_1)) = \mathbb{E}(f(X))$  and  $\sigma(R_{f_d}(X; \theta_2(m_1), m_1)) = \sigma(R_f(X))$ .

Now we need to prove that the above change-loss insurance and dual change-loss insurance are optimal solutions to the problem 5.1 for third degree risk averse agents. There is no difference in the proof of these two cases. Thus, we only show the proof for the change-loss insurance. In this case, we have  $\mathbb{E}(R_{f_c}(X; \theta_1(p_1), p_1)) = \mathbb{E}(R_f(X))$  and  $\pi(f_c(X; \theta_1(p_1), p_1)) = \pi(f(X))$ . Note that the functions  $R_f(x)$  and  $R_{f_c}(x; \theta_1(p_1), p_1)$  cross at most twice.

If they only cross once, Lemma 3 in [Ohlin \(1969\)](#) shows that these two non-negative stochastic variables are ranked in convex order. Because  $\mathbb{E}(R_f(X)^2) = \mathbb{E}(R_{f_c}(X; \theta_1(p_1), p_1)^2)$ , by Theorem 3.A.42 in [Shaked and Shanthikumar \(2007\)](#),  $R_f(X)$  and  $R_{f_c}(X; \theta_1(p_1), p_1)$  have the same distribution. As  $\pi(f_c(X; \theta_1(p_1), p_1)) = \pi(f(X))$ , we obtain the results.

If  $R_f(x)$  and  $R_{f_c}(x; \theta_1(p_1), p_1)$  cross twice, we are in the case depicted in Fig 3.2 in [Chi \(2017\)](#). Let  $0 < x_1 < x_2 < \infty$  be the two intersection points of  $R_f(x)$  and  $R_{f_c}(x; \theta_1(p_1), p_1)$ . We obtain:

$$\begin{cases} w_f(x) \leq w_{f_c}(x) & x \leq x_1 \\ w_f(x) \geq w_{f_c}(x) & x_1 < x \leq x_2 \\ w_f(x) < w_{f_c}(x) & x > x_2. \end{cases}$$

Let  $F_f$  and  $F_{f_c}$  be the cumulative distribution functions of  $w_f(X)$  and  $w_{f_c}(X)$ ,

respectively, and set  $z_i = w_f(x_i)$ ,  $i = 1, 2$ . Note that  $z_1 > z_2$  and that

$$\begin{cases} F_f(z) > F_{f_c}(z) & z < z_2 \\ F_f(z) \leq F_{f_c}(z) & z_2 \leq z < z_1 \\ F_f(z) \geq F_{f_c}(z) & z \geq z_1. \end{cases}$$

Then,  $F_f$  and  $F_{f_c}$  cross twice and the last sign of  $F_f(z) - F_{f_c}(z)$  is a  $+$ . In conjunction with

$$\begin{cases} \pi(f_c(X; \theta_1(p_1), p_1)) = \pi(f(X)) \\ \mathbb{E}(R_{f_c}(X; \theta_1(p_1), p_1)) = \mathbb{E}(R_f(X)) \\ \sigma(R_{f_c}(X; \theta_1(p_1), p_1)) = \sigma(R_f(X)) \end{cases} \implies \begin{cases} \mathbb{E}(w_f(X)) = \mathbb{E}(w_{f_c}(X)) \\ \sigma(w_f(X)) = \sigma(w_{f_c}(X)) \end{cases},$$

by Theorem 3.A.66 in [Shaked and Shanthikumar \(2007\)](#), we have  $w_f(X) \leq_{3-cx} w_{f_c}(X)$ . The same proof also gives  $w_f(X) \leq_{3-cx} w_{f_d}(X)$ . The equivalence of the orders  $\leq_{3-cx}$  and  $\leq_{3-cv}$  gives our results.  $\square$

In the concave set  $\mathcal{Q}$ , we obtain the optimal contracts in the same form of [Chi \(2017\)](#), though we have different optimization criteria. Further, our optimal contract form depends on the coefficient of variation of retained loss while theirs depends on the ceded loss part.

**Remark 3.** We write the ceded loss function of dual change-loss insurance as follows:

$$f_d(x; \theta, d) = x \mathbb{1}_{x \leq d} + ((1 - \theta)x + \theta d) \mathbb{1}_{x > d}.$$

We can see that the dual change-loss insurance has full coverage of small loss and changes into a change-loss insurance plus a constant protection for large loss. Thus, in contrast to the stop-loss insurance favored by risk averters, when the coefficient variation of retained loss is less than full loss, decreasing the compensation for large loss can attract risk lovers. But in another case, only both the full protection for small loss and less payment for large loss can satisfy risk lovers.

## 5.5 Background risk

We confine the analysis to a new set of ceded loss functions, defined as

$$\mathcal{K} = \{f(x) \mid 0 \leq f(x) \leq x, 0 \leq f'(x) \leq 1\}.$$

In contrast to the set  $\mathcal{S}$ , we further assume the marginal indemnity to be less than one. In the set  $\mathcal{K}$ , both  $f(x)$  and  $R_f(x)$  are increasing, which make the insurer and insured pay more for a larger realization of loss, thus reducing ex post moral hazard. As the retained loss function of concave contracts are all increasing and the set  $\mathcal{K}$  includes the dual limited limited stop-loss insurance that is excluded out of  $\mathcal{Q}$ , the

## 5.5. Background risk

---

set  $\mathcal{K}$  is larger than the set  $\mathcal{Q}$ . Then, we seek optimal insurance policy among the set  $\mathcal{K}$ .

Denote  $Y$  the background risk. With an insurance contract, the final wealth of the insured  $w_f(X, Y)$  changes into the form of:

$$w_f(X, Y) = w_0 - Y - R_f(X) - \pi(f(X)).$$

The optimization problem is formulated as

$$f^* = \arg \max_{f \in \mathcal{K}} \mathbb{E}(U(w_f(X, Y))), \quad (5.6)$$

where we still make the utility function  $U(x)$  satisfy  $U'''(x) \geq 0$ . Then, we show the optimal contract form in the two cases that the insurable risk  $X$  and background risk  $Y$  are independent and dependent.

### 5.5.1 The independent case

When the insurable risk  $X$  and background risk  $Y$  are independent, the addition of independent  $Y$  doesn't change the stochastic ordering. Thus, the same arguments as in section 4.1 derive the following result.

**Proposition 20.** *For any admissible insurance policy  $f \in \mathcal{K}$ , if  $X$  and  $Y$  are independent, there exists a dual limited stop-loss insurance  $f_d$  such that*

$$w_f(X, Y) \leq_{3-cv} w_{f_d}(X, Y).$$

**Remark 4.** *The independent background risk doesn't change the optimal contract form. This observation has also been given by [Gollier \(1996\)](#), where they show that with an independent background risk, the stop-loss contract is still the optimal insurance policy for risk-averse insured, but point out that the independent background risk will change the optimal level of the deductible. Similarly, in our case, the independent background risk also modifies the optimal level of the upper limit and deductible.*

### 5.5.2 The dependent case

When the background risk  $Y$  is correlated with the insurable risk  $X$ , we show the optimal insurance policy changes into the stop-loss insurance. To proceed, we introduce necessary definitions of directionally convex function and directionally convex order. A function  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  is said to be directionally convex if for any  $\mathbf{x}_i \in \mathbb{R}^n, i = 1, \dots, 4, \mathbf{x}_1 \leq \mathbf{x}_2 \leq \mathbf{x}_4, \mathbf{x}_1 \leq \mathbf{x}_3 \leq \mathbf{x}_4$  and  $\mathbf{x}_1 + \mathbf{x}_4 = \mathbf{x}_2 + \mathbf{x}_3$ , such that

$$\phi(\mathbf{x}_2) + \phi(\mathbf{x}_3) \leq \phi(\mathbf{x}_1) + \phi(\mathbf{x}_4).$$

Denote  $\mathbf{Z} = (Z_1, \dots, Z_n)$  and  $\mathbf{Y} = (Y_1, \dots, Y_n)$  the two random vectors. Then, the directionally convex order is defined as:

**Definition 7.** (*Shaked and Shanthikumar (2007)*) Random vector  $Z$  is said to be smaller than  $Y$  in the directionally convex order, denoted as  $Z \leq_{\text{dir-cx}} Y$ , if and only if

$$\mathbb{E}(\phi(Z)) \leq \mathbb{E}(\phi(Y)), \quad (5.7)$$

for all directionally convex functions  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ .

Then, we can give the following result.

**Proposition 21.** For any admissible insurance policy  $f \in \mathcal{X}$ , if  $X$  and  $Y$  are comonotonic, there exists a stop-loss insurance  $f_d$  such that

$$w_f(X, Y) \leq_{3-cv} w_{f_d}(X, Y).$$

*Proof.* The proof in Proposition 1 has shown that there exists a stop-loss insurance policy with retained loss function

$$R_{f_d}(x) = x - (x - p_0)_+,$$

such that

$$R_{f_d}(X) \leq_{cx} R_f(X).$$

As the  $R_{f_d}(x)$  and  $R_f(x)$  are increasing functions and  $X$  and  $Y$  are comonotonic, both  $(Y, R_{f_d}(X))$  and  $(Y, R_f(X))$  are comonotonic vectors. By Lemma 3.12.13 of Müller and Stoyan (2002), we have

$$(Y, R_{f_d}(X)) \leq_{\text{dir-cx}} (Y, R_f(X)).$$

Denote the function

$$\phi(u, v) = U'(z + u)\mathbf{1}_{\{v > z\}}.$$

As  $\frac{\partial^2 \phi}{\partial u^2}(u, v) = U'''(z + u)\mathbf{1}_{\{v > z\}} \geq 0$ ,  $\frac{\partial^2 \phi}{\partial u \partial v}(u, v) = \frac{\partial^2 \phi}{\partial v^2}(u, v) = 0$ , the  $\phi(u, v)$  is a directionally convex function. Then, we obtain

$$\mathbb{E}(U'(z + Y)\mathbf{1}_{\{R_{f_d}(X) > z\}}) \leq \mathbb{E}(U'(z + Y)\mathbf{1}_{\{R_f(X) > z\}}).$$

On both sides, we integrate  $z$  over the interval  $(0, \infty)$  and add  $\mathbb{E}(U(Y))$ . We have

$$\int_0^\infty \mathbb{E}(U'(z + Y)\mathbf{1}_{\{R_{f_d}(X) > z\}})dz + \mathbb{E}(U(Y)) \leq \int_0^\infty \mathbb{E}(U'(z + Y)\mathbf{1}_{\{R_f(X) > z\}})dz + \mathbb{E}(U(Y)).$$

We exchange the order of the expectation and the integration, and obtain

$$\mathbb{E} \left( \int_0^\infty U'(z + Y)\mathbf{1}_{\{R_{f_d}(X) > z\}}dz \right) + \mathbb{E}(U(Y)) \leq \mathbb{E} \left( \int_0^\infty U'(z + Y)\mathbf{1}_{\{R_f(X) > z\}}dz \right) + \mathbb{E}(U(Y)).$$

## 5.6. Conclusion

---

Then, we have

$$\mathbb{E}(U(R_{f_d}(X) + Y)) \leq \mathbb{E}(U(R_f(X) + Y)).$$

Further, we obtain

$$R_{f_d}(X) + Y \leq_{3-cv} R_f(X) + Y$$

and then

$$-R_f(X) - Y \leq_{3-cv} -R_{f_d}(X) - Y.$$

As  $\pi(f_d(X; p_0)) = \pi(f(X))$ , we obtain the results. □

**Remark 5.** *When background risk is comonotonic with insurable risk, even risk lovers also need take much more care of large loss, due to the large risk exposure.*

## 5.6 Conclusion

This paper investigates the optimal design of insurance policy when both of risk lovers and risk averters are covered. We find a new contract, a dual limited stop-loss insurance, is favored by both of these two type of persons. When restricted to concave types, the optimal contract is a change-loss insurance or a dual change-loss insurance, which depends on the coefficient of variation of the retained loss. In the presence of background risk, the optimal contract reduces to the stop-loss insurance policy. The issue of these contracts can attract a wider range of insureds, thus making the insurance company more competitive and profitable.

This paper can be extended to study the optimal contract design under higher order degree risk, such as the well-documented fourth degree risk (temperance), thus providing some new contracts favored by a larger population of insureds. The premium principle is worth being generalized. In this paper, we make use of expected value premium principle. However, this premium principle is not quite realistic. It is significant to study the optimal contracts under more general premium principles. These extentions will be left for future research.



# Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

---

## Contents

---

<b>6.1</b>	<b>Introduction</b>	<b>103</b>
<b>6.2</b>	<b>Stochastic Gradient Boosting Frequency-Severity Model</b>	<b>106</b>
6.2.1	The Dependent Frequency-Severity Model	106
6.2.2	Stochastic Gradient Boosting	107
6.2.3	The Dependent-FSBoost Model	108
6.2.4	Tuning Parameters Choice and $\delta$ Estimation	113
<b>6.3</b>	<b>Simulation Study</b>	<b>113</b>
6.3.1	Simple Case	114
6.3.2	Complex Case	117
6.3.3	Examination of $\delta$	120
<b>6.4</b>	<b>Application</b>	<b>121</b>
6.4.1	Data	121
6.4.2	Model Comparison	124
6.4.3	Model Interpretation	125
<b>6.5</b>	<b>Conclusion</b>	<b>129</b>

---

## 6.1 Introduction

Insurance claim modeling is a topic of great concern in non-life insurance. The model helps the insurer accurately estimate potential loss, so as to make appropriate actuarial decisions. For each insurance policy, the model enables insurer to set the premium. It is essential to charge the policyholders with fair premium. For example, in auto insurance, if insurance companies charge too little for young drivers and too much for old drivers, young drivers will be attracted while old drivers will switch to competitors. This adverse selection issue makes insurers lose profitable and gain underpriced policies, both resulting in economic losses. For the entire insurance company, the model impacts the decision making on the level of risk capital.

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

---

The underestimation of loss will induce insurance company to fall short of risk capital, thus raising bankruptcy risk. The overestimation will lead insurance company to be overly conservative, thereby reducing liquid capital and hampering business expansion. Therefore, the accurate insurance claim model is of great significance for the competency and profits of insurance company.

Frequency-severity model is a widely used standard model, which separately models the claim frequency and average claim severity. The claim frequency examines the number of claims and the average claim severity takes account of the average amount of claims conditional on occurrence. The claim frequency and severity are highly dependent on the characteristics of an individual policy, for instance, in auto insurance, the age, gender and motor vehicle record points of the policyholder, per captial income or population density of the policyholder's residential area, age and model of the vehicle, etc. Thus, the predictive models are needed. The traditional frequency-severity model chooses generalized linear models (GLM) as the marginal regression models for claim frequency and severity. The frequency part employs a Poisson or negative binomial regression and the severity part makes use of a gamma or inverse Gaussian regression.

Though the frequency-severity model is widely used in industry or in actuarial studies, there are two obvious limitations. First, the structure of the models is restricted to a linear form. However, in practice, there is a variety of nonlinear and complex interaction effects. For example, in auto insurance, the nonlinear relation between claim size and insured's age is well documented, see, for instance, [Frees, Shi, and Valdez \(2009\)](#). The generalized additive model (GAM) developed in [Hastie and Tibshirani \(1986\)](#) and popularized by [Wood \(2006\)](#) overcomes the restrictive linear form of the GLM model, by modeling the continuous variable with smooth functions. However, the GAM model is also restricted to an additive form and hence the complex interaction effects can not be automatically identified and modeled. In practice, it is difficult to manually find all the interaction effects, especially when many variables are involved. Missing some important interactions will adversely affect prediction accuracy. Second, the model assumes an independent relation between the claim frequency and severity. However, in practice, the claim frequency and severity are often dependent. For example, in auto insurance, the claim frequency and severity are often negatively correlated, see, for instance, [Gschlößl and Czado \(2007\)](#). Home insurance claims due to flood tend to be both large and frequent in the affected area. [Frees, Gao, and Rosenberg \(2011\)](#) point out that claim frequency has a significant effect on claim severity for outpatient expenditures. [Erhardt and Czado \(2012\)](#); [Frees, Gao, and Rosenberg \(2011\)](#); [Garrido, Genest, and Schulz \(2016\)](#); [Gschlößl and Czado \(2007\)](#) capture the dependence by treating claim frequency as a predictor variable in the regression model for average claim severity. However, as indicated in [Shi, Feng, and Ivantsova \(2015\)](#), the predictor method with marginal GLM model only can measure a linear relation between claim frequency and severity. [Czado, Kastenmeier, Brechmann, and Min \(2012\)](#); [Krämer, Brechmann, Silvestrini, and Czado \(2013\)](#); [Shi, Feng, and Ivantsova \(2015\)](#) employ the parametric copulas to model the joint distribution of claim frequency and aver-

## 6.1. Introduction

---

age claim severity. However, the popular parametric copulas, such as elliptical and Archimedean copulas, are restricted to certain correlation structures, thus being unable to fully capture the nonlinear relation between claim frequency and severity. In total, there is a need to develop a data-driven dependent frequency-severity model, which can learn the optimal model structure from the data and can flexibly capture the nonlinear dependence between claim frequency and severity.

Boosting is one of the most successful ensemble learning methods, which additively combine a large number of weak prediction models called weak learners to enhance prediction performance. Freund and Schapire (1997) first introduce a boosting algorithm named AdaBoost for classification. Breiman (1999); Breiman et al. (1998) give an important observation that the AdaBoost algorithm can be viewed as a functional gradient descent algorithm. Friedman, Hastie, Tibshirani, et al. (2000) lay out the ground work to reveal that the AdaBoost and other boosting algorithms are additive models, i.e., an additive combination of weak learners, and together with the previous observed connection to the optimization method, to propose a gradient boosting algorithm. The gradient boosting algorithm can be treated as an estimation method for the additive model that combines weak learners. Based on this new perspective, many boosting regression models are developed. For example, Friedman, Hastie, and Tibshirani (2001); Friedman (2001, 2002) give the boosting regression models with least-squares, least absolute deviation and Huber loss functions and Kriegler and Berk (2010); Ridgeway (1999a,b) develop boosting Poisson regression, boosting proportional hazards regression and boosting quantile regression models. There has been an extensive literature on boosting models. See Bühlmann, Hothorn, et al. (2007); Schapire and Freund (2012) for more comprehensive review.

In insurance domain, there are a few papers that make use of boosting models. Noll, Salzmann, and Wuthrich (2018) show that the boosting Poisson model performs better than the GLM model in predicting claim frequency. Yang, Qian, and Zou (2018) develop a gradient boosting Tweedie compound Poisson model, where they apply gradient boosting algorithm to estimate the Tweedie compound Poisson model and manifest that the model makes more accurate premium prediction than the GLM and GAM Tweedie compound Poisson models.

In this paper, we develop a stochastic gradient boosting frequency-severity model (Dependent-FSBoost), in which we employ gradient boosting algorithm to estimate the marginal regression model of claim frequency and severity, which are in the form of additive models with regression tree as weak learners, and we capture the dependence between claim frequency and severity, by treating claim frequency as a predictor in the regression model for average claim severity. The Dependent-FSBoost model is capable of fitting a flexible nonlinear relation between claim frequency (severity) and predictors and capturing complex interactions among predictors. As a byproduct of data-driven model, the Dependent-FSBoost model can fully capture the nonlinear dependence between claim frequency and severity. Besides these advantages, the data-driven model inherits all the desirable features of boosting models, such as the data-driven model structure, high prediction accuracy, feature

selection, high capacity of avoiding overfitting, etc. In a simulation study, we compare our model with the GLM and GAM frequency-severity models, and show that our model makes more accurate prediction of the frequency and severity distribution, which in turn leads to a more precise estimation of the potential loss. We also confirm that the frequency-severity model can be significantly improved by adding claim frequency as a predictor in the regression model for average claim severity, when the claim frequency and severity are correlated. These results are also confirmed in a case study of French auto insurance policy.

The rest of this paper is organized as follows. Section 2 reviews the dependent frequency-severity model and stochastic gradient boosting algorithm, and introduces our Dependent-FSBoost model. Section 3 investigates the performance of our model in a simulation study. Then, section 4 applies the model to analyze a French auto insurance claim data.

## 6.2 Stochastic Gradient Boosting Frequency-Severity Model

Lots of distributions from the exponential dispersion family can model the claim frequency and severity, such as the binomial, Poisson, negative binomial for claim frequency, and the gamma, inverse Gaussian and lognormal for claim severity. For simple statement, we illustrate the Dependent-FSBoost model with the gamma and Poisson cases and other members can also be used in the same manner. We begin with the introduction of dependent frequency-severity model and of stochastic gradient boosting algorithm. Then, we show the implementation of the Dependent-FSBoost model.

### 6.2.1 The Dependent Frequency-Severity Model

The aggregate loss is expressed as

$$S = \sum_{j=1}^N Y_j,$$

where  $N$  is the number of claims, and  $Y_j, j = 1, \dots, N$  is the  $j^{\text{th}}$  claim amount. Conditional on  $N$ , the claim amount  $Y_1, \dots, Y_N$  are i.i.d.. For  $N > 0$ , denote by  $\bar{Y} = \frac{Y_1 + \dots + Y_N}{N}$  the average claim severity. Then, the aggregate loss can be written as

$$S = N\bar{Y}.$$

The claim frequency  $N$  is modeled as a Poisson distribution with parameter  $\lambda > 0$ ,

$$f_N(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda} \quad \text{for } n = 0, 1, \dots,$$

## 6.2. Stochastic Gradient Boosting Frequency-Severity Model

---

and the average claim size  $\bar{Y}$  conditional on  $N$  is modeled via a gamma distribution

$$f_{\bar{Y}|N}(x|\mu_N, \delta) = \frac{1}{s\Gamma\left(\frac{1}{\delta}\right)} \left(\frac{s}{\mu_N\delta}\right)^{\frac{1}{\delta}} e^{-\frac{s}{\mu_N\delta}} \quad \text{for } s > 0,$$

where mean parameter  $\mu_N > 0$  depends on  $N$  and dispersion parameter  $\delta > 0$ .

Denote by  $\mathbf{x}$  the vector of predictors of an individual policy. These information can be used to predict the  $E(N)$  and  $E(\bar{Y}|N)$ , i.e.,  $\lambda$  and  $\mu_N$ . Suppose that the parameters  $\lambda$  and  $\mu_N$  are predicted by the following two marginal regression models:

$$\log(\lambda) = F_N(\mathbf{x}; \boldsymbol{\alpha}) \quad \text{and} \quad \log(\mu_N) = F_{\bar{Y}|N}(\mathbf{x}, N; \boldsymbol{\beta}),$$

where  $F_N : \mathbb{R}^p \rightarrow \mathbb{R}$  and  $F_{\bar{Y}|N} : \mathbb{R}^p \times \mathbb{N} \rightarrow \mathbb{R}$  are two regression functions, and  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  denote the vector of regression parameters, and  $N$  is treated as a predictor in the regression model for the average claim severity so as to capture the dependence between claim frequency and severity. In the GLM or GAM models, the functions  $F_N$  and  $F_{\bar{Y}|N}$  are restricted to a linear or additive form. In our model, we will employ stochastic gradient boosting algorithm to estimate the  $F_N$  and  $F_{\bar{Y}|N}$ . The  $F_N$  and  $F_{\bar{Y}|N}$  will be in the fully flexible additive form that uses regression trees as additive components, different from the GAM that is additive in the predictor variables.

### 6.2.2 Stochastic Gradient Boosting

Before we proceed, we briefly review stochastic gradient boosting algorithm in [Friedman \(2002\)](#). Denote  $\mathbf{x} = (x_1, \dots, x_p)$  the set of predictor variables and  $y$  the response variable. Given the training sample  $\{y_i, \mathbf{x}_i\}_1^d$  and the loss function  $\Psi(y, f(\mathbf{x}))$ , the algorithm is to estimate the function  $\hat{f}(\mathbf{x})$  that minimizes the loss over the training sample,

$$\hat{f}(\mathbf{x}) = \arg \min_{f(\mathbf{x})} \frac{1}{d} \sum_{i=1}^d \Psi(y_i, f(\mathbf{x}_i)),$$

where  $f(\mathbf{x})$  is constrained to the form of a sum of weak learners as

$$f(\mathbf{x}) = h(\mathbf{x}; \mathbf{a}_0) + \sum_{m=1}^M \beta_m h(\mathbf{x}; \mathbf{a}_m), \quad (6.1)$$

where  $h(\mathbf{x}; \mathbf{a}_m)$  is a weak learner with parameters  $\mathbf{a}_m$ ,  $\beta_m \in \mathbb{R}$  is the expansion coefficient,  $M$  is the number of weak learners, usually choosing regression or decision trees.

The algorithm works in a forward stagewise manner. Let  $f_0(\mathbf{x})$  be an initial

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

---

estimate of  $\hat{f}(\mathbf{x})$  as a constant

$$f_0(\mathbf{x}) = h(\mathbf{x}; \mathbf{a}_0) = \arg \min_{\rho} \sum_{i=1}^d \Psi(y_i, \rho).$$

Denote  $f_{m-1}(\mathbf{x})$  the estimate at the  $(m-1)^{\text{th}}$  step. Then, at the  $m^{\text{th}}$  step, the algorithm randomly selects subsample of size  $\tilde{d} < d$ ,  $\{\tilde{y}_i, \tilde{\mathbf{x}}_i\}_1^{\tilde{d}}$ , and computes the negative gradient

$$\tilde{z}_i = - \frac{\partial \Psi(\tilde{y}_i, f(\tilde{\mathbf{x}}_i))}{\partial f(\tilde{\mathbf{x}}_i)} \Big|_{f(\tilde{\mathbf{x}}_i) = f_{m-1}(\tilde{\mathbf{x}}_i)},$$

and then fits the weak learner  $h(\mathbf{x}; \mathbf{a}_m)$  to minimize the following least square sum

$$\mathbf{a}_m = \arg \min_{\mathbf{a}} \sum_{i=1}^{\tilde{d}} (\tilde{z}_i - h(\tilde{\mathbf{x}}_i; \mathbf{a}))^2.$$

The optimal  $\beta_m$  is the solution to

$$\beta_m = \arg \min_{\beta} \sum_{i=1}^{\tilde{d}} \Psi(\tilde{y}_i, f_{m-1}(\tilde{\mathbf{x}}_i) + \beta h(\tilde{\mathbf{x}}_i; \mathbf{a}_m)).$$

Then, the estimate is updated as

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \nu \beta_m h(\mathbf{x}; \mathbf{a}_m),$$

where  $0 < \nu \leq 1$  is the shrinkage factor that controls the learning rate. [Friedman \(2001\)](#) points out that small  $\nu$  reduces overfitting and increases predictive performance. It is worth noting that if using  $\tilde{d} = d$  at each iteration, the algorithm does not introduce the sample randomness, thus reducing to the standard gradient boosting algorithm. [Friedman \(2002\)](#) shows that incorporating the sample randomness into the gradient boosting can substantially improve both of the computation speed and predictive performance.

### 6.2.3 The Dependent-FSBoost Model

Now, we can begin to use the stochastic gradient boosting algorithm to estimate the functions  $F_N$  and  $F_{\bar{Y}|N}$ . Denote by  $\{n_i, s_i, \mathbf{x}_i\}$  the claim frequency, the average claim severity and the vector of predictor variables for the  $i^{\text{th}}$  contract. From  $d$

## 6.2. Stochastic Gradient Boosting Frequency-Severity Model

independent insurance contracts, we have the joint log-likelihood in the form of

$$\begin{aligned} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \delta | \{n_i, s_i, x_i\}_{i=1}^d) &= \sum_{i=1}^d \log f_N(n_i | \lambda_i) + \sum_{i=1}^d \log f_{\bar{Y}|N}(s_i | \mu_{n_i}, \delta) \\ &= \sum_{i=1}^d \log \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i} + \sum_{i=1}^d \log \frac{1}{s_i \Gamma\left(\frac{1}{\delta}\right)} \left(\frac{s_i}{\mu_{n_i} \delta}\right)^{\frac{1}{\delta}} e^{-\frac{s_i}{\mu_{n_i} \delta}}. \end{aligned} \quad (6.2)$$

The log-likelihood function (5.2) could be decomposed into two pieces:

$$\left\{ \begin{array}{l} l_1(\boldsymbol{\alpha}) = \sum_{i=1}^d \log \frac{e^{n_i F_N(\mathbf{x}_i; \boldsymbol{\alpha})}}{n_i!} e^{-e^{F_N(\mathbf{x}_i; \boldsymbol{\alpha})}} \\ l_2(\boldsymbol{\beta}, \delta) = \sum_{i=1}^d \log \frac{1}{s_i \Gamma\left(\frac{1}{\delta}\right)} \left(\frac{s_i}{\delta e^{F_{\bar{Y}|N}(\mathbf{x}_i, n_i; \boldsymbol{\beta})}}\right)^{\frac{1}{\delta}} e^{-\frac{s_i}{\delta e^{F_{\bar{Y}|N}(\mathbf{x}_i, n_i; \boldsymbol{\beta})}}} \end{array} \right.$$

For the time being, we assume dispersion parameter  $\delta$  is known. We use the negative log-likelihood functions  $-l_1(\boldsymbol{\alpha})$  and  $-l_2(\boldsymbol{\beta}, \delta)$  as two loss functions. Then, the stochastic gradient boosting algorithm estimates the functions  $F_N(x; \boldsymbol{\alpha})$  and  $F_{\bar{Y}|N}(x, N; \boldsymbol{\beta})$  that minimize the loss over the data  $\{n_i, s_i, \mathbf{x}_i\}_{i=1}^d$ ,

$$\hat{f}(\mathbf{x}) = \arg \min_{f(\mathbf{x})} \sum_{i=1}^d \Psi_1(n_i, f(\mathbf{x}_i)) \quad \text{and} \quad \hat{g}(\mathbf{x}, n) = \arg \min_{g(\mathbf{x}, n)} \sum_{i=1}^d \Psi_2(s_i, g(\mathbf{x}_i, n_i)),$$

where

$$\left\{ \begin{array}{l} \Psi_1(n_i, f(\mathbf{x}_i)) = -\log \frac{e^{n_i f(\mathbf{x}_i)}}{n_i!} e^{-e^{f(\mathbf{x}_i)}} \\ \Psi_2(s_i, g(\mathbf{x}_i, n_i)) = -\log \frac{1}{s_i \Gamma\left(\frac{1}{\delta}\right)} \left(\frac{s_i}{\delta e^{g(\mathbf{x}_i, n_i)}}\right)^{\frac{1}{\delta}} e^{-\frac{s_i}{\delta e^{g(\mathbf{x}_i, n_i)}}} \end{array} \right.,$$

and the functions  $f(\mathbf{x})$  and  $g(\mathbf{x}, n)$  are confined to the form of (5.1) as a sum of weak learners.

Then, the algorithm solves the above optimization problems in a forward stage-wise manner. The initial estimates are chosen as

$$\left\{ \begin{array}{l} f_0(\mathbf{x}) = \arg \min_{\rho} \sum_{i=1}^d \Psi_1(n_i, \rho) \\ g_0(\mathbf{x}, n) = \arg \min_{\rho} \sum_{i=1}^d \Psi_2(s_i, \rho) \end{array} \right.$$

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

---

Denote  $f_{m-1}(\mathbf{x})$  and  $g_{m-1}(\mathbf{x}, n)$  the estimates at the  $(m-1)$ <sup>th</sup> step. At the  $m$ <sup>th</sup> step, the algorithm randomly selects subsample of size  $\tilde{d} < d$ ,  $\{\tilde{n}_i, \tilde{s}_i, \tilde{\mathbf{x}}_i\}_1^{\tilde{d}}$  and computes the negative gradient

$$\begin{cases} \tilde{z}_i^f = \tilde{n}_i - e^{f_{m-1}(\tilde{\mathbf{x}}_i)} \\ \tilde{z}_i^g = \frac{\tilde{s}_i e^{-g_{m-1}(\tilde{\mathbf{x}}_i, \tilde{n}_i)} - 1}{\delta} \end{cases}$$

Then, the algorithm fits the two weak learners  $h^f(\mathbf{x}; \mathbf{a}_m^f)$  and  $h^g(\mathbf{x}, n; \mathbf{a}_m^g)$  to minimize the following least square sum

$$\begin{cases} \mathbf{a}_m^f = \arg \min_{\mathbf{a}} \sum_{i=1}^{\tilde{d}} (\tilde{z}_i^f - h^f(\tilde{\mathbf{x}}_i; \mathbf{a}))^2 \\ \mathbf{a}_m^g = \arg \min_{\mathbf{a}} \sum_{i=1}^{\tilde{d}} (\tilde{z}_i^g - h^g(\tilde{\mathbf{x}}_i, \tilde{n}_i; \mathbf{a}))^2 \end{cases}, \quad (6.3)$$

where we use K-terminal node regression trees as weak learners, i.e.,

$$\begin{cases} h^f(\mathbf{x}; \mathbf{a}_m^f) = \sum_{k=1}^K \bar{n}_k \mathbf{1}_{\{\mathbf{x} \in U_{k,m}\}} \\ h^g(\mathbf{x}, n; \mathbf{a}_m^g) = \sum_{k=1}^K \bar{s}_k \mathbf{1}_{\{\mathbf{x} \in V_{k,m}\}} \end{cases},$$

where

$$\begin{cases} \bar{n}_k = \frac{\sum_{i=1}^{\tilde{d}} \tilde{n}_i \mathbf{1}_{\{\tilde{\mathbf{x}}_i \in U_{k,m}\}}}{\sum_{i=1}^{\tilde{d}} \mathbf{1}_{\{\tilde{\mathbf{x}}_i \in U_{k,m}\}}} \\ \bar{s}_k = \frac{\sum_{i=1}^{\tilde{d}} \tilde{s}_i \mathbf{1}_{\{(\tilde{\mathbf{x}}_i, \tilde{n}_i) \in V_{k,m}\}}}{\sum_{i=1}^{\tilde{d}} \mathbf{1}_{\{(\tilde{\mathbf{x}}_i, \tilde{n}_i) \in V_{k,m}\}}} \end{cases},$$

and  $\{U_{k,m}\}_{k=1}^K$  and  $\{V_{k,m}\}_{k=1}^K$  are disjoint regions of the  $\mathbf{x}$  and  $(\mathbf{x}, n)$  spaces, respectively, representing terminal nodes of regression trees. In this case, the parameters  $\mathbf{a}_m^f$  and  $\mathbf{a}_m^g$  are the splitting variables and split points of regression trees, which determine the regions  $\{U_{k,m}\}_{k=1}^K$  and  $\{V_{k,m}\}_{k=1}^K$ . Then, the optimization problem (5.3) is solved by a greedy algorithm with a least squared splitting criterion as in [Friedman \(2001\)](#).

Once the weak learners  $h^f(\mathbf{x}; \mathbf{a}_m^f)$  and  $h^g(\mathbf{x}, n; \mathbf{a}_m^g)$  are obtained, the optimal

## 6.2. Stochastic Gradient Boosting Frequency-Severity Model

---

expansion coefficients  $\beta_m^f$  and  $\beta_m^g$  are the solutions to

$$\begin{cases} \beta_m^f = \arg \min_{\beta} \sum_{i=1}^{\tilde{d}} \Psi_1 \left( \tilde{n}_i, f_{m-1}(\tilde{\mathbf{x}}_i) + \beta \sum_{k=1}^K \tilde{n}_k \mathbf{1}_{\{\tilde{\mathbf{x}}_i \in U_{k,m}\}} \right) \\ \beta_m^g = \arg \min_{\beta} \sum_{i=1}^{\tilde{d}} \Psi_2 \left( \tilde{s}_i, g_{m-1}(\tilde{\mathbf{x}}_i, \tilde{n}_i) + \beta \sum_{k=1}^K \tilde{s}_k \mathbf{1}_{\{(\tilde{\mathbf{x}}_i, \tilde{n}_i) \in V_{k,m}\}} \right) \end{cases}.$$

One can improve the quality of the fit by changing a single expansion coefficient  $\beta_m^f$  ( $\beta_m^g$ ) into an optimal coefficient  $\gamma_{k,m}^f$  ( $\gamma_{k,m}^g$ ) for each region  $U_{k,m}$  ( $V_{k,m}$ ). These optimal coefficients are the solutions to

$$\begin{cases} \gamma_{k,m}^f = \arg \min_{\gamma} \sum_{\tilde{\mathbf{x}}_i \in U_{k,m}} \Psi_1(\tilde{n}_i, f_{m-1}(\tilde{\mathbf{x}}_i) + \gamma) \\ \gamma_{k,m}^g = \arg \min_{\gamma} \sum_{(\tilde{\mathbf{x}}_i, \tilde{n}_i) \in V_{k,m}} \Psi_2(\tilde{s}_i, g_{m-1}(\tilde{\mathbf{x}}_i, \tilde{n}_i) + \gamma) \end{cases}.$$

The explicit solutions are given by

$$\begin{cases} \gamma_{k,m}^f = \log \left( \frac{\sum_{i=1}^{\tilde{d}} \tilde{n}_i}{\sum_{i=1}^{\tilde{d}} e^{f_{m-1}(\tilde{\mathbf{x}}_i)}} \right) \\ \gamma_{k,m}^g = \log \left( \frac{\sum_{i=1}^{\tilde{d}} \tilde{s}_i e^{-g_{m-1}(\tilde{\mathbf{x}}_i, \tilde{n}_i)}}{\tilde{d}} \right) \end{cases}.$$

Then, the estimate is updated as

$$\begin{cases} f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \nu \sum_{k=1}^K \gamma_{k,m}^f \mathbf{1}_{\mathbf{x} \in U_{k,m}} \\ g_m(\mathbf{x}, n) = g_{m-1}(\mathbf{x}, n) + \nu \sum_{k=1}^K \gamma_{k,m}^g \mathbf{1}_{(\mathbf{x}, n) \in V_{k,m}} \end{cases},$$

where we set  $\nu = 0.03$  as [Friedman \(2001\)](#) observes that the shrinkage factor  $\nu \leq 0.1$  leads to much better generalization error.

We summary the Dependent-FSBoost algorithm as follows:

---

### The Dependent-FSBoost Algorithm

---

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

---

1. Initialize  $f_0(\mathbf{x})$  and  $g_0(\mathbf{x}, n)$

$$\begin{cases} f_0(\mathbf{x}) = \arg \min_{\rho} \sum_{i=1}^d \Psi_1(n_i, \rho) \\ g_0(\mathbf{x}, n) = \arg \min_{\rho} \sum_{i=1}^d \Psi_2(s_i, \rho) \end{cases}.$$

2. For  $m = 1$  to  $M$  do

1. Generate a random subsample  $\{\tilde{n}_i, \tilde{s}_i, \tilde{\mathbf{x}}_i\}_1^{\tilde{d}}$ .

2. Compute the negative gradient  $(\tilde{z}_1^f, \dots, \tilde{z}_{\tilde{d}}^f)$  and  $(\tilde{z}_1^g, \dots, \tilde{z}_{\tilde{d}}^g)$

$$\begin{cases} \tilde{z}_i^f = \tilde{n}_i - e^{f_{m-1}(\tilde{\mathbf{x}}_i)} \\ \tilde{z}_i^g = \frac{\tilde{s}_i e^{-g_{m-1}(\tilde{\mathbf{x}}_i, \tilde{n}_i)} - 1}{\delta}, i = 1, \dots, \tilde{d}. \end{cases}$$

3.  $K$ -terminal node regression trees fit two datasets  $\{\tilde{z}_i^f, \tilde{\mathbf{x}}_i\}_{i=1}^{\tilde{d}}$  and  $\{\tilde{z}_i^g, (\tilde{\mathbf{x}}_i, \tilde{n}_i)\}_{i=1}^{\tilde{d}}$  with a least squared splitting criterion and obtain two partitions  $\{U_{k,m}\}_{k=1}^K$  and  $\{V_{k,m}\}_{k=1}^K$ .

4. Compute the optimal coefficients for each region  $U_{k,m}, V_{k,m}, k = 1, \dots, K$

$$\begin{cases} \gamma_{k,m}^f = \log \left( \frac{\sum_{i=1}^{\tilde{d}} \tilde{n}_i}{\sum_{i=1}^{\tilde{d}} e^{f_{m-1}(\tilde{\mathbf{x}}_i)}} \right) \\ \gamma_{k,m}^g = \log \left( \frac{\sum_{i=1}^{\tilde{d}} \tilde{s}_i e^{-g_{m-1}(\tilde{\mathbf{x}}_i, \tilde{n}_i)}}{\tilde{d}} \right) \end{cases}, k = 1, \dots, K.$$

5. Update  $f_m(\mathbf{x})$  and  $g_m(\mathbf{x}, n)$  as

$$\begin{cases} f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \nu \sum_{k=1}^K \gamma_{k,m}^f \mathbf{1}_{\mathbf{x} \in U_{k,m}} \\ g_m(\mathbf{x}, n) = g_{m-1}(\mathbf{x}, n) + \nu \sum_{k=1}^K \gamma_{k,m}^g \mathbf{1}_{(\mathbf{x}, n) \in V_{k,m}} \end{cases}.$$

end

3. Return  $f_M(\mathbf{x})$  and  $g_M(\mathbf{x}, n)$ .

---

### 6.3. Simulation Study

---

#### 6.2.4 Tuning Parameters Choice and $\delta$ Estimation

In this section, we show the choice of tuning parameters including the number of trees  $M$  and the depth of the trees  $T$ , and also the estimation of dispersion parameter  $\delta$ . A suitable number of trees  $M$  avoids over-fitting and improves out-of-sample prediction. The value  $T$  controls the degree of interaction among the explanatory variables. A tree of depth  $T$  has maximum of interaction order as  $T - 1$ . In our model, we adopt cross-validation method to determine the parameters  $(M, T)$ . For the dispersion parameter  $\delta$ , we obtain the estimation by maximizing profile log-likelihood.

The  $K$ -fold cross-validation method splits the data into  $K$  equal-sized folds. Let  $\kappa(i) : \{1, \dots, n\} \rightarrow \{1, \dots, K\}$  be an index function that indicates the fold to which the  $i^{\text{th}}$  observation is allocated by the randomization. We calculate the loss of the  $k^{\text{th}}$  fold data by the model that is trained using the remaining  $K - 1$  fold. We proceed this process for  $k = 1, \dots, K$ . Denote by  $\hat{f}_{-k}(x; M, L)$  the model fitted with the  $k^{\text{th}}$  fold data removed. Then, we combine the cross-validation estimate of the loss on each  $k^{\text{th}}$  fold data as

$$\text{CV}(M, L) = \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \hat{f}_{-\kappa(i)}(x_i; M, L)).$$

The optimal  $(M, L)$  pair is chosen to minimize the loss, i.e.,

$$(\hat{M}, \hat{L}) = \arg \min_{M, L} \text{CV}(M, L).$$

The dispersion parameter  $\delta$  is estimated by minimizing the loss:

$$\hat{\delta} = \arg \min_{\delta} \Psi_2(\delta),$$

where  $\Psi_2(\delta)$  is the empirical loss estimated by the Dependent-FSBoost model for a fixed  $\delta$ . To reduce the computation, we find the optimal  $\delta$  by making a simple grid search over  $S$  grid points  $\{\delta_1, \dots, \delta_S\}$ , i.e.,

$$\hat{\delta} = \arg \min_{\delta \in \{\delta_1, \dots, \delta_S\}} \Psi_2(\delta).$$

### 6.3 Simulation Study

In this section, we compare our data-driven frequency-severity model with the GLM frequency-severity model and the GAM frequency-severity model in two simulation studies. We consider each model in the cases of dependence and independence between the frequency and severity. For simple statement, denote the six models by GLM, Dependent-GLM, GAM, Dependent-GAM, FSBoost, Dependent-FSBoost, where Dependent-GLM, Dependent-GAM, Dependent-FSBoost represent the models in which the frequency and severity are dependent. We make comparison

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

---

among them in the prediction accuracy of claim frequency and severity components. Further, we also investigate the impact of  $\delta$  on the predictive performance of the Dependent-FSBoost model.

In the simulation studies, we use one set of samples for training and another one for testing. Denote  $\{\bar{n}_i, \bar{s}_i\}_{i=1}^{\bar{d}}$  the testing sample with known true parameters  $\{\bar{\lambda}_i, \bar{\mu}_i, \bar{\delta}\}$ . Let  $\{\hat{\lambda}_i, \hat{\mu}_i\}$  be the predicted parameters. We use the following out-of-sample loss and parameter estimation error to measure the prediction accuracy of claim frequency and severity:

- out-of-sample loss for frequency part:

$$\text{Frequency Loss} = - \sum_{i=1}^{\bar{d}} \log \frac{\hat{\lambda}_i^{\bar{n}_i}}{\bar{n}_i!} e^{-\hat{\lambda}_i},$$

- average relative error of  $\lambda$  estimation:

$$\frac{1}{\bar{d}} \sum_{i=1}^{\bar{d}} \frac{|\hat{\lambda}_i - \bar{\lambda}_i|}{\bar{\lambda}_i},$$

- out-of-sample loss for severity part:

$$\text{Severity Loss} = - \sum_{i=1}^{\bar{d}} \log \frac{1}{\bar{s}_i \Gamma\left(\frac{1}{\bar{\delta}}\right)} \left(\frac{\bar{s}_i}{\hat{\mu}_i \bar{\delta}}\right)^{\frac{1}{\bar{\delta}}} e^{-\frac{\bar{s}_i}{\hat{\mu}_i \bar{\delta}}},$$

- average relative error of  $\mu$  estimation:

$$\frac{1}{\bar{d}} \sum_{i=1}^{\bar{d}} \frac{|\hat{\mu}_i - \bar{\mu}_i|}{\bar{\mu}_i}.$$

For the FSBoost and Dependent-FSBoost models, we adopt five-fold cross-validation method to select the optimal  $(M, L)$  pair among the combinations of tuning parameters  $M \in \{100, 200, 300, 400, 500\}$  and  $L \in \{1, 2, 3, 4, 5\}$ . We search the optimal  $\delta$  in the range  $(1, 3)$  with step length 0.1, i.e., the set  $\{1, 1.1, \dots, 3\}$ .

### 6.3.1 Simple Case

In this simulation study, we demonstrate the capacity of the Dependent-FSBoost model in capturing nonlinear effect, complex interactions and nonlinear dependence between claim frequency and severity. The sample  $\{n_i, s_i, \mathbf{x}_i\}_{i=1}^d$  is generated according to

$$n_i \sim \mathcal{P}(\lambda_i), s_i \sim \mathcal{G}(\mu_{n_i}, \delta), x_{i,j} \sim \text{Unif}(0, 1), i = 1, \dots, n, j = 1, \dots, 4,$$

### 6.3. Simulation Study

where  $\lambda_i = e^{F_1(\mathbf{x}_i)}$ ,  $\mu_{n_i} = e^{F_2(\mathbf{x}_i, n_i)}$ ,  $\delta = 2$ , and

$$\begin{cases} F_1(\mathbf{x}_i) = \frac{\pi}{15}(3x_{i,1}^2 + 2(1 - x_{i,2})^2 + 10x_{i,1}x_{i,2}) \\ F_2(\mathbf{x}_i, n_i) = \ln(n_i + 3)e^{x_{i,3}^2 - 2(1 - x_{i,4})^2} + \ln(n_i + 5)e^{\frac{1}{2}x_{i,3}x_{i,4}}. \end{cases}$$

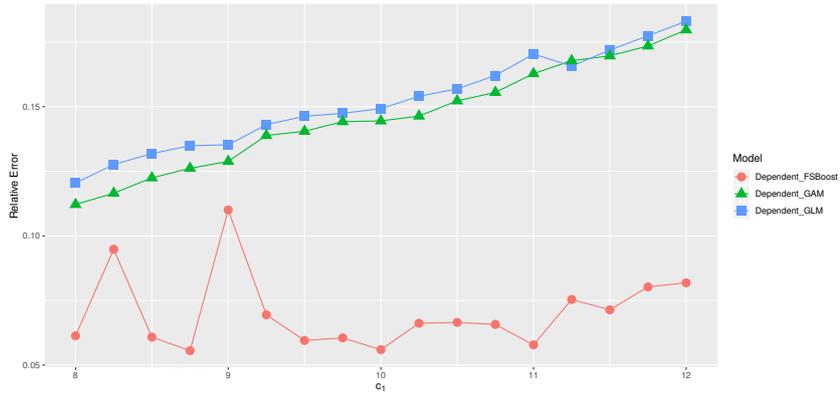
We generate sample of size 10000 for training and another sample of equal size for testing. The resulting out-of-sample loss and parameter estimation error on the testing sample are listed in Table 1, which are averaged over 30 independent replications. Note that the independent model and dependent model share the same claim frequency model. Thus, we only list the claim frequency result for the dependent case. We can find that the dependent model performs better than the corresponding independent model. In the dependent models, the Dependent-FSBoost model outperforms the Dependent-GAM and Dependent-GLM in terms of the smallest out-of-sample loss and parameter estimation errors.

In contrast to the GLM, Dependent-GLM, GAM and Dependent-GAM models, the FSBoost and Dependent-FSBoost models are able to capture the complex interaction effects. Denote by  $c_1$  and  $c_2$  the coefficients of the cross-product terms  $x_{i,1}x_{i,2}$  and  $x_{i,3}x_{i,4}$ , respectively. In Figure 5.1, we range  $c_1$  from 8 to 12 and  $c_2$  from 0.3 to 0.7 to increase the impact of interaction term. We can see that the FSBoost and Dependent-FSBoost models keep relatively stable predictive performance, while the parameter estimation error in the GLM, Dependent-GLM, GAM and Dependent-GAM models show an obvious increasing trend, due to the lack of capacity in capturing interaction effect. Further, we fix  $x_{i,j}$ ,  $j = 3, 4$  in the training sample and change the frequency from 0 to 20. Then, we calculate the value  $\hat{\mu}^* = \frac{1}{10000} \sum_{i=1}^{10000} \log(\hat{\mu}_i)$  and display the change of the  $\hat{\mu}^*$  with respect to the frequency in Figure 5.2. Both the Dependent-GAM and Dependent-FSBoost models can capture nonlinear dependence between claim severity and frequency and the Dependent-FSBoost model performs better, whereas the Dependent-GLM model can only measure linear relation.

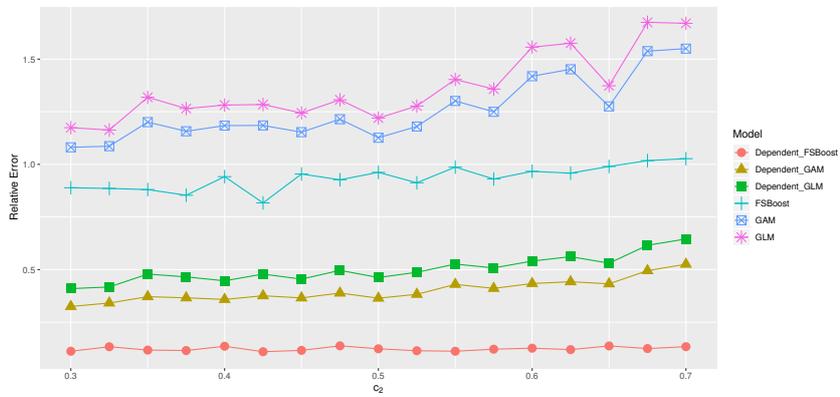
**Table 1** The out-of-sample loss and parameter estimation error

	Frequency Loss	Severity Loss	$\lambda$	$\mu$
GLM	-	55062.75 (3671.05)	-	1.3813 (0.0908)
Dependent-GLM	18447.92 (89.07)	46796.23 (413.18)	0.1517 (0.0026)	0.4844 (0.0187)
GAM	-	50811.05 (1402.98)	-	1.2805 (0.0750)
Dependent-GAM	18380.78 (86.64)	46109.83 (236.17)	0.1450 (0.0019)	0.3969 (0.0147)
FSBoost	-	46896.12 (176.43)	-	0.9310 (0.0357)
Dependent-FSBoost	18107.46 (74.91)	45559.32 (177.73)	0.0693 (0.0157)	0.1229 (0.0077)

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims



(a)



(b)

Figure 6.1: Relative error of parameter estimation in the interaction case

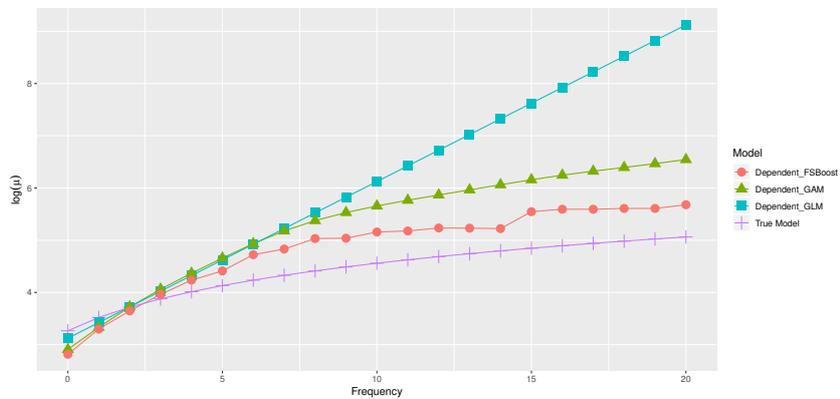


Figure 6.2: The change trend of  $\hat{\mu}^*$  w.r.t. claim frequency

### 6.3. Simulation Study

---

#### 6.3.2 Complex Case

We examine the predictive performance of six models with a more complex case. Since the true target function can vary greatly over different problems, we compare the models on a variety of randomly generated functions, by making use of the "random function generator" in Friedman (2001).

The "random function generator" generates the function in the form of a linear expansion of functions  $\{g_k\}_{k=1}^{20}$  :

$$F(\mathbf{x}) = \sum_{k=1}^{20} a_k g_k(\mathbf{z}_k)$$

The coefficients  $\{a_k\}_{k=1}^{20}$  are generated from a uniform distribution  $\text{Unif}(0, 1)$ . The variable  $\mathbf{z}_k$  is a  $m_k$ -sized subset of the  $p$ -input variables  $\mathbf{x}$  as

$$\mathbf{z}_k = \{x_{\omega(k)}\}_{k=1}^{m_k},$$

where  $\omega(k)$  is an independent random permutation of the integers  $\{1, 2, \dots, p\}$ . The size of each subset  $m_k$  is randomly selected as  $\lfloor 2.5 + r_k \rfloor$ , where  $r_k$  is generated from an exponential distribution with mean 2. Then, the expected number of input variables for each  $g_k(\mathbf{z}_k)$  is between four and five. Each function  $g_k(\mathbf{z}_k)$  is an  $m_k$ -dimensional Gaussian function

$$g_k(\mathbf{z}_k) = e^{-\frac{1}{2}(\mathbf{z}_k - \mathbf{u}_k)^T V_k (\mathbf{z}_k - \mathbf{u}_k)},$$

where each mean vector  $\mathbf{u}_k$  is generated from standard normal distribution  $N(0, \mathbf{I}_{m_k})$ . The  $m_k \times m_k$  covariance matrix  $V_k$  is defined by

$$V_k = U_k D_k U_k^T,$$

where  $U_k$  is a random orthonormal matrix,  $D_k = \text{diag}\{d_1^k, \dots, d_{m_k}^k\}$  and the square roots of the eigenvalues  $\{\sqrt{d_j^k}\}_{j=1}^{m_k}$  are generated from a uniform distribution  $\text{Unif}(0.1, 2)$ .

We set the number of predictors to be  $p = 10$  and generate the data  $\{n_i, s_i, \mathbf{x}_i, \mathbf{y}_i\}_{i=1}^d$  according to

$$n_i \sim \mathcal{P}(\lambda_i), s_i \sim \mathcal{G}(\mu_{n_i}, \delta), \mathbf{x}_i \sim N(0, \mathbf{I}_p), \mathbf{y}_i \sim N(0, \mathbf{I}_p), i = 1, \dots, n \quad (6.4)$$

where  $\lambda_i = 1.2e^{F_1(\mathbf{x}_i)}$ ,  $\mu_{n_i} = e^{\log(n_i+5)F_2(\mathbf{y}_i)}$ ,  $\delta = 2$ , and  $F_1(\mathbf{x}_i), F_2(\mathbf{y}_i)$  are the functions generated from the above "random function generator".

We also generate sample of size 10000 for training and another sample of equal size for testing. Table 2 lists the parameter estimation error on the testing sample, based on 30 independent replications. Figure 5.3 shows ten out-of-sample loss for claim frequency and severity. The results have no significant difference to the simple case: Dependent model still provides better predictive performance than the

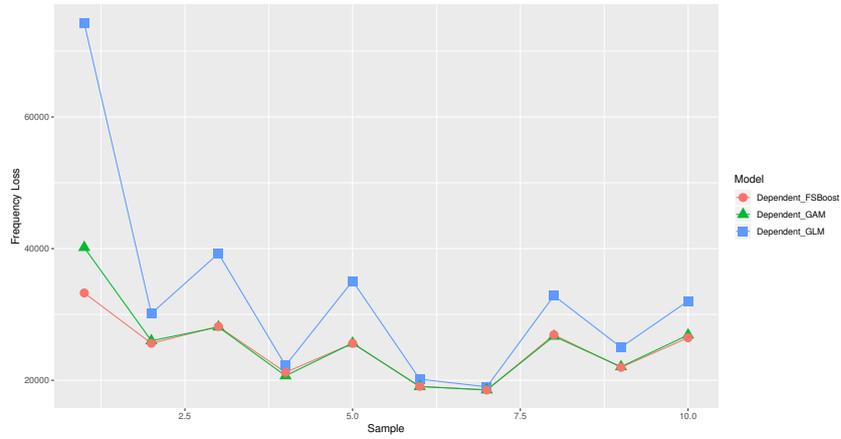
## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

---

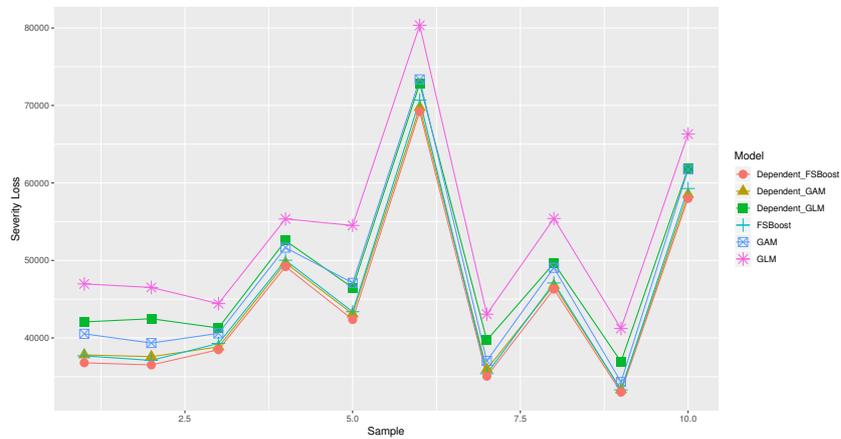
corresponding independent model, except the GLM case. The Dependent-FSBoost model has the best capacity in predicting claim frequency and severity.

In fact, there is no surprise that the dependent model performs better than the independent one since we correlate the claim frequency and severity in the data generation process. One may wonder that whether the dependent model keeps excellent predictive performance when no dependence exists between claim frequency and severity. In this case, the Dependent-FSBoost model keeps almost the same predictive performance as FSBoost model, due to the automatical feature selection of the data-driven model. We generate the sample according to the setting (5.4), except where we make  $\mu_{n_i} = e^{F_2(\mathbf{y}_i)}$ . Table 3 lists the parameter estimation error on the testing sample over 30 independent replications. Figure 5.4 shows ten out-of-sample loss for claim severity. The results confirm our conclusion. Surprisingly, the dependent-GLM and dependent-GAM models also keep the same prediction accuracy as the independent ones. Thus, we can safely use the dependent models even in some cases that there is no knowledge of the dependence between claim frequency and severity.

### 6.3. Simulation Study



(a)



(b)

Figure 6.3: The out-of-sample loss in the dependent case

**Table 2** The parameter estimation error in the dependent case

	$\lambda$	$\mu$
GLM	-	4.5489 (1.9179)
Dependent-GLM	0.5045 (0.1633)	5.5260 (4.8389)
GAM	-	3.9330 (4.1238)
Dependent-GAM	0.2355 (0.0467)	1.1572 (0.6273)
FSBoost	-	1.9463 (1.6482)
Dependent-FSBoost	0.2134 (0.0408)	0.5692 (0.1361)

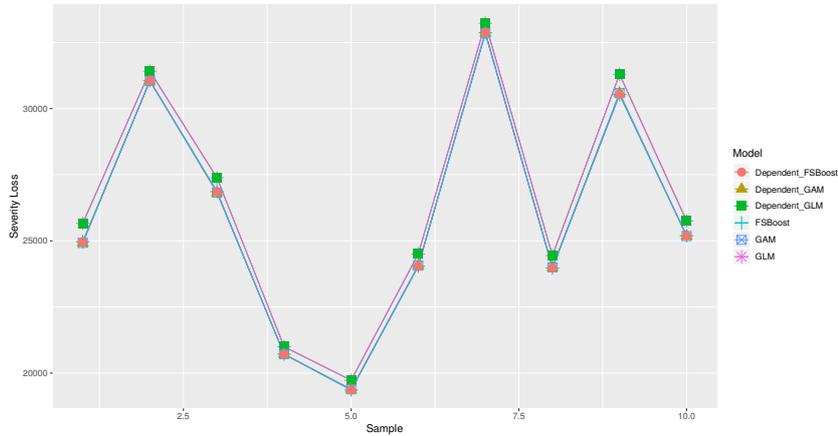


Figure 6.4: The out-of-sample loss in the independent case

Table 3 The parameter estimation error in the independent case

	$\mu$
GLM	0.4941 (0.1520)
Dependent-GLM	0.4944 (0.1519)
GAM	0.2338 (0.0355)
Dependent-GAM	0.2344 (0.0354)
FSBoost	0.2251 (0.0271)
Dependent-FSBoost	0.2263 (0.0264)

### 6.3.3 Examination of $\delta$

We can further confirm our results by comparing the estimation accuracy of dispersion parameter  $\delta$ , based on the knowledge that more accurate model gives more precise  $\delta$ . Table 4 reports estimation results of  $\delta$  for 30 sets of training samples, which are generated according to the setting (5.4). The results are consistent with our expectation: The dependent model provide closer estimates to the true value than the corresponding independent model, and the Dependent-FSBoost model makes the best estimate of  $\delta$ .

In our FSBoost and Dependent-FSBoost models, we make use of a rough grid method to find the optimal  $\delta$ . Thus, we need make robust check to see that if the misspecified dispersion parameter  $\delta$  has significant effect on the estimation accuracy of  $\mu$ . We expect that the estimation accuracy of  $\mu$  is insensitive to the value of  $\delta$ , thus ensuring the estimation accuracy of  $\mu$  even with not quite accurate  $\delta$ . We generate 20 sets of training samples according to the setting (5.4), each sample of size 10000. We fit the Dependent-FSBoost model with 11 values of  $\delta \in \{1.5, 1.6, \dots, 2.5\}$  and observe the difference of estimation error of  $\mu$ . Figure 5.5 displays the results. We can see that the value of  $\delta$  has no significant effect on estimation accuracy of  $\mu$ .

## 6.4. Application

**Table 4** The estimation of parameter  $\delta$

	$\delta$
GLM	16.4009 (5.2675)
Dependent-GLM	8.2895 (2.9613)
GAM	12.2253 (8.0950)
Dependent-GAM	4.1126 (1.0491)
FSBoost	2.7233 (0.2344)
Dependent-FSBoost	2.2933 (0.1388)

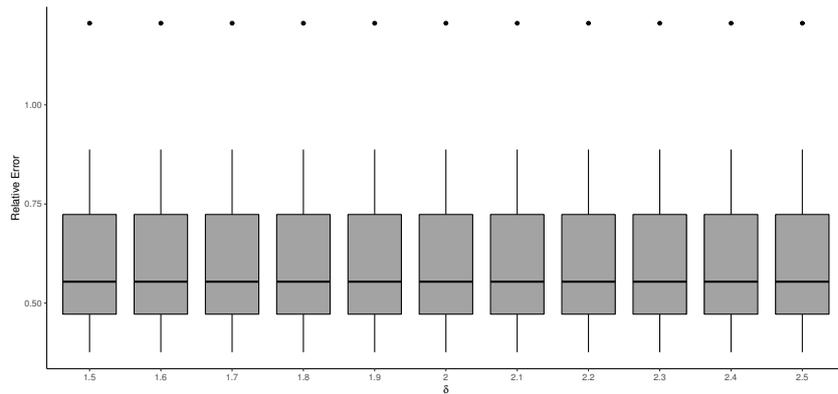


Figure 6.5: The distribution of estimation error of parameter  $\mu$  when varying parameter  $\delta$  from 1.5 to 2.5 in the Dependent-FSBoost model

## 6.4 Application

### 6.4.1 Data

We consider a French motor third-party liability dataset, where the data "freMTPL2freq" and "freMTPL2sev" are included in the R package "CASdatasets". We make data preprocess as in [Noll, Salzmann, and Wuthrich \(2018\)](#), except deleting the records that have positive claim counts but no claim amount, and partitioning different categorization levels on variables "VehAge" and "DrivAge". After data preprocess, the dataset contains 668897 records, each of which consists of claim counts, exposure, averaged claim severity, and 9 rating variables. Table 5 summaries the dataset. There are 24944 (3.73%) policies that have positive claim counts. Table 6 displays the claim frequency distribution and respective mean of averaged claim severity. Only several policies have claim counts larger than 3 and the mean of averaged severity shows an increasing trend when claim count ranges from 0 to 3.

We plot the average claim frequency and severity per French region in Figure 5.6. In the French map, more blue color denotes a higher value. We can see that the accidents occur most often in regions Champagne-Ardenne, Bourgogne, and Corse while serious accidents happen most frequently in regions Picardie and Centre. In Figure 5.7 and 5.8, the usage of old cars tend to incur more accidents and

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

Variable	Type	Description
ClaimNb	Numeric	Number of claims during the exposure period
Exposure	Numeric	The period of exposure in years
ClaimSev	Numeric	The average cost of the claim for a policy
Area	Categorical	The area code (1-6)
VehPower	Categorical	The power of the car (6 classes)
VehAge	Categorical	The vehicle age in years ([0,1],[1,4],[4,10],[10,∞))
DrivAge	Categorical	The driver age in years ([18,21],[21,25],[25,35],[35,45],[45,55],[55,70],[70,∞))
BonusMalus	Numeric	Bonus/malus (50-150)
VehBrand	Categorical	The car brand (B1-B14)
VehGas	Categorical	The car gas (Diesel or regular)
LogDensity	Numeric	The log-density of inhabitants in the driver's city (number of inhabitants per km <sup>2</sup> )
Region	Categorical	The policy region in France (22 classes)

	0	1	2	3	4	5	6	8	9	11	16
number of claims											
number of policies	643953	23570	1299	62	5	2	1	1	1	2	1
mean of averaged severity	0	2177.12	2932.36	4115.35	2203.49	3559.01	1608.93	3103.22	2039.41	1966.92	2220.59

higher claim payments. Young drivers drive more recklessly than middle-age and old drivers, leading to more crashes and more severe loss. The Pearson and Spearman correlation coefficients between frequency and severity are 0.0056 and 0.0313, where we only calculate the records with positive claim counts and which suggests a weak positive linear or monotonic relation. We adjust the frequency as the frequency/exposure and the coefficients change into 0.01115 and 0.1123, which implies a relatively stronger positive association. Note that there are interactions among rating factors. From Table 7, we can see that the effect of the vehicle age on the claim frequency are greater for the young drivers and the effect reduces when the driver age raises. The effect of the vehicle age on the claim severity are more significant for the young and old drivers and the effect is relatively weak for the middle-age drivers.

## 6.4. Application

---

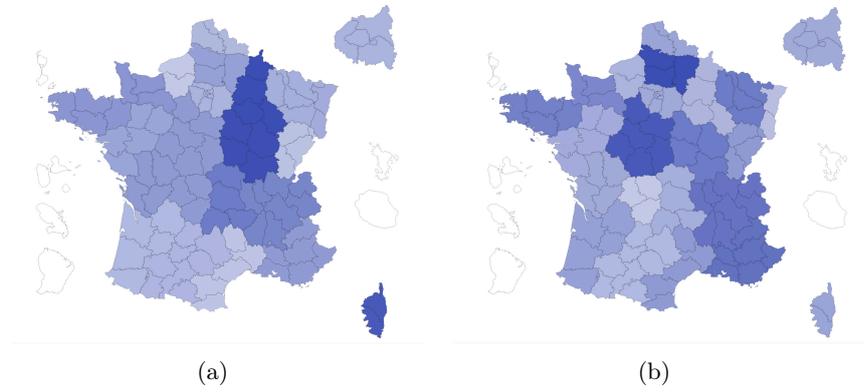


Figure 6.6: Average claim frequency and severity per region

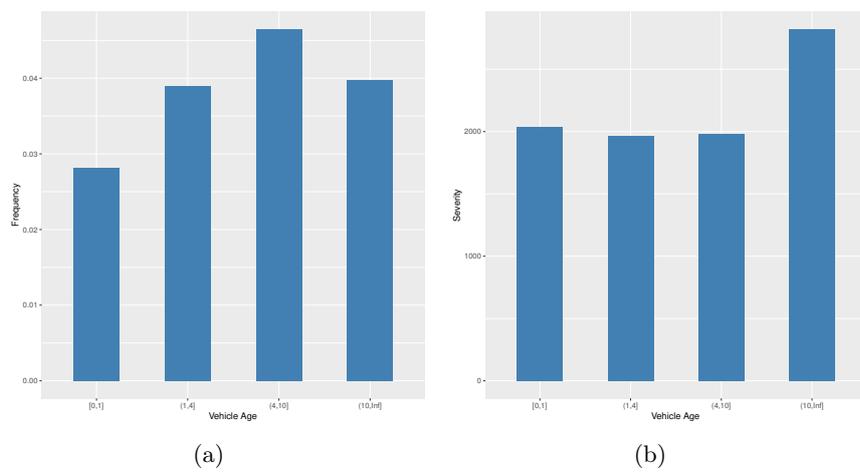


Figure 6.7: Histogram of claim frequency and severity per vehicle age group

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

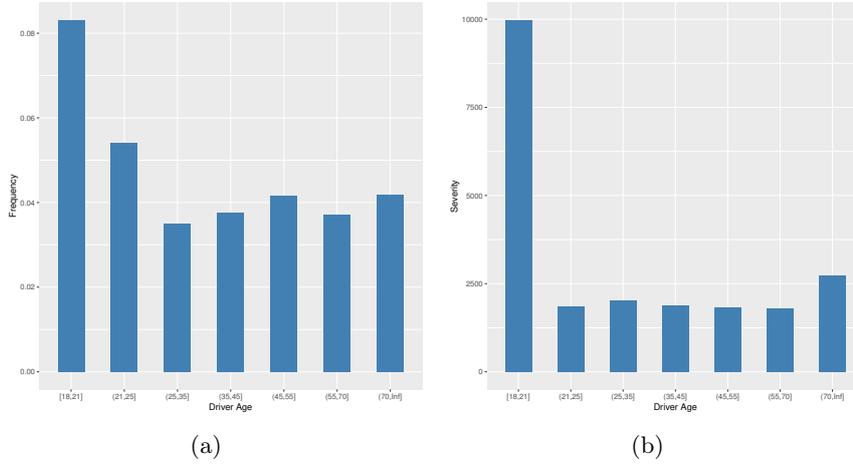


Figure 6.8: Histogram of claim frequency and severity per driver age group

**Table 7** The average frequency/exposure (severity) per group of vehicle age and driver age

	[18,21]	(21,25]	(25,35]	(35,45]	(45,55]	(55,70]	(70,∞)
[0,1]	0.3582 (11041.08)	0.1210 (3328.73)	0.0912 (1728.09)	0.1088 (1888.64)	0.0825 (1978.76)	0.0723 (1669.11)	0.0864 (1914.51)
(1,4]	0.3392 (2865.33)	0.1720 (1954.50)	0.1017 (2007.92)	0.1013 (1554.53)	0.1034 (1934.74)	0.0863 (2038.04)	0.0778 (3038.81)
(4,10]	0.3912 (1913.02)	0.2596 (1759.62)	0.1549 (1975.80)	0.1222 (2131.24)	0.1330 (1846.17)	0.1004 (1913.70)	0.0914 (2388.30)
(10,∞]	0.5243 (17734.18)	0.2355 (1461.27)	0.1197 (2287.19)	0.1143 (1847.24)	0.1398 (1676.51)	0.0911 (1477.60)	0.0828 (3359.60)

### 6.4.2 Model Comparison

We make use of the 2/3 dataset as a training set and the remaining 1/3 as a testing set, where frequency distributions are kept close to the whole set. Then, we fit six models GLM, Dependent-GLM, GAM, Dependent-GAM, FSBoost, Dependent-FSBoost with the training set. Recall that frequency/exposure is more correlated with the average severity. Thus, in the dependent models, we use the frequency/exposure instead of frequency as a predictor variable. Table 8 lists the out-of-sample loss for the frequency and severity parts. We can find that the dependent model is still more competitive than the corresponding independent model. The Dependent-FSBoost has the most excellent predictive performance.

**Table 8** The out-of-sample loss

	Frequency loss	Severity loss
GLM	-	82170.00
Dependent-GLM	37664.53	81999.60
GAM	-	81424.14
Dependent-GAM	34697.80	81305.76
FSBoost	-	78371.91
Dependent-FSBoost	34155.54	78355.77

## 6.4. Application

---

To make further examination of the predictive performance, we compute the pure premium prediction of each model on the testing set. In the dependent case, we calculate the expected pure premium  $E(S|x)$  by

$$E(S|\mathbf{x}) = E(NE(\bar{Y}|N, \mathbf{x})|\mathbf{x}) = E(Ne^{F_{\bar{Y}|N}(\mathbf{x}, N; \boldsymbol{\beta})}).$$

For the Dependent-GLM model, we use the closed-form formula (5) in [Garrido, Genest, and Schulz \(2016\)](#). For the Dependent-GAM and Dependent-FSBoost model, we exploit truncated sum of 1000 terms, in conjunction with the estimated function  $F_{\bar{Y}|N}(\mathbf{x}, N; \boldsymbol{\beta})$  and probability mass function  $f_N(n|\lambda)$  of  $N$ . In the independent case, the computation of  $E(S|x)$  reduces to

$$E(S|x) = E(N|x)E(\bar{Y}|x) = e^{F_N(\mathbf{x}; \boldsymbol{\alpha}) + F_{\bar{Y}}(\mathbf{x}; \boldsymbol{\beta})}.$$

Then, we compare these models by measuring the differences between predicted pure premium and real loss on the testing data. As high proportions of zeros exist in real losses, we compute the Pearson, Spearman and Kendall coefficients only for the records with positive losses. Table 9 summarizes these results. We can also see that the dependent models are more favored. The Dependent-FSBoost model outperforms other dependent models, in terms of the strongest correlation between the predicted pure premium and positive losses.

**Table 9**      **The correlation coefficients between predicted pure premium and real loss**

	Pearson	Kendall	Spearman
GLM	0.0002	0.0068	0.0098
Dependent-GLM	0.0071	0.0269	0.0392
GAM	0.0003	0.0825	0.1210
Dependent-GAM	0.0121	0.0907	0.1334
FSBoost	0.0007	0.0921	0.1348
Dependent-FSBoost	0.0357	0.0950	0.1388

### 6.4.3 Model Interpretation

In this section, we introduce two useful tools from [Friedman \(2001\)](#) for model interpretation, including variable importance and partial dependence plots. These two tools give the importance of each variable on predicting the frequency (severity) and also the specific dependence between the predictor variable and the frequency (severity).

#### 6.4.3.1 Variable Importance

Variable importance measures the importance of each variable on the prediction. For a single  $J$ -terminal node tree  $T$ , [Breiman, Friedman, Olshen, and Stone \(1984\)](#)

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

---

propose the following importance measure of variable  $x_l$

$$\hat{I}_{x_l}(T) = \sum_{j=1}^{J-1} \hat{\iota}_j^2 \mathbb{1}_{v_j}(x_l),$$

where the summation is over all  $J - 1$  internal nodes,  $v_j$  is the splitting variable associated with node  $j$ , and  $\hat{\iota}_j$  denotes the reduce of squared error as a result of the split. [Friedman \(2001\)](#) generalizes the variable importance measure to a collection of trees  $\{T_m\}_1^M$  in the gradient boosting algorithm, by taking average over all of the trees as

$$\hat{I}_{x_l} = \frac{1}{M} \sum_{m=1}^M \hat{I}_{x_l}(T_m). \quad (6.5)$$

The variable importance measure is biased as an independent variable  $x_l$  to response variable  $y$  can also be selected as a splitting variable and thus make the  $\hat{I}_{x_l}^2$  not be zero. [Sandri and Zuccolotto \(2008\)](#) and [Sandri and Zuccolotto \(2010\)](#) correct the bias by defining the following adjusted variable importance measure

$$\bar{I}_{x_l} = \frac{1}{S} \sum_{s=1}^S (\hat{I}_{x_l}^s - \hat{I}_{z_l}^s),$$

where  $x_l, z_l$  are the  $l^{\text{th}}$  variables of  $\bar{Y}, Z_s$ , respectively, and  $Z_s, s = 1, \dots, S$  are generated by randomly permuting  $S$  times the rows of original data  $\bar{Y}$ , and  $\hat{I}_{x_l}^s$  and  $\hat{I}_{z_l}^s$  are computed by the formula (5.5) on the data  $[\bar{Y}, Z_s]$ .

Figure 5.9 shows the relative importance of the 9 rating variables in predicting frequency and severity. We can find that the VehBrand and BonusMalus are two most important variables in the frequency part and the VehBrand dominates the overall prediction. For the severity part, the DrivAge, BonusMalus, LogDensity and VehPower are most influential and the DrivAge exerts dominant effect.

## 6.4. Application

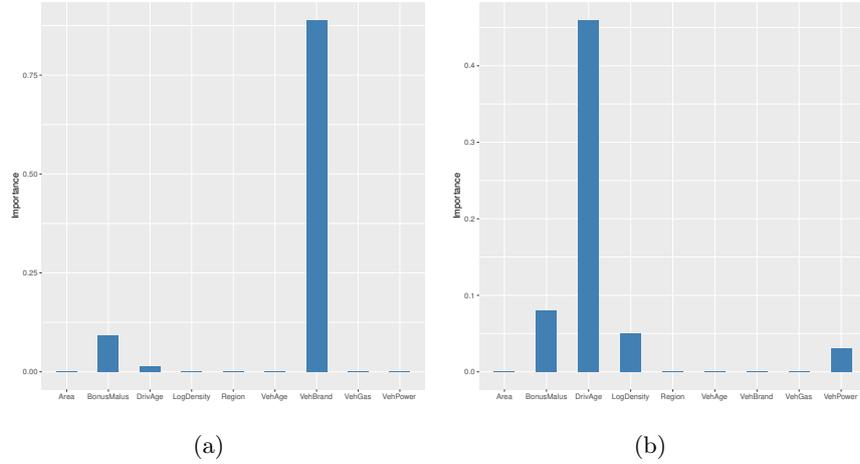


Figure 6.9: Variable importances of the 9 rating variables in the prediction of claim frequency and severity

### 6.4.3.2 Partial Dependence Plots

Partial dependence plots developed in [Friedman \(2001\)](#) is a powerful visualization tool to look into the main effect and interaction effects. Let  $\mathbf{z}_k$  be the target subset of the predictor variables  $\mathbf{x}$  and  $\mathbf{z}_{\setminus k}$  be the complement subset,

$$\mathbf{z}_k \cup \mathbf{z}_{\setminus k} = \mathbf{x}.$$

For main effect and second-order interaction effects, the size of the variable subset  $\mathbf{z}_k$  is one or two. Given the fitted function  $\hat{F}(\mathbf{x})$  and training data  $\{y_i, \mathbf{x}_i\}_{i=1}^N$ , the partial dependence of  $\hat{F}(\mathbf{x})$  on the chosen variable subset  $\mathbf{z}_k$  is computed as

$$\bar{F}(\mathbf{z}_k) = \frac{1}{N} \sum_{i=1}^N \hat{F}(\mathbf{z}_k, \mathbf{z}_{i,\setminus k}),$$

where  $\mathbf{z}_{i,\setminus k}$  is the particular values of the variable subset  $\mathbf{z}_{i,\setminus k}$  from the  $i^{\text{th}}$  record of training data  $\mathbf{x}_i$ . Then, the function  $\bar{F}(\mathbf{z}_k)$  is plotted against  $\mathbf{z}_k$ .

Figure 5.10 shows partial dependence plots, which indicate main effects of the two most important variables for frequency and severity parts. From the top two panels, we can see that the car brands B7-B9 induce a larger amount of accidents and the frequency is positively correlated with the bonus-malus level. The average occurrence of accidents increase quickly when the bonus-malus level is near to 60 and 100. The bonus-malus system penalizes the policyholders with one or more accidents by premium surcharges or malus and rewards claim-free policyholders by awarding a discount or bonus. In France, the bonus-malus level less than 100 means bonus and larger than 100 denotes malus. The change from bonus to malus represents the incidence of accidents, thus explaining the sudden change of frequency at bonus-malus level 100. The bonus-malus level 60 can be understood as the enough

## Chapter 6. Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims

bonus level that encourages policyholders to drive more carefully. The bottom two panels show that the young drivers induce more accidents and the severity exhibits dramatic increases at bonus-malus level 100. The latter change is also explained by the mechanism of bonus-malus system.

Figure 5.11 displays interaction effect of the BonusMalus and VehBrand variables on the frequency and of the BonusMalus and DrivAge variables on the severity. We can see that the change trend of frequency for each vehicle brand is analogous. However, the change of frequency for B7-B9 proceeds on a relatively high level. The severity case is similar, where the change pattern of severity for each driver age group is of no significant difference and the change of severity for young drivers is also kept at a relatively high level.

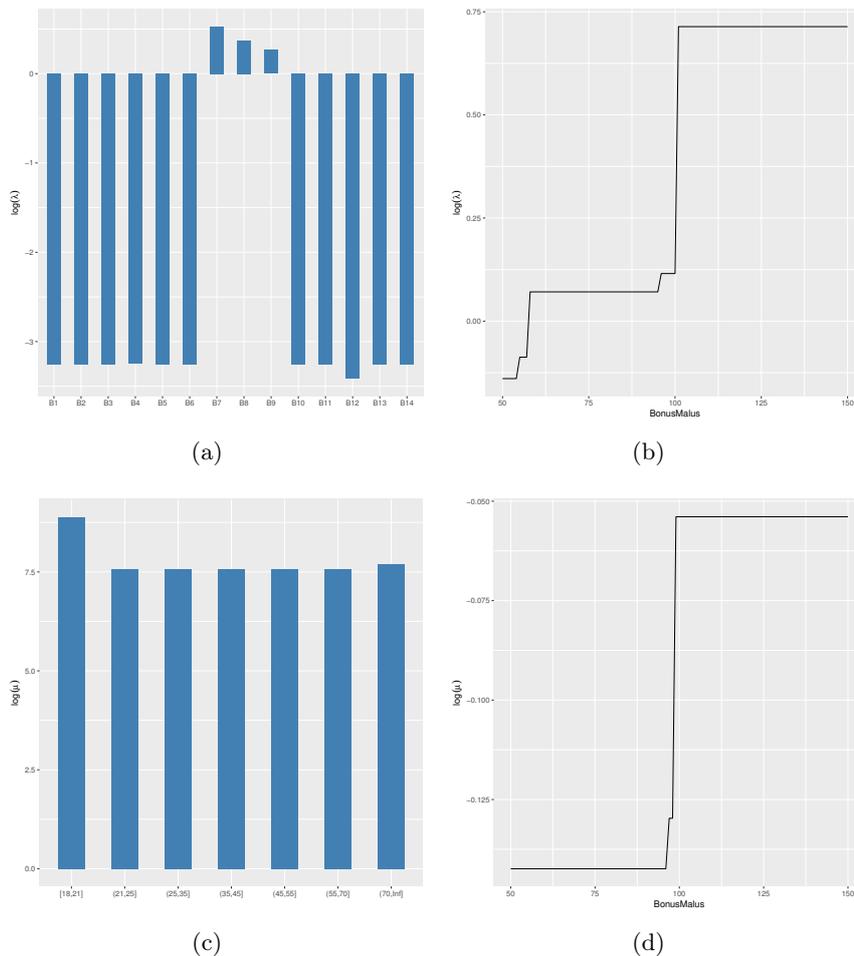


Figure 6.10: Partial dependence plots of the two most important variables in the frequency and severity parts

## 6.5. Conclusion

---

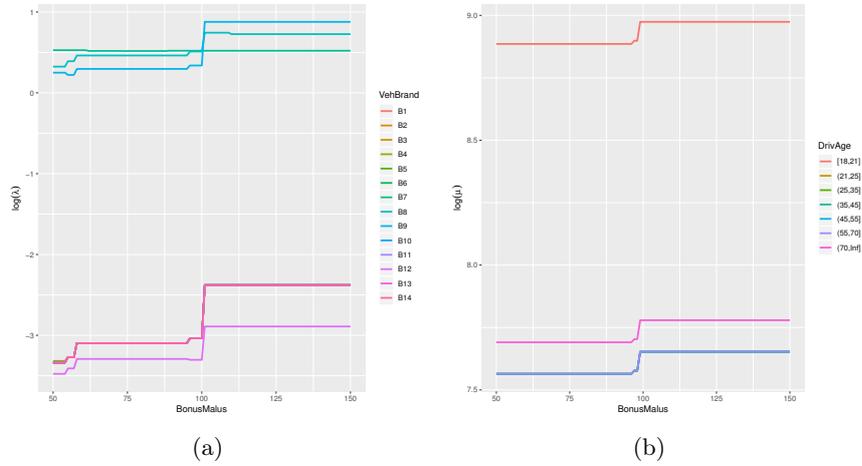


Figure 6.11: Interaction effect of the two most important variables in the frequency and severity parts

## 6.5 Conclusion

This paper develops a stochastic gradient boosting frequency-severity model and demonstrates its advantages over the currently popular frequency-severity models, mainly on the strong capacity of fitting a flexible nonlinear relation between claim frequency (severity) and predictors, and of capturing complex interactions among predictors and nonlinear dependence between claim frequency and severity.

Our model can also be extended to capture some other features of the claim data. For example, our model can accommodate with the hurdle and zero-inflated modeling framework to handle the overdispersion and zero inflation in claim counts. The model combined with the hurdle modeling framework in [Shi, Feng, and Ivantsova \(2015\)](#) can be one solution. Another solution is to inflate the gradient boosting Poisson model in the frequency part to the zero-inflated gradient boosting Poisson model. These extensions will be our future works.



# General conclusion

---

The design, pricing and hedging of insurance contracts is vital to profit-making and competency of insurance company, and also efficiency of insurance market. On this aspect, rapid advances of theoretical tools in some other fields can provide many new insights. This thesis makes use of some theoretical tools in financial engineering, decision theory, machine learning, to improve the design, pricing and hedging of insurance contracts. Chapter 3 develops closed-form pricing formulas for participating life insurance contracts, based on matrix Wiener-Hopf factorization, where multiple risk sources, such as credit, market, and economic risks, are considered. The pricing method proves to be accurate and efficient. We also introduce dynamic and semi-static hedging strategies to assist insurance company to reduce risk exposure arising from the issue of participating contracts. Chapter 4 discusses the optimal contract design when the insured is third degree risk averse. The results show that dual limited stop-loss, change-loss, dual change-loss, and stop-loss can be optimal contracts favored by both of risk averters and risk lovers in different settings. Chapter 5 develops a stochastic gradient boosting frequency-severity model, which improves the important and popular GLM and GAM frequency-severity models. This model fully inherits advantages of gradient boosting algorithm, overcoming the restrictive linear or additive forms of the GLM and GAM frequency-severity models, through learning the model structure from data. Meanwhile, our model can also capture the flexible nonlinear dependence between claim frequency and severity. In total, our studies provide some further insights on the insurance domain, by making use of the tools in other fields.

Chapter 3 incorporates credit, market (jump) and economic (regime switching) risks into the pricing of the participating contracts. These risks have been well documented. The default of the insurer is assumed to happen when the reference portfolio value falls below a proportion of minimum guarantee. The market and economic risks are characterized by making the reference portfolio evolve as a regime switching double exponential jump diffusion model. Based on matrix Wiener-Hopf factorization, we deduce closed-form pricing formulas for participating contracts up to Laplace or Laplace-Fourier transform. Then, we obtain the price by performing numerical Laplace and Fourier inversion and by implementing the matrix Wiener-Hopf factorization. We compare our pricing method with Monte Carlo simulation and show that our pricing method is accurate and efficient. Besides the pricing of standard contracts, we also design new participating contracts, whose minimum guaranteed rate is linked to market interest rates. This design reduces risk exposure

of the contracts to the fluctuations of market interest rate, which circumvents an issue of a dramatic narrowing in the safety margin when low interest rate persists for long. We show that our pricing method can also price the floating contracts, both accurately and efficiently. We compare the standard contracts with floating contracts and find that floating contracts transfer interest rate risk into other risk sources, thus leading floating contracts to bear higher default risk. However, the floating contracts are more worthy than standard contracts, which makes floating contracts be good products to attract risk-seeking investors. To hedge multiple risks in the participating contracts, we introduce dynamic and semi-static hedging strategies and show that both of them can significantly reduce risk. The dynamic hedging performs better in the hedging of extreme risk. Finally, we conclude that ignorance of economic risk makes the contract features only understood from a short-term viewpoint. We also point out that the interest rate term structure can be further modeled as a regime switching HJM model, and then between the dynamics of the investment funds and interest rates, a regime switching correlation structure can be incorporated.

Chapter 4 investigates the optimal insurance design when covering both of risk averters and risk lovers. Previous works only consider the insured who are risk averse. Though a majority of insured are risk averse, there is still a significant minority of them who are risk lovers. The prudence of risk averters and risk lovers have been well documented. Thus, we only assume the insured to be prudent (third degree risk averse), without more restrictions on the first and second order preferences, ensuring that both of the risk averters and risk lovers are covered. First, we show the optimal contract for risk averters and risk lovers is a stop-loss insurance and a limited full insurance, respectively. Then, when considering one contract favored by both of risk averters and risk lovers, we show that the optimal contract is a new type of contract, a dual limited stop-loss contract, by making use of some tools in the 3-cv order, an equivalent notion to third degree risk. Further, we restrict the contracts to concave types. In this case, we point out that the optimal contract changes into a dual change-loss insurance policy or a change-loss insurance policy, depending upon the coefficient of variation of the retained loss. Finally, we introduce background risk into the contract design and find that the optimal contract is a stop-loss insurance policy when background risk and insurable risk are comonotonic. We discuss the contract design under third degree risk in different settings. The further work is to discuss the contract design under higher order degree risk.

Chapter 5 improves the frequency-severity model that is widely used in non-life insurance. One major limitation of traditional GLM or GAM frequency-severity models is the restricted linear or additive form, which is too rigid and thus limits predictive capacity. The success of machine learning methods have promoted a variety of applications in various fields. Under this background, we develop a stochastic gradient boosting frequency-severity model (Dependent-FSBoost), which is a data-driven model inheriting all the advantages of boosting algorithm, capable of fitting a flexible non-linear relation between claim frequency (severity) and predictors, and capturing complex interactions among predictors. The dependence structure be-

---

tween the claim frequency and severity is well documented. We incorporate such a dependence structure by treating frequency as a predictor in the gradient boosting regression model for average claim severity. As a byproduct of the data-driven model, the nonlinear dependence between claim frequency and severity can be fully captured. The model can use all the members of exponential dispersion family and we illustrate with the gamma and Poisson for the claim frequency and severity part, respectively. In our model, we use grid search to find the optimal dispersion parameter of gamma model and adopt cross-validation method to select the tuning parameters. We exploit two simulation studies to compare our model with GLM and GAM frequency-severity models, each model with the cases of dependence and independence between claim frequency and severity. Within simple simulation study, we demonstrate the capacity of the Dependent-FSBoost model in capturing nonlinear effect, complex interactions and nonlinear dependence between claim frequency and severity. In both of two simulation studies, we consider the out-of-sample loss and relative error of parameter estimation as model performance criterion. We find that dependent models perform better than the independent ones and the Dependent-FSBoost model has the best and robust predictive performance. Further, we also show that the Dependent-FSBoost model keeps the robust and accurate prediction when no dependence exists between claim frequency and severity, and also reveal that prediction accuracy of the model is insensitive to the dispersion parameter of gamma model. Then, we apply our model to analyze a real dataset, a French motor third-party liability dataset, which has complex interactions among rating variables. We show that the Dependent-FSBoost model has the least out-of-sample loss. We also compare six models in the prediction of pure premium and find that dependent models are still more favored and the Dependent-FSBoost model outperforms other dependent models, in terms of the strongest correlation between the predicted pure premium and real loss. Finally, we introduce two useful tools, variable importance and partial dependence plots, to explain our model, which give the importance of each variable on predicting the frequency (severity) and also the specific dependence between the predictor variable and the frequency (severity). We conclude that our data-driven model is superior to other state-of-the-art models. The further work is to accommodate our model with the hurdle and zero-inflated modeling framework to handle the overdispersion and zero inflation in claim counts.



# Bibliography

- ANDRÉ, E., O. LE COURTOIS, AND X. SU (2018): “Optimal Insurance under Third Degree Risk,” *Working Paper*. (Cited on pages 15 and 29.)
- ARROW, K. J. (1963): “Uncertainty and the welfare economics of medical care,” *The American economic review*, 53(5), 941–973. (Cited on pages 12, 27, 85, 86, 88, 89 and 91.)
- ASMUSSEN, S., F. AVRAM, AND M. R. PISTORIUS (2004): “Russian and American put options under exponential phase-type Lévy models,” *Stochastic Processes and their Applications*, 109(1), 79–111. (Cited on pages 11, 26 and 57.)
- BACINELLO, A. R. (2001): “Fair pricing of life insurance participating policies with a minimum interest rate guaranteed,” *ASTIN Bulletin: The Journal of the IAA*, 31(2), 275–297. (Cited on page 56.)
- (2003a): “Fair valuation of a guaranteed life insurance participating contract embedding a surrender option,” *Journal of risk and insurance*, 70(3), 461–487. (Cited on page 56.)
- (2003b): “Pricing guaranteed life insurance participating policies with annual premiums and surrender option,” *North American Actuarial Journal*, 7(3), 1–17. (Cited on page 56.)
- BAKSHI, G., C. CAO, AND Z. CHEN (1997): “Empirical performance of alternative option pricing models,” *The Journal of finance*, 52(5), 2003–2049. (Cited on pages 11 and 26.)
- BALL, C. A., AND W. N. TOROUS (1983): “A simplified jump process for common stock returns,” *Journal of Financial and Quantitative analysis*, 18(1), 53–65. (Cited on pages 10 and 26.)
- BALLOTTA, L. (2005): “A Lévy process-based framework for the fair valuation of participating life insurance contracts,” *Insurance: Mathematics and Economics*, 37(2), 173–196. (Cited on page 57.)
- BARLOW, M. T., L. ROGERS, AND D. WILLIAMS (1980): “Wiener-Hopf factorization for matrices,” in *Séminaire de Probabilités XIV 1978/79*, pp. 324–331. Springer. (Cited on page 58.)
- BERNARD, C., X. HE, J.-A. YAN, AND X. Y. ZHOU (2015): “Optimal insurance design under rank-dependent expected utility,” *Mathematical Finance*, 25(1), 154–186. (Cited on pages 27 and 86.)

- BERNARD, C., O. LE COURTOIS, AND F. QUITTARD-PINON (2005): “Market value of life insurance contracts under stochastic interest rates and default risk,” *Insurance: Mathematics and Economics*, 36(3), 499–516. (Cited on pages 7, 22, 56 and 57.)
- (2006): “Development and pricing of a new participating contract,” *North American Actuarial Journal*, 10(4), 179–195. (Cited on pages 56, 57 and 73.)
- BERNARD, C., AND W. TIAN (2009): “Optimal reinsurance arrangements under tail risk measures,” *Journal of risk and insurance*, 76(3), 709–725. (Cited on pages 27 and 86.)
- BOLLEN, N., S. GRAY, AND R. WHALEY (2000): “Regime switching in foreign exchange rates: Evidence from currency option prices,” *Journal of Econometrics*, 94, 239–276. (Cited on pages 11 and 26.)
- BORCK, K. (1960): *An attempt to determine the optimum amount of stop loss reinsurance*. (Cited on pages 85, 88 and 91.)
- BOYLE, P. P., AND E. S. SCHWARTZ (1977): “Equilibrium prices of guarantees under equity-linked contracts,” *Journal of Risk and Insurance*, pp. 639–660. (Cited on pages 6, 22 and 56.)
- BREIMAN, L. (1999): “Prediction games and arcing algorithms,” *Neural computation*, 11(7), 1493–1517. (Cited on page 105.)
- BREIMAN, L., ET AL. (1998): “Arcing classifier (with discussion and a rejoinder by the author),” *The annals of statistics*, 26(3), 801–849. (Cited on page 105.)
- BREIMAN, L., J. FRIEDMAN, R. OLSHEN, AND C. STONE (1984): “Classification and Regression Trees,” . (Cited on page 125.)
- BRENNAN, M. J., AND E. S. SCHWARTZ (1979): “Alternative investment strategies for the issuers of equity linked life insurance policies with an asset value guarantee,” *Journal of Business*, pp. 63–93. (Cited on pages 6, 22 and 56.)
- BRENNAN, M. J., E. S. SCHWARTZ, ET AL. (1976): “The pricing of equity-linked life insurance policies with an asset value guarantee,” *Journal of Financial Economics*, 3(3), 195–213. (Cited on pages 6, 22 and 56.)
- BRIYS, E., AND F. DE VARENNE (1994): “Life insurance in a contingent claim framework: pricing and regulatory implications,” *The Geneva Papers on Risk and Insurance Theory*, 19(1), 53–72. (Cited on pages 9, 24 and 56.)
- BRIYS, E., AND F. DE VARENNE (1997): “On the risk of insurance liabilities: debunking some common pitfalls,” *Journal of Risk and Insurance*, pp. 673–694. (Cited on pages 9, 24 and 56.)

## Bibliography

---

- BUFFINGTON, J., AND R. J. ELLIOTT (2002): “American options with regime switching,” *International Journal of Theoretical and Applied Finance*, 5(05), 497–514. (Cited on pages 11 and 26.)
- BÜHLMANN, P., T. HOTHORN, ET AL. (2007): “Boosting algorithms: Regularization, prediction and model fitting,” *Statistical Science*, 22(4), 477–505. (Cited on page 105.)
- CAI, J., AND K. S. TAN (2007): “Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures,” *ASTIN Bulletin: The Journal of the IAA*, 37(1), 93–112. (Cited on pages 12, 27 and 85.)
- CAI, J., K. S. TAN, C. WENG, AND Y. ZHANG (2008): “Optimal reinsurance under VaR and CTE risk measures,” *Insurance: Mathematics and Economics*, 43(1), 185–196. (Cited on pages 12, 27, 85, 86 and 88.)
- CARIBONI, J., AND W. SCHOUTENS (2007): “Pricing credit default swaps under Lévy models,” *Journal of Computational Finance*, 10(4), 71. (Cited on pages 11 and 26.)
- CARR, P., AND L. WU (2003): “The finite moment log stable process and option pricing,” *The journal of finance*, 58(2), 753–777. (Cited on pages 11 and 26.)
- CHENG, C., AND J. LI (2018): “Early default risk and surrender risk: Impacts on participating life insurance policies,” *Insurance: Mathematics and Economics*, 78, 30–43. (Cited on page 56.)
- CHI, Y. (2017): “Insurance choice under third degree stochastic dominance,” *Insurance: Mathematics and Economics*. (Cited on pages 12, 27, 87, 89, 95, 97 and 98.)
- CHI, Y., AND X. S. LIN (2014): “Optimal reinsurance with limited ceded risk: A stochastic dominance approach,” *ASTIN Bulletin: The Journal of the IAA*, 44(1), 103–126. (Cited on page 86.)
- CHI, Y., AND K. S. TAN (2011): “Optimal reinsurance under VaR and CVaR risk measures: a simplified approach,” *ASTIN Bulletin: The Journal of the IAA*, 41(2), 487–509. (Cited on pages 12, 27, 86, 87 and 95.)
- CHI, Y., AND W. WEI (2018): “Optimum insurance contracts with background risk and higher-order risk attitudes,” *ASTIN Bulletin: The Journal of the IAA*, 48(3), 1025–1047. (Cited on pages 12, 27 and 87.)
- CHI, Y., AND M. ZHOU (2017): “Optimal reinsurance design: A mean-variance approach,” *North American Actuarial Journal*, 21(1), 1–14. (Cited on pages 12, 27, 85, 86 and 89.)

- CHOLLETE, L., A. HEINEN, AND A. VALDESOGO (2009): “Modeling international financial returns with a multivariate regime-switching copula,” *Journal of financial econometrics*, 7(4), 437–480. (Cited on pages 11 and 26.)
- COLEMAN, T., Y. KIM, Y. LI, AND M. PATRON (2007): “Robustly hedging variable annuities with guarantees under jump and volatility risks,” *Journal of Risk and Insurance*, 74(2), 347–376. (Cited on pages 7 and 22.)
- COLEMAN, T. F., Y. LI, AND M.-C. PATRON (2006): “Hedging guarantees in variable annuities under both equity and interest rate risks,” *Insurance: Mathematics and Economics*, 38(2), 215–228. (Cited on pages 7 and 22.)
- CONT, R. (2001): “Empirical properties of asset returns: stylized facts and statistical issues,” . (Cited on pages 10 and 26.)
- CONT, R., AND E. VOLTCHKOVA (2005a): “A finite difference scheme for option pricing in jump diffusion and exponential Lévy models,” *SIAM Journal on Numerical Analysis*, 43(4), 1596–1626. (Cited on pages 11 and 26.)
- (2005b): “Integro-differential equations for option prices in exponential Lévy models,” *Finance and Stochastics*, 9(3), 299–325. (Cited on pages 11 and 26.)
- CRAINICH, D., L. EECKHOUDT, AND A. TRANNOY (2013): “Even (mixed) risk lovers are prudent,” *American Economic Review*, 103(4), 1529–35. (Cited on pages 12, 27, 87 and 90.)
- CUI, Z., J. L. KIRKBY, AND D. NGUYEN (2017): “Equity-linked annuity pricing with cliquet-style guarantees in regime-switching and stochastic volatility models with jumps,” *Insurance: Mathematics and Economics*, 74, 46–62. (Cited on page 57.)
- CUMMINS, J. D., AND O. MAHUL (2004): “The demand for insurance with an upper limit on coverage,” *Journal of Risk and Insurance*, 71(2), 253–264. (Cited on pages 12, 27 and 85.)
- CZADO, C., R. KASTENMEIER, E. C. BRECHMANN, AND A. MIN (2012): “A mixed copula model for insurance claims and claim sizes,” *Scandinavian Actuarial Journal*, 2012(4), 278–305. (Cited on pages 13, 28 and 104.)
- DANA, R.-A., AND M. SCARSINI (2007): “Optimal risk sharing with background risk,” *Journal of Economic Theory*, 133(1), 152–176. (Cited on page 86.)
- DECK, C., AND H. SCHLESINGER (2014): “Consistency of higher order risk preferences,” *Econometrica*, 82(5), 1913–1943. (Cited on pages 12, 27, 87 and 90.)
- DOHERTY, N. A., AND H. SCHLESINGER (1983): “The optimal deductible for an insurance policy when initial wealth is random,” *Journal of Business*, pp. 555–565. (Cited on page 86.)

## Bibliography

---

- DUFFIE, D., J. PAN, AND K. SINGLETON (2000): “Transform analysis and asset pricing for affine jump-diffusions,” *Econometrica*, 68(6), 1343–1376. (Cited on pages 11 and 26.)
- EKERN, S. (1980): “Increasing Nth degree risk,” *Economics Letters*, 6(4), 329–333. (Cited on pages 47, 48, 87 and 90.)
- ELLIOTT, R. J., L. CHAN, AND T. K. SIU (2005): “Option pricing and Esscher transform under regime switching,” *Annals of Finance*, 1(4), 423–432. (Cited on pages 11 and 26.)
- ELLIOTT, R. J., AND T. K. SIU (2016): “Pricing regime-switching risk in an HJM interest rate environment,” *Quantitative Finance*, 16(12), 1791–1800. (Cited on page 81.)
- ENGEL, C., AND J. D. HAMILTON (1990): “Long swings in the dollar: Are they in the data and do markets know it?,” *The American Economic Review*, pp. 689–713. (Cited on pages 11 and 26.)
- ERHARDT, V., AND C. CZADO (2012): “Modeling dependent yearly claim totals including zero claims in private health insurance,” *Scandinavian Actuarial Journal*, 2012(2), 106–129. (Cited on pages 13, 28 and 104.)
- FAN, K., Y. SHEN, T. K. SIU, AND R. WANG (2015): “Pricing annuity guarantees under a double regime-switching model,” *Insurance: Mathematics and Economics*, 62, 62–78. (Cited on page 57.)
- FARD, F. A., AND T. K. SIU (2013): “Pricing participating products with Markov-modulated jump–diffusion process: An efficient numerical PIDE approach,” *Insurance: Mathematics and Economics*, 53(3), 712–721. (Cited on page 57.)
- FREES, E. W., J. GAO, AND M. A. ROSENBERG (2011): “Predicting the frequency and amount of health care expenditures,” *North American Actuarial Journal*, 15(3), 377–392. (Cited on pages 13, 28 and 104.)
- FREES, E. W., P. SHI, AND E. A. VALDEZ (2009): “Actuarial applications of a hierarchical insurance claims model,” *ASTIN Bulletin: The Journal of the IAA*, 39(1), 165–197. (Cited on page 104.)
- FREUND, Y., AND R. E. SCHAPIRE (1997): “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of computer and system sciences*, 55(1), 119–139. (Cited on page 105.)
- FRIEDMAN, J., T. HASTIE, AND R. TIBSHIRANI (2001): *The elements of statistical learning*, vol. 1. Springer series in statistics New York. (Cited on page 105.)
- FRIEDMAN, J., T. HASTIE, R. TIBSHIRANI, ET AL. (2000): “Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors),” *The annals of statistics*, 28(2), 337–407. (Cited on page 105.)

- FRIEDMAN, J. H. (2001): “Greedy function approximation: a gradient boosting machine,” *Annals of statistics*, pp. 1189–1232. (Cited on pages 53, 105, 108, 110, 111, 117, 125, 126 and 127.)
- (2002): “Stochastic gradient boosting,” *Computational statistics & data analysis*, 38(4), 367–378. (Cited on pages 105, 107 and 108.)
- FUSAI, G., G. GERMANO, AND D. MARAZZINA (2016): “Spitzer identity, Wiener-Hopf factorization and pricing of discretely monitored exotic options,” *European Journal of Operational Research*, 251(1), 124–134. (Cited on page 57.)
- FUSAI, G., AND A. MEUCCI (2008): “Pricing discretely monitored Asian options under Lévy processes,” *Journal of banking & finance*, 32(10), 2076–2088. (Cited on pages 11 and 26.)
- GARRIDO, J., C. GENEST, AND J. SCHULZ (2016): “Generalized linear models for dependent frequency and severity of insurance claims,” *Insurance: Mathematics and Economics*, 70, 205–215. (Cited on pages 104 and 125.)
- GOLLIER, C. (1996): “Optimum insurance of approximate losses,” *Journal of Risk and Insurance*, pp. 369–380. (Cited on pages 86 and 99.)
- GOLLIER, C., AND H. SCHLESINGER (1996): “Arrow’s theorem on the optimality of deductibles: a stochastic dominance approach,” *Economic Theory*, 7(2), 359–363. (Cited on pages 27 and 86.)
- GREEN, R., G. FUSAI, AND I. ABRAHAMS (2010): “The wiener-hopf technique and discretely monitored path-dependent option pricing,” *Mathematical Finance*, 20(2), 259–288. (Cited on page 57.)
- GROSEN, A., AND P. L. JØRGENSEN (2000): “Fair valuation of life insurance liabilities: the impact of interest rate guarantees, surrender options, and bonus policies,” *Insurance: Mathematics and Economics*, 26(1), 37–57. (Cited on pages 9, 24 and 56.)
- (2002): “Life insurance liabilities at market value: an analysis of insolvency risk, bonus policy, and regulatory intervention rules in a barrier option framework,” *Journal of risk and insurance*, 69(1), 63–91. (Cited on pages 9, 24, 56 and 57.)
- GSCHLÖSSL, S., AND C. CZADO (2007): “Spatial modelling of claim frequency and claim size in non-life insurance,” *Scandinavian Actuarial Journal*, 2007(3), 202–225. (Cited on pages 13, 28 and 104.)
- GUIDOLIN, M., AND A. TIMMERMANN (2005): “Economic implications of bull and bear regimes in UK stock and bond returns,” *The Economic Journal*, 115(500), 111–143. (Cited on pages 11 and 26.)

## Bibliography

---

- (2007): “Asset allocation under multivariate regime switching,” *Journal of Economic Dynamics and Control*, 31(11), 3503–3544. (Cited on pages 11 and 26.)
- (2008): “International asset allocation under regime switching, skew, and kurtosis preferences,” *Review of financial studies*, 21(2), 889–935. (Cited on pages 11 and 26.)
- GUO, X. (2001): “An explicit solution to an optimal stopping problem with regime switching,” *Journal of Applied Probability*, 38(2), 464–481. (Cited on pages 11 and 26.)
- HADAR, J., AND W. R. RUSSELL (1969): “Rules for ordering uncertain prospects,” *The American economic review*, 59(1), 25–34. (Cited on page 87.)
- HAMILTON, J. D. (1989): “A new approach to the economic analysis of nonstationary time series and the business cycle,” *Econometrica: Journal of the Econometric Society*, pp. 357–384. (Cited on pages 11 and 26.)
- HANOCH, G., AND H. LEVY (1969): “The Efficiency Analysis of Choices Involving Risk,” *The Review of Economic Studies*, 36(3), 335–346. (Cited on page 87.)
- HARDY, M. R. (2001): “A regime-switching model of long-term stock returns,” *North American Actuarial Journal*, 5(2), 41–53. (Cited on pages 11 and 26.)
- HASTIE, T., AND R. TIBSHIRANI (1986): “Generalized Additive Models,” *Statistical Science*, 1(3), 297–318. (Cited on page 104.)
- HE, C., J. S. KENNEDY, T. F. COLEMAN, P. A. FORSYTH, Y. LI, AND K. R. VETZAL (2006): “Calibration and hedging under jump diffusion,” *Review of Derivatives Research*, 9(1), 1–35. (Cited on pages 77 and 78.)
- HIEBER, P. (2017): “Cliquet-style return guarantees in a regime switching Lévy model,” *Insurance: Mathematics and Economics*, 72, 138–147. (Cited on page 57.)
- HUBERMAN, G., D. MAYERS, AND C. W. SMITH JR (1983): “Optimal insurance policy indemnity schedules,” *The Bell Journal of Economics*, pp. 415–426. (Cited on page 87.)
- JARROW, R. A., AND E. R. ROSENFELD (1984): “Jump risks and the intertemporal capital asset pricing model,” *Journal of Business*, pp. 337–351. (Cited on pages 10 and 26.)
- JEAN, W. H. (1980): “The geometric mean and stochastic dominance,” *The Journal of Finance*, 35(1), 151–158. (Cited on pages 48 and 87.)
- JENSEN, B., P. L. JØRGENSEN, AND A. GROSEN (2001): “A finite difference approach to the valuation of path dependent life insurance liabilities,” *The Geneva Papers on Risk and Insurance Theory*, 26(1), 57–84. (Cited on page 56.)

- JIANG, Z., AND M. R. PISTORIUS (2008): “On perpetual American put valuation and first-passage in a regime-switching model with jumps,” *Finance and Stochastics*, 12(3), 331–355. (Cited on pages 57, 58, 62 and 64.)
- JOBERT, A., AND L. C. ROGERS (2006): “Option pricing with Markov-modulated dynamics,” *SIAM Journal on Control and Optimization*, 44(6), 2063–2078. (Cited on page 57.)
- JORION, P. (1988): “On jump processes in the foreign exchange and stock markets,” *The Review of Financial Studies*, 1(4), 427–445. (Cited on pages 10 and 26.)
- KALUSZKA, M. (2001): “Optimal reinsurance under mean-variance premium principles,” *Insurance: Mathematics and Economics*, 28(1), 61–67. (Cited on pages 27 and 85.)
- (2005): “Optimal reinsurance under convex principles of premium calculation,” *Insurance: Mathematics and Economics*, 36(3), 375–398. (Cited on pages 27, 85 and 86.)
- KALUSZKA, M., AND A. OKOLEWSKI (2008): “An extension of Arrow’s result on optimal reinsurance contract,” *Journal of Risk and Insurance*, 75(2), 275–288. (Cited on pages 27 and 85.)
- KASSBERGER, S., R. KIESEL, AND T. LIEBMANN (2008): “Fair valuation of insurance contracts under Lévy process specifications,” *Insurance: Mathematics and Economics*, 42(1), 419–433. (Cited on page 57.)
- KÉLANI, A., AND F. QUITTARD-PINON (2017): “Pricing and hedging variable annuities in a Lévy market: a risk management perspective,” *Journal of Risk and Insurance*, 84(1), 209–238. (Cited on pages 7 and 22.)
- KIMBALL, M. S. (1993): “Standard risk aversion,” *Econometrica: Journal of the Econometric Society*, pp. 589–611. (Cited on page 86.)
- KIMBALL, M. S., ET AL. (1990): “Precautionary Saving in the Small and in the Large,” *Econometrica*, 58(1), 53–73. (Cited on page 86.)
- KLAR, B. (2002): “A note on the L-class of life distributions,” *Journal of Applied Probability*, 39(1), 11–11. (Cited on page 92.)
- KOU, S. G. (2002): “A jump-diffusion model for option pricing,” *Management science*, 48(8), 1086–1101. (Cited on pages 11 and 26.)
- KOU, S. G., AND H. WANG (2003): “First passage times of a jump diffusion process,” *Advances in applied probability*, 35(2), 504–531. (Cited on page 70.)
- (2004): “Option pricing under a double exponential jump diffusion model,” *Management science*, 50(9), 1178–1192. (Cited on pages 11 and 26.)

## Bibliography

---

- KRÄMER, N., E. C. BRECHMANN, D. SILVESTRINI, AND C. CZADO (2013): “Total loss estimation using copula-based regression models,” *Insurance: Mathematics and Economics*, 53(3), 829–839. (Cited on pages 13, 28 and 104.)
- KRIEGLER, B., AND R. BERK (2010): “Small area estimation of the homeless in Los Angeles: An application of cost-sensitive stochastic gradient boosting,” *The Annals of Applied Statistics*, pp. 1234–1255. (Cited on page 105.)
- LE COURTOIS, O., AND H. NAKAGAWA (2013): “On surrender and default risks,” *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(1), 143–168. (Cited on page 56.)
- LE COURTOIS, O., AND F. QUITTARD-PINON (2008): “Fair Valuation of Participating Life Insurance Contracts with Jump Risk,” *Geneva Risk and Insurance Review*, 33(2), 106–136. (Cited on page 57.)
- LE COURTOIS, O., F. QUITTARD PINON, AND X. SU (2018): “Pricing and Hedging Defaultable Participating Contracts with Regime Switching and Jump Risk,” *Working Paper*. (Cited on pages 14 and 29.)
- LE COURTOIS, O., AND X. SU (2018): “Structural Pricing of CoCos and Deposit Insurance with Regime Switching and Jumps,” *Working Paper*. (Cited on pages 57, 58, 62, 64 and 65.)
- LIN, X. S., K. S. TAN, AND H. YANG (2009): “Pricing annuity guarantees under a regime-switching model,” *North American Actuarial Journal*, 13(3), 316–332. (Cited on pages 7, 22 and 57.)
- LONDON, R., H. MCKEAN, L. ROGERS, AND D. WILLIAMS (1982a): “A martingale approach to some Wiener-Hopf problems, I,” in *Séminaire de Probabilités XVI 1980/81*, pp. 41–67. Springer. (Cited on page 58.)
- LONDON, R., H. P. MCKEAN, L. ROGERS, AND D. WILLIAMS (1982b): “A martingale approach to some Wiener-Hopf problems, II,” in *Séminaire de Probabilités XVI 1980/81*, pp. 68–90. Springer. (Cited on page 58.)
- LU, Z., L. LIU, AND S. MENG (2013): “Optimal reinsurance with concave ceded loss functions under VaR and CTE risk measures,” *Insurance: Mathematics and Economics*, 52(1), 46–51. (Cited on pages 86 and 88.)
- MENEZES, C., C. GEISS, AND J. TRESSLER (1980): “Increasing downside risk,” *The American Economic Review*, 70(5), 921–932. (Cited on page 90.)
- MERTON, R. C. (1976): “Option pricing when underlying stock returns are discontinuous,” *Journal of financial economics*, 3(1-2), 125–144. (Cited on pages 10, 11 and 26.)

- MIJATOVIĆ, A., AND M. PISTORIUS (2011): “Exotic derivatives under stochastic volatility models with jumps,” in *Advanced Mathematical Methods for Finance*, pp. 455–508. Springer. (Cited on pages 58 and 66.)
- MILTERSEN, K. R., AND S.-A. PERSSON (2003): “Guaranteed investment contracts: Distributed and undistributed excess return,” *Scandinavian Actuarial Journal*, 2003(4), 257–279. (Cited on page 56.)
- MÜLLER, A., AND D. STOYAN (2002): *Comparison methods for stochastic models and risks*, vol. 389. Wiley New York. (Cited on page 100.)
- NAIK, V., AND M. LEE (1990): “General equilibrium pricing of options on the market portfolio with discontinuous returns,” *The Review of Financial Studies*, 3(4), 493–521. (Cited on pages 11 and 26.)
- NOLL, A., R. SALZMANN, AND M. V. WUTHRICH (2018): “Case Study: French Motor Third-Party Liability Claims,” . (Cited on pages 8, 23, 105 and 121.)
- NOUSSAIR, C. N., S. T. TRAUTMANN, AND G. VAN DE KUILEN (2013): “Higher order risk attitudes, demographics, and financial decisions,” *Review of Economic Studies*, 81(1), 325–355. (Cited on page 86.)
- OHLIN, J. (1969): “On a class of measures of dispersion with application to optimal reinsurance,” *ASTIN Bulletin: The Journal of the IAA*, 5(2), 249–266. (Cited on pages 43, 91, 93, 94, 96 and 97.)
- PAN, J. (2002): “The jump-risk premia implicit in options: Evidence from an integrated time-series study,” *Journal of financial economics*, 63(1), 3–50. (Cited on pages 11 and 26.)
- PROTTER, P. E. (2005): “Stochastic differential equations,” in *Stochastic integration and differential equations*, pp. 249–361. Springer. (Cited on page 42.)
- RAVIV, A. (1979): “The design of an optimal insurance policy,” *The American Economic Review*, 69(1), 84–96. (Cited on pages 12, 27 and 86.)
- RIDGEWAY, G. (1999a): “The state of boosting,” *Computing Science and Statistics*, pp. 172–181. (Cited on page 105.)
- RIDGEWAY, G. K. (1999b): “Generalization of boosting algorithms and applications of bayesian inference for massive datasets,” Ph.D. thesis. (Cited on page 105.)
- ROGERS, L., ET AL. (1994): “Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains,” *The Annals of Applied Probability*, 4(2), 390–413. (Cited on page 58.)
- ROGERS, L., AND Z. SHI (1994): “Computing the invariant law of a fluid model,” *Journal of Applied Probability*, 31(4), 885–896. (Cited on page 58.)

## Bibliography

---

- ROTHSCHILD, M., AND J. E. STIGLITZ (1970): “Increasing risk: I. A definition,” *Journal of Economic theory*, 2(3), 225–243. (Cited on page 87.)
- SANDRI, M., AND P. ZUCCOLOTTO (2008): “A bias correction algorithm for the Gini variable importance measure in classification trees,” *Journal of Computational and Graphical Statistics*, 17(3), 611–628. (Cited on page 126.)
- (2010): “Analysis and correction of bias in total decrease in node impurity measures for tree-based algorithms,” *Statistics and Computing*, 20(4), 393–407. (Cited on page 126.)
- SCHAPIRE, R. E., AND Y. FREUND (2012): *Boosting: Foundations and algorithms*. MIT press. (Cited on page 105.)
- SCHOUTENS, W., AND J. CARIBONI (2010): *Lévy processes in credit risk*, vol. 519. John Wiley & Sons. (Cited on page 33.)
- SHAKED, M., AND J. G. SHANTHIKUMAR (2007): *Stochastic Orders*. Springer Science & Business Media. (Cited on pages 43, 44, 45, 46, 47, 90, 92, 94, 95, 97, 98 and 100.)
- SHI, P., X. FENG, AND A. IVANTSOVA (2015): “Dependent frequency–severity modeling of insurance claims,” *Insurance: Mathematics and Economics*, 64, 417–428. (Cited on pages 13, 28, 104 and 129.)
- SIU, C. C., S. C. P. YAM, AND H. YANG (2015): “Valuing equity-linked death benefits in a regime-switching framework,” *ASTIN Bulletin: The Journal of the IAA*, 45(2), 355–395. (Cited on page 57.)
- SIU, T. K. (2005): “Fair valuation of participating policies with surrender options and regime switching,” *Insurance: Mathematics and Economics*, 37(3), 533–552. (Cited on pages 7, 22 and 57.)
- SIU, T. K., J. W. LAU, AND H. YANG (2008): “Pricing participating products under a generalized jump-diffusion model,” *International Journal of Stochastic Analysis*, 2008. (Cited on page 57.)
- SU, X., AND M. BAI (2018): “Stochastic Gradient Boosting Frequency-Severity Model of Insurance Claims,” *Working Paper*. (Cited on pages 15 and 29.)
- TAN, K. S., C. WENG, AND Y. ZHANG (2009): “VaR and CTE criteria for optimal quota-share and stop-loss reinsurance,” *North American Actuarial Journal*, 13(4), 459–482. (Cited on pages 12 and 27.)
- TANKOV, P. (2003): *Financial modelling with jump processes*. Chapman and Hall/CRC. (Cited on pages 11, 26, 39, 40 and 41.)
- TANSKANEN, A. J., AND J. LUKKARINEN (2003): “Fair valuation of path-dependent participating life insurance contracts,” *Insurance: Mathematics and Economics*, 33(3), 595–609. (Cited on page 56.)

- 
- VALCHEV, S. (2004): “Stochastic volatility Gaussian Heath-Jarrow-Morton models,” *Applied Mathematical Finance*, 11(4), 347–368. (Cited on page 81.)
- WHITMORE, G. A. (1970): “Third-degree stochastic dominance,” *The American Economic Review*, 60(3), 457–459. (Cited on page 87.)
- WOOD, S. N. (2006): “Generalized Additive Models: An Introduction with R,” . (Cited on page 104.)
- WUTHRICH, M. V., AND C. BUSER (2017): “Data analytics for non-life insurance pricing,” . (Cited on pages 8 and 23.)
- YANG, Y., W. QIAN, AND H. ZOU (2018): “Insurance premium prediction via gradient tree-boosted Tweedie compound Poisson models,” *Journal of Business & Economic Statistics*, 36(3), 456–470. (Cited on pages 8, 23 and 105.)
- YOUNG, V. R. (1999): “Optimal insurance under Wang’s premium principle,” *Insurance: Mathematics and Economics*, 25(2), 109–122. (Cited on pages 12, 27 and 86.)
- (2014): “Premium principles,” *Wiley StatsRef: Statistics Reference Online*. (Cited on pages 12 and 27.)
- ZHANG, Q., AND X. GUO (2004): “Closed-form solutions for perpetual American put options with regime switching,” *SIAM Journal on Applied Mathematics*, 64(6), 2034–2049. (Cited on pages 11 and 26.)
- ZHOU, C., W. WU, AND C. WU (2010): “Optimal insurance in the presence of insurer’s loss limit,” *Insurance: Mathematics and Economics*, 46(2), 300–307. (Cited on pages 27 and 86.)
- ZHOU, X. Y., AND G. YIN (2003): “Markowitz’s mean-variance portfolio selection with regime switching: A continuous-time model,” *SIAM Journal on Control and Optimization*, 42(4), 1466–1482. (Cited on pages 11 and 26.)