



Prévention et assurance : contributions aux approches actuarielle, cognitive et dynamique

Sarah Bensalem

► To cite this version:

Sarah Bensalem. Prévention et assurance : contributions aux approches actuarielle, cognitive et dynamique. Gestion et management. Université de Lyon, 2020. Français. NNT : 2020LYSE1300 . tel-03331257

HAL Id: tel-03331257

<https://theses.hal.science/tel-03331257>

Submitted on 1 Sep 2021

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 : 2020LYSE1300

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale ED 486
SCIENCES ECONOMIQUES ET DE GESTION

Spécialité de doctorat : Sciences Économiques
Discipline : Actuariat

Soutenue publiquement le 17/12/2020, par :
Sarah Myriam BENSALEM

**Prévention et assurance :
Contributions aux approches
actuarielle, cognitive et dynamique.**

Devant le jury composé de :

Etner, Johanna, Professeure à l'Université Paris-Nanterre	Rapporteure
Hillairet, Caroline, Professeure à l'ENSAE-Paris Tech	Rapporteure
Barrieu, Pauline Professeure à London School of Economics	Examinateuse
Havet, Nathalie, Maitre de conférence à l'Université Lyon 1	Examinateuse
Loisel, Stéphane, Professeur à l'Université Lyon 1	Examinateur
Picard, Pierre, Professeur à l'École Polytechnique	Président du jury
Rullière, Jean-Louis, Professeur à l'Université Lyon 1	Directeur.rice de thèse
Kazi-Tani, Nabil, Maitre de conférence à l'Université Lyon 1	Co-directeur.rice de thèse

Remerciements

Faire un doctorat se rapproche de vivre à bord de montagnes russes. L'aboutissement de chaque chapitre représentant un sommet du manège. La montée est progressive et parfois lente, l'arrivée au sommet est excitante et la chute est inattendue. Ces années peuvent parfois paraître longues mais au final tout passe très vite. Même si je suis la seule à présenter ce travail de doctorat, je ne l'ai certainement pas été pendant ces trois années. Les quelques lignes qui vont suivre sont donc pour rendre hommage aux personnes qui m'auront accompagnées.

Je commencerai par remercier mes deux directeurs de thèse Jean-Louis Rullière et Nabil Kazi-Tani, qui m'auront guidée du premier mot de mon mémoire de master au dernier de ce manuscrit de thèse. Je les remercie de m'avoir fait confiance et de m'avoir laissée beaucoup d'autonomie. Je suis particulièrement reconnaissante pour leurs conseils avisés et leurs relectures attentives. J'ai beaucoup appris à leurs côtés.

Je remercie les membres du jury qui par la diversité de leurs compétences et de leurs parcours, me font l'honneur d'évaluer mon travail. Pauline Barrieu, Nathalie Havet, que j'ai pu cotoyer pendant ces trois années de thèse au laboratoire SAF, Stéphane Loisel, également directeur du laboratoire SAF et avec qui les échanges auront été formateurs, ainsi que Pierre Picard. Je remercie particulièrement Johanna Etner et Caroline Hillairet, pour leur travail de relecture minutieuse du manuscrit et pour leurs corrections et suggestions afin d'améliorer la qualité de ce document.

Je remercie également Jean-Yves Lesueur, professeur à l'Université de Lyon et Idris Kharroubi, professeur à Sorbonne Université qui m'ont suivi lors de trois ans à l'occasion de mes comités de suivi de thèse. Ils m'ont tous les deux formulés de précieux conseils pour améliorer et continuer mon travail. J'ai apprécié de pouvoir retrouver à ces occasions Idris que j'ai eu la chance d'avoir comme professeur de méthodes de Monte-Carlo en master 2 à Paris-Dauphine. Mon projet de réaliser une thèse de doctorat a été en partie motivé par lui et par Dylan Possamaï, professeur à l'ETH Zürich, qui m'a fait découvrir le monde du Principal-Agent et que j'ai toujours plaisir à revoir.

Je voudrais remercier les membres du laboratoire SAF avec qui j'ai pu passer des moments conviviaux et avec qui la discussion est facile et intéressante. Stéphane Loisel, directeur du laboratoire, Alexis Bienvenue, Caroline Champagne (que j'espère croiser à l'Alpe!), Denis Clot

(merci pour les conseils pédagogiques notamment !), Pierre-Olivier Goffard, Xavier Millaud, Denys Pommeret, Christian Robert, Didier Rullière, Yahia Sahli, Pierre Théron. J'aimerais remercier plus particulièrement les membres de la chaire Prevent' Horizon : Marine Brocard, Nathalie Havet et Jean-Yves Lesueur. Merci également à l'équipe administrative de l'ISFA, Djana Moktari, Alice Fumex et ceux qui ont la double casquettes recherche-administratif Nicolas Leboisne et Anne Eyraud-Loisel.

Je voudrais également remercier les doctorants que j'ai pu côtoyer à différentes occasions, Arthur, Pierrick (à qui je dois l'achat de verres à vin), Nesrine, Marwa, Rihem, Carine, Nathalia, Tachfine (merci pour le comparateur !), Pierre C., Behzad (bientôt je serai ta boss ;)). Les post-doc Karim, Frank et Patrick (also known as Thesaurus Pat). Une mention spéciale pour mes compagnons de galère sans qui ces trois années n'auraient pas été les mêmes. Maxou pour les photos de ton chat et pour les références toujours bien placées (la meilleure étant "y'a pot"), Morgane ma soeur de thèse qui comprendra mieux que personne ce que j'ai vécu pendant ces années (chargée de relecture inestimable et partenaire du Run-in Lyon), Pierre co-représentant de choc, je n'aurais pas pu avoir meilleur co-bureau (et capitaine de la ISFAST team), Romain qui aura toujours une idée farfelue à explorer ou à partager (s'il-te-plait apprends à tenir une tasse correctement...) et Steve, copain officiel de pause, Google Scholar simplifié (je n'oublierais jamais plus Kenkel) et partenaire au Molkky.

I would like to thank Nico the Chilean guy. I really enjoy the work dynamic that we have and I hope we can keep the best work group ever alive. Next step is coming to Chile to try real empanadas and see guanacos !

Merci à Emma ma co-MASEF mangeuse de frites. Merci pour nos conversations messages qui m'auront beaucoup fait rire même si j'aurais préféré les faire en direct. Rdv au Diable des Lombards, je te payerai un burger après.

Merci également à Jessica ma copine toulousaine rencontrée pendant une école d'été, tes messages m'ont fait sourire pendant ces trois ans, je suis très contente de t'avoir rencontrée !

Merci à ma voisine Pauline avec qui les discussions sur le balcon ou sur le pallier sont toujours cool. Vielen Dank an meine Nachbarin Ronith für all diese Gespräche auf Deutsch, mit denen ich meine persönliche Kultur vervollständigen konnte.

Merci à mes amis de toujours, la team Champo. Marionnette de l'amour pour tes goûts musicaux et tes idées de dingue. Pipo pour tes photos qui m'ont fait voyager et pour être toujours partante. Mes deux lyonnais préférés : Cyrillou pour ton rire parfait et les apéro lyonnais, Guillaume pour m'avoir écouter raconter des bêtises mais pas que et pour tes pulls à message. Juju/Chuchu pour ta lourdeur (of course) et pour nos conversations qui débouchent sur le GIF qui résume tout. Léo pour tes marcels, le Beaufort et mon trône. Peffenou partenaire de cours de maths, pour trouver un créneau dans ton emploi du temps de ministre et pour ton houmous parfait. Kévinou de m'avoir pardonnée de t'avoir traumatisé au Sept Laux. Mais aussi, celles qui complètent notre groupe : Ambre, Auriane, Camille ma copine de dancefloor

(Starliiiight, tout est dit), Mathilde et Marianne. Merci pour les weekend ski, les soirées, les repas, les Zoom du confinement, les cueillettes de fraises ratées et bien plus encore. Vous êtes les meilleurs amis qu'une personne peut avoir.

J'aimerais remercier aussi des personnes qui m'ont apportée "mathématiquement", Violaine, Samy et Selim mais aussi Emmanuel Russ, professeur à l'Université Grenoble Alpes qui m'a soutenue et accompagnée dans mon projet de candidature en deuxième année de master. Merci également à Yvette et Antoine, qui ont toujours un mot gentil lors d'étapes importantes dans mon parcours.

Je voudrais remercier de tout coeur la famille Masset. Marie-Hélène et Olivier pour leur gentillesse infinie depuis que je les ai rencontrés en terrain tronchois, pour les vacances à Milon et Hardelot qui auront participé à l'écriture de cette thèse et pour les montées à cheval. Philippe et Guillaume merci pour les conversations intéressantes et animées et pour les échanges de découvertes musicales. Et je profite de ce paragraphe pour remercier Sarah, avec qui je partage plus qu'un prénom, pour nos discussions sur des sujets de société qui nous tiennent toutes les deux à cœur.

A ma famille. Mes parents qui m'ont transmis l'envie d'apprendre, de travailler et de toujours aller plus loin. Merci pour tout ce que vous faites et d'être là. Inès, ma grande soeur, petit Coraya venu de la mer, partenaire de running, de trail et co-radio Londres, Sébastien avec qui j'ai passé de super weekend lyonnais à marcher dans la ville (même par temps de canicule) et à découvrir des resto. Riadh, mon grand frère, converti à Francis Ponge et décrypteur de l'info, Clémence avec qui j'ai passé de super moments et qui a toujours un message à m'envoyer. Mes neveux, Anton ("Mais non Tata, non") dont la naissance à marquer le début de ce travail de thèse et Auguste le petit nouveau qui en aura marqué la fin (littéralement vu que tu pointes ton nez pile le jour où ce manuscrit se termine). A mon gros patapouf Magnum, mangeur de thon invétéré qui donne son amour avec parcimonie mais en abondance quand la porte du frigo va s'ouvrir.

A Seb qui est toujours présent pour moi et prêt à me faire rire. Partenaire de vie et de confinement qui soutient toutes mes idées (bokashiiiiii) ou presque. D'un nain de jardin aux karaokés impromptus, you make loving fun.

Résumé

Cette thèse de doctorat porte sur la modélisation de l'effort de prévention et de sa relation avec l'assurance de marché. Chacun des chapitres la composant, tente de capturer différents aspects de cette problématique, de l'étude d'un critère conforme aux pratiques actuarielles à celui du côté de l'offre en assurance, en passant par l'inclusion de biais de perception du risque et par une approche de la prévention en temps dynamique.

Le chapitre 1 modélise la relation entre un assureur et un assuré sous la forme d'un jeu de Stackelberg. Dans ce jeu, l'assureur joue en premier en proposant un contrat d'assurance sous la forme d'un facteur de chargement. L'assuré joue ensuite en choisissant le taux de couverture et son effort de prévention optimaux. L'assuré comme l'assureur ont pour but de minimiser leurs mesures de risque respectives qui sont toutes deux cohérentes. Les effets respectifs de l'auto-assurance et l'auto-protection, sur la minimisation du risque seront étudiés. Dans chaque cas, il sera montré que les choix optimaux de l'assuré existent et le contrat optimal pour l'assureur sera caractérisé. De plus, il sera montré que si la mesure de risque de l'agent décroît plus rapidement que l'espérance de sa perte, alors l'effort optimal est croissant avec le facteur de chargement avec une discontinuité potentielle lorsque la couverture optimale passe de complète à nulle. Cependant, si la diminution de la mesure du risque de l'agent est plus rapide que celle de l'espérance de la perte, l'effort optimal peut être croissant ou décroissante en fonction du facteur de chargement.

Le chapitre 2 étudie la relation entre auto-assurance et assurance de marché également sous la forme d'un problème d'optimisation pour un agent. De manière similaire au chapitre 1, cet agent doit déterminer le taux de couverture et l'effort de prévention qui réduiront de manière optimale sa mesure de risque. Dans ce chapitre, la mesure de risque considérée est dite de distorsion et est définie à partir d'une fonction de distorsion non concave. Ceci permet de tenir compte de biais cognitifs individuels potentiels dans la perception du risque. La caractérisation de la solution optimale pour l'agent permet d'apporter une nouvelle conclusion dans la relation entre auto-assurance et assurance de marché. L'auto-assurance n'est plus seulement substituable à l'assurance de marché, elle peut être également complémentaire à celle-ci, suivant la sensibilité de l'effort de prévention au prix de l'assurance.

Le chapitre 3 se concentre sur l'auto-protection en proposant un problème de maximisation

d'utilité espérée en version dynamique. Ceci se présente sous forme d'un problème de contrôle stochastique dans lequel l'agent choisit sa couverture assurantielle et son effort de prévention qui est dynamique. Le problème peut être séparé en deux sous-problèmes, le premier est une optimisation en l'effort et le second en la couverture assurantielle. Comme l'individu veut obtenir la richesse finale la plus importante possible, il cherche à maximiser l'espérance de l'utilité exponentielle de cette richesse. La richesse de l'agent peut être vue comme la solution d'une équation différentielle stochastique rétrograde à saut, cette équation admet une unique solution et qui est de plus explicite. En particulier, on obtient que l'effort optimal d'auto-protection est constant. La distribution initiale du processus de perte, quand il n'y a pas d'effort, est donnée par un processus de Poisson composé qui est notamment un processus de Lévy. Obtenir un effort optimal constant signifie donc que la propriété de Lévy des processus est préservée par la maximisation d'une espérance d'utilité exponentielle. L'analyse du problème en la couverture assurantielle donne une condition suffisante pour obtenir l'existence d'un niveau de couverture optimal. L'individu pourra alors souscrire à une assurance en fournissant un effort de prévention qui lui permettra de maximiser sa satisfaction ou bien choisir de ne pas souscrire au contrat mais en prenant toutefois part à des actions d'auto-protection.

Mots-clés : Prévention, Mesures de risque cohérentes, Mesures de risque de distorsion non concave, Programmation dynamique, Utilité de continuation, Équations différentielles stochastiques rétrogrades.

Abstract

This PhD dissertation studies the modeling of prevention effort and its relation to market insurance. The three chapters capture different aspects of this problematic, from the investigation of a criteria more in line with actuarial practices to a study of the insurer's point of view and by examining the effect of risk perception biases and the effect of a continuous-time framework.

Chapter 1 models the relationship between an insurer and an insured in the form of a Stackelberg game. In this game, the insurer plays first by offering an insurance contract in the form of safety loading. The insured then plays by choosing the optimal non-proportional insurance share and prevention effort. Both the insured and the insurer aim to minimize their respective risk measures, which are both coherent. The respective effects of self-insurance and self-protection actions on risk minimization will be studied. In each case, it will be shown that the optimal choices of the insured exist and the optimal contract for the insurer will be characterized. In addition, we show that if the buyer's risk measure decreases faster in effort than his expected loss, optimal effort is non-decreasing in the safety loading with a potential discontinuity when optimal coverage switches from full to zero. On the contrary, if the decrease of the buyer's risk measure is slower than the expected loss, optimal effort may or may not be non-decreasing in the safety loading.

Chapter 2 also studies the relationship between self-insurance and market insurance in the form of an optimization problem for an economic agent. Similarly to Chapter 1, this agent determines the non-proportional insurance share and the prevention effort that will optimally reduce his risk measure. In this chapter, the risk measure considered is a distortion risk and is defined from a non-concave distortion function. This makes it possible to take into account potential individual cognitive biases in the perception of risk. The characterization of the optimal solution for the agent allows to bring a new conclusion in the relation between self-insurance and market insurance. Self-insurance is no longer only substitutable for market insurance, it can also be complementary to the latter, depending on the sensitivity of the prevention effort to the price of the insurance.

Chapter 3 focuses on self-protection in the form of a problem of expected utility maximization in time-dynamic version. This takes the form of a stochastic control problem in which the

agent chooses his insurance coverage and his prevention effort which is time-dynamic. The problem can be separated into two sub-problems, the first is the optimization in the effort and the second is in the insurance coverage. As the individual wants to obtain the greatest possible final wealth, he seeks to maximize the expectation of the exponential utility of this wealth. The wealth of the agent can be seen as the solution of a backward stochastic differential equation with jumps, we show that this equation admits a unique solution that is also explicit. In particular, we obtain that the optimal prevention effort is constant. The initial distribution of the loss process, when there is no effort, is given by a particular Lévy process: a compound Poisson process. The fact that the optimal effort is constant says that the Lévy property is preserved under exponential utility maximization. The analysis of the insurance coverage optimization gives a sufficient condition to obtain the existence of an optimal level of coverage. The individual can then take out insurance by providing a prevention effort that will allow him to maximize his satisfaction or choose not to take out the contract but nevertheless take part in self-protection actions.

Keywords: Prevention, Coherent Risk measures, Non-concave distortion risk measures, Dynamic programming, Continuation utility, Backward SDEs .

Table des matières

Liste des figures	xii
Introduction Générale	1
1 Contexte et enjeux	1
1.1 Les définitions de la prévention	1
1.2 La prévention et l'assurance	4
1.3 Prévention et aléa moral	5
2 Les mesures de risque	6
2.1 Axiomes et mesures de risque usuelles	7
2.1.1 Value-at-Risk	8
2.1.2 Tail Value-at-Risk	9
2.1.3 Mesures de risque par distorsion	9
2.2 Lien entre les mesures de risque et les fonctions d'utilité	11
3 La théorie de l'utilité espérée	11
3.1 De la naissance du critère d'utilité espérée à une théorie axiomatique du choix sous incertitude	12
3.2 Représentation de von Neumann Morgenstern	12
3.3 Représentation de Savage	17
3.4 Les types de fonctions d'utilité	18
3.5 Limites et critiques	19
3.5.1 L'expérimentation comme remise en cause	19
3.5.2 Biais comportementaux	21
3.5.3 Théories alternatives	24
4 Ordre stochastique et équations différentielles rétrogrades	30
4.1 Dominance stochastique	30
4.2 Équations différentielles stochastiques rétrogrades à saut	31
5 Modèles de prévention en assurance	35
5.1 Le modèle d'Ehrlich et Becker	35
5.2 Distorsion des probabilités et prévention	38
6 Contributions et résultats principaux	40
6.1 Prevention efforts, insurance demand and price incentives under coherent risk measures	40
6.2 Self-insurance and Non-concave Distortion Risk Measures	42
6.3 A Continuous-Time Model of Self-Protection	44
Bibliographie	46

1 Prevention efforts, insurance demand and price incentives under coherent risk measures	54
1 Introduction	55
2 Model and Assumptions	60
3 The insurance buyer's optimization problem	64
3.1 Optimal insurance cover	64
3.2 Non-increasing G	65
3.3 Self-Protection	68
4 The insurance seller's optimization problem	72
4.1 Solving the problem of IS	72
4.2 Adverse Selection	73
5 Case studies	74
5.1 The Pareto Case	74
5.1.1 Self-insurance	75
5.1.2 Adverse selection in the case of self-insurance	76
5.1.3 Self-protection	77
5.2 Discrete distributions	80
5.2.1 Self-insurance: the discrete case	80
5.2.2 Self-protection: the discrete case	83
6 Numerical analysis	84
6.1 The Pareto case	84
6.2 Approximation of the compound Poisson distribution	85
7 Conclusion	88
Bibliographie	90
2 Self-insurance and non-concave distortion risk measures	94
1 Introduction	94
2 Model and assumptions	98
3 The optimization problem	103
3.1 Optimal insurance coverage	103
3.2 Optimal self-insurance effort	104
4 Case study	110
5 Conclusion	115
Bibliographie	116
3 A Continuous-Time Model of Self-Protection	119
1 Introduction	119
2 Model and Assumptions	123
3 Solving the insurance problem	125
3.1 Optimal prevention effort	125
3.2 Optimal insurance cover	127
4 Example: Linear cost and linear effect on the frequency	130
Bibliographie	134
Conclusion Générale	136

Bibliographie	139
A Preuves du Chapitre 1	141
1 Proofs of section 2	141
2 Proofs of section 3	141
2.1 Proofs of subsection 3.2, G non-increasing	141
2.2 Proofs of subsection 3.3, self-protection	144
3 Proofs of section 5	146
Bibliographie	151
B Preuves du Chapitre 3	152
1 Dynamic programming principle	152
2 Proof of Proposition 3.1.	154
Bibliographie	159

Liste des figures

0.1	Exemples de fonctions de distorsion non concaves. A gauche, la fonction $\exp(-(-\log(u))^\gamma)$, $\gamma = 0.67$. A droite, la fonction $\frac{u^\gamma}{(u^\gamma + (1-u)^\gamma)^{1/\gamma}}$, $\gamma = 1.5$.	43
1.1	Local minimizer of L_N on the set \mathcal{N} , as a function of θ .	70
1.2	Comparison of optimal insurance demand and effort levels for different safety loadings.	71
1.3	Optimal efforts $e_I(\theta)$, e_N and boundary effort e_θ for the Pareto distribution, as functions of θ .	84
1.4	Objective functions $L_I^\theta(e_I(\theta))$ and $L_N(e_N)$ for the Pareto distribution, as functions of θ .	85
1.5	Estimated Kolmogorov distance and error bound from Lemma 6.1, as functions of p .	87
1.6	Cumulative distribution functions of X (black curve) and X_p (red curve).	87
2.1	An example of an inverse S-shaped distortion function: $\psi(u) = \exp(-(-\log(u))^\alpha)$, with $0 < \alpha < 1$.	101
2.2	Illustration of a non-monotonic function G crossing $1 + \theta$ multiple times for a fixed θ .	105
2.3	Illustration of the change in the \mathcal{I} and \mathcal{N} sets as θ rises.	106
2.4	Illustration of G for small claims with $p = 0.06$, $\eta = 0.4$, $\lambda = 10$ and $\tau(e) = \sqrt{e + \frac{1}{100}}$.	112
2.5	Illustration of G for small claims with $p = 0.08$, $\eta = 0.2$, $\lambda = 150$ and $\tau(e) = \sqrt{e}$.	113
2.6	Illustration of G for small claims with $p = 0.05$, $\eta = 0.4$, $\lambda = 8$ and $\tau(e) = \sqrt{e}$.	113
2.7	Illustration of the values of $e_I(\theta)$ and e_N for small claims with $p = 0.06$, $\eta = 0.4$, $\lambda = 10$ and $\tau(e) = \sqrt{e + \frac{1}{100}}$.	114
2.8	Illustration of the values of $e_I(\theta)$ and e_N for small claims with $p = 0.06$, $\eta = 0.4$, $\lambda = 10$ and $\tau(e) = \sqrt{e + \frac{1}{100}}$.	115

Introduction Générale

1 Contexte et enjeux

1.1 Les définitions de la prévention

"Prendre les mesures nécessaires pour éviter un mal, un danger ; ensemble de mesures prises contre certains risques", telle est la définition de la prévention dans le dictionnaire. Elle permet par sa généralité d'appréhender l'étendue du sujet de la prévention qui concerne tous les domaines, de la santé à la sécurité routière, en passant par l'aviation civile (par exemple le soutien européen de la recherche industrielle en faveur de moteurs moins polluants dans le cadre de l'ACARE (*Advisory Council for Aeronautics Research in Europe*)). La simplicité de cette définition montre également l'intérêt du monde de l'assurance pour cette notion, dont le cœur de métier est la gestion des risques. Ainsi il est opportun pour les organismes d'assurance de proposer des programmes de prévention auprès de leurs assurés. Devant la globalité de la définition, il semble impératif de spécifier les actions de prévention possibles pour pouvoir les appliquer techniquement. Une des rares définitions des actions de prévention a été proposée par l'Organisation Mondiale de la Santé (OMS) (voir [47] et [61]), dans laquelle trois types de prévention sont distingués

Prévention primaire : "mesures applicables à une maladie ou groupe de maladies pour en bloquer les causes avant qu'elles n'agissent sur l'homme ; en d'autres termes, pour empêcher la survenue de la maladie". En d'autres termes, il s'agit des actions permettant de prévenir la survenue d'un risque. Le port du casque à vélo ou lors de la pratique sportive du ski pour éviter les traumatismes crâniens, le lavage réguliers des mains pour empêcher la contamination au virus SARS-CoV-2 (responsable de la maladie COVID-19) peuvent être cités comme exemples.

Prévention secondaire : "mesures destinées à interrompre un processus morbide en cours pour prévenir de futures complications et séquelles, limiter les incapacités et éviter le décès". Cela signifie qu'il faut repérer la survenue du risque le plus tôt possible pour en réduire les conséquences. Les dépistages de la population générale des cancers du sein et de la prostate en sont des exemples. Un autre exemple est la surveillance dans le cadre d'un dépistage ciblé lorsque l'individu présente des antécédents familiaux (par exemple la prévention secondaire d'un infarctus du myocarde chez un patient ayant déjà eu un syndrome coronarien aigu. En

effet, une personne ayant déjà vécu un événement coronarien a un risque augmenté de survenue d'un infarctus du myocarde par rapport à la population générale. En associant des traitements médicamenteux, comme l'aspirine ou les bêtabloquants, à des mesures hygiéno-diététiques, on peut réduire voire supprimer ce sur-risque).

Prévention tertiaire : "ensemble de mesures visant à permettre aux personnes handicapées de recouvrer leurs fonctions initiales ou d'utiliser au maximum les capacités qui leur restent ; la réadaptation comprend à la fois des interventions individuelles et des actions sur l'environnement." Il s'agit ici de l'ensemble des actions permettant de réduire les conséquences et les rechutes d'une maladie, notamment d'un point de vue curatif.

Cette distinction faite par l'OMS est axée vers la santé et ne peut convenir telle quelle à tous les domaines d'application possibles de la prévention. En particulier, elle ne convient pas au monde économique qui veut l'appliquer à tout type de risque. Ainsi, Ehrlich et Becker [36] suggèrent une définition et une séparation différentes

Auto-assurance : ensemble des mesures applicables à la diminution de la taille d'un sinistre, la taille pouvant être monétaire ou bien être exprimée en terme de gravité.

Auto-protection : ensemble des mesures applicables à la diminution de la probabilité d'occurrence d'un risque.

Même si la définition proposée par Ehrlich et Becker s'applique à tous les types de risques et en particulier au cadre de l'assurance, elle peut être mise en parallèle avec la définition de l'OMS. Ceci a été fait par de nombreux économistes, on peut citer notamment les travaux de Devaux et Eeckhoudt [32] ainsi que ceux de Courbage et Rey [24]. L'auto-assurance est souvent associée aux préventions secondaire et tertiaire tandis que l'auto-protection est considérée comme le pendant de la prévention primaire.

Un exemple d'action d'auto-assurance serait la prévention en santé du pied diabétique. En effet, il s'agit d'une complication classique mais grave du diabète qui est la première cause d'amputation non traumatique dans le monde. Pour éviter l'amputation chez l'individu, il a été démontré que l'éducation du patient à l'auto-surveillance quotidienne des pieds (hygiène régulière, apparition de plaies) et un suivi bimensuel par un pédicure podologue diplômé d'État pour traiter et soigner les lésions entraînent une réduction du risque. Une illustration de l'auto-protection en santé serait le recours à la chirurgie prophylactique lorsqu'une mutation génétique des gènes BRCA1 et BRCA2 chez une femme est repérée. Chez une femme porteuse d'une mutation du BRCA1 ou du BRCA2, le risque de cancer du sein varie de 40% à 80% au cours de la vie (cette variation s'explique par différents résultats trouvés dans des études scientifiques menées, le type de gène concerné, l'histoire familiale de cancer du sein et l'âge). Par rapport à la population générale, ces femmes ont un risque accru de développer un cancer du sein à un âge jeune, habituellement avant la ménopause et un cancer du sein bilatéral. Ainsi, une mastectomie totale bilatérale prophylactique peut être recommandée.

Ce travail de thèse propose de modéliser l'effort de prévention selon la distinction proposée par Ehrlich et Becker dans différents cadres théoriques. Ceux-ci seront détaillés au fil de cette introduction. Ce manuscrit est composé de trois chapitres. Les chapitres 1 et 2 se placent dans un cadre statique, de manière similaire à [36] et proposent de nouveaux critères de préférences individuels. Au lieu de considérer le gain de l'agent économique, il s'agit de réinterpréter dans le domaine des pertes, les résultats avancés par Ehrlich et Becker. Le chapitre 1 traite de l'auto-assurance et de l'auto-protection dans le cadre d'une souscription à contrat d'assurance et lorsque le critère de préférence donné par une mesure de risque de cohérente (voir Section 2). Dans ce chapitre, le côté de l'offre en assurance est également étudié en modélisant la perte propre à l'assureur qu'il cherche à minimiser au travers d'une mesure de risque cohérente. Le chapitre 2 se concentre sur le côté de la demande en assurance et sur sa relation avec l'effort d'auto-assurance. Dans ce contexte, l'individu évalue son risque à l'aide d'une mesure de distorsion définie à partir d'une fonction de distorsion non concave (voir Section 2.1.3). Le chapitre 3 cherche à étendre les résultats présentés dans [36] sur la relation entre demande en assurance et effort d'auto-protection dans un cadre dynamique. La signature du contrat d'assurance marque le début de la période de temps considérée qui s'étend jusqu'à la fin du contrat ou jusqu'à la renégociation des termes de ce contrat. Le critère de préférence de l'individu est donné par une fonction d'utilité exponentielle (voir Section 3.4) et de manière similaire à [36], il veut maximiser sa satisfaction.

Cette introduction est constituée de six parties, la première visant à introduire le contexte de la prévention en assurance avec notamment des exemples d'initiatives retrouvées dans l'industrie. Cette partie présente également la problématique de l'aléa moral en assurance, qui sera retrouvée dans les trois chapitres constituant ce travail de thèse. Ensuite, il conviendra d'introduire et d'expliquer en détails les outils théoriques indispensables pour comprendre les résultats présentés dans ce travail de thèse de la Section 2 à la Section 4. Les chapitres 1 et 2 font appel à la notion de mesures de risque, présentée en Section 2, le chapitre 3 se place quant à lui dans le cadre de l'utilité espérée, détaillée dans la Section 3. Le chapitre 2 repose également sur la notion de biais comportementaux qui est approfondie dans la Section 3.5.2 et qui a permis entre autres de définir des théories alternatives à l'utilité espérée (voir Section 3.5.3) qui ont inspiré ce chapitre 2. Le chapitre 3 s'appuie sur la théorie du contrôle stochastique avec en particulier les équations différentielles stochastiques à saut, dont les fondements théoriques seront exposés dans la Section 4.

Une fois les outils théoriques présentés, la cinquième partie de l'introduction présentera les principaux modèles théoriques de la prévention en assurance dans la littérature économique. La partie 5 commencera par détailler le modèle d'Ehrlich et Becker (voir [36]) considéré comme un des modèles de référence mettant en relation prévention et demande en assurance. Ce modèle sera à la base de ce travail de thèse, chaque chapitre le constituant visant à étendre et à vérifier les conclusions retrouvées dans [36] sous différentes hypothèses. Différentes extensions des

travaux d'Ehrlich et Becker existent, la Section 5.2 propose un état des lieux de la littérature. Enfin une dernière partie présentera plus en détail les chapitres et les apports de la thèse.

1.2 La prévention et l'assurance

L'assurance est un système de gestion des risques dans la mesure où par définition elle prévient un agent des conséquences économiques liées à la survenance d'un risque particulier. Ainsi un des rôles d'une assurance est de protéger les individus, les entreprises ou les associations. En ce sens, la prévention doit être une partie intégrante du métier de l'assureur car elle permet d'éviter un sinistre ou d'en limiter les conséquences. Ceci peut se faire au travers de l'éducation, de la sensibilisation des assurés ou encore par la mise en place de programmes de prévention. Cependant, les propositions alliant ces deux domaines sont encore peu développées dans le monde de l'assurance et beaucoup de questions se posent à ce sujet : quelles actions de prévention peuvent être mises en place ? Est-il pertinent d'agir sur tous les risques ?... La modélisation de la prévention et l'estimation de l'impact d'un tel programme sont d'autant plus difficiles qu'il existe peu de modèles liant prévention et assurance dans la théorie économique et actuarielle.

Parmi les initiatives pratiques déjà en place, on peut noter l'association des membres de la Fédération Française de l'Assurance (FFA) au travers de l'Assurance Prévention qui finance et pilote des actions de sensibilisation pour les risques quotidiens avec par exemple, les risques routiers, les accidents domestiques et les risques en santé. Parmi les différentes actions de cette association, on peut noter la distribution de kits "Prévention Covid" (contenant deux masques, un flacon de solution hydro alcoolique, un paquet de lingettes désinfectantes) en partenariat avec la Gendarmerie Nationale au cours de l'été 2020, ou encore l'organisation d'une tournée d'été avec une piste d'éducation routière visant les conducteurs de deux-roues motorisés, en partenariat avec des moniteurs CRS. On peut noter aussi une autre association mise en place par la FFA, l'association Mission Risques Naturels qui a pour objectif de participer à une meilleure connaissance des risques naturels en contribuant d'un point de vue technique aux politiques de prévention des catastrophes naturelles, dont la fréquence ne cesse d'augmenter au cours du temps. Les assurances ne se contentent pas d'agir collectivement mais elles le font aussi individuellement auprès de leurs assurés, en particulier dans le domaine de la santé. AG2R a ainsi mené une campagne d'information et de dépistages des pathologies dentaires auprès des apprentis boulangers. La MGEN et Harmonie Mutuelle ont respectivement mis en place la prise en charge d'une partie des dépenses sportives de leurs clients sous certaines conditions. On peut également noter que Generali propose au travers d'un programme de gagner des points en adoptant des comportements sains qu'il est possible d'échanger contre des bons d'achat.

La mise en place d'actions réglementaires en France a également donné un élan au développement de programmes de prévention, notamment dans le cadre des contrats collectifs d'assu-

rance santé avec depuis janvier 2016, la mise en place de l'Accord National Interprofessionnel (ANI) qui impose qu'au moins 2% des cotisations soit consacré à la prévention santé. Les produits d'assurance santé proposés aux salariés des entreprises incluent donc des programmes de prévention dans les domaines de la nutrition ou encore du fitness. Ces programmes représentent alors des incitations aux assurés à modifier leurs comportements et améliorer leur hygiène de vie et leur santé. Malgré cette obligation imposée aux employeurs de proposer une assurance santé complémentaire à leurs salariés via l'ANI, celle-ci ne porte en réalité que sur la couverture d'un panier de biens minimal. L'assuré a donc le choix du taux de couverture assurantuelle. Ainsi cette décision est soumise à des biais comportementaux notamment dans la perception de la santé individuelle et du risque de maladies (voir [19]). Les individus surestiment leurs risques de tomber malade ce qui semble paradoxal par rapport aux données gouvernementales qui indiquent que les individus ne font pas assez d'actes préventifs en santé (voir [79]). De multiples biais comportementaux peuvent modifier la perception individuelle d'un état et notamment d'un risque. La section 3.5.2 sera consacrée au développement des biais majeurs. La perception des probabilités de risque est donc importante à prendre en compte dans la modélisation des préférences individuelles. Ceci sera proposé dans le chapitre 2 en intégrant la caractérisation mathématique des biais de perception dans la modélisation des décisions d'assurance et de prévention.

1.3 Prévention et aléa moral

Il est intéressant de remarquer que les définitions présentées Section 1.1 peuvent également être considérées en présence d'asymétrie d'information. Les trois chapitres de ce travail de thèse considèrent un individu voulant souscrire à un contrat d'assurance pour faire face à une perte potentielle, en parallèle duquel il a la possibilité de faire de la prévention ce qui se traduit par entreprendre un effort. La relation contractuelle d'assurance (respectivement de réassurance) peut être vue comme un modèle Principal-Agent avec l'assureur (respectivement le réassureur) comme Principal et l'assuré (respectivement l'assureur) comme Agent. Le chapitre 1 s'inspire notamment de cette théorie pour une approche de la problématique de l'assureur qui propose un contrat à l'individu. Bien que le comportement de l'assuré influence la probabilité d'occurrence et la sévérité d'un sinistre (voir [112]), l'assureur ne peut observer directement les actions de l'assuré une fois le contrat signé. Le comportement de l'assuré est donc soumis à ce qu'on appelle l'aléa moral (voir [6], [78] et [7]). Pour l'assureur voyant le sinistre de son assuré, il est impossible de dissocier ce qui est dû au comportement de l'assuré de ce qui ne l'est pas. Pour l'assuré, ceci est positif dans le sens où il ne supporte plus seul le risque grâce au contrat. Cependant, il y a également un aspect négatif car l'effort lui est coûteux et est invisible aux yeux de l'assureur. Il peut donc choisir de se comporter de manière à augmenter sa prise de risque par rapport à la situation où il supportait entièrement les conséquences négatives d'un sinistre. Les trois chapitres de cette thèse prennent en compte la question de l'aléa moral en considérant que les actions de prévention réalisées par l'assuré ne sont pas observables par

l'assureur.

Ce phénomène d'aléa moral peut se manifester de deux manières distinctes. L'aléa moral ex ante survient entre la signature du contrat et la survenue du sinistre. Il est typiquement représenté par l'absence ou la diminution de l'effort de prévention. L'aléa moral ex post survient après la survenue du sinistre et correspond à la mauvaise déclaration d'un sinistre, par exemple dans le cas d'une dégradation de biens. L'aléa moral ex post peut être limité grâce à l'envoi d'experts par l'assureur pour évaluer le sinistre. Les problèmes d'asymétrie d'information en assurance ont été notamment étudiés dans le modèle de Rothschild et Stiglitz (voir [90]). Le problème de la moralité aléatoire des assurés est un problème récurrent pour les assureurs. Une solution possible dans le cadre de la prévention, serait l'utilisation d'objets connectés comme certains programmes le proposent déjà. Par exemple les assureurs américains John Hancock Financial et Aetna qui, en 2018, ont offert des montres connectées à leurs clients pour mesurer leur progrès lors d'un programme de prévention avec de l'activité physique. Ces objets permettraient à l'assureur d'observer alors l'effort de prévention (hygiène de vie, bilans de santé) et de fait réduire le risque d'aléa moral. On peut également noter la mise en place d'un espace Services Santé au sein de l'entreprise Airbus, qui a été conçu par l'institution de prévoyance du secteur Ipeca. Trois types de services santé à destination des salariés sont proposés : une cabine de téléconsultation médicale, un centre optique permettant un essayage virtuel de montures avant un essayage réel et un espace ostéopathie sur rendez-vous.

2 Les mesures de risque

Dans cette section sera présentée la théorie des mesures de risque. Les chapitres 1 et 2 considèrent des agents économiques qui veulent minimiser leurs pertes, les critères de préférences sont ainsi donnés par des fonctions de pertes potentielles appelées mesures de risque. Dans ces deux chapitres, l'individu qui souscrira au contrat d'assurance minimisera sa mesure de risque en choisissant son niveau de couverture assurantielle et son niveau d'effort de prévention. Le chapitre 1 étudiera également le côté de l'offre en assurance au travers d'un agent économique qui proposera un contrat d'assurance. Cet agent, appelé assureur, cherchera à minimiser sa prise de risque en basant le contrat sur le prix de l'assurance exprimé par le taux de chargement dans le cadre d'une assurance proportionnelle (c'est-à-dire qu'il s'agit d'un contrat d'assurance dans lequel une proportion du sinistre réalisé sera prise en charge par l'assureur, le reste étant à la charge de l'assuré).

En assurance et en finance, les acteurs des marchés voient leur capital exposé au risque. Il est donc important de quantifier le risque d'une position afin de décider si il est acceptable ou non. Suite à des sous-évaluations des réserves chez les différents intervenants, des réglementations ont été mises en place depuis quelques années, respectivement nommées Bâle II pour les établissements financiers et Solvabilité II pour les compagnies d'assurance. Ces réglementations

ont introduit diverses classes d'outils mathématiques appelés mesures de risque. Les mesures de risque sont par la suite devenues un outil de management du risque très utilisé notamment grâce à la possibilité d'analyser les axiomes qu'elles respectent ainsi que l'interprétation économique qu'il est possible d'en tirer (voir [9] et [44]). C'est pourquoi, elles ont été choisies comme critère pour les préférences individuelles dans les chapitres 1 et 2. Les mesures de risque apparaissent comme un outil actuariel indispensable à intégrer dans la modélisation de la prévention. Après avoir donné la définition et les propriétés d'une mesure de risque, cette partie en développera les principaux exemples avant de s'intéresser au lien entre les mesures de risque et l'espérance d'utilité. Le lecteur peut se référer à [43], [39] et [70] pour plus de détails sur les mesures de risque.

2.1 Axiomes et mesures de risque usuelles

On considère l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. Soit L^1 l'espace des fonctions mesurables à valeurs dans \mathbb{R} dont la valeur absolue est intégrable au sens de Lebesgue.

Définition 2.1. *Une mesure de risque est une application ρ définie sur L^1 et prenant ses valeurs dans \mathbb{R} . On dit qu'une mesure de risque est*

1. *monotone si $\forall X, Y \in L^1, X \leq Y$ presque-sûrement implique $\rho(X) \leq \rho(Y)$.*
2. *invariante par translation si $\forall X \in L^1$ et $\forall m \in \mathbb{R}$, $\rho(X + m) = \rho(X) + m$.*
3. *convexe si $\forall X, Y \in L^1$ et $\forall \lambda \in [0, 1]$, $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$.*
4. *positivement homogène si $\forall X \in L^1$ et $\forall \lambda \in \mathbb{R}^+$, $\rho(\lambda X) = \lambda\rho(X)$.*
5. *invariante en loi si $\forall X, Y \in L^1, X \stackrel{d}{=} Y$ implique $\rho(X) = \rho(Y)$ où $\stackrel{d}{=}$ représente l'égalité au sens de la distribution de probabilités.*

En combinant ces différentes propriétés, on peut obtenir différents types de mesures de risque. Par exemple, une mesure de risque est dite monétaire si elle possède les propriétés 1. et 2., elle est dite convexe si elle vérifie les propriétés de 1. à 3. et elle est appelée mesure de risque cohérente si elle vérifie les propriétés de 1. à 4. (voir [9]). On peut remarquer que si une mesure de risque vérifie au moins les propriétés 1., 2. et 4. alors elle est normalisée i.e. $\rho(0) = 0$.

Ces différentes propriétés peuvent être interprétées économiquement, ce qui illustre l'intérêt des mesures de risque. La propriété 1. indique que si une perte Y est plus importante qu'une perte X dans tous les états de la nature, il semble alors naturel de lui associer une valeur de risque plus importante. La propriété 2. donne une unité à la quantité $\rho(X)$, dans l'ensemble de cette thèse, cette unité est considérée comme monétaire (euros, dollars, etc.) car il s'agit de la quantité qui doit être ajoutée à la position risquée X pour la rendre acceptable du point de vue de l'agent. La propriété 3. peut se traduire comme la préférence des agents pour la diversification pour tenir compte de diverses sources de risque. La propriété 4. indique que la mesure de risque est indépendante de la taille de la position et doit être indépendante de la

devise dans laquelle elle est exprimée. On peut noter que cette propriété a été particulièrement critiquée, suggérant qu'elle pousse les individus à concentrer les risques. Il a même été suggéré que pour les valeurs de λ grandes, il faudrait avoir $\rho(\lambda X) > \lambda\rho(X)$ pour pénaliser la concentration des risques. En plus de cela, la propriété 5. indique que la mesure de risque doit être également indépendante des actifs sous-jacents et ne dépend que des distributions de probabilité des pertes. On peut remarquer qu'en combinant la propriété 3. et la propriété 4., on obtient la sous-additivité des risques. Pour plus de détails sur les propriétés et sur le contexte réglementaire, le lecteur peut se référer à [43] (Chapitre 4).

Dans le chapitre 1, les mesures de risque considérées sont cohérentes et donc en particulier convexes. Dans le chapitre 2, les mesures de risque considérées ne respectent pas la propriété de convexité. Ainsi, d'un point de vue actuariel, le chapitre 2 prend en compte les situations où la diversification du risque n'est pas réalisable. Ceci s'illustre lorsque des pertes comonotones sont étudiées.

La suite de cette section est consacrée à la présentation des principales classes de mesures de risque. Après avoir décrit succinctement la Value-at-Risk (VaR) et la Tail Value-at-Risk (TVaR) nous nous attarderons sur les mesures de risque de distorsion qui constituent l'outil principal des chapitres 1 et 2.

2.1.1 Value-at-Risk

La mesure VaR a été introduite par J.P. Morgan au début des années 1990 pour mesurer le risque de marché et est depuis au cœur de la méthodologie appelée RiskMetrics. Par la suite, la VaR a rapidement pris de l'importance dans la gestion des risques des institutions financières, devenant même, avec les accords de Bâle et Solvabilité, la mesure de référence d'un point de vue réglementaire. Le principe est le suivant, pour mesurer le risque d'une position X , on peut s'intéresser à la distribution des quantiles de queue de X . Il s'agit des quantiles issus de la fonction de survie \bar{F}_X de la variable aléatoire X .

Définition 2.2. *Pour une variable aléatoire X , la fonction de survie \bar{F}_X est définie par*

$$\bar{F}_X(x) := \mathbb{P}[X > x]. \quad (2.1)$$

Définition 2.3. *Pour $u \in (0, 1)$, le quantile de queue de niveau u d'une variable aléatoire X , noté $\bar{q}_X(u)$ est défini par*

$$\inf \{x \in \mathbb{R}, \bar{F}_X(x) \leq u\} = \bar{F}_X^{-1}(u). \quad (2.2)$$

La VaR s'intéresse donc aux probabilités d'événement rares et extrêmes.

Définition 2.4. *On appelle Value-at-Risk de niveau $u \in (0, 1)$, le quantile de queue de niveau u*

$$\text{VaR}(X, u) = \bar{q}_X(u). \quad (2.3)$$

La VaR s'interprète économiquement comme le montant minimal de capital qu'on ajoute à la position X et qu'on investit dans un actif sans risque pour garder la probabilité de perte en dessous du niveau u . Les valeurs typiques que l'on retrouve dans l'industrie pour le niveau u sont $u = 0.05$ et $u = 0.01$. Ainsi la VaR permet de contrôler la probabilité d'une perte cependant, elle ne permet pas de capturer la sinistralité de cette perte lorsqu'elle survient. De plus, bien que cette mesure de risque soit invariante par translation et homogène, elle n'est pas sous-additive. Elle peut ainsi pénaliser la diversification au lieu de l'encourager. Comme elle n'est pas cohérente, sa comparaison en pratique à une autre mesure de risque appelée Tail VaR est motivée.

2.1.2 Tail Value-at-Risk

La mesure des risques grâce à la TVaR a été popularisée par Artzner, Delbaen, Eber et Heath (voir [8] et [9]). La TVaR permet d'étudier ce qui se passe en moyenne lorsque ces événements extrêmes surviennent.

Définition 2.5. *On appelle Tail Value-at-Risk de niveau $u \in (0, 1)$, la quantité notée $\text{TVaR}(X, \alpha)$*

$$\text{TVaR}(X, u) = \frac{1}{1-u} \int_0^u \text{VaR}(X, t) dt. \quad (2.4)$$

La TVaR est ainsi définie comme la moyenne des VaR de niveau inférieur à u , comme la VaR elle est homogène et invariante par translation. Il existe de nombreuses variantes de la TVaR avec des noms différents, elles sont toutes égales pour les risques modélisés par des variables aléatoires admettant une fonction de densité par rapport à la mesure de Lebesgue (voir [4] pour plus de détails sur la relation entre ces mesures de risques). De plus, pour les risques dont la fonction de répartition est continue, la TVaR est sous-additive et donc cohérente (et est alors appelée indifféremment TVaR, Average VaR ou Expected Shortfall).

La VaR et la TVaR ont été intensivement utilisées pour une variété de problèmes différents, comme la détermination des primes d'assurance (voir [107]) ou encore l'allocation du capital (voir [98]). On doit leur utilisation dans le calcul des fonds propres notamment à Acerbi (voir [3]). Il existe un débat sur l'utilisation de la TVaR plutôt que de la VaR, notamment à cause de la cohérence de la TVaR et de sa capacité à mieux représenter les risques de queue. Cependant la VaR est plus facile à estimer notamment pour les distributions à queues épaisses.

2.1.3 Mesures de risque par distorsion

Les mesures de risque par distorsion trouvent leur origine dans la théorie duale de Yaari (voir [111] et voir Section 3.5.3). Il s'agit de mesurer les risques en appliquant une fonction dite de distorsion à la fonction de survie, pour transformer une distribution en une nouvelle. Ceci permet en particulier de pouvoir intégrer les biais comportementaux individuels (voir Section 3.5.2) ce qui sera notamment fait dans les chapitres 1 et 2. Dans le chapitre 1, la fonction de

distorsion est donnée pour un individu averse au risque qui surévalue toutes les probabilités de sinistre, elle est choisie concave. Dans le chapitre 2, la fonction de distorsion est donnée non concave et peut être en particulier prise comme décrite par Kahneman et Tversky, en forme de S inversé. C'est-à-dire que l'individu surestime les probabilités faibles et sous-estime les probabilités moyennes et élevées. Les mesures de risque par distorsion sont définies sur les variables aléatoires réelles.

Définition 2.6. *On appelle fonction de distorsion, toute fonction croissante $\psi : [0, 1] \rightarrow [0, 1]$ telle que $\psi(0) = 0$ et $\psi(1) = 1$.*

Définition 2.7. *On appelle mesure de risque par distorsion la quantité*

$$\rho(X) := \int_0^1 \bar{q}_X(1-u)d\psi(u), \quad (2.5)$$

avec ψ une fonction de distorsion.

Remarque 2.1. *La VaR et la TVaR peuvent être définies selon la Définition 2.7 avec les fonctions de distorsion respectives $\psi(u) = \mathbf{1}_{u \geq p}$ et $\psi(u) = \min\left\{\frac{u}{p}, 1\right\}$ avec p une constante comprise entre 0 et 1.*

Remarque 2.2. *On peut remarquer que dans le cas particulier où ψ est la fonction de répartition de la loi uniforme sur $[0, 1]$ de telle sorte que ψ est la fonction identité sur $[0, 1]$, $\rho(X) = \mathbb{E}[X]$. Ceci illustre la première utilisation des mesures de risque en actuariat, pour le développement des primes de risque.*

Sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sans atome, on peut prouver que les mesures de risque par distorsion sont convexes si et seulement si la fonction de distorsion les définissant est concave. Ces mesures de risque vérifient les propriétés de 1. à 4., elles sont donc cohérentes.

Exemple 2.1. *Il existe des mesures de risque qui sont cohérentes mais qui ne sont pas des mesures de risque de distorsion. La mesure de risque suivante ρ en est un exemple (voir [54]) :*

$$\rho(X) := \mathbb{E}[X] + \delta \mathbb{E}[(\mathbb{E}[X] - X)_-^{1/p}], \quad \delta \in (0, 1), \quad 1 < p < +\infty.$$

où $(\mathbb{E}[X] - X)_- = -\min\{\mathbb{E}[X] - X, 0\}$.

Les chapitres 1 et 2 de ce manuscrit reposent sur l'utilisation de mesures de risque par distorsion pour donner les critères des agents économiques. Dans le premier chapitre, les mesures de risque utilisées pour l'assuré et pour l'assureur sont caractérisées par des fonctions de distorsion concaves, ces mesures de risque sont donc cohérentes. Dans le second chapitre, la mesure de risque utilisée pour l'assuré est obtenue à partir de fonctions de distorsion non concaves, qui permet de prendre en compte en particulier les fonctions de distorsion de probabilités proposées par Kahneman et Tversky (voir Section 3.5.3). Dans ces deux chapitres, elles respectent au minimum les propriétés 1., 2., 4. et 5. de la définition 2.1. Ces propriétés permettent ainsi de décomposer la perte à laquelle fait face l'agent économique pour obtenir un problème d'optimisation en deux variables qui est décomposable et d'optimiser chaque variable séparément.

2.2 Lien entre les mesures de risque et les fonctions d'utilité

Bien que les mesures de risque représentent une théorie à part entière, il est possible de réinterpréter celle-ci en termes d'utilité espérée. La différence majeure entre ces deux approches réside dans le fait que les mesures de risque ont une unité monétaire à la différence de l'utilité espérée.

Soit ρ une mesure de risque cohérente. Pour passer de la théorie de l'utilité espérée aux mesures de risque, il faut commencer par considérer les pertes d'un individu plutôt que ses gains. Ainsi pour une position risquée X , la relation suivante

$$U(X) := -\rho(-X), \quad (2.6)$$

définit une fonction d'utilité dite monétaire. Il est important de remarquer que la famille des fonctions d'utilité monétaire ne contient pas les utilités espérées. L'unique utilité espérée étant également monétaire est linéaire et donc représente les préférences d'un agent neutre au risque. Denuit, Dhaene, Goovaerts, Kaas et Laeven proposent dans [72] des correspondances explicites entre les différentes mesures de risques et les fonctions d'utilité dans la théorie de l'utilité espérée. Ce lien entre les fonctions d'utilité et les mesures de risque permet de réinterpréter les résultats des chapitres 1 et 2 en termes d'utilité.

3 La théorie de l'utilité espérée

Dans cette section sera décrit la théorie de l'utilité espérée et ses remises en cause. La théorie de l'utilité espérée peut être considérée comme le véritable point de départ de la théorie de la décision individuelle en présence d'incertitude. Elle est largement employée dans la modélisation économique. Le chapitre 3 de ce manuscrit repose sur cette approche. En effet, l'individu considéré dans ce modèle choisit à la fois son niveau de couverture assurantielle et son niveau d'effort de prévention lui permettant de maximiser sa satisfaction traduite en termes d'utilité espérée.

Le critère de l'utilité espérée a été défini par les travaux de Bernoulli [13]. En effet, ce dernier a identifié que les individus n'estiment pas leurs gains à l'aide de l'espérance mathématique mais grâce à une fonction particulière et subjective. Les travaux de Samuelson [92], von Neumann et Morgenstern [104] ainsi que ceux de Savage [73] ont permis de rendre l'utilité espérée comme critère canonique dans la théorie de la décision avec incertitude. Malgré le caractère normatif du modèle de l'utilité espérée, différentes limites ont été mises en évidence et ont ouvert la voie à une remise en cause de ce modèle au profit de modèles alternatifs, qui seront développés à la fin de cette section.

3.1 De la naissance du critère d'utilité espérée à une théorie axiomatique du choix sous incertitude

A l'origine de la théorie de l'utilité espérée se trouve la résolution du paradoxe de Saint Petersbourg par Daniel Bernoulli. Le paradoxe de Saint Petersbourg correspond à un jeu de lancer de pièce à pile ou face, jusqu'à ce que le côté "face" apparaisse. Le joueur reçoit une unité monétaire si "face" apparaît du premier coup, en notant n le nombre de lancers nécessaires pour finir le jeu, le gain du joueur sera donné par 2^{n-1} unités monétaires si "face" apparaît au n -ème coup. La probabilité que le côté "face" apparaisse au bout de n lancers est de $(\frac{1}{2})^n$. Daniel Bernoulli (voir [13]) souligne alors que l'espérance mathématique de gain étant infiniment grande, un individu rationnel devrait accepter de payer une somme importante pour participer à ce jeu alors que personne ne semble vouloir participer pour un prix modéré. Il émet alors la solution suivante : il existe une évaluation subjective des gains monétaires qui repose sur l'utilité qu'un objet rapporte. Il existe donc une fonction d'utilité cardinale subjective qui dépend de la psychologie de l'individu. Cette fonction, croissante et concave, va remplacer le gain lui-même dans le calcul de l'espérance et se base sur les conséquences certaines d'une loterie. Le travail de Bernoulli reste cependant incomplet car il ne propose pas une manière de mesurer l'utilité et de plus, la rationalité de l'espérance d'utilité n'est pas prouvée.

Malgré cette découverte précoce, une construction axiomatique précise ne fut proposée par von Neumann et Morgenstern qu'en 1944 (voir [104]). Ils définissent une relation dite de préférence (précisée dans ce qui suit) sur des distributions de probabilités à laquelle une série d'axiomes est associée. Ces axiomes permettent ainsi de poser les bases d'un choix rationnel en situation incertaine et de représenter le critère de l'utilité espérée comme un critère portant sur les préférences. La fonction d'utilité présentée par von Neumann et Morgenstern diffère donc de celle de Bernoulli qui repose sur la psychologie de l'agent économique. Par la suite Savage (voir [73]) précisera les conditions sous lesquelles les préférences définies par un profil de paiement sur un espace de fonctions donné, admettent une représentation sous la forme de l'espérance mathématique d'une fonction.

3.2 Représentation de von Neumann Morgenstern

Cette partie a pour but de présenter les fondements théoriques de l'utilité espérée définie par von Neumann et Morgenstern. Après avoir précisé les axiomes de la théorie de l'utilité espérée, le théorème de représentation ainsi que les propriétés seront détaillés avant de mentionner les différents types de fonctions d'utilité. Pour plus de détails sur ces notions, le lecteur peut se référer à [43].

La notion de préférence étant le fondement des travaux de von Neumann et Morgenstern, il est essentiel de l'introduire pour comprendre l'axiomatisation donnée dans [104]. Etant donné \mathcal{A} un ensemble non vide, un choix possible pour un agent économique est représenté comme

un élément $a \in \mathcal{A}$. Face à deux choix possibles a et b , l'agent peut avoir une préférence pour l'un par rapport à l'autre. Cela se formalise par ce qui est appelé une relation de préférence définie comme suit

Définition 3.1. *Une relation de préférence sur \mathcal{A} est une relation binaire, notée \succ avec les propriétés suivantes*

1. *Asymétrie : si $a \succ b$ alors $b \not\succ a$.*
2. *Transitivité négative : si $a \succ b$ et $c \in \mathcal{A}$ alors c vérifie $a \succ c$ ou $c \succ b$ ou les deux.*

Remarque 3.1. *La propriété de transitivité peut être interprétée de la manière suivante. S'il existe une préférence claire entre deux choix a et b , lorsqu'un troisième choix c est ajouté alors il existe encore un choix qui est le moins désirable (b dans le cas où $c \succ b$) ou le plus désirable (a dans le cas où $a \succ c$).*

Il est également possible de définir une relation de préférence dite faible, représentée par $a \succeq b \iff b \not\succ a$ et une relation d'indifférence donnée par $a \sim b \iff a \succeq b$ et $b \succeq a$. Ainsi $a \succeq b$ peut être interprété soit comme a est préféré à b soit il n'existe pas de préférence claire entre ces deux choix.

Remarque 3.2. *Il est possible de vérifier que 1. et 2. sont équivalentes aux deux propriétés suivante de l'ordre de préférence faible \succeq*

3. *Complétude : pour tous a, b dans \mathcal{A} , $a \succeq b$ ou $b \succeq a$ ou les deux sont vérifiées.*
4. *Transitivité : si $a \succeq b$ et $b \succeq c$ alors $a \succeq c$.*

Inversement, la négation de \succeq entraîne un ordre de préférence faible pour une relation de préférence faible qui est complète et transitive.

Remarque 3.3. *On peut remarquer que la relation d'indifférence est une relation d'équivalence car elle est réflexive, symétrique et transitive.*

Passons maintenant à ce qui va permettre de définir une fonction d'utilité

Définition 3.2. *On appelle représentation numérique de la relation de préférence \succ , une fonction $U : \mathcal{A} \rightarrow \mathbb{R}$ telle que*

$$a \succ b \iff U(a) > U(b).$$

Remarque 3.4. *Il est possible de définir de manière similaire une représentation numérique de la relation de préférence faible*

$$a \succeq b \iff U(a) \geq U(b).$$

Remarque 3.5. *Il est important de noter qu'une telle représentation n'est pas unique. En effet, pour toute fonction f strictement croissante $f(U(a))$ est une représentation numérique d'une relation de préférence.*

Le théorème suivant permet de caractériser pour quelles relations de préférences il existe une représentation numérique. La preuve peut être trouvée dans [43].

Théorème 3.1. *Une relation de préférence admet une représentation numérique si \mathcal{A} est dénombrable.*

Reprenons l'ensemble \mathcal{A} des choix possibles pour un agent économique et considérons l'espace \mathcal{P} de toutes les distributions de probabilités possibles pour tous les choix offerts à l'agent, il est supposé convexe. Dans la théorie du choix, les éléments μ de \mathcal{P} sont appelés des loteries. Ils peuvent également être nommés perspectives risquées.

Définition 3.3. *Une représentation U de la relation de préférence \succ sur \mathcal{P} est dite représentation de von Neumann Morgenstern (VNM) si elle est de la forme*

$$U(\mu) = \int u(x)\mu(dx), \quad \forall \mu \in \mathcal{M}$$

où u est une fonction réel.

Toute représentation VNM est affine dans le sens suivant, pour tous $\mu, \nu \in \mathcal{P}$ et $\beta \in [0, 1]$,

$$U(\beta\mu + (1 - \beta)\nu) = \beta U(\mu) + (1 - \beta)U(\nu).$$

Présentons maintenant les axiomes qu'une telle relation de préférence respectent

Axiome d'indépendance : *Pour tout $\mu, \nu \in \mathcal{P}$,*

$$\mu \succ \nu \implies \beta\mu + (1 - \beta)\eta \succ \beta\nu + (1 - \beta)\eta,$$

pour tout $\eta \in \mathcal{P}$, pour tout $\beta \in [0, 1]$.

Cet axiome s'interprète comme la conservation de la préférence entre deux loteries par toute combinaison convexe avec une troisième loterie indépendante notée β ici. Ceci s'explique par le fait que $\beta\mu + (1 - \beta)\eta$ et $\beta\nu + (1 - \beta)\eta$ peuvent être considérées comme des loteries particulières dite loteries composées. En effet, en regardant β comme une probabilité, on obtient alors une distribution de probabilités sur des loteries (et non sur des choix) et les deux loteries composées peuvent alors être vues comme des loteries en deux étapes. En premier lieu, les loteries μ et η (ou ν et η) sont tirées avec respectivement probabilité β et $1 - \beta$. En second lieu, la loterie tirée se réalise. Ceci est équivalent à jouer directement à la loterie $\beta\mu + (1 - \beta)\eta$ ou $\beta\nu + (1 - \beta)\eta$. L'unique différence entre ces deux loteries réside dans le tirage de μ ou ν avec probabilité β . Ainsi si $\mu \succ \nu$ alors il semble raisonnable de préférer la loterie composée avec μ à celle avec ν .

Axiome archimédien¹ : *Pour tout triplet (μ, ν, η) tel que $\mu \succ \eta \succ \nu$, il existe $\beta, \delta \in (0, 1)$ tels que*

$$\beta\mu + (1 - \beta)\nu \succ \eta \succ \delta\mu + (1 - \delta)\nu.$$

1. Le nom de cet axiome trouve son origine dans sa similitude avec le principe d'Archimède en algèbre et en analyse réelle.

Cet axiome montre que si une première loterie μ est préférée à une deuxième η , il n'existe pas de troisième loterie ν , à laquelle on préfère η , pour laquelle il existe une combinaison probabiliste ν qui serait préférée à η .

Maintenant que les axiomes d'indépendance et archimédien ont été présentés, le théorème de représentation de l'utilité espérée peut être écrit. Cet énoncé donne les conditions d'existence d'une représentation affine numérique pour une relation de préférence satisfaisant les axiomes précédents.

Théorème 3.2. *Supposons que \succ est une relation de préférence sur \mathcal{P} satisfaisant l'axiome d'indépendance et l'axiome archimédien. Il existe alors une représentation affine numérique notée U de \succ . De plus, cette représentation est unique à une transformation affine positive près.*

Remarque 3.6. *Il est important de noter que la propriété affine de la représentation numérique n'implique pas qu'il s'agit d'une représentation VNM. Il existe deux cas dans lesquels les représentations numériques affines seront toujours des représentations VNM. Le premier cas concerne toute relation de préférence sur l'ensemble \mathcal{P} , lorsque cet ensemble est égal à l'ensemble de toutes les distributions de probabilités qui s'écrivent comme combinaison finie et convexe de masses de Dirac. De telles distributions sont alors dites simples. Le deuxième cas correspond à toute relation de préférence sur \mathcal{P} , égal à un ensemble fini. En effet, sur un ensemble fini, toutes les distributions de probabilités sont simples.*

Dans le cas où l'ensemble \mathcal{P} est infini, la représentation VNM peut ne pas exister comme l'exemple suivant le montre.

Exemple 3.1. *Sur l'ensemble des entiers strictement positifs, considérons l'ensemble de distributions de probabilité $\mu \in \mathcal{P}$ pour lesquelles $U(\mu) := \lim_{n \rightarrow \infty} n^2 \mu(n)$ est un nombre fini réel. Il est possible de montrer que U est affine et induit une relation de préférence sur \mathcal{P} tout en satisfaisant les axiomes d'indépendance et archimédien. Cependant U n'est pas une représentation VNM.*

Il est possible d'obtenir une représentation VNM en rajoutant une hypothèse de continuité à la relation de préférence \succ , au sens topologique.

Définition 3.4. *Soit \mathcal{A} un espace topologique. Une relation de préférence \succ est dite continue si pour tout $a \in \mathcal{A}$, les ensembles*

$$\mathcal{B}_\succ(a) := \{b \in \mathcal{A} | b \succ a\} \quad \text{et} \quad \mathcal{B}_\preceq(a) := \{b \in \mathcal{A} | a \succ b\}$$

sont des sous-ensembles ouverts de \mathcal{A} .

Considérons l'espace \mathcal{P} , de toutes les distributions de probabilité sur un espace métrique séparable, muni de la topologie faible et de la tribu borélienne associée. Le théorème suivant donne les conditions d'existence d'une représentation VNM sur un tel ensemble.

Théorème 3.3. *Soit \mathcal{P} l'espace de toutes les distributions de probabilité muni de la topologie faible et soit \succ une relation de préférence continue sur \mathcal{P} satisfaisant l'axiome d'indépendance. Il existe alors une représentation VNM pour tout fonction u à valeur réelles qui est bornée et continue*

$$U(\mu) = \int u(a)\mu(da).$$

U et u sont unique à une transformation affine près.

Ce théorème étant restreint aux fonctions bornées, il ne présente pas un cadre assez général. Il est possible d'en élargir les conditions. Plusieurs approches sont possibles et ne seront pas détaillées ici. Deux sont présentées dans [43].

Avant de présenter la représentation de Savage, il convient de présenter une notion qui sera retrouvée au chapitre 3. Pour une relation de préférence donnée \succ sur \mathcal{P} qui admet une représentation VNM

$$U(\mu) = \int u d\mu,$$

avec u une fonction réelle strictement croissante et continue.

Définition 3.5. *On appelle équivalent certain de la loterie $\mu \in \mathcal{P}$ par rapport à une fonction d'utilité u , l'unique quantité réelle notée $EC(\mu)$ et vérifiant*

$$u(EC(\mu)) = U(\mu) = \int u d\mu.$$

Remarque 3.7. *Le théorème des valeurs intermédiaires appliqué à u donne l'existence de l'équivalent certain. Cette quantité s'interprète de la manière suivante, un individu est indifférent entre la loterie μ et le montant certain $EC(\mu)$.*

L'équivalent certain peut être interprété comme la borne supérieure de tout montant μ qui serait acceptable pour un agent économique utilisant la fonction d'utilité u . A partir de l'équivalent certain, on peut notamment définir la prime de risque, notée $p(\mu)$ par

$$p(\mu) := \int x\mu(dx) - EC(\mu),$$

avec $\int x\mu(dx)$ l'espérance, qui peut être également interprété, d'un point de vue assurantiel, comme la prime actuarielle. La prime de risque correspond alors à la quantité que l'individu est prêt à payer pour remplacer la partie risquée d'une loterie par son espérance mathématique. Ce nombre est toujours positif, voire strictement si la distribution de la loterie ne présente aucun risque.

3.3 Représentation de Savage

La théorie de l'utilité espérée en présence d'incertitude a d'abord été développée au travers des travaux de deux mathématiciens, Ramsey (voir [85]) et de Finetti [29]. Ramsey offre une méthode permettant de déterminer conjointement l'utilité et les probabilités. Quant à de Finetti, il travailla sur la conception des probabilités dites subjectives en définissant cette notion ainsi qu'en précisant les significations des lois. Inspiré par ces travaux et en se basant sur la représentation de von Neumann et Morgenstern, Savage (voir [73]) va développer une théorie axiomatique de l'utilité espérée subjective. Il s'agit de construire une fonction d'utilité et une mesure de probabilité subjective basées sur les préférences individuelles.

Sur un espace mesurable $(\Omega, \mathcal{F})^2$, avec \mathcal{F} la famille de filtrations associée, soit \mathcal{X} l'ensemble des fonctions mesurables et bornées notées X . Dans ce cadre-ci, les fonctions X sont considérées comme des actifs financiers. De plus, on suppose que \mathcal{X} est muni de la relation de préférence \succ .

Remarque 3.8. *Il est important de remarquer qu'aucune mesure de probabilité a priori n'est fixé sur l'espace mesurable car on considère cette fois-ci les décisions en présence d'incertitude et non de risque.*

Une hypothèse supplémentaire de monotonie sur la relation de préférence \succ est faite

$$Y \succeq X \text{ si } Y(\omega) \geq X(\omega) \text{ pour tout } \omega \in \Omega.$$

Savage formula une série d'axiomes supplémentaires pour garantir l'existence d'une représentation numérique dite représentation de Savage décrite de la manière suivante

$$U(X) = \mathbb{E}^{\mathbb{P}}[u(X)] = \int u(X(\omega))\mathbb{P}(d\omega), \quad \text{pour tout } X \in \mathcal{X}, \quad (3.1)$$

où \mathbb{P} est une mesure de probabilité sur (Ω, \mathcal{F}) et u une fonction à valeur dans \mathbb{R} . La mesure de probabilité \mathbb{P} est implicite à la relation de préférence, elle est donc subjective dans le sens où elle reflète les probabilités que l'individu assigne lui-même aux événements. L'hypothèse de monotonie précédente est équivalente à supposer que u est une fonction croissante.

Remarque 3.9. *Il est possible de retrouver la représentation VNM à partir de la représentation de Savage. En effet, soit $\mu_{\mathbb{P},X}$ une distribution de X sous \mathbb{P} , la relation de préférence \succ sur \mathcal{X} donnée par (3.1), induit une relation de préférence sur $\mathcal{A}_{\mathbb{P}} := \{\mu_{\mathbb{P},X} | X \in \mathcal{X}\}$ avec la représentation VNM donnée par $U_{\mathbb{P}}(\mu_{\mathbb{P},X}) := U(X) = \mathbb{E}^{\mathbb{P}}[u(X)] = \int u d\mu_{\mathbb{P},X}$.*

2. On considère maintenant l'ensemble des mesures de probabilités, et non plus les distributions de probabilités

3.4 Les types de fonctions d'utilité

Différents types de fonctions d'utilité sont décrits dans la littérature économique. Ces fonctions sont classées selon un paramètre propre à l'individu et à son attitude face au risque. Cette notion a été introduite par Pratt [82] et traduit le comportement humain vis-à-vis de l'incertitude. Trois types de comportements peuvent être distingués

1. l'aversion au risque pour une personne qui préfère accepter un paiement certain plutôt que de participer à une loterie, de moyenne égale au paiement certain.
2. la neutralité face au risque pour une personne qui est indifférente entre avoir un paiement certain et participer à une loterie, de moyenne égale au paiement certain.
3. la recherche du risque pour une personne qui préfère participer à la loterie plutôt que d'avoir le paiement certain même quand celui-ci est supérieur au gain moyen de la loterie.

Ces comportements se traduisent d'un point de vue mathématique au travers de la forme de la fonction d'utilité. L'aversion pour le risque se traduira par une fonction concave, l'attrait pour le risque se traduira quant à lui par une fonction convexe et pour finir, l'indifférence au risque se retrouve au travers d'une fonction linéaire.

Définition 3.6. *Pour une fonction d'utilité u deux fois différentiable, continue, de dérivées continues et strictement croissante. Alors*

$$\gamma_A(x) := -\frac{u''(x)}{u'(x)},$$

est appelé le coefficient d'aversion absolue au risque d'Arrow-Prat de niveau x .

Il existe également un coefficient d'aversion relative au risque de niveau x , noté γ_R , défini comme $\gamma_R(x) = x\gamma_A(x)$. On peut alors donner une classification des fonctions d'utilité selon les valeurs de ces coefficients.

1. Fonction CARA (*constant absolute risk aversion*)

$$u(x) = e^{-\gamma_A x}.$$

2. Fonction HARA (*hyperbolic absolute risk aversion*) ou CRRA (*constant relative risk aversion*)

$$u(x) = \frac{x^{1-\gamma_R}}{1-\gamma_R}$$

avec $\gamma_R > 0$, $\gamma_R \neq 1$.

3. Fonction DARA (*decreasing absolute risk aversion*)

$$u(x) = \ln(x).$$

4. Fonction IARA (*increasing absolute risk aversion*).

Remarque 3.10. Pour $\gamma_R = 1$, $u(x) = \ln(x)$ définit une fonction CRRA et une fonction DARA.

Remarque 3.11. Les fonctions dites DARA ou IARA sont considérées comme plus cohérentes avec les faits empiriques.

Dans le chapitre 3, les préférences de l'agent sont données par une fonction d'utilité CARA exponentielle. Grâce aux propriétés de la fonction exponentielle, on pourra alors décrire le problème de maximisation de l'utilité espérée de l'agent grâce à l'approche de [26] et de [?].

3.5 Limites et critiques

Comme montré dans ce qui précède, la théorie de l'utilité espérée repose sur le respect des axiomes d'indépendance et archimédien pour une relation de préférence. Cependant, de nombreux articles ont remis en cause le fait que les individus agissent selon ces préceptes. Dans cette section, il s'agira d'aborder les principales limites de cette théorie, notamment au niveau expérimental ainsi qu'au travers des biais comportementaux qui ont été mis en évidence par Kahneman (voir [56]). Ensuite, il sera question de présenter des théories alternatives qui ont été développées par la suite. Ces limites sont à l'origine des modélisations proposées dans les chapitres 1 et 2. Un des objectifs de ces chapitres est de proposer un nouveau critère de choix dans le domaine des risques en incluant pour le chapitre 2, les avancées apportées par l'étude des biais comportementaux dans l'évaluation des perspectives risquées.

3.5.1 L'expérimentation comme remise en cause

En 1953, Allais propose une critique de la théorie de l'utilité espérée de von Neumann et Morgenstern, dans un discours prononcé à une conférence de l'American Economic Society, qui fut ensuite publié sous forme d'article scientifique (voir [5]). Ce discours, communément appelé le paradoxe d'Allais, offre une réfutation empirique de l'axiome d'indépendance dans le cadre de VNM. Il souligne la préférence des individus pour la sécurité au voisinage de la certitude. L'expérimentation d'Allais montre que la présence d'un gain qui devient très incertain dans l'alternative proposée conduit un grand nombre de personnes à ne pas se conformer à cet axiome. Donnons une illustration de ce phénomène grâce à un jeu en deux étapes. Premièrement, des individus se retrouvent face aux deux loteries suivantes

$$\begin{aligned}\mu_1 &= 0.20\delta_{500} + 0.79\delta_{400} + 0.01\delta_0, \\ \nu_1 &= \delta_{400}.\end{aligned}$$

Avec la loterie μ_1 , un individu gagne 500 avec une probabilité de 0.20, 400 avec une probabilité de 0.79 et 0 avec une probabilité de 0.01. Avec la loterie ν_1 , un individu est sûr de gagner 400. La majorité des personnes faisant face à ces deux loteries préféreront avoir 400 de manière

sûre plutôt que de participer à la loterie incertaine, ayant pourtant une espérance de gain plus grande, égale à 416.

Deuxièmement, les mêmes personnes se retrouvent alors devant les deux loteries suivantes

$$\begin{aligned}\mu_2 &= 0.21\delta_{400} + 0.79\delta_0, \\ \nu_2 &= 0.20\delta_{500} + 0.80\delta_0.\end{aligned}$$

Avec la loterie μ_2 , un individu gagne 400 avec une probabilité de 0.21 et 0 avec une probabilité de 0.79. Avec la loterie ν_2 , un individu gagne 500 avec une probabilité de 0.21 et 0 avec une probabilité de 0.80. Dans ce cas-ci, les individus tendent à préférer la loterie plus risquée ν_2 , ce qui est en adéquation avec le fait que son espérance d'utilité est plus élevée que celle de μ_2 (elles sont respectivement égales à 84 et 100).

Allais ne remet pas en cause la théorie de l'utilité espérée dans son ensemble, il démontre néanmoins expérimentalement que lorsque le risque est extrême, le joueur se focalise davantage sur la prime de risque. Les implications du paradoxe d'Allais donneront lieu à de multiples développements en théorie de la décision et en économie comportementale. Ces résultats expérimentaux ont été confirmés par Kahneman et Tversky (voir [59]).

En 1961, Ellsberg (voir [38]) identifia un autre jeu remettant en cause l'axiome d'indépendance dans le cadre de Savage. Le jeu est comme suit. Un individu se retrouve en face de deux urnes, contenant chacune 100 balles qui sont soit rouges, soit noires. Dans la première urne, la proportion de balles rouges est connue et est égale à p . Dans la seconde urne, la proportion \tilde{p} de boules rouges n'est pas connue.

L'individu se retrouve face au choix suivant, il doit élire une urne et tirer une boule dedans. Dans le premier cas, si cette boule est rouge alors il gagne 100 euros, si elle est noire alors il gagne 0 euro. Ellsberg montre que les agents vont alors choisir la première urne pour laquelle ils connaissent la distribution de boules rouges. Dans le deuxième cas, si la boule tirée est noire alors l'individu gagne 100 euros, sinon il gagne 0 euro. Il est alors mis en évidence que les individus choisissent encore une fois la première urne pour laquelle la distribution de boules rouges est connue.

Ceci est en contradiction avec l'axiome d'indépendance présenté par Savage [73]. En effet, pour toute probabilité subjective \tilde{p} de tirer une boule rouge dans la deuxième urne, le premier choix implique $p > \tilde{p}$, le second choix implique $1 - p > 1 - \tilde{p}$ alors que les individus devraient préférer l'inverse.

Depuis les paradoxes d'Allais et d'Ellsberg, de nombreuses analyses ont montré des contradictions des axiomes de la théorie de l'utilité espérée. C'est notamment le cas des travaux

de Kahneman et Tversky (voir [57] et [59]), montrant que les individus ne traitent pas parfaitement l'information disponible. Les auteurs mettent en évidence que la prise de décision risquée répond à des mécanismes cognitifs simplistes qui permettent aux individus d'accepter une réponse simple mais pas toujours correcte. De plus, le jugement des probabilités répond également à des stratégies cognitives qui mènent à de nombreuses erreurs. Ceci mènera au développement d'une théorie alternative appelée théorie des perspectives qui sera développée dans la partie suivante. Des expérimentations en laboratoire (voir [80]) et sur le terrain (voir [30]), ont permis de mettre en évidence trois types de déviations par rapport à la théorie de l'utilité espérée. Tout d'abord, au niveau des préférences individuelles, celles-ci interviennent directement dans la fonction d'utilité (ou dans son actualisation) en affectant un résultat. Ensuite, au niveau des probabilités subjectives, la perception d'un individu des distributions de probabilités des états de la nature est erronée. L'évaluation subjective des probabilités conduit donc l'individu à avoir des croyances biaisées. Enfin l'individu fait face à des biais systématiques dans son processus de prise de décision. Il va préférer édicter des règles simples dans sa réflexion plutôt que de procéder à des calculs complexes.

3.5.2 Biais comportementaux

Cette partie a pour but de présenter l'une des principales limites de la théorie de l'utilité espérée qui est à l'origine de la modélisation du risque proposée dans le chapitre 2 de ce travail de thèse. L'utilité espérée repose sur le fait que les individus sont rationnels, logiques par rapport à la réalité et qu'ils sont impartiaux dans leurs processus de pensées. Cependant, la plupart, voire l'ensemble, des individus ne perçoivent pas la réalité de manière objective. En effet, Kahneman et Tversky ont montré (voir [100]) que les individus répondent à différentes déviations systématiques appelées biais cognitifs notamment dans le domaine économique. Ces biais sont inconscients et se traduisent par une évaluation différente de faits similaires par des individus différents. L'interprétation de ces biais fait débat dans la littérature, pour certains auteurs, ils se reflètent dans les erreurs de raisonnement ou de jugement d'un agent. Pour Kahneman et Tversky, ils peuvent toutefois conduire à des jugements raisonnables et donc ne sont pas entièrement négatifs.

Pour expliquer comment les biais s'introduisent dans le raisonnement humain, Simon (voir [95]) introduit la notion d'heuristique. Il présente cela comme des règles qui permettent de considérer un problème en omettant sa complexité, donnant une réponse satisfaisante bien qu'approximative. En 1974, la notion d'heuristique trouve une nouvelle définition avec Kahneman et Tversky (voir [100]). Il s'agit alors de toutes les règles qui régissent le traitement de l'information par le cerveau. Elles sont automatiques, intuitives et rapides et sont considérées comme des raccourcis mentaux. Elles simplifient le processus de réflexion pour une prise de décision rapide face à l'incertitude et préserver les efforts des individus. Dans son livre [56], Kahneman explique le fonctionnement du cerveau humain lors du processus de prise de

décision. Il existe deux systèmes opposés

Système 1 : il est basé sur l'intuition. Il est automatique et rapide, demande peu d'effort et est involontaire. Il associe les idées par relation de cause à effet et crée un flux de pensées. Bien que ces associations soient cohérentes, elles ne représentent pas une réflexion logique. Ces associations reposent en réalité sur l'expérience personnelle de l'individu. Ce système est donc celui qui donne naissance aux biais cognitifs de l'individu.

Système 2 : il est basé sur l'évaluation. Il est analytique et plus lent que le système 1, qu'il contrôle. Il est utilisé pour la résolution de problèmes complexes et rentre en jeu par le biais de la concentration et de l'effort. Selon le contrôle qu'il exerce sur le système 1, le système 2 est alors parfois enclin à accepter les conclusions de ce dernier. L'individu questionnera alors moins ses intuitions et n'évitera pas les biais du système 1.

Différentes heuristiques ont été mises en avant dans la littérature, la référence principale étant le livre de Kahneman [56]. Dans la suite seront présentés les principales heuristiques et les principaux biais associés développés dans la littérature. Seuls ceux en lien avec la formulation du critère de préférence du chapitre 2 seront détaillés. Tout d'abord, deux types d'heuristiques se distinguent

Heuristiques statistiques : les individus ont une compréhension partielle des statistiques et s'en servent pour poser des jugements et prendre des décisions.

Heuristiques non-statistiques : au contraire, les individus négligent l'information statistique et privilégient l'information individualisante.

Les heuristiques les plus détaillées dans la littérature sont de nature non-statistiques.

L'heuristique de représentativité, décrit par Kahneman et Tversky (voir [100]), est probablement la plus connue. Il s'agit d'une règle permettant de catégoriser des événements par classe en estimant la probabilité pour un événement d'appartenir à une classe donnée. On peut noter par exemple, le classement social des individus selon les stéréotypes véhiculés dans la société. Plusieurs biais cognitifs sont associés à l'heuristique de représentativité.

Le premier est le biais d'ignorance du taux de base qui consiste à ignorer la distribution de probabilités sur un échantillon. Ceci se retrouve dans l'expérimentation suivante faite par Kahneman et Tversky. Les participants à l'étude sont divisés en deux groupes égaux et deux traitements sont effectués. Le premier traitement consiste à présenter à un groupe le problème suivant : sur 100 personnes, il y a 30% d'ingénieurs et 70% d'avocats, quelle est la probabilité qu'un individu au hasard soit ingénieur ? A l'autre moitié, les distributions des ingénieurs et des avocats étaient inversées. Dans ce cas-là, les individus tiennent compte de la distribution pour placer l'individu tiré au hasard parmi les ingénieurs. Ensuite, un deuxième traitement est proposé. Cette fois-ci, en plus des distributions de probabilités, les participants reçoivent une

description d'un individu qui correspond au stéréotype d'un ingénieur. Peu importe la distribution sur les 100 personnes, les deux groupes estiment qu'il y a une plus grande probabilité que l'individu soit ingénieur.

Le second biais associé à l'heuristique de représentativité est l'insensibilité à la taille de l'échantillon. Kahneman et Tversky montrent dans [99], que les personnes croient que les caractéristiques d'un processus aléatoires sont présentes à la fois dans l'entièreté d'une séquence et dans chaque sous-séquence. Par exemple, on considère deux maternités, une grande avec 45 naissances par jour et l'autre petite avec 15 naissances par jour. Il y a en moyenne 50% de naissance de garçons dans les deux maternités. Lorsqu'il est demandé à des individus d'estimer quel hôpital compte le plus de jours avec un taux de naissance masculine qui dépasse les 60%, les individus pensent que le nombre de jours est équivalent dans les deux hôpitaux, alors que celui-ci sera supérieur dans le petit hôpital du fait du petit échantillonnage. La mauvaise compréhension du hasard influe donc les individus et les empêche de tenir compte de la taille de l'échantillon. Un autre exemple développé par Rogers (voir [86]) sur les lancers de pièces valide l'existence de ce biais. Ceci est également confirmé par DellaVigna (voir [30]) qui montre que pour deux événements se réalisant successivement, un agent a tendance à conclure que celui qui se réalise en deuxième suit la même loi de probabilité que le premier.

Un troisième biais lié à l'heuristique de représentativité est celui de l'insensibilité à la prévisibilité. Les jugements tiennent peu compte de la fiabilité des informations à la disposition de l'individu. Les personnes ont tendance à être sur-confiantes par rapport à leurs décisions, ce qui entraîne une illusion de validité qui est produite par une bonne combinaison entre les informations disponibles et le résultat prédit.

Le dernier biais présenté en lien avec l'heuristique de représentativité est le biais d'optimisme. Ce biais a été notamment proposé par Shepperd, Carroll, Grace et Terry (voir [20]). Il décrit le fait que les individus ont tendance à surestimer la probabilité d'un événement positif et à sous-estimer la probabilité d'un événement négatif par rapport à la distribution de probabilités objectives et par rapport aux autres individus. Une illustration populaire de ceci serait donc la phrase "cela n'arrive qu'aux autres". L'origine de ce biais se trouve dans la distorsion que les individus appliquent aux probabilités objectives. Jacquemet, Rullière et Vialle (voir [51]) ont proposé une expérimentation pour mettre en évidence ce biais : un individu a la possibilité de prendre un risque ou non et de mesurer comment ses choix sont influencés par de l'optimisme ou du pessimisme. Deux traitements sont alors effectués. Le premier consiste en une loterie individuelle rendant l'information publique ou non. Le deuxième représente également une loterie mais avec d'autres individus, une personne sur dix est tirée et son choix de prendre un risque ou non est rendu public aux autres individus. Pour des individus rationnels, aucune différence ne devrait être observée entre les deux traitements. Empiriquement, il est trouvé que les individus ont une illusion de contrôle dans le second traitement car la probabilité d'être contrôlé pour un individu, est sous-estimée par rapport à celle du reste du groupe alors

que dans la loterie individuelle, la probabilité de contrôle est traitée comme une variable non contrôlable.

Une autre heuristique décrite par Kahneman et Tversky dans [100], est l'heuristique de disponibilité. Il s'agit de la capacité d'un individu à évaluer la probabilité qu'un événement se produise en se basant sur la facilité avec laquelle il est présent dans la mémoire. Les individus vont se rappeler mieux d'un événement qu'ils considèrent comme fréquent du fait de sa disponibilité dans la mémoire. Ceci donne lieu à plusieurs biais également. Par exemple, Slovic, Fishoff et Liechtenstein (voir [96]) ont montré dans leur étude expérimentale que les individus pensaient que la mort par homicide est plus fréquente que la mort par cancer de l'estomac bien que ce soit l'inverse en réalité. La médiatisation des homicides les rend plus disponibles dans la mémoire. Ceci est le biais des souvenirs qui influe à penser qu'un événement facilement accessible en mémoire a lieu plus fréquemment. Ceci rejoint également le biais de recherche en mémoire montré par Tversky et Kahneman (voir [58]). Les individus soumis à ce biais ont l'impression que plus il est facile de penser à un élément, plus il est plus fréquent.

Une troisième heuristique décrite par Kahneman et Tversky, est l'heuristique d'ancrage et d'ajustement. Il apparaît quand un individu se focalise sur un aspect particulier d'un événement ce qui influence son évaluation des probabilités. L'agent est alors influencé par l'information qui est présentée en premier ou qu'il connaît déjà et ne pourra se défaire de sa première impression sur l'événement. L'origine de cette heuristique n'a pas de consensus dans la littérature et pourrait être multifactorielle avec la personnalité, l'humeur et l'expérience de l'individu entre autres. Cet effet d'ancrage implique notamment le biais d'évaluation subjective d'événement (voir par exemple [81] et [100]). Les individus évaluent une distribution de probabilités en basant leurs réponses sur la réponse la plus accessible qui est le plus souvent la moyenne pour eux.

La dernière heuristique présentée ici sera l'heuristique de la pseudo-certitude. Cette heuristique fut identifiée par Tversky et Kahneman (voir [101]) via une série de questions à deux étapes. Ils ont mis en évidence que les individus ont tendance à percevoir certains résultats comme des événements certains alors qu'ils ne le sont objectivement pas. Dans des processus de décisions à étapes multiples, les personnes agissent comme si les étapes étaient indépendantes les unes des autres. Ainsi, face à une étape donnée, elles ne prennent pas en compte l'incertitude des étapes précédentes alors que la probabilité du résultat final est conditionnée à la réalisation des étapes précédentes.

3.5.3 Théories alternatives

De nombreuses théories alternatives à la théorie de l'utilité espérée ont été développées (voir [97] et [17] pour des revues de littératures complètes). Elles se placent également dans un contexte de risque et ont cherché à généraliser l'espérance d'utilité. Trois d'entre-elles seront

développées ici, la théorie des perspectives de Kahneman et Tversky (voir [59]), la théorie de dépendance des rangs de Quiggin (voir [84]) ainsi que la théorie duale de Yaari (voir [111]). Le choix de ne développer que ces trois théories réside dans le fait que le chapitre 2 reprend l'élément essentiel de la théorie dite des perspectives cumulées (voir [102]) : la distorsion des probabilités perçues par les individus. Cette théorie des perspectives cumulées trouve naissance dans la révision de la théorie des perspectives de Kahneman et Tversky suite à la sortie de la théorie de Quiggin. Quant à la théorie duale de Yaari, elle est à l'origine de la représentation d'une classe de mesures de risque développée Section 2.1.3. Cette classe permet de définir les critères de préférences des agents économiques considérés dans les chapitres 1 et 2.

L'une des théories majeures qui a fait suite à la théorie de l'utilité espérée, est la théorie des perspectives développée par Kahneman et Tversky en 1979. Cette théorie se veut représentative du processus mental de la prise de décision. Elle intègre à la fois une théorie économique et une théorie cognitive sur le traitement de l'information. Ceci se fait au travers des deux phases consécutives suivantes :

Phase d'édition : il s'agit de la phase de préparation qui vise à trier, organiser, reformuler et simplifier les perspectives offertes, pour en faciliter l'évaluation. L'individu va alors avoir une représentation simple de la perspective.

Phase d'évaluation : il s'agit de la phase d'évaluation de l'utilité subjective des perspectives simplifiées. Celle qui aura la plus grande utilité subjective sera choisie, selon un principe de maximisation similaire à celui de la théorie de l'utilité.

La phase d'édition est certainement la caractéristique qui fait la particularité de cette théorie de prise de décision. En effet, cette phase suppose que le problème n'est pas présenté objectivement à l'individu mais que ce dernier l'interprète selon son expérience personnelle. Ainsi deux problèmes présentant des résultats ayant la même distribution de probabilités, pourront ne pas être identiques pour l'individu. Malgré l'importance de cette phase, elle rend difficile la représentation mathématique de la théorie des perspectives. Elle sera par la suite abandonnée dans les extensions de ce modèle (i.e., voir [102]). La phase d'évaluation est la phase la plus importante dans cette théorie ; elle explique les résultats expérimentaux retrouvés dans [59]. Elle fait intervenir une fonction de la valeur subjective et une fonction de pondération des probabilités.

La théorie des perspectives se base sur des changements de l'état de richesse initiale de l'individu et non finale comme dans la théorie de l'utilité espérée. Cette richesse initiale est définie comme le point de référence qui permet de placer les gains ou les pertes qui sont alors relatifs par rapport à cet état initial de la richesse de l'individu. De plus, Kahneman et Tversky identifient que l'individu présente une attirance pour le risque dans le domaine des gains alors qu'il présente une aversion pour le risque dans le domaine des pertes. Il existe ainsi une fonction

qui détermine la valeur subjective d'une perspective risquée pour l'individu. Celle-ci transforme la valeur objective au travers de l'attitude face au risque de l'individu. La forme de la fonction valeur subjective diffère donc suivant la zone dans laquelle on la regarde. Pour traduire la recherche du risque dans les gains, elle est concave au-dessus du point de référence. Pour traduire l'aversion au risque dans les pertes, elle est convexe en dessous du point de référence. Il est également intéressant de noter que la pente est plus importante dans le domaine des pertes que dans celui des gains.

Parlons à présent, de la fonction de pondération des probabilités qui intervient dans la théorie des perspectives. Contrairement à la pondération de l'utilité par sa probabilité, la fonction de pondération multiplie la valeur d'un résultat par un poids. Ce poids est entièrement subjectif et se fait à partir de probabilités objectives qui sont déformées par l'individu. L'expérimentation montre que dans les pertes, cette fonction de pondération a les propriétés suivante

- Pour une probabilité nulle, elle est nulle.
- Pour une probabilité certaine, elle prend la valeur 1.
- Elle surpondère les probabilités faibles.
- Elle sous-pondère les probabilités moyennes et élevées.

Dans les gains, les petites probabilités sont sous-évaluées et les probabilités moyennes et fortes sont surévaluées.

Bien que cette théorie soit fondamentale de par les nouvelles considérations qu'elle apporte à la théorie de la décision, elle présente une limite importante. En effet, elle ne respecte pas la dominance stochastique du premier ordre (voir Section 4 pour la définition) qui est pourtant une propriété fondamentale en économie. La dominance stochastique demande qu'un changement de probabilité, d'un résultat mauvais à un résultat moins mauvais pour un individu, améliore la valeur de la perspective de l'individu. C'est ainsi qu'une révision appelée théorie des perspectives cumulées (voir [102]) a été proposée, en associant la théorie de Kahneman et Tversky à la théorie proposée par Quiggin en 1982, développée ci-après.

En 1982, Quiggin (voir [84]) formalisa une nouvelle théorie de la décision en présence d'incertitude qui généralise la théorie de l'utilité espérée. Ce modèle, dit de dépendance des rangs (RDEU), a pour but d'expliquer le comportement observé dans le paradoxe d'Allais ainsi que l'observation contradictoire montrant que les individus participent à la fois à des loteries (recherche du risque) tout en s'assurant contre les pertes (aversion au risque). L'idée fondamentale est que les agents surévaluent seulement les événements extrêmes de faibles probabilités plutôt que les événements de faibles probabilités. Il faut ainsi appliquer une transformation à la fonction de répartition des probabilités plutôt qu'aux probabilités elles-mêmes. Cette transformation se traduit par un poids associé à chaque événement puis par le classement de ces événements entre eux, suivant le poids associé.

Une préférence \succeq sur \mathcal{P} peut être représentée par la théorie RDEU s'il existe une fonction d'utilité u et une fonction de répartition de probabilité $\psi : [0, 1] \rightarrow [0, 1]$ telle que $\psi(0) = 0$ et $\psi(1) = 1$ et telle que la fonction $U : \mathcal{P} \rightarrow \mathbb{R}$ représente \succeq , avec $U(\mu)$ construit de la façon suivante

- Les événements de μ , sont rangés x_1, \dots, x_n tels que $x_1 \succ x_2 \succ \dots \succ x_n$.
- $U(\mu)$ est égal à

$$U(\mu) = \psi(\mu_1)u(x_1) + \sum_{i=2}^n \left(\psi\left(\sum_{j=1}^i \mu_j\right) - \psi\left(\sum_{k=1}^{i-1} \mu_k\right) \right) u(x_i).$$

En pratique, cela signifie que lorsqu'un individu est face à plusieurs événements, le poids μ_1 est associé à l'événement qui présente l'issue la plus favorable pour l'individu. Ensuite pour les événements suivant i , le poids associé dépend de la probabilité d'avoir au moins quelque chose d'aussi bien que l'événement i moins la probabilité d'avoir quelque chose de meilleur que l'événement i .

Remarque 3.12. *Lorsque la fonction ψ est égale à la fonction identité, on retrouve l'espérance d'utilité.*

De nombreux articles se sont intéressés à la forme de la fonction de pondération (voir [18], [108], [109]). Celle-ci présente une forme dite en S inversé, c'est-à-dire que les probabilités faibles sont surévaluées tandis que les probabilités moyennes et élevées sont sous-évaluées (voir [1]). Pour mieux caractériser la forme en S inversé de la fonction de pondération, Tversky et Wakker (voir [103]) ont proposé deux notions

- La sous-additivité inférieure : un intervalle de probabilités faibles (proches de 0) a plus d'impact qu'un intervalle de probabilités moyennes.
- La sur-additivité supérieure : un intervalle de probabilités élevées (proches de 1) a plus d'impact qu'un intervalle de probabilités moyennes.

Ainsi les individus sont plus sensibles aux probabilités extrêmes qu'aux probabilités moyennes, dans le sens où un changement de probabilités proche de la probabilité nulle ou proche de la probabilité certaine, sera perçu par un agent comme plus important qu'un changement de probabilités dans les valeurs médianes.

Comme mentionné précédemment, la théorie de dépendance des rangs a aidé Kahneman et Tversky (voir [102]) à reviser leur théorie des perspectives. C'est ainsi que la théorie des perspectives cumulées fut créée. Les caractéristiques fondamentales de la théorie des perspectives sont conservées. Les gains et les pertes sont toujours considérés de manière séparée et la valeur de chaque résultat est multipliée par un poids de décision. Considérer les gains et les pertes de manière séparée, permet de prendre en compte les cas où l'individu a une attitude différente face à ces deux types de résultat. Contrairement à la théorie classique, les probabilités

individuelles ne sont plus considérées mais c'est la fonction de pondération de la théorie de dépendance des rangs qui est utilisée à la place. Elle est appliquée aux gains et aux pertes de manière séparée pour ensuite sommer les deux évaluations ensemble. Cette théorie respecte donc la dominance stochastique du premier ordre à la différence de la théorie des perspectives classiques.

La formulation de cette théorie est la suivante. Pour une fonction valeur f , soit ψ^+ la fonction de pondération pour les gains et ψ^- la fonction de pondération pour les pertes. Soit $(x_i)_{i \in \{1, \dots, n\}}$ une famille de résultats qu'on suppose classés de la façon suivante

$$x_1 \leq \dots \leq x_k \leq 0 \leq x_{k+1} \leq \dots \leq x_n,$$

les k premiers résultats sont donc des pertes et les $n - k$ derniers sont donc des gains. La valeur de la perspective $(x_1, p_1; \dots, x_n, p_n)$ est donnée par la formule

$$\sum_{i=1}^k \pi_i^- f(x_i) + \sum_{i=k+1}^n \pi_i^+ f(x_i),$$

avec π_i^- et π_i^+ les poids de décisions définis par

$$\begin{aligned} \pi_1^- &= \psi^-(x_1), & \pi_i^- &= \psi^-(p_1 + \dots + p_i) - \psi^-(p_1 + \dots + p_{i-1}), \quad 2 \leq i \leq k, \\ \pi_n^+ &= \psi^+(x_n), & \pi_i^+ &= \psi^+(p_i + \dots + p_n) - \psi^+(p_{i+1} + \dots + p_n), \quad k+1 \leq i \leq n-1. \end{aligned}$$

Remarque 3.13. *Il est possible de retrouver l'espérance d'utilité en prenant $\psi^-(p) = \psi^+(p) = p$ pour tout $p \in [0, 1]$.*

De manière similaire à la théorie de dépendance en rang, les probabilités cumulées représentent la probabilité d'avoir un résultat ou quelque chose de meilleur (respectivement moins bien) que celui-ci en ce qui concerne les gains (respectivement les pertes). Les poids de décision sont obtenus en prenant la différence entre les valeurs transformées des probabilités cumulées. Kahneman et Tversky ont ainsi montré par l'expérimentation (voir [102]) que cette nouvelle fonction de pondération est sensible aux changements de probabilités dans les extrêmes (proche de 0 ou de 1) et est relativement insensible aux changements dans les valeurs intermédiaires de probabilité. Les formes des deux fonctions de pondération respectivement pour les gains et pour les pertes présentent des différences. Bien qu'elles aient toutes les deux des formes de S inversé, celle des pertes est plus élevée et moins incurvée. Ces résultats ont aussi été retrouvés par Wakker (voir [105]).

Peu de temps après la théorie de dépendance en rang, une autre théorie se développa pour pallier les problèmes que présente la théorie de l'utilité espérée. En 1987, Yaari (voir [111]) proposa la théorie duale du choix. Alors que la théorie de l'utilité espérée est linéaire pour les probabilités et non pour les paiements, la théorie duale inverse ces deux propriétés, elle est

linéaire pour les paiements et non pour les probabilités. Les perspectives risquées sont ainsi évaluées en utilisant la somme pondérée des quantiles. Ceci permet de respecter la dominance stochastique du premier ordre. En effet, l'individu faisant face à des perspectives risquées va préférer celles qui auront une valeur plus élevée dans tous les états de la nature. Cette représentation respecte également une autre propriété appelée insensibilité aux perspectives comonotones. Si un individu est indifférent entre deux perspectives qui donnent leurs meilleurs et leurs pires résultats dans les mêmes états de la nature, alors il sera indifférent à toute combinaison convexe de ces perspectives.

Yaari commença par définir une fonction de distribution dite décumulée, qui correspond en actuariat à la fonction de survie d'une variable aléatoire X

$$S_X(t) := \mathbb{P}[X > u], \quad u \in [0, 1].$$

Ainsi, en restreignant la valeur de X , on obtient que les paris sont normalisés. Ils ne peuvent résulter en une perte plus grande que la richesse de l'individu et aucun pari ne peut offrir un prix plus grand qu'un nombre prédéterminé.

Plusieurs axiomes régissent la théorie duale. On considère une relation de préférence notée \gtrsim où $>$ donne la préférence strict et \sim donne l'indifférence.

1. **Axiome de neutralité** : pour deux loteries X et Y , $S_X = S_Y$ implique $X \sim Y$.
2. **Axiome de l'ordre faible complet** : la relation de préférence \succeq est réflexive, transitive et connectée.
3. **Axiome de continuité** : pour S_1, S_2, S_3, S_4 telles que $S_1 > S_2$ il existe $\epsilon > 0$ tel que $\|S_1 - S_3\| < \epsilon$ et $\|S_2 - S_4\| < \epsilon$ implique $S_3 > S_4$, avec $\|S\| = \int |S(t)|dt$, est la norme L_1 .
4. **Axiome de monotonie** : si $S_X(t) \geq S_Y(t)$ pour tout $t \in [0, 1]$ alors $S_X \gtrsim S_Y$.
5. **Axiome d'indépendance** : pour S_1, S_2 et S_3 et β un réel tel que $0 \leq \beta \leq 1$ alors $S_1 \gtrsim S_2$ implique $(\beta S_1^{-1} + (1 - \beta)S_3^{-1})^{-1} \gtrsim (\beta S_2^{-1} + (1 - \beta)S_3^{-1})^{-1}$.

On peut remarquer plusieurs points à partir de ces axiomes. L'axiome 4. est défini par rapport à la dominance stochastique du premier ordre. Les axiomes de 1. à 4. permettent d'obtenir l'axiome 5. L'axiome 5. donne l'insensibilité aux perspectives comonotones mentionnée précédemment. Ces cinq axiomes permettent de donner le théorème de représentation de la théorie duale.

Théorème 3.4. *Une relation de préférence \gtrsim satisfait les axiomes de (1) à (5) si et seulement s'il existe une fonction ψ continue et croissante, définie sur $[0, 1]$ telle que pour tout X et Y , on a*

$$X \gtrsim Y \iff \int_0^1 \psi(S_X(t))dt \geq \int_0^1 \psi(S_Y(t))dt.$$

De plus, ψ est unique à une transformation affine positive près.

La principale limite de la théorie duale de Yaari est qu'elle ne s'applique qu'aux perspectives risquées qui sont des variables aléatoires univariées. Les choix sur des perspectives multidimensionnelles ne sont donc pas pris en compte alors que l'utilité espérée offre cette possibilité. Une version multivariée de cette théorie fut proposée par Yaari (voir [110]). Cette extension repose sur deux hypothèses : l'indépendance et la séparabilité des composantes du risque. Ces deux hypothèses restrictives ramènent donc un problème multivarié en un problème univarié. Une autre extension de la théorie duale dans le cas multivarié a été proposée récemment par Galichon et Henry (voir [45]) en enlevant ces deux hypothèses restrictives.

4 Ordre stochastique et équations différentielles rétrogrades

Les différents chapitres de ce travail de thèse font appel à d'autres outils mathématiques qu'il convient de définir. Dans les chapitres 1 et 2, les pertes auxquelles font face les agents économiques étudiés sont modélisées par des variables aléatoires respectant l'hypothèse de dominance stochastique du premier ordre (voir Section 4.1). Dans le chapitre 3, la perte de l'agent est modélisée par une famille de processus stochastiques à saut largement utilisée en actuariat, appelée processus de Poisson composé. Le dernier chapitre fait également appel à la théorie des équations différentielles stochastiques rétrogrades (EDSR) (voir Section 4.2) qui sont des équations différentielles stochastiques dont on connaît la condition terminale, avec la particularité ici qu'elles incluent un terme dit à saut du fait du processus de Poisson composé.

4.1 Dominance stochastique

La notion de dominance stochastique (voir [15]) permet de comparer des distributions entre elles. La dominance stochastique a une grande importance en économie où elle a été utilisée pour résoudre des problèmes de choix en présence d'incertitude (voir [48]). Elle a aussi été mise en oeuvre en finance pour déterminer les choix de portefeuille (voir [42]). Le lecteur peut se référer à [63] pour une revue de littérature sur l'utilisation de la dominance stochastique. Il existe plusieurs ordres de dominance stochastique, seul le premier ordre sera défini dans ce qui suit, étant celui qui est retenu dans les chapitres 1 et 2 de cette thèse.

Définition 4.1. Soient F_1 et F_2 deux fonctions de répartition. F_1 domine F_2 suivant la dominance stochastique du premier ordre si

$$F_1(x) \leq F_2(x), \text{ pour tout } x.$$

Cette définition montre l'importance de la dominance stochastique du premier ordre. En effet, le quantile de queue de F_1 sera supérieur à celui de F_2 . Ainsi en considérant des variables aléatoires de perte, on veut choisir le quantile de queue le plus petit. La dominance stochastique du premier ordre peut donc donner une relation de préférence seulement avec une information partielle et elle est indépendante des préférences vis-à-vis du risque de l'agent.

Remarque 4.1. *Comme mentionné précédemment, la théorie des perspectives, développée Section 3, ne respecte pas la dominance stochastique du premier ordre. Cependant, la théorie de l'utilité espérée comme la théorie de dépendance en rang la respectent. Il sera montré dans les chapitres 1, 2 que les familles de variables aléatoires étudiées dans ces deux chapitres respectent cette propriété.*

4.2 Équations différentielles stochastiques rétrogrades à saut

Les équations différentielles stochastiques rétrogrades peuvent être vues comme des équations différentielles stochastiques pour lesquelles la condition terminale est donnée. Dans le cas d'une équation différentielle ordinaire, connaître la condition terminale est équivalent à inverser le temps. Dans le cas stochastique, les choses sont plus complexes car il faut que les solutions soient adaptées à la filtration de l'espace de probabilités. Sur un espace de probabilités filtré $(\Omega, \mathcal{F}, \mathbb{P})$, il s'agit donc de trouver un couple, noté $(Y_t, Z_t)_{t \in [0, T]}$, de processus adaptés par rapport à la filtration \mathcal{F} du mouvement brownien $(W_t)_{t \in [0, T]}$, vérifiant l'équation différentielle stochastique

$$-dY_t = f(t, Y_t, Z_t)dt - Z_t dW_t, \quad 0 \leq t \leq T,$$

avec la condition terminale $Y_T = \xi$ où ξ est une variable de carré intégrable. Cette équation doit être comprise au sens intégral

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s, \quad 0 \leq t \leq T.$$

La théorie des EDSR a été introduite par Bismut (voir [14]) dans le cas où le générateur f est linéaire en Y et en Z . Pardoux et Peng (voir [77]) ont traité le cas où f est non-linéaire mais lipschitzienne en les deux variables, avec un résultat d'existence et d'unicité des solutions. Depuis cette théorie a bénéficié d'un certain nombre d'apports au cours du temps du fait de ses nombreuses applications mathématiques (aussi bien en théorie du contrôle stochastique qu'en économie ou en finance) et la condition d'intégrabilité de ξ ainsi que les hypothèses concernant f ont pu être affaiblies.

Depuis plusieurs années, la théorie du contrôle stochastique est utilisée pour résoudre des problèmes actuariels. De manière générale, les processus à saut sont utilisés en actuariat pour modéliser les sinistres (voir [74]). C'est pourquoi la suite de cette présentation se concentrera sur les EDSR dirigées par un mouvement brownien et par une mesure aléatoire compensée, appelées EDSR à saut. Pour plus de détails sur les EDSR le lecteur peut se référer à [37] et à [31] pour les EDSR à saut dans le cas d'un générateur lipschitzien. Dans le cadre non-Lipschitzien, Lepeltier et San Martin (voir [66] et [67]) ont étendu les résultats pour les EDSR dirigés par des mouvements browniens et des générateurs à croissance linéaire ou quadratique. Le cas des générateurs monotones a été établi par Fan et Jiang (voir [41]). D'autres extensions peuvent être mentionnées, notamment dans le cas des EDSR quadratiques avec les travaux de

Barrieu et El Karoui (voir [11]). Pour plus de détails sur les EDSR quadratiques, le lecteur peut se référer au livre (voir [34]). Pour des applications à la finance des EDSR à saut, le lecteur peut se référer au livre [89].

On considère l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, généré par un mouvement brownien W à valeur dans \mathbb{R} et un horizon de temps fini T . On considère une mesure μ aléatoire admettant un compensateur ν , c'est-à-dire une mesure aléatoire telle que pour toute variable aléatoire prévisible U , le processus M défini par

$$M_t := \int_0^t \int_0^\infty U_s(x)(\mu - \nu)(ds, dx)$$

est une martingale locale. On notera par la suite $\tilde{\mu} := \mu - \nu$. On définit également les espaces suivants

$$\begin{aligned} \mathcal{S}^2 &:= \left\{ \phi \text{ est une } \mathbb{F}\text{-semi-martingale optionnelle à trajectoires continues à droite avec une limite à gauche (càd} \right. \\ &\quad \left. \| \phi \|_{\mathcal{S}^2}^2 := \mathbb{E} \left[\sup_{t \in [0, T]} |\phi_t|^2 \right] < \infty \right\}, \\ \mathbb{H}^2 &:= \left\{ U \text{ est un processus tel que } \| U \|_{\mathbb{H}^2}^2 := E \left[\int_0^T d \langle \int_0^t \int_0^\infty U_s(x) \tilde{\mu}(ds, dx) \rangle_t \right] < \infty \right\}, \\ \mathcal{H}^2 &:= \left\{ \phi \text{ est une } \mathbb{F}\text{-semi-martingale optionnelle à trajectoires càdlàg et à valeurs dans } \mathbb{R} \text{ telle que} \right. \\ &\quad \left. \| \phi \|_{\mathcal{H}^2}^2 := \mathbb{E} \left[\int_0^T |\phi_t|^2 dt \right] < \infty \right\}. \end{aligned}$$

Soit l'EDSR à saut suivante

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s, U_s(.)) ds - \int_t^T Z_s dW_s - \int_t^T \int_{\mathbb{R}} U_s(z) \tilde{\mu}(ds, dz), \quad 0 \leq t \leq T. \quad (4.1)$$

Pour une condition terminale ξ et un générateur f donnés, il s'agit de trouver un triplet $(Y, Z, U) \in \mathcal{S}^2 \times \mathbb{H}^2 \times \mathcal{H}^2$ satisfaisant (4.1) \mathbb{P} -presque-sûrement. Les processus Z et U sont des processus de contrôle pour que le processus adapté Y respecte la condition terminale de l'EDSR à saut. Pour cela, donnons les trois hypothèses suivantes, considérées comme standard dans la théorie

- La valeur terminale ξ est à valeurs dans l'espace $\mathbb{L}^2(\mathbb{R})$.
- Le générateur $f : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R} \times L_{\mathbb{P}}^2(\mathbb{R}) \rightarrow \mathbb{R}$ est prévisible et lipschitz continu au sens

$$\left| f(\omega, t, y, z, u) - f(\omega, t, y', z', u') \right|^2 \leq K \left(|y - y'|^2 + |z - z'|^2 + \int_{\mathbb{R}} |u(x) - u'(x)|^2 \tilde{\mu}(dx) \right),$$

presque-sûrement et presque-partout $(\omega, t) \in \Omega \times [0, T]$, pour tout $(y, z, u), (y', z', u') \in \mathbb{R} \times \mathbb{R} \times L_{\mathbb{P}}^2(\mathbb{R})$.

$$- \mathbb{E} \left[\int_0^T |f(t, 0, 0, 0)|^2 dt \right] < \infty.$$

Plusieurs résultats d'existence et d'unicité de la solution à (4.1) ont été établis. Dans le livre de Delong [31], la preuve repose sur l'utilisation d'estimée a priori des processus Y , Z , U , inspirée de la méthodologie proposée par El Karoui, Peng et Quenez [37]. Des preuves similaires peuvent être trouvées dans les travaux de Barles, Buckdahn et Pardoux (voir [10]), Royer (voir [91]) dans le cas d'une mesure de Poisson et également dans l'article [12] de Becherer dans le cas d'une mesure aléatoire générale. Plus récemment, dans Papapantoleon, Possamaï et Saplaouras [76], les auteurs proposent également une preuve de l'existence et de l'unicité d'une solution d'une EDSR à saut, basée également sur des estimations a priori dans le cas d'un générateur dit Lipschitz-stochastique et pour un horizon de temps possiblement infini et non borné. L'existence d'une solution pour les EDSR dirigées par un mouvement Brownien et une mesure de Poisson avec générateurs continus satisfaisant une condition faible de monotonie, a été étudiée par Rong (voir [88]). Une extension des résultats d'existence de solution au cas des EDSR quadratiques à saut a été proposée par Kazi-Tani, Possamaï et Zhou (voir [60]).

Le chapitre 3 de ce travail de thèse, présente une modélisation de l'auto-protection au travers d'une EDSR à saut dont la mesure aléatoire est une mesure des sauts d'un processus de Poisson composé, qui modélise le risque auquel fait face l'agent. La modélisation des sinistres sous la forme de processus de Poisson composé est la plus simple et la plus utilisée dans la théorie du risque (voir [75], [50], [49] et [93]). Dans ce qui suit, la construction et seules les principales propriétés des processus de Poisson composé, pertinentes à ce travail de thèse seront présentées. Pour plus de détails sur cette famille de processus, le lecteur peut se référer au livre de Cont et Tankov [22].

On se place sur un espace de probabilité fixé $(\Omega, \mathcal{F}, \mathbb{P})$. La base d'un processus de Poisson composé, est un processus de Poisson. Un processus de Poisson noté $N := (N_t)_{t \leq 0}$ est un processus de comptage qui compte le nombre de temps aléatoire entre 0 et t , dont les incrément sont indépendants et stationnaires. Pour construire formellement un tel processus, on considère une suite $X := (X_i)_{i \geq 1}$ de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) suivant une loi exponentielle de paramètre λ

$$\mathbb{P}[X_i \geq x] = e^{-\lambda x}.$$

On note T la suite de variables aléatoires définie par $T = 0$ et pour tout $n \leq 1$

$$T_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N}.$$

T_n représente le temps du n -ième saut de N et les éléments de X représentent les intervalles de temps entre les sauts. Ainsi le processus de Poisson N de paramètre λ s'écrit

$$N_t = \sum_{n \geq 1} \mathbb{1}_{t \geq T_n},$$

Les trajectoires de N sont constantes par morceaux, càdlàg. A chaque temps $t > 0$, N_t suit une loi de Poisson de paramètre λt et

$$\mathbb{P}[N_t = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Un processus de Poisson est un processus de Lévy, c'est-à-dire que pour tout $t > s$, les incrémentums $N_t - N_s$ sont indépendants du passé du processus jusqu'au temps s et à la même loi que le processus N_{t-s} .

Un processus de Poisson composé est une généralisation d'un processus de Poisson pour lequel les sauts peuvent avoir des tailles différentes. L'intervalle entre deux temps de sauts suit encore une loi exponentielle mais les tailles de sauts ont une loi arbitraire. Formellement, pour un processus de Poisson N de paramètre λ , soit $(Z_i)_{i \geq 1}$ une suite de variables aléatoires réelles de fonction de distribution G , indépendants de N , le processus J défini par

$$J_t = \sum_{i=1}^{N_t} Z_i,$$

est appelé un processus de Poisson composé. Ses trajectoires sont constantes par morceaux et càdlàg, la loi des tailles des sauts est donnée par G . J est également un processus de Lévy. Il existe une mesure aléatoire associée au processus de Poisson composé J , notée μ_J donnée pour tout ensemble mesurable $A \subset \mathbb{R} \times [0, T]$ par

$$\mu_J(A) = \sum_{t \in [0, T]} \mathbb{1}_{\Delta Z_t \in A},$$

avec $\Delta Z_t = Z_t - Z_{t-}$ et à partir de laquelle on définit la mesure aléatoire compensée notée $\tilde{\mu}_J$ par

$$\tilde{\mu}_J(dt, dz) = \mu_J(dt, dz) - \lambda dt dG(z),$$

Lorsqu'un risque est modélisé par un processus de Poisson composé J , chaque survenance de ce risque, c'est-à-dire chaque sinistre lié à ce risque, est modélisé par un saut de J . Ainsi la variable aléatoire Z_i représente le coût du i -ème sinistre.

Une telle modélisation sera utilisée dans le chapitre 3 pour les sinistres potentiels que l'agent veut prévenir. De plus, dans ce chapitre, le résultat d'unicité de la solution à l'EDSR à saut repose sur le théorème de représentation martingale appliqué à un processus de Poisson composé et sur une approche basée sur le principe de la programmation dynamique (voir [26]). Le principe de la programmation dynamique est fondamental dans la théorie du contrôle stochastique. Dans le chapitre 3, l'effort d'auto-protection sera assimilé à un contrôle stochastique dans l'optimisation de la fonction valeur de l'agent. L'idée derrière ce principe est qu'un contrôle optimal sur la période de temps $[t, T]$ peut être recollé en deux contrôles optimaux, l'un sur la période $[t, \tau]$ et l'autre sur la période $[\tau, T]$ quel que soit le temps d'arrêt τ . Dans ce chapitre, le taux de couverture assurantielle optimal sera déterminé dans un deuxième temps

grâce à la forme particulière de la fonction d'utilité de l'agent. Ce taux optimal sera obtenu notamment grâce à l'utilisation d'estimée a priori (voir [37]) pour la solution de l'EDSR à saut considérée.

5 Modèles de prévention en assurance

La prévention est un sujet intensivement étudié en économie au travers notamment de la théorie de la décision. La littérature sur le sujet s'étend dans plusieurs directions différentes, des spécificités de la relation entre prévention et choix assurantiels aux effets de la richesse et de l'aversion au risque en passant par l'étude de l'impact des différents types de risques. Cette section vise à présenter différents modèles théoriques mettant en relation prévention et marché de l'assurance. Une première partie se concentrera sur un modèle développé sous la théorie de l'utilité espérée, celui d'Ehrlich et Becker (voir [36]). Une seconde partie détaillera d'autres modélisations de la prévention dans un contexte assurantiel sous différentes théories alternatives à l'utilité espérée. En particulier, le modèle d'Ehrlich et Becker sera considéré comme point d'origine dans le développement des extensions présentées, tout comme celui de ce travail de thèse. Le lecteur pourra se référer à la revue de littérature économique sur la prévention faite par Courbage, Rey et Treich [25] ou encore aux travaux de Eeckhoudt, Gollier et Schlesinger [35] (Chapitre 9) et ceux de Gollier, Hammitt et Treich [46].

5.1 Le modèle d'Ehrlich et Becker

Le modèle théorique le plus couramment étudié dans la littérature de la prévention liée au marché assurantiel est le modèle d'Ehrlich et Becker (voir [36]). Comme mentionné précédemment, ce modèle s'applique dans un cadre général à tout type de risque et a pour but d'étudier la relation et les arbitrages entre le marché de l'assurance et les comportements préventifs, individuels ou collectifs. Ehrlich et Becker introduisent alors les concepts d'auto-assurance et d'auto-protection (définis Section 1) dans le cadre théorique de l'utilité espérée, et étudient leurs relations respectives avec le marché de l'assurance. Dans ce qui suit, une présentation des caractéristiques du modèle théorique et des résultats principaux sera faite avant de s'attarder sur les conclusions qui peuvent en être tirées.

Soit un agent économique, averse au risque, qui maximise une fonction d'utilité de type VNM, notée U . Il fait face à deux états possibles de la nature notés 0 et 1, de probabilités respectives p et $1 - p$. L'état 1 correspond à l'apparition d'un sinistre, l'état 0 représente l'événement contraire. Ce modèle étant atemporel, toutes les décisions de prévention et d'assurance sont prises au début de la période considérée et tous les résultats sont observés à la fin de cette période. Face au risque d'avoir un sinistre, l'individu peut souscrire à une assurance en choisissant son taux de couverture, noté α et en échange duquel il verse une prime d'assurance, notée Π . En plus de ce contrat, il peut adopter un comportement préventif vis-à-vis du risque,

pour en réduire soit la taille, soit la probabilité de survenue. Ce comportement est modélisé par une constante d'effort e , qui a un coût supposé linéaire, cet effort est supposé observable par l'assureur avec lequel l'individu souscrit à un contrat. Seules les conséquences monétaires pour l'agent sont étudiées exprimées comme la perte du revenu de l'agent notée L . On note son revenu dans l'état 0 I_0 et I_1 dans l'état 1 lorsque l'événement risqué se réalise. Comme l'utilité du décideur est basée uniquement sur des conséquences monétaires, le modèle est uni-dimensionnel.

Pour comprendre les mécanismes des activités d'auto-assurance et d'auto-prévention, Ehrlich et Becker ont tout d'abord regardé les deux types de prévention en absence de contrat d'assurance. Ils concluent qu'un agent averse au risque fournira un effort d'auto-assurance et un effort d'auto-protection. Les niveaux optimaux respectifs d'efforts dépendent toutefois des coûts de ces activités de prévention.

Ehrlich et Becker examinent ensuite l'arbitrage entre marché de l'assurance et auto-assurance. L'effort d'auto-assurance a un effet direct sur le montant de la perte L , qui devient alors une fonction décrite par $L := L(e) = I_1 - I_0$ avec e représentant l'effort auto-assurance. Plus l'individu va investir en auto-assurance, plus la perte va diminuer, ainsi L est une fonction décroissante de e . L'individu va alors chercher à résoudre le problème suivant

$$\max_{(\alpha, e)} U^*$$

où $U^* := (1-p)U(I_1 - e - \alpha\Pi) + pU(I_0 - e + \alpha) = (1-p)U(I_1 - e - \alpha\Pi) + pU(I_1 - L(e) - e + \alpha)$ représente l'utilité espérée de l'agent. En notant U_0 l'utilité dans l'état 0 et U_1 l'utilité dans l'état 1 et α^* , e^* les niveaux optimaux, les conditions du premier ordre (CPO) sont les suivantes

$$\begin{aligned} & -(1-p)\partial_\alpha U_1\Pi + p\partial_\alpha U_0 = 0, \\ & -(1-p)\partial_e U_1 - p\partial_e U_0(\partial_e L(e^*) + 1) = 0, \end{aligned}$$

et sous les conditions sur second ordre (CSO) suivantes

$$\begin{aligned} & -(1-p)\partial_\alpha^2 U_1\Pi^2 + p\partial_\alpha^2 U_0 < 0, \\ & -(1-p)\partial_e^2 U_1 - p\partial_e^2 U_0(\partial_e L(e^*) + 1)^2 - p\partial_e U_0\partial_e^2 L(e^*) < 0, \end{aligned}$$

les CPO sont satisfaites si $\partial_e L(e^*) + 1 < 1$ c'est-à-dire si l'investissement en auto-assurance augmente le revenu net dans l'état 1. Les CSO sont satisfaites si les utilités sont convexes et la perte est concave, c'est-à-dire si les utilités marginales du revenu et la productivité marginale de l'auto-assurance sont décroissantes. Pour étudier la relation entre auto-assurance et contrat d'assurance, les auteurs proposent de regarder l'évolution de l'effort optimal en auto-assurance e^* et la demande optimale en assurance α^* par rapport à la prime Π . En appliquant la règle de Cramer aux CPO et CSO, ils obtiennent alors

$$\frac{de^*}{d\Pi} > 0, \quad \frac{d\alpha^*}{d\Pi} < 0.$$

Ceci permet de conclure que l'auto-assurance est toujours substituable à la souscription d'un contrat d'assurance pour un individu. En effet, le contrat d'assurance comme l'auto-assurance implique simplement un transfert de richesse entre les deux états de la nature, du plus favorable vers celui qui l'est le moins. Pour un événement assuré donné, la prime étant indépendante de l'effort, si le degré de protection donné par l'assureur augmente, alors un agent rationnel réduira son investissement dans les activités d'auto-assurance. Ceci induira alors une réduction des conséquences de cet événement. Si l'agent est de plus averse au risque, alors il va chercher à égaliser les profits marginaux de son contrat d'assurance et de l'auto-assurance. Dès lors que le prix de la couverture assurantielle augmentera, le profit marginal réduira et donc l'individu ajustera en investissant dans des actions d'auto-assurance. Plus l'aversion au risque augmentera, plus l'effort d'auto-assurance augmentera, réduisant ainsi la demande en assurance.

Dans le cas de l'auto-protection, l'effort de l'agent agit sur la distribution de probabilités des états du monde et non sur la perte monétaire. Ainsi la probabilité de sinistre p , devient une fonction décrite par $p := p(e)$ avec e représentant, cette fois-ci, l'investissement en auto-protection. Plus l'effort d'auto-protection est important, plus la probabilité d'occurrence du sinistre diminue, ainsi, p est décroissante avec e . L'effort est observable par l'assureur qui l'inclut donc dans le calcul de la prime. De plus, pour inciter l'agent à faire de l'auto-protection, la prime est supposée décroissante avec l'effort. L'individu va alors chercher à résoudre le problème suivant

$$\max_{(\alpha, e)} U^*$$

où $U^* := (1 - p(e))U(I_1 - e - \alpha\Pi(e)) + p(e)U(I_0 - e + \alpha) = (1 - p(e))U(I_1 - e - \alpha\Pi(e)) + p(e)U(I_1 - e + \alpha)$ représente l'utilité espérée de l'agent. En notant U_0 l'utilité dans l'état 0 et U_1 l'utilité dans l'état 1 et α^* , e^* les valeurs optimales, les conditions du premier ordre (CPO) sont les suivantes

$$\begin{aligned} & - (1 - p(e^*))\partial_\alpha U_1 \Pi + p(e^*)\partial_\alpha U_0 = 0, \\ & - \partial_e p(e^*)(U_1 - U_0) - (1 - p(e^*))\partial_e U_1 (1 + \alpha^* \partial_e \Pi(e^*)) - p(e^*)\partial_e U_0 = 0, \end{aligned}$$

et les conditions sur second ordre (CSO) s'écrivent

$$\begin{aligned} & - (1 - p(e))\partial_\alpha^2 U_1 \Pi^2 + p(e)\partial_\alpha^2 U_0 < 0, \\ & - \partial_e^2 p(e)(U_1 - U_0) + (1 - p(e))\partial_e^2 U_1 (1 + \alpha^* \partial_e \Pi(e^*))^2 + \\ & p(e^*)\partial_e^2 U_0 + 2\partial_e p(e^*)(\partial_e U_1 (1 + \alpha^* \partial_e \Pi(e^*)) - \partial_e U_0) - (1 - p(e^*))\partial_e U_1 \alpha^* \partial_e^2 \Pi(e^*) < 0. \end{aligned}$$

Les CSO sont satisfaites si les utilités sont convexes, la probabilité d'occurrence et la prime d'assurance sont concaves. Pour étudier la relation entre auto-protection et contrat d'assurance, les auteurs proposent de regarder l'évolution de l'effort optimal en auto-protection e^*

et la demande optimale en assurance α^* par rapport à la prime Π , en appliquant la règle de Cramer aux CPO et CSO, ils obtiennent

$$\frac{de^*}{d\Pi} < 0, \quad \frac{d\alpha^*}{d\Pi} < 0$$

L'effort réduit la richesse dans les deux états de la nature en n'affectant pas la perte potentielle. La différence entre les utilités dans ces deux états va alors diminuer et ainsi le gain marginal de l'auto-protection va se réduire. La relation entre l'auto-protection et le marché de l'assurance présente donc une indétermination qui va dépendre de l'élasticité (aussi appelée sensibilité) du prix de l'assurance par rapport à l'auto-protection. En effet, l'auto-protection peut être substituable à un contrat assurantiel car une augmentation de la protection proposée par l'assureur peut également entraîner une diminution générale de l'effort d'auto-protection. De plus, lorsque le degré de protection offert par l'assureur augmente, l'agent va alors augmenter son investissement dans les activités d'auto-protection. En ce sens l'auto-protection et le contrat assurantiel seront complémentaires : la diminution (respectivement l'augmentation) de la demande de l'un impliquera la diminution (respectivement l'augmentation) de l'autre. L'indétermination de l'effet du contrat d'assurance sur l'auto-protection disparaît lorsque l'assureur propose un taux de chargement de la prime d'assurance suffisamment proche de 0, c'est-à-dire pour une prime d'assurance proche de la prime actuarielle qui représente fidèlement le risque de l'assuré (voir [69]).

5.2 Distorsion des probabilités et prévention

Le modèle d'Ehrlich et Becker présenté Section 5.1 est considéré comme l'un des premiers ayant mis en relation assurance et prévention, avec notamment la définition de deux types de prévention. Malgré l'apport significatif de ce modèle, il présente certaines limites. Une première limite de ce modèle est qu'il ne considère que deux états de la nature, un état sans sinistre et un état où le sinistre a lieu et les résultats ne sont pas directement transposables à un cas d'étude avec plusieurs états possibles de la nature. Dans [65], Lee considère un modèle avec plusieurs états de la nature dans lequel un agent économique averse au risque peut choisir de participer à des activités qui ont une action simultanée d'auto-assurance et d'auto-protection (*Self-insurance-cum-protection*). L'auteur montre que l'augmentation du degré d'aversion au risque n'implique pas qu'un individu investira plus dans des actions de prévention, lorsque la réduction marginale de la perte est supérieure à l'augmentation marginale du coût de la prévention. Il fournit cependant une condition suffisante sur la fonction de répartition de la perte pour obtenir ce résultat.

Une deuxième limite du modèle décrit dans [36] est qu'il ne considère que des pertes suivant une loi de Bernoulli. Jewitt [53] tenta de pallier cela en proposant une distribution de perte plus générale qui respecte la propriété dite de *single-crossing*. Cette propriété donne des conditions suffisantes sur les fonctions d'utilités et les distributions de probabilités pour obtenir des

statiques comparatives, en se basant sur l'approche de Diamond et Stiglitz (voir [33]), pour laquelle l'auto-protection préserve l'utilité moyenne pour toute fonction d'utilité. Il en est de même pour les travaux de Jullien, Salanié et Salanié [55] ainsi que ceux de Shogren et Crocker [94] et ceux de Meyer et Meyer [71].

Une autre limite du modèle décrit par Ehrlich et Becker, est qu'il a pour cadre la théorie de l'utilité espérée. Ainsi l'agent considéré est uniquement rationnel et ne souffre d'aucun biais de perception de la réalité. Il a cependant été montré, notamment grâce aux travaux de Kahneman et Tversky (voir Section 3.5.2) qu'un certain nombre de biais perturbent les perceptions du risque et des probabilités, remettant donc en cause la validité globale de l'utilité espérée. Ceci a motivé les extensions sous d'autres théories de décision en présence d'incertitude (voir Section 3.5.3), ce qui est détaillé ci-après.

Konrad et Skaperdas (voir [62]) étudient des actions d'auto-assurance et d'auto-protection avec la théorie de la dépendance en rang (voir Section 3.5.3) avec notamment une étude de l'impact de l'aversion au risque. Plusieurs des statiques comparatives obtenues avec l'utilité espérée sont encore valables dans ce cadre-ci. En particulier, pour les agents averses au risque, l'effort d'auto-assurance augmente. L'aversion au risque se traduit à la fois dans la forme de la fonction d'utilité et dans la fonction de transformation des probabilités. L'indétermination dans le cadre de l'auto-protection est également retrouvée. Cependant pour les risques de faibles probabilités et de probabilités larges une détermination qualitative est possible. Il est à noter que la possibilité de souscrire à une assurance n'est pas prise en compte dans ce modèle. Bleichrodt et Eeckhoudt (voir [16]) proposent également de regarder l'effet de la théorie de la dépendance en rang sur les problématiques d'auto-protection sous le concept de propension à payer pour réduire la probabilité d'un sinistre en santé lorsque l'individu utilise une fonction d'utilité univariée. Deux articles ont montré, sous la théorie de l'utilité espérée, l'équivalence entre la prévention selon Ehrlich et Becker et la notion de propension à payer, qui correspond au montant qu'il est prêt à payer pour réduire la perte ou la taille d'un sinistre : celui de Chiu (voir [21]) et ensuite celui de Dachraoui, Dionne, Eeckhoudt et Godfroid ([28]) dans le cadre de l'auto-protection. Bleichrodt et Eeckhoudt montrent que l'introduction du chargement des probabilités augmente la propension à payer pour réduire le risque lorsque l'agent sous-évalue sa probabilité d'être en bonne santé, l'effet inverse étant observé lorsque l'agent surévalue la probabilité d'être en bonne santé. Cependant, la pondération des probabilités donne des estimées peu stables de la propension à payer obtenue empiriquement. Langlais (voir [64]) a également étudié le lien entre aversion au risque et propension à payer dans un cadre d'utilité non-espérée, avec des hypothèses minimales sur les préférences individuelles de dominances du premier et du second ordre. Dans ce cadre, la réduction du risque passe par trois effets : un effet à la moyenne, un effet sur le risque pur et un effet sur la richesse de l'individu. Suivant la variation de chacun de ces effets, la propension à payer d'un individu averse au risque peut être plus faible que celle d'un individu neutre au risque. Plus récemment, une étude de la

perception du risque de santé dans le cadre de l'auto-protection a été faite par Etner and Jeleva (voir [40]). Elle souligne l'importance fondamentale de la fonction de transformation des probabilités pour comprendre les choix d'auto-protection des individus.

Une autre extension récente à souligner du modèle d'Ehrlich et Becker, est celle de Courbage (voir [23]), sous la théorie duale de Yaari (voir Section 3.5.3). Les résultats présentés dans [36], sont encore valables sous la théorie duale. L'auto-assurance et l'assurance de marché sont substituables et peuvent parfois être complémentaires lorsque la fiabilité des actions d'auto-assurance n'est pas garantie. La relation entre assurance de marché et auto-protection reste encore indéterminée. Cette indétermination s'explique cette fois-ci par le rôle de la fonction de transformation des probabilités. Courbage a toutefois montré que dans le cas particulier où la prime d'assurance n'est pas tarifée en fonction de l'efficacité des actions d'auto-protection, l'assurance de marché et l'auto-protection sont substituables.

6 Contributions et résultats principaux

Ce travail de thèse s'inscrit dans la continuité des extensions du modèle d'Ehrlich et Becker, présentées Section 5.2. Il se place d'abord dans une logique de test de robustesse sous un critère conforme aux pratiques actuarielles puis ensuite dans une logique de prolongement d'un modèle statique à un modèle dynamique. Ce manuscrit se compose de trois chapitres, chaque chapitre correspond à un article. L'article constituant le premier chapitre a été accepté à la revue *Insurance : Mathematics and Economics*.

6.1 Prevention efforts, insurance demand and price incentives under coherent risk measures

Le chapitre 1 étudie un modèle d'équilibre à une période, entre un assureur et un agent qui souhaite souscrire à un contrat d'assurance proportionnelle. Dans ce type de contrat, l'assureur prend en charge une part de la perte de l'agent contre une prime d'assurance, le reste étant à la charge de l'assuré. En plus de ce contrat d'assurance, l'agent a la possibilité d'entreprendre des actions de prévention contre son risque. La terminologie reprise dans ce chapitre est celle définie par Ehrlich et Becker dans [36] : pour un sinistre donné, on distingue deux types de prévention, l'auto-assurance qui permet d'en réduire la taille et l'auto-protection qui permet d'en réduire la probabilité d'occurrence. L'interaction assureur-assuré est modélisée comme un jeu de Stackelberg à une période dans lequel, l'assureur joue en premier en proposant un prix et l'agent joue en second en choisissant sa couverture assurantielle ainsi que son effort de prévention, qui lui est coûteux. Le coût de l'effort est modélisé par une fonction croissante et strictement convexe en l'effort.

Cette formulation s'inspire de la théorie des contrats, en particulier des problèmes Principal-Agent. L'assureur est alors vu comme le Principal et l'assuré comme l'Agent et la résolution

du jeu se fait par induction rétrograde. C'est-à-dire que le problème de l'agent est résolu en premier et la solution est implémentée dans le problème de l'assureur pour le résoudre ensuite. Les deux parties cherchent à minimiser leurs pertes respectives au travers de mesures de risque cohérentes (voir Section 2.1).

Pour étudier le problème de prévention de manière unifiée, la distribution de perte de l'agent est donnée par une famille de mesures de probabilités ordonnées stochastiquement et indexée par l'effort de prévention. L'assureur est considéré comme étant en situation de monopole sur le marché assurantiel. Il propose à l'agent un prix pour l'assurance qui est entièrement déterminé par un facteur de chargement. Ainsi l'effort de prévention de l'agent n'est pas observable par l'assureur. Le chapitre 1 se place donc dans le cadre de l'aléa moral. Ce type de problème est difficile à résoudre dans le cas général (voir [87], [52] ou encore [27] Chapitre 2). Le chapitre 1 montre que le problème admet un unique équilibre qu'il est possible de caractériser en précisant la distribution de perte de l'agent et son critère de préférence. Ceci se fait en étudiant localement les solutions au problème d'optimisation puis en comparant ces solutions locales entre elles.

Le contrat proposé par l'assureur étant entièrement caractérisé par le facteur de chargement, le problème d'optimisation de ce dernier consiste alors à trouver le chargement qui réduira de manière optimale sa mesure de risque. L'agent optimise quant à lui sa mesure de risque en choisissant de manière simultanée la couverture assurantielle et l'effort de prévention. Les distributions de probabilités de cette famille forment une famille de mesures de probabilités qui est décroissante pour la dominance stochastique du premier ordre (voir Section 4.1). Dans le cadre de l'auto-assurance, l'effort de prévention agit directement sur les distributions de probabilités de la perte. Ainsi, on suppose qu'il agit directement sur la fonction quantile issue de la fonction de survie, de manière à ce que l'impact marginal de l'effort sur la fonction quantile soit décroissant. L'hypothèse de décroissance de l'impact marginal est standard en économie. Cette hypothèse induit que les mesures de risque considérée dans ce chapitre ainsi que l'espérance mathématique des pertes sont des fonctions convexes, continues et croissantes en l'effort.

Le chapitre 1 commence par résoudre le problème d'optimisation de l'agent qui veut souscrire à un contrat d'assurance et faire de la prévention. Le premier résultat concerne le taux de couverture, il existe une condition suffisante sous laquelle, le niveau optimal de couverture assurantielle est égal à 0 ou 1. Cette condition, étant que la prime d'assurance soit positivement homogène, est vérifiée notamment pour les primes d'assurance sous forme de taux de chargement. Il est alors possible de caractériser le niveau optimal de couverture en fonction de l'effort et de distinguer deux intervalles distincts dépendants du facteur de chargement. Sur chaque intervalle et de manière homogène entre l'auto-assurance et l'auto-protection, on pourra alors étudier un problème d'optimisation de fonctions convexes sur des ensembles compacts. Un niveau optimal d'effort de prévention existe alors pour chaque type de prévention

suivant la nomenclature de [36].

En considérant le problème d'auto-assurance, cet article a pour but de montrer que si la mesure de risque de l'assuré décroît plus vite en l'effort que l'espérance de la perte, alors l'effort optimal est croissant avec le facteur de chargement et présente une discontinuité lorsque l'individu choisit de prendre une couverture complète. L'augmentation du prix de l'assurance induit une diminution de la demande en assurance de marché et une augmentation de l'effort de prévention. L'auto-assurance est dite alors substituable à l'assurance de marché. Du point de vue de l'auto-protection, si la perte moyenne décroît plus vite en l'effort que la mesure de risque de l'assuré alors l'effort optimal peut croître ou décroître avec le prix de l'assurance traduisant alors l'indétermination de la relation entre assurance et auto-protection en fonction de la sensibilité au prix de l'assurance.

Une illustration de ces deux problèmes de prévention sera donnée dans le cas d'une distribution de pertes de Pareto. Les résultats théoriques de ce chapitre s'appliquent à une large famille de distributions de sinistre, la distribution de Pareto a été choisie pour son utilisation standard en actuariat. Dans ce cas particulier, l'assureur choisira de fixer le prix de l'assurance comme égal au prix maximal pour lequel l'assuré préfère une couverture complète à l'alternative de n'avoir aucun contrat. Le cas des distributions discrètes sera également traité. Pour étudier l'auto-assurance, il faudra supposer que l'impact marginal de l'effort est plus important sur les petites pertes que sur les pertes extrêmes. Il sera alors démontré que l'effort optimal est croissant avec le prix de l'assurance. Dans le cas de l'auto-protection, des conditions suffisantes seront énoncées pour obtenir un effort optimal croissant ou non monotone par rapport au facteur de chargement.

6.2 Self-insurance and Non-concave Distortion Risk Measures

Le second chapitre de cette thèse se concentre sur l'arbitrage entre assurance de marché et auto-assurance. On considère un agent économique faisant face à un risque contre lequel il veut souscrire à une police d'assurance proportionnelle et contre lequel il peut entreprendre des actions de prévention dans le but de réduire la taille des sinistres. Le critère de préférences de l'agent est donné par une mesure de risque de distorsion définie par une distorsion non concave (voir Section 2.1.3). Ce type de fonctions de distorsion prend en compte les biais cognitifs régissant la perception des risques des individus. Comme mentionné Section 3.5.2 et Section 3.5.3, il a été montré que les individus transforment les probabilités objectives en probabilités subjectives de manière non concave et dépendante de leur aversion au risque.

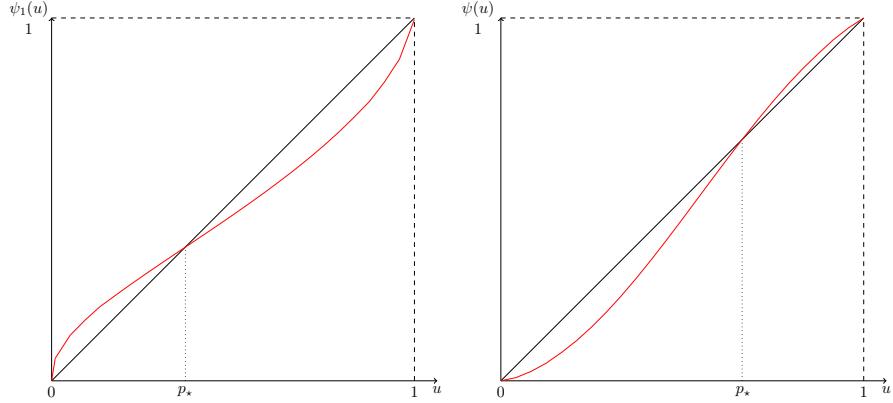


FIGURE 0.1 – Exemples de fonctions de distorsion non concaves. A gauche, la fonction $\exp(-(-\log(u))^\gamma)$, $\gamma = 0.67$. A droite, la fonction $\frac{u^\gamma}{(u^\gamma + (1-u)^\gamma)^{1/\gamma}}$, $\gamma = 1.5$.

Figure 0.1 est une illustration de deux fonctions de distorsion possibles. A gauche, la fonction correspond à une forme dite en S inversé. L'intervalle de probabilités est divisé en deux : une première partie comprenant les probabilités faibles où la fonction est concave, une deuxième partie comprenant les probabilités moyennes et élevées où la fonction est convexe. La valeur de la borne délimitant ces deux intervalles est noté p^* . Comme décrit dans [102], [108] et [2], p^* est un point fixe de la fonction de distorsion ψ . A droite, la fonction a une forme dite en S. Pour les probabilités faibles et moyennes, la fonction est convexe, pour les probabilités élevées, elle est concave. Comme souligné dans [2] et [106], il existe une relation entre la forme de la fonction de distorsion et le caractère optimiste ou pessimiste d'un individu. Pour une fonction de distorsion donnée, la concavité de celle-ci correspond à une surpondération des probabilités et la convexité signifie que l'individu sous-pondère les probabilités. La forme de la fonction révèle donc la sensibilité de l'agent aux événements extrêmes et comment cette sensibilité change quand on approche les probabilités moyennes. Pour un agent optimiste, la pente de la fonction de distorsion augmente quand on s'éloigne de 0 puis s'aplatit dans les valeurs moyennes et élevées, ce qui signifie qu'une légère augmentation de la probabilité de risque a un plus grand impact sur les probabilités faibles que sur les probabilités élevées. Au contraire, pour un agent pessimiste, les probabilités élevées sont plus déformées que les petites probabilités. De nombreux exemples de fonctions de distorsion sont décrites dans [2] et [83].

On peut souligner de plus que sous l'hypothèse d'invariance par translation, le principe de diversification est équivalent à la convexité de la mesure de risque, elle-même équivalente à la concavité de la fonction de distorsion (voir Section 2). Ainsi, considérer des fonctions de distorsion non concaves permet de prendre en compte les situations dans lesquelles l'agent n'a pas la possibilité de diversifier ses sources de risque.

Le modèle proposé est similaire à celui du chapitre 1 : dans un cadre statique d'aléa moral, l'agent minimise sa mesure de risque en choisissant de manière simultanée sa couverture as-

surantielle et son effort d'auto-assurance. Cet effort lui est coûteux et le coût est donné par une fonction supposée croissante et strictement convexe. En échange de la couverture assurantielle qu'il choisit, il paye une prime d'assurance dont le prix est entièrement déterminé par un facteur de chargement. De même, la distribution de perte de l'agent respecte les mêmes hypothèses que dans le chapitre 1.

La mesure de risque considérée dans ce chapitre 2 est également positivement homogène et le contrat d'assurance étudié est linéaire, ainsi la résolution du problème d'optimisation de l'agent suit la même méthodologie que celle développée dans le chapitre 1.

On retrouve l'existence d'un effort optimal d'auto-assurance qui s'écrit comme une fonction du facteur de chargement de la prime d'assurance. Il existe une valeur particulière du prix de l'assurance qui engendre un saut dans la fonction de l'effort optimal. A la différence du chapitre 1, ce saut peut être positif ou négatif. Ainsi en prenant comme critère de préférences des mesures de risque de distorsion définies par des fonctions de distorsion non concaves, une indétermination dans l'arbitrage entre assurance et auto-assurance apparaît. Lorsque le saut est positif, une augmentation du prix de l'assurance implique une diminution de la demande en assurance de marché et une augmentation de l'effort d'auto-assurance. L'assurance de marché et la prévention sont alors substituables comme décrit dans la littérature économique. Lorsque ce saut est négatif, une augmentation du prix de l'assurance implique une diminution de la demande en assurance de marché et une diminution de l'effort d'auto-assurance. La relation devient donc complémentaire. Ce résultat est retrouvé uniquement dans le modèle théorique [68] qui considère le comportement d'un assuré dans le contexte d'une assurance complémentaire santé obligatoire en France et où l'effort peut être compensé financièrement ou via des conseils dans le cadre d'un programme de prévention.

Une illustration de ce problème de prévention est donnée dans le cas d'une distribution de perte de Weibull. Les résultats théoriques de ce chapitre s'applique à une large famille de distributions de sinistres, la distribution de Weibull est retenue car elle est standard en pratique actuarielle. En précisant les différents paramètres du modèle, on obtient une comparaison graphique des niveaux d'efforts optimaux. Il est également possible de déterminer graphiquement la valeur maximale du facteur de chargement que l'assureur peut proposer à l'agent pour qu'il accepte de souscrire à la police d'assurance.

6.3 A Continuous-Time Model of Self-Protection

Le chapitre 3 s'intéresse à l'arbitrage entre auto-protection et assurance de marché en temps continu. On considère un agent voulant souscrire à un contrat d'assurance linéaire pour se prémunir d'une perte, contre laquelle il peut également exercer un effort de prévention. Cet effort lui est coûteux, le coût étant décrit par une fonction croissante et strictement convexe. Cette modélisation de la demande en assurance et de l'effort de prévention peut être vue comme un problème de contrôle stochastique. La perte est donnée par une classe générale contenant

en particulier tous les processus de Poisson composés qui sont classiquement utilisés dans les modèles actuariels et en pratique. Les processus de pertes considérés sont décrits par un processus de comptage, modélisant les arrivées des sinistres et une distribution de probabilités unidimensionnelle, modélisant le coût ou la gravité de chaque sinistre.

L'agent cherche à maximiser l'espérance d'utilité de la valeur terminale de sa richesse, son utilité étant donnée par une fonction exponentielle. Cette maximisation porte sur les deux paramètres que contrôle l'agent : sa couverture assurantielle et son effort de prévention. La couverture assurantielle est indépendante du temps et est choisie pour toute la durée du contrat à la signature de ce dernier. L'effort de prévention lui est modélisé par un processus dynamique. Pour étudier l'auto-protection, il est supposé que seul le processus de comptage est impacté par l'effort. Le problème de contrôle est donc formulé dans un sens faible car l'agent ne contrôle que la distribution de probabilité du processus de perte. De plus, son effort n'est pas observable par l'assureur ainsi le chapitre 3 traite également de l'aléa moral.

Le problème de contrôle est résolu en utilisant le principe de la programmation dynamique (voir Section 4.2) et la fonction valeur de l'agent est représentée grâce à une équation différentielle stochastique rétrograde à saut. Le problème présente alors une solution unique et explicite en déterminant l'équivalent certain dynamique de l'agent et en prouvant que l'effort optimal est constant par rapport au temps pour une large classe de processus de pertes. En particulier, cet effort constant est caractérisé comme l'unique valeur minimisant un Hamiltonien explicite, donnant dans des cas particuliers l'effort optimal de l'agent. Avoir un effort optimal constant signifie qu'il est fixé par l'agent au temps 0 et qu'il ne change pas avec le temps ou avec les occurrences des pertes. Ainsi un agent qui maximise une espérance d'utilité exponentielle n'a jamais d'incitation à modifier dynamiquement son effort. Ceci implique que l'agent ne devrait maximiser son utilité que dans la sous-classe des processus de Lévy, rendant ainsi le processus de pertes homogène et markovien. Ainsi, obtenir un effort optimal constant implique que la propriété de Lévy des processus est préservée par la maximisation d'une espérance d'utilité exponentielle. Enfin pour déterminer le niveau de couverture optimal, il s'agit d'étudier la dépendance de l'EDSR par rapport à ce dernier. Le chapitre 3 montre l'existence d'une couverture linéaire optimale, qui n'est pas nécessairement nulle ou complète.

Une illustration dans le cas d'un coût linéaire et de l'effet linéaire de l'effort sur la fréquence est proposée. Dans ce cadre, il est possible d'expliquer la solution de l'EDSR à saut et l'effort optimal d'auto-protection. On obtient également explicitement la valeur optimale de la couverture assurantielle qui dépendra alors du prix de l'assurance. Suivant ce prix, l'agent peut choisir de souscrire ou non à l'assurance et quand il le fera, il pourra prendre une couverture intermédiaire qui pourra ne pas être complète.

Bibliographie

- [1] ABDELLAOUI, M. Parameter-free elicitation of utility and probability weighting functions. *Management Science* 46, 11 (2000), 1497–1512.
- [2] ABDELLAOUI, M., L'HARIDON, O., AND ZANK, H. Separating curvature and elevation : A parametric probability weighting function. *Journal of Risk and Uncertainty* 41, 1 (2010), 39–65.
- [3] ACERBI, C. Coherent representations of subjective risk aversion.
- [4] ACERBI, C., AND TASCHE, D. On the coherence of expected shortfall. *Journal of Banking & Finance* 26, 7 (2002), 1487–1503.
- [5] ALLAIS, M. Le comportement de l'homme rationnel devant le risque : Critique des postulats et axiomes de l'école américaine. *Econometrica : Journal of the Econometric Society* 21 (1953), 503–546.
- [6] ARROW, K. J. Uncertainty and the welfare economics of medical care. *The American Economic Review* 53, 5 (1963), 941–973.
- [7] ARROW, K. J. The economics of moral hazard : Further comment. *The American Economic Review* 58, 3 (1968), 537–539.
- [8] ARTZNER, P., DELBAEN, F., EBER, J.-M., AND HEATH, D. Thinking coherently. *RISK* 10 (1997), 68–71.
- [9] ARTZNER, P., DELBAEN, F., EBER, J.-M., AND HEATH, D. Coherent measures of risk. *Math. Finance* 9, 3 (1999), 203–228.
- [10] BARLES, G., BUCKDAHN, R., AND PARDOUX, E. Backward stochastic differential equations and integral–partial differential equations. *Stochastics : An International Journal of Probability and Stochastic Processes* 60, 1-2 (1997), 57–83.
- [11] BARRIEU, P., AND EL KAROUI, N. Monotone stability of quadratic semimartingales with applications to unbounded general quadratic BSDEs. *The Annals of Probability* 41, 3B (2013), 1831–1863.
- [12] BECHERER, D. Bounded solutions to backward SDEs with jumps for utility optimization and indifference hedging. *The Annals of Applied Probability* 16, 4 (2006), 2027–2054.

- [13] BERNOULLI, D. Specimen theoriae novae de mensura sortis, Commentarii Academiae Scientiarum Imperialis Petropolitanae (5, 175-192, 1738). *Econometrica* 22 (1954), 23–36. Translated by L. Sommer.
- [14] BISMUT, J.-M. *Analyse convexe et probabilités*. PhD thesis, Faculté des sciences de Paris, 1973.
- [15] BLACKWELL, D. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics* 24, 2 (1953), 265–272.
- [16] BLEICHRODT, H., AND EECKHOUDT, L. Willingness to pay for reductions in health risks when probabilities are distorted. *Health Economics* 15, 2 (2006), 211–214.
- [17] CAMERER, C. F. *Recent Tests of Generalizations of Expected Utility Theory*. Springer Netherlands, Dordrecht, 1992, pp. 207–251.
- [18] CAMERER, C. F., AND HO, T.-H. Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty* 8, 2 (1994), 167–196.
- [19] CARMAN, K. G., AND KOOREMAN, P. Probability perceptions and preventive health care. *Journal of Risk and Uncertainty* 49, 1 (2014), 43–71.
- [20] CARROLL, P., GRACE, J., AND TERRY, M. Exploring the causes of comparative optimism. *Psychologica Belgica* (2002), 65–98.
- [21] CHIU, W. H. On the propensity to self-protect. *The Journal of Risk and Insurance* 67, 4 (2000), 555–577.
- [22] CONT, R., AND TANKOV, P. *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, 2004.
- [23] COURBAGE, C. Self-insurance, self-protection and market insurance within the dual theory of choice. *The Geneva Risk and Insurance Review* 26, 1 (2001), 43–56.
- [24] COURBAGE, C., AND REY, B. Decision thresholds and changes in risk for preventive treatment. *Health Economics* 25, 1 (2016), 111–124.
- [25] COURBAGE, C., REY, B., AND TREICH, N. Prevention and precaution. In *Handbook of Insurance*, G. Dionne, Ed. Huebner International Series on Risk, Insurance, and Economic Security (New York : Springer), New York, 2013, ch. 8, pp. 185–204.
- [26] CVITANIĆ, J., POSSAMAÏ, D., AND TOUZI, N. Dynamic programming approach to principal-agent problems. *Finance and Stochastics* 22, 1 (2018), 1–37.
- [27] CVITANIĆ, J., AND ZHANG, J. *Contract theory in continuous-time models*. Springer, 2012.
- [28] DACHRAOUI, K., DIONNE, G., EECKHOUDT, L., AND GODFROID, P. Comparative Mixed Risk Aversion : Definition and Application to Self-Protection and Willingness to Pay. *Journal of Risk and Uncertainty* 29, 3 (December 2004), 261–276.

- [29] DE FINETTI, B. La prévision : Ses lois logiques, ses sources subjectives. *Annales de l’Institut Henri Poincaré* 17 (1937), 1–68.
- [30] DELLA VIGNA, S. Psychology and Economics : Evidence from the Field. *Journal of Economic Literature* 47, 2 (June 2009), 315–372.
- [31] DELONG, L. *Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications*. Springer, London, 2013.
- [32] DERVAUX, B., AND EECKHOUDT, L. Prévention en économie et en médecine : À propos de quelques malentendus. *Revue économique* 55, 5 (2004), 849–856.
- [33] DIAMOND, P., AND STIGLITZ, J. Increases in risk and in risk aversion. *Journal of Economic Theory* 8, 3 (1974), 337–360.
- [34] DOS REIS, G. *Some Advances on Quadratic BSDE : Theory, Numerics, Applications*. LAP Lambert Academic Publishing, Saarbrücken, 2011.
- [35] EECKHOUDT, L., GOLLIER, C., AND SCHLESINGER, H. *Economic and Financial Decisions Under Risk*. Economic and Financial Decisions Under Risk. Princeton University Press, 2005.
- [36] EHRLICH, I., AND BECKER, G. S. Market insurance, self-insurance, and self-protection. *Journal of political Economy* 80, 4 (1972), 623–648.
- [37] EL KAROUI, N., PENG, S., AND QUENEZ, M.-C. Backward stochastic differential equations in finance. *Mathematical Finance* 7, 1 (1997), 1–71.
- [38] ELLSBERG, D. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics* 75, 4 (1961), 643–669.
- [39] EMBRECHTS, P., AND HOFERT, M. Risk measures and dependence modeling. In *Handbook of Insurance*, G. Dionne, Ed. Huebner International Series on Risk, Insurance, and Economic Security (New York : Springer), New York, 2013, ch. 6, pp. 135–166.
- [40] ETNER, J., AND JELEVA, M. Risk perception, prevention and diagnostic tests. *Health Economics* 22, 2 (2013), 144–156.
- [41] FAN, S.-J., AND JIANG, L. Uniqueness result for the bsde whose generator is monotonic in y and uniformly continuous in z . *Comptes Rendus Mathematique* 348, 1 (2010), 89 – 92.
- [42] FISHBURN, P. C. Mean-risk analysis with risk associated with below-target returns. *American Economic Review* 67, 2 (1977), 116–26.
- [43] FÖLLMER, H., AND SCHIED, A. *Stochastic Finance : An Introduction in Discrete Time*. De Gruyter Textbook. De Gruyter, 2016.

- [44] FRITTELLI, M., AND ROSAZZA GIANIN, E. Putting order in risk measures. *Journal of Banking & Finance* 26, 7 (July 2002), 1473–1486.
- [45] GALICHON, A., AND HENRY, M. Dual theory of choice with multivariate risks. *Journal of Economic Theory* 147, 4 (2012), 1501 – 1516. Inequality and Risk.
- [46] GOLLIER, C., HAMMITT, J. K., AND TREICH, N. Risk and choice : A research saga. *Journal of Risk and Uncertainty* 47, 2 (Oct 2013), 129–145.
- [47] GRUENBERG, E. M. The prevention of mental disease. *The ANNALS of the American Academy of Political and Social Science* 286, 1 (1953), 158–166.
- [48] HANOCH, G., AND LEVY, H. The efficiency analysis of choices involving risk. *Review of Economic Studies* 36, 3 (1969), 335–346.
- [49] HIPP, C., AND PLUM, M. Optimal investment for insurers. *Insurance : Mathematics and Economics* 27, 2 (2000), 215 – 228.
- [50] HØJGAARD, B., AND TAKSAR, M. Optimal proportional reinsurance policies for diffusion models. *Scandinavian Actuarial Journal* 1998, 2 (1998), 166–180.
- [51] JACQUEMET, N., RULLIÁLRE, J.-L., AND VIALLE, I. Monitoring optimistic agents. *Journal of Economic Psychology* 29, 5 (2008), 698 – 714. Affect, Motivation and Decision Making.
- [52] JEWITT, I. Justifying the first-order approach to principal–agent problems. *Econometrica* 56, 5 (1988), 1177–1190.
- [53] JEWITT, I. Choosing between risky prospects : The characterization of comparative statics results, and location independent risk. *Management Science* 35, 1 (1989), 60–70.
- [54] JOUINI, E., SCHACHERMAYER, W., AND TOUZI, N. Optimal risk sharing for law invariant monetary utility functions. *Math. Finance* 18, 2 (2008), 269–292.
- [55] JULLIEN, B., SALANIÉ, B., AND SALANIÉ, F. Should more risk-averse agents exert more effort ? *The Geneva Papers on Risk and Insurance Theory* 24, 1 (Jun 1999), 19–28.
- [56] KAHNEMAN, D. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.
- [57] KAHNEMAN, D., AND TVERSKY, A. Subjective probability : A judgment of representativeness. *Cognitive Psychology* 3 (1972), 430 – 454.
- [58] KAHNEMAN, D., AND TVERSKY, A. On the psychology of prediction. *Psychological Review* 80 (1973), 237–251.
- [59] KAHNEMAN, D., AND TVERSKY, A. Prospect theory : An analysis of decision under risk. *Econometrica* 47, 2 (1979), 263–91.
- [60] KAZI-TANI, N., POSSAMAÏ, D., AND ZHOU, C. Quadratic BSDEs with jumps : a fixed-point approach. *Electronic Journal of Probability* 20, 66 (2015), 1–28.

- [61] KENKEL, D. Prevention. In *Handbook of Health Economics*, A. J. Culyer and J. P. Newhouse, Eds. Elsevier, 2000, ch. 31, pp. 1675–1720.
- [62] KONRAD, K., AND SKAPERDAS, S. Self-insurance and self-protection : A nonexpected utility analysis. *The Geneva Risk and Insurance Review* 18, 2 (1993), 131–146.
- [63] KROLL, Y., AND LEVY, H. Stochastic dominance : A review and some new evidence. *Research in Finance* 2 (1980), 163–227.
- [64] LANGLAIS, E. Willingness to pay for risk reduction and risk aversion without the expected utility assumption. *Theory and Decision* 59, 1 (2005), 43–50.
- [65] LEE, K. Risk aversion and self-insurance-cum-protection. *Journal of Risk and Uncertainty* 17, 2 (Nov 1998), 139–151.
- [66] LEPELTIER, J.-P., AND SAN MARTÍN, J. Backward stochastic differential equations with continuous coefficient. *Statistics & Probability Letters* 32, 4 (1997), 425–430.
- [67] LEPELTIER, J.-P., AND SAN MARTÍN, J. Existence for BSDE with superlinear-quadratic coefficient. *Stochastics : An International Journal of Probability and Stochastic Processes* 63, 3-4 (1998), 227–240.
- [68] LESUEUR, J.-Y. Health insurance coverage vs self-prevention : What impact of the insurer’s healthcare prevention programs ? working paper or preprint, Feb. 2018.
- [69] LOUBIÈRES, S., PARENT, A., AND TALLON, J.-M. L’économie de la prévention. Enjeux et problématiques. Université Paris1 Panthéon-Sorbonne (Post-Print and Working Papers) hal-00279338, HAL, 2004.
- [70] MCNEIL, A. J., FREY, R., AND EMBRECHTS, P. *Quantitative risk management : concepts, techniques and tools*. 2005.
- [71] MEYER, D. J., AND MEYER, J. A diamond-stiglitz approach to the demand for self-protection. *Journal of Risk and Uncertainty* 42, 1 (Feb 2011), 45–60.
- [72] MICHEL, D., JAN, D., MARC, G., ROB, K., AND ROGER, L. Risk measurement with equivalent utility principles. *Statistics & Risk Modeling* 24, 1/2006 (July 2006), 1–25.
- [73] MICHEL, V. Savage (leonard j.) - the foundations of statistics. *Revue Économique* 7, 6 (1956), 1042–1043.
- [74] MIKOSCH, T. *Non-Life Insurance Mathematics*. Springer, Berlin, Heidelberg, 2009.
- [75] MOORE, K. S., AND YOUNG, V. R. Optimal insurance in a continuous-time model. *Insurance : Mathematics and Economics* 39, 1 (2006), 47 – 68.
- [76] PAPAPANTOLEON, A., POSSAMAÏ, D., AND SAPLAOURAS, A. Existence and uniqueness for BSDEs with jumps : the whole nine yards. *arXiv preprint arXiv :1607.04214* (2016).

- [77] PARDOUX, E., AND PENG, S. Adapted solution of a backward stochastic differential equation. *System and Control Letters* 14, 1 (1990), 55–61.
- [78] PAULY, M. V. The economics of moral hazard : Comment. *The American Economic Review* 58, 3 (1968), 531–537.
- [79] PERETTI-WATEL, P., SEROR, V., DU ROSCOÄT, E., AND BECK, F. La prévention en question : attitudes à l'égard de la santé, perceptions des messages préventifs et impact des campagnes. *Documents de l'Institut national de prévention et d'éducation pour la santé* (2009).
- [80] PLOTT, C., AND SMITH, V., Eds. *Handbook of Experimental Economics Results*, 1 ed., vol. 1, Part 2. Elsevier, 2008.
- [81] PLOUS, S. Thinking the unthinkable : The effects of anchoring on likelihood estimates of nuclear war1. *Journal of Applied Social Psychology* 19, 1 (1989), 67–91.
- [82] PRATT, J. W. Risk aversion in the small and in the large. *Econometrica* 32, 1/2 (1964), 122–136.
- [83] PRELEC, D. The probability weighting function. *Econometrica* 66, 3 (1998), 497–528.
- [84] QUIGGIN, J. A theory of anticipated utility. *Journal of Economic Behavior & Organization* 3, 4 (December 1982), 323–343.
- [85] RAMSEY, F. P. Truth and probability. In *The Foundations of Mathematics and other Logical Essays*, R. B. Braithwaite, Ed. McMaster University Archive for the History of Economic Thought, 1926, ch. 7, pp. 156–198.
- [86] ROGERS, P. The cognitive psychology of lottery gambling : A theoretical review. *Journal of Gambling Studies* 14 (1998), 111–134.
- [87] ROGERSON, W. P. The first-order approach to principal-agent problems. *Econometrica* 53, 6 (1985), 1357–1367.
- [88] RONG, S. On solutions of backward stochastic differential equations with jumps and applications. *Stochastic Processes and their Applications* 66, 2 (1997), 209–236.
- [89] RONG, S. *Theory of Stochastic Differential Equations with Jumps and Applications*. Springer US, 2005.
- [90] ROTHSCHILD, M., AND STIGLITZ, J. Equilibrium in competitive insurance markets : An essay on the economics of imperfect information. *The Quarterly Journal of Economics* 90, 4 (1976), 629–649.
- [91] ROYER, M. Backward stochastic differential equations with jumps and related non-linear expectations. *Stochastic Processes and their Applications* 116, 10 (2006), 1358–1376.

- [92] SAMUELSON, P. A note on measurement of utility. *Review of Economic Studies* 4, 2 (1937), 155–161.
- [93] SCHMIDL, H. On minimizing the ruin probability by investment and reinsurance. *Ann. Appl. Probab.* 12, 3 (08 2002), 890–907.
- [94] SHOGREN, J. F., AND CROCKER, T. D. Risk, self-protection, and ex ante economic value. *Journal of Environmental Economics and Management* 20, 1 (1991), 1 – 15.
- [95] SIMON, H. A behavioral model of rational choice. *The Quarterly Journal of Economics* 69, 1 (1955), 99–118.
- [96] SLOVIC, P., FISCHHOFF, B., AND LICHTENSTEIN, S. Why study risk perception? *Risk Analysis* 2, 2 (1982), 83–93.
- [97] SLOVIC, P., LICHTENSTEIN, S., AND FISCHHOFF, B. Decision making. In *Stevens' Handbook of Experimental Psychology, second edition (in preparation)*, R. Atkinson, R. Herrnstein, G. Lindzey, and R. Luce, Eds. Wiley, New York, 1988, pp. 673–738.
- [98] TSANAKAS, A. Dynamic capital allocation with distortion risk measures. *Insurance : Mathematics and Economics* 35, 2 (2004), 223 – 243. Papers presented at the 7th IME Conference.
- [99] TVERSKY, A., AND KAHNEMAN, D. Belief in the law of small numbers. *Psychological Bulletin* 76 (1971), 105–110.
- [100] TVERSKY, A., AND KAHNEMAN, D. Judgment under uncertainty : Heuristics and biases. *Science* 185, 4157 (1974), 1124–1131.
- [101] TVERSKY, A., AND KAHNEMAN, D. Rational choice and the framing of decisions. *The Journal of Business* 59, 4 (1986), S251–S278.
- [102] TVERSKY, A., AND KAHNEMAN, D. Advances in prospect theory : Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 4 (Oct 1992), 297–323.
- [103] TVERSKY, A., AND WAKKER, P. Risk attitudes and decision weights. *Econometrica* 63, 6 (1995), 1255–80.
- [104] VON NEUMANN, J., AND MORGENTERN, O. *Theory of games and economic behavior*. Princeton University Press, 1947.
- [105] WAKKER, P. *Transforming Probabilities without Violating Stochastic Dominance*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989, pp. 29–47.
- [106] WAKKER, P. P. *Prospect Theory : For Risk and Ambiguity*. Cambridge University Press, 2010.
- [107] WANG, S. S. A class of distortion operators for pricing financial and insurance risks. *The Journal of Risk and Insurance* 67, 1 (2000), 15–36.

- [108] WU, G., AND GONZALEZ, R. Curvature of the probability weighting function. *Management Science* 42, 12 (1996), 1676–1690.
- [109] WU, G., AND GONZALEZ, R. Nonlinear decision weights in choice under uncertainty. *Management Science* 45, 1 (1999), 74–85.
- [110] YAARI, M. E. *Univariate and multivariate comparisons of risk aversion : a new approach*, vol. 3. Cambridge University Press, 1986, pp. 173–188.
- [111] YAARI, M. E. The dual theory of choice under risk. *Econometrica* 55, 1 (1987), 95–115.
- [112] ZWEIFEL, P., AND EISEN, R. *Insurance Economics*, 2012 ed. Springer, 2012.

Chapitre 1

Prevention efforts, insurance demand and price incentives under coherent risk measures

Ce chapitre reprend un article co-écrit avec Nicolás Hernández Santibáñez et Nabil Kazi-Tani et publié dans le journal Insurance : Mathematics and Economics¹.

Abstract

This paper studies an equilibrium model between an insurance buyer and an insurance seller, where both parties' risk preferences are given by convex risk measures. The interaction is modeled through a Stackelberg type game, where the insurance seller plays first by offering prices, in the form of safety loadings. Then the insurance buyer chooses his optimal proportional insurance share and his optimal prevention effort in order to minimize his risk measure. The loss distribution is given by a family of stochastically ordered probability measures, indexed by the prevention effort. We give special attention to the problems of self-insurance and self-protection, and show that if the buyer's risk measure decreases faster in effort than his expected loss, optimal effort is non-decreasing in the safety loading with a potential discontinuity when optimal coverage switches from full to zero. On the contrary, if the decrease of the buyer's risk measure is slower than the expected loss, optimal effort may or may not be non-decreasing in the safety loading. In self-insurance, we consider also an adverse selection setting, where the type of the insurance buyers is given by his loss probability, and study the screening and shutdown contracts. In case of Pareto distributed losses, the seller sets the highest possible price under which the buyer still prefers full insurance over no insurance. We also analyze the case of discrete distributions: on the one hand, for self-insurance, under the assumption that the marginal impact of the effort is higher on small losses than it is on catastrophic losses, the optimal effort is non-decreasing in the safety loading. On the other hand,

1. <https://doi.org/10.1016/j.insmatheco.2020.05.006>

in the case of self-protection, more conditions are needed, in particular, we obtain sufficient conditions for the optimal effort to be non-decreasing or non-monotone in the safety loading.

Key words: Prevention, Self-insurance, Self-protection, Coherent risk measures, Stackelberg game.

1 Introduction

The fact that rational agents undertake actions to lower their risk exposure is now standard in economic analysis. Insurance buyers can try to reduce the size of the potential claims that they face (self-insurance) or reduce the probability of a claim arrival (self-protection).

In the particular case of health insurance, the increased knowledge of individual health state, through connected devices for instance, make the efforts of risk reduction easier. Insurers themselves start to incentivize prevention efforts: the companies John Hancock or Aetna decided in 2018 to offer connected watches or bracelets to their clients, in exchange for their commitment to a physical activity program, whose progress is measured by the connected device.

The first systematic study of prevention effort, in the form of self-insurance or self-protection, was made in the seminal paper of Ehrlich and Becker [22]. Their comparative statics analysis showed that the following two conclusions hold true:

- (A) Market insurance and self-insurance are substitutes, in the sense that an increased insurance price implies more prevention effort to reduce the losses size.
- (B) Market insurance and self-protection can be complements, in the sense that a more expensive market insurance entails less effort to reduce the probability of losses.

Since then, several papers tested the robustness of these results to various changes in the model (see the paper [14] and the literature below for a short overview). This allowed a better understanding of prevention in the case of observable effort, while making the distinction between self-protection and self-insurance. In the case of moral hazard, little is known on optimal choices under these two forms of risk reduction. Under expected utility, Hölstrom [29] shows that in case of self-protection, the optimal insurance contract is a deductible, also called unlimited layer, while in case of self-insurance, Winter [48] explains that low losses are fully covered and large losses are partially covered.

One of the objectives of this paper is to revisit the question whether risk reduction is a substitute or a complement for insurance but under moral hazard. We want to contribute to this line of research in at least three aspects: first, we model the interaction between the insurance buyer and seller as a game and we construct an equilibrium consisting of optimal prevention and insurance parameters for the buyer and optimal price strategy for the seller.

Second, as criteria for both market participants, we take general coherent risk measures in the sense of [2]. The advantage in using coherent risk measures for the objective functions lies in the possibility to identify a simple condition that demarcates situations where only substitution is possible from those where complementarity can arise. And last, the impact of the prevention effort is modeled via a family of stochastically ordered probability measures which are indexed by the prevention effort: this allows to deal with self-insurance and self-protection as particular cases, and to give sufficient conditions so that the conclusions (A) and (B) hold true. Let us now describe each of these aspects and the related assumptions.

Summary of the approach.

A Stackelberg game. The way that we formulate the problem is inspired by contract theory, and in particular by the Principal–Agent problem. Indeed, we consider the insurance buyer as the Agent, whose action is given by a vector (α, e) , where α is the proportion of losses paid by the insurer and e is the effort made by the Agent to reduce his risk. Throughout the paper, we will sometimes refer to the insurance buyer as the Agent, and to the insurance seller as the Principal. The contract proposed by the Principal is completely summarized by the price offered for the cover α , which is given in the form of a safety loading (Equation (3.1)). The effort e will only impact the losses distribution, and not the loss outcomes directly. As a consequence, the choice α will be directly observable by the Principal, whereas the effort action e will be unobservable. This unobservable action case is referred to as the *moral hazard* situation ([43],[29],[44]).

The problem that we consider is a one period model. Static Principal-Agent problems involving moral hazard are known to be difficult to solve (see Chapter 2 in [17] or the articles [44, 31] for a discussion on the relaxed formulations of the problem), we are however able to completely solve the Agent's problem in the particular case that we deal with, thanks to the singular form of the optimal coverage α , which will be equal to either 0 or 1. Then, the remaining optimization in the variable e will be non convex, but still solvable by analyzing the problem "locally" and then carefully pasting the solutions together (subsections 3.2 and 3.3).

As already noticed and discussed in [20], the corner type solution for the optimal coverage mainly comes from the positive homogeneity of the criteria (see Definition 2.1) and the linearity of the insurance contract. The linearity of the insurance contract is not a restrictive assumption: it corresponds to some contracts found in practice, in particular in health insurance. It is also extremely common in reinsurance, where this type of contract is referred to as "quota-share". Notice that our analysis readily extends to the relation between an insurer (this time the Agent) and a reinsurer (the Principal), where the effort of the agent in that case consists for instance in underwriting policy control, or in prevention campaigns for the clients.

Risk Measures. In standard contract theory literature (see [43, 29, 44, 45, 17, 15]), the formal optimization problems consist in expected utility maximization. This is also the case in [22] and in most of the papers that followed ([14]). In the present paper however, the criterion of both the Agent and the Principal will be given by a law invariant coherent risk measure (see Definition 2.1). Risk measures are a popular risk assessment tool, in particular because of the associated axiomatic analysis and its interpretation ([2, 26]). It is a common tool in insurance and reinsurance practice, since it appears in the European regulation texts: indeed, the quantile function, also called Value-at-Risk, is the risk measure that European insurers and reinsurers have to compute, according to the Solvency II rules. Note that the quantile risk measure is not coherent (which was one of the initial motivations of [2]), and it is quite often in practice compared to the Average Value-at-Risk (AVaR), also called Tail Value-at-Risk, or Expected Shortfall (see [25] for the definitions, and the differences between these measures). AVaR is one example of a coherent risk measure, which is also a distortion risk measure (see (2.1)).

The results obtained in this paper are valid if the risk measures used by the Agent and the Principal are law invariant and coherent. This class of risk measures is more general than the class of distortion risk measures. We give in Example 2.1 a class of law invariant coherent risk measures, which are not distortion risk measures. Note that if ρ is a coherent risk measure, then $U(X) := -\rho(-X)$ defines a monetary utility function. Our results can then be recast in terms of law invariant monetary utility functions, but we state them in the language of risk measures, which is closer to actuarial practice. Note that the class of monetary utility functions do not contain expected utilities as a particular case. On the contrary, the only expected utility function which is also *cash-additive* (see Definition 2.1) is linear, corresponding to a risk neutral agent. The general risk sharing problem has been solved in [33], when economic agents are law invariant monetary utility maximizers, and in [41] in the case of quasiconvex risk measures. As showed in [4] and [33], the resolution of the risk sharing problem in the monetary case reduces to the *inf-convolution* of the given risk measures. In our case however, since the agent's effort is not observable by the Principal, we have to follow another approach, mainly the backward induction, which consists in solving the Agent's problem first and then solving the Principal's problem knowing the optimal responses of the Agent. This is doable here since, contrary to the setting in [4] and [33], we fix the form of insurance contracts, which are linear. We refer the reader to [25, 23] for more details on risk measures and insurance applications.

Controlling the distributions. Let us now discuss the type of loss distributions that we consider. In the literature on the interaction between market insurance, self-insurance and self-protection, the losses distributions are discrete, concentrated on two points: the insured suffers a loss $L > 0$ with probability p or no loss with probability $1 - p$, where $p \in (0, 1)$. In our setting (see Section 2 for more details), we will consider a family of random variables $(X_e)_{e \in (0, +\infty)}$, indexed by a prevention effort parameter e , and such that the distributions of

$(X_e)_{e \in (0, +\infty)}$ form a family of probability measures which is decreasing in the sense of first order stochastic dominance (FSD), that is,

$$e_1 < e_2 \implies E[f(X_{e_1})] > E[f(X_{e_2})], \quad (1.1)$$

for every non-decreasing function $f : \mathbb{R} \rightarrow \mathbb{R}$. In other words, an increase of the effort lowers the severity of the loss, where severity is measured by FSD. We will write $X \leq_{(\text{mon})} Y$ to say that X is dominated by Y in the sense of FSD, also called the *monotone* order. We will use the following property of the FSD (see Theorem 2.68 in [25])

$$X \leq_{(\text{mon})} Y \iff \bar{q}_X(u) \leq \bar{q}_Y(u), \quad \forall u \in (0, 1),$$

where \bar{q}_X denotes the tail quantile function associated to the distribution of X . Since we will use a family of distributions, indexed by some prevention effort, we will assume that the effort impacts the tail quantile of the loss distribution, in such a way that the marginal impact will be decreasing. Note that in Equation (2.2), we do not make any convexity nor concavity assumption on f , this means that the situation in which the distributions of $(X_e)_{e \in (0, +\infty)}$ are decreasing for the monotone convex order is included in our analysis.

Using this framework, we are able to deal with self-insurance and self-protection in a unified way. In particular, we provide sufficient conditions for Statements (A) and (B) to hold true.

We give several explicit examples of probability distributions, which are in line with standard actuarial practice. We analyze in more details the case where, for every e , X_e has a distribution given by a mixing between a Dirac mass at 0 and a Pareto distribution with some parameter $k(e)$ depending of the effort. For more details on standard actuarial assumptions, we refer the reader to [6].

Related Literature.

As already mentioned above, there is a large literature on self-insurance and self-protection, and the very brief overview that we give here is by no means exhaustive. We refer the interested reader to [27], [21] (Chapter 9) and to the paper [14], which provides a survey on the economics of prevention.

Even if we started with conclusions (A) and (B) from the seminal paper of Erlich and Becker [22] as a motivation, our analysis differs from [22] in various ways: first, we model risk preferences through risk measures (or equivalently, through monetary utility functions), which are different from expected utility theory. Second, the buyer's effort is not observable by the seller in our framework, corresponding to the moral hazard case. And third, we consider a unique risk-averse seller, unlike in [22], where the supply side of insurance is given by multiple risk-neutral insurers in a competitive market: this is sometimes referred to as *monopoly pricing*. Our model is also equivalent to a unique seller maximizing a monetary utility function,

applied to a net profit. So our seller's behavior is close to profit maximizing. Despite these differences, we still want to analyze whether risk reduction is a substitute or a complement for insurance.

Most of the literature focuses on binary risks. Some contributions already extended that simple framework: [32], [46], [34], examine the effects of agents' risk preferences on their optimal effort, for more general loss distributions, in particular for distributions satisfying the so-called single-crossing property. This property appears also in [3], that provides sufficient conditions on the utility functions and the probability distributions, allowing to obtain comparative statics results.

When the agent is subject to both self-insurance and self-protection activities, [38] studied the case of discrete distributions. The author showed that even if the marginal reduction of a loss dominates the marginal increase of the effort's cost, more risk averse agents do not necessarily exert more effort: he provides a sufficient condition on the probability distribution for this result to hold, this condition is comparable to the one we obtain in Subsection 5.2 for Statement (A) to hold true. [42] analyzes the interaction between self-protection activities, prudence and risk aversion, under the assumption that a change in the level of self-protection is not mean preserving, but rather mean utility preserving, for any utility function. [13] considers random vectors instead of one-dimensional random variables. The marginal distributions are still bi-variate, and the effect of the dependence structure between the marginals on prevention decisions is studied.

In all these papers, agents maximize their expected utility. Several authors studied the robustness of assertions (A) and (B), or the relationship between risk aversion and prevention efforts, under other criteria. [35] considered the class of Rank Dependent Expected Utilities (RDEU), which allows for probability weighting. In the language of risk measures, in the RDEU framework, an agent facing a risk X minimizes the quantity $\rho(u(X))$ where ρ is a distortion risk measure (see (2.1)) and u a non-decreasing and concave utility function. It is shown in [35] that more risk averse agents exert more self-insurance effort. In [24], self-protection activities in a health insurance context are studied, in a RDEU framework, with non concave distortions. When RDEU is considered in the particular case of $u(x) = x$, we obtain the so-called Yaari's Dual Theory (DT). [12] analyzed the interaction between market insurance, self-insurance, and self-protection in the DT framework, and proved that the main results of [22] carry over to that case. Let us mention the papers [37] and [5], that focus on the relation between risk aversion and the Willingness to Pay (WTP) for general risk preferences (RDEU in the case of [5]).

Finally, let us mention some literature on reinsurance arrangements. [9] studies the Stackelberg equilibrium between an insurer and a monopolistic reinsurer, who chooses the optimal pricing density of the contract. [11] extends the previous work to more general premiums offered by the reinsurer and a distortion risk measure for the insurer. [10] studies the dynamic continuous-

time formulation of the problem, under utility maximization criteria for both sides. In all these works the insurer's only action is to choose the desired level of insurance, which is of course observed by the reinsurer. This entails some differences between their execution of the backward induction approach and ours, due to the absence of moral hazard in their setting. Technically, in these works the agent does not control the distribution of the losses so they study a strong formulation of the Stackelberg game, whereas we consider a weak formulation. Also, the two-dimensionality of our setting does not allow to use directly the standard optimality conditions for the problem of the insurance buyer since we have to be careful about corner solutions.

Main contributions.

The current paper extends the problems of self-insurance and self-protection to the case of a Stackelberg game between the insurance buyer and seller, under law invariant coherent risk measures, and in a situation of moral hazard. We prove that the formulated game always admits an equilibrium that we characterize.

The focus lies on the impact of an increased price on optimal insurance demand and prevention effort. We give emphasis to the fact that whether Statements (A) and (B) hold true or not essentially depends on the relative impact of the effort on the losses distribution through the risk measure and on the price (Assumptions 3.1 and 3.2).

The methodology described in this paper can be applied to a large class of insurance claim distributions. To illustrate this, we analyze in detail the case of the Pareto distribution, which is standard in actuarial science and we give some insight about the case of discrete distributions, which is standard in insurance economics.

Structure of the paper

Section 2 discusses the problem and our main assumptions. In Section 3, we completely solve the buyer's problem. Section 4 is dedicated to the seller's problem. Section 5 illustrates the proposed methodology in particular cases, where we specify the claim distribution family and the class of risk measures. Section 6 provides numerical analysis for a Pareto-type distribution family, then we show that the form of probability distribution that we consider can be used as a proxy for the Compound Poisson distribution, which is common in actuarial models. In order to facilitate reading, technical proofs are provided in the Appendix A.

2 Model and Assumptions

Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a fixed probability space. For a given random variable X , \bar{F}_X denotes its survival function defined by $\bar{F}_X(x) := \mathbb{P}(X > x)$ and \bar{q}_X denotes its tail quantile function,

defined as the inverse of \bar{F}_X :

$$\bar{q}_X(u) := \inf\{x \in \mathbb{R} \mid \bar{F}_X(x) \leq u\}.$$

We assume that the insured can make some effort, modeled by a parameter $e \in \mathbb{R}^+$, in order to reduce the losses, or the probability to suffer losses. Our assumptions allow us to deal with self-insurance and self-protection in a unified way, using stochastic orders. Let us introduce the probability distributions of insurance claims in more details.

The losses distributions. Let $(X_e)_{e \in [0, +\infty)}$ be a family of non-negative random variables, representing losses, such that the distributions μ_e of X_e are decreasing in the sense of FSD. In particular, this implies that for each $u \in (0, 1)$, the function $e \mapsto \bar{q}_{X_e}(u)$ is non-increasing. To simplify the notations, we will write $Q(u, e) := \bar{q}_{X_e}(u)$. We also assume that the marginal impact of the effort is non-increasing.

Assumption 2.1. *For every $u \in (0, 1)$, $e \mapsto Q(u, e)$ is a convex function.*

The objectives. We model the interaction between the insurance buyer (IB) and the insurance seller (IS) as a sequential game: IS plays first by offering prices, then IB plays second by choosing his optimal insurance cover and his optimal effort parameter. We assume that both players minimize a law invariant coherent risk measure, whose definition is recalled below for completeness.

Definition 2.1. *$\rho : L^1 \rightarrow \mathbb{R}$ is a law invariant coherent risk measure if*

1. *ρ is monotone, in the sense that $X \leq Y$ almost surely implies that $\rho(X) \leq \rho(Y)$.*
2. *ρ is cash-additive: $\forall m \in \mathbb{R}$, $\rho(X + m) = \rho(X) + m$.*
3. *ρ is convex : For each X and Y in L^1 and for all $\lambda \in [0, 1]$, $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$.*
4. *ρ is positively homogeneous: $\forall \lambda \in \mathbb{R}^+$, $\rho(\lambda X) = \lambda\rho(X)$.*
5. *ρ is law-invariant: $X \stackrel{d}{=} Y$ implies that $\rho(X) = \rho(Y)$, where $\stackrel{d}{=}$ denotes equality in distribution.*

If ρ satisfies conditions 1. to 4. in Definition 2.1, then it is called a coherent risk measure, as introduced in [2]. Let us briefly discuss the economic interpretations behind these properties: property 1. is the most obvious, indeed if a loss Y is bigger than X in all states of nature, we want to associate to Y a bigger risk value. The second property is a rather strong requirement mathematically, economically though, it says that the number $\rho(X)$ has a unit, in our case measured in capital (Euros, Dollars, etc.). The third property corresponds to the intuitive idea that agents should prefer diversification, and under property 4., it is equivalent to the subadditivity of risks. The positive homogeneous property says that the way we measure the

risk should be independent of the size of the position: this is obviously a strong economic assumption. On the other hand, it can be interpreted by saying that the risk measurement should be independent of the currency in which the losses are evaluated. Finally, property 5., which is of probabilistic nature, says that the risk measurement procedure should be independent of the particular underlying assets, but that only the distribution of losses matters. For more details and properties of risk measures, in particular for relations with stochastic dominance (see also Remark 2.1), or regulatory issues, we refer the reader to [25] (Chapter 4).

An important class of law-invariant coherent risk measures is given by the distortion risk measures, defined by

$$\rho(X) := \int_0^1 \bar{q}_X(u) d\psi(u), \quad (2.1)$$

where ψ is a distortion function, i.e. a non-decreasing function from $[0, 1]$ to $[0, 1]$ such that $\psi(0) = 0$ and $\psi(1) = 1$. If the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ has no atoms, then one can prove that a risk measure of the form (2.1) is convex if and only if the distortion function ψ is concave.

Example 2.1. Let us give an example of coherent risk measure, which is not a distortion risk measure:

$$\rho(X) := \mathbb{E}[X] + \delta \mathbb{E}[(\mathbb{E}[X] - X)_-^p]^{1/p}, \quad \delta \in (0, 1), \quad 1 < p < +\infty,$$

satisfies all properties of Definition 2.1, and cannot be written in the form (2.1). We refer the reader to [33] (Section 3.4) for more details on the so-called semi-deviation risk measure.

The next lemma is fundamental for our study. Its proof is deferred to Appendix 1.

Lemma 2.1. Let ρ be a risk measure satisfying the properties 1,2,3 and 5 in Definition 2.1. If the distributions of $(X_e)_{e \in [0, +\infty)}$ satisfy Assumption 2.1, then

1. The function $e \mapsto \rho(X_e)$ is convex.
2. The function $e \mapsto \rho(X_e)$ is continuous, non-increasing and $\rho(X_e) \geq \mathbb{E}[X_e] \geq 0$.

Remark 2.1. The fact that $e \mapsto \rho(X_e)$ is non-increasing, essentially comes from the fact that the class of risk measures that we consider respects first order stochastic dominance.

Remark 2.2. For the particular distortion function $\psi(u) = u$, the risk measure defined in (2.1) reduces to $\rho(X) = \mathbb{E}(X)$. Then by Lemma 2.1, we have that $e \mapsto \mathbb{E}(X_e)$ is a continuous, non-increasing and convex function.

The sequential game. As mentioned above, we model the interaction between IB and IS through a game, where IS plays first and IB plays second. We solve this game using backward induction.

IB has initial wealth denoted ω_0 and faces a risk of loss X_e for which he subscribes a proportional insurance contract with IS. In this insurance program, he chooses his level of insurance denoted $\alpha \in [0, 1]$ and he pays a premium denoted Π . The prevention effort $e \in \mathbb{R}^+$ made by IB has a monetary cost $c(e)$, where c is a non-decreasing and strictly convex function.² We suppose that IB uses a law-invariant and coherent risk measure ρ_1 . The goal of IB is to minimize the risk measure associated to his total loss, which is given by:

$$L_1(\alpha, e) := \rho_1((1 - \alpha)X_e + \Pi(\alpha X_e) + c(e) - \omega_0).$$

Using the positive homogeneity and cash-additivity properties of ρ_1 , we obtain that IB's objective function simplifies to

$$L_1(\alpha, e) = (1 - \alpha)\rho_1(X_e) + \Pi(\alpha X_e) + c(e) - \omega_0.$$

Therefore IB wants to solve the following optimization problem:

$$\inf_{(\alpha, e) \in [0, 1] \times (0, \infty)} L_1(\alpha, e). \quad (2.2)$$

Remark 2.3. *From this point forward, we consider the initial wealth ω_0 to be equal to 0. This is without loss of generality, by the cash-additivity of ρ_1 .*

On the other hand, IS wants to find the best premium for the insurance coverage that IB is willing to accept. We suppose that IS uses a law-invariant and coherent risk measure ρ_2 and her goal is to minimize her risk measure associated to her loss, given by

$$L_2(\Pi) := \rho_2(\alpha^* X_{e^*} - \Pi(\alpha^* X_{e^*})),$$

where α^* and e^* are functions of the premium Π and correspond to the insurance demanded by IB and the effort he performs under it. Therefore, IS solves the following optimization problem

$$\inf_{\Pi} L_2(\Pi), \quad (2.3)$$

where the set of premiums over which IS optimizes will be specified in the next section.

2. If c is merely convex, all our analysis holds true by choosing the smallest effort realizing the minimum when needed.

3 The insurance buyer's optimization problem

3.1 Optimal insurance cover

We start the resolution of IB's problem by studying his optimal insurance coverage. In the literature this coverage is either equal to 0 or 1 (see [12], [40], [47], or see [20] for empirical results). In the following proposition, we give a sufficient condition on the pricing functional Π under which the optimal value for α is either 0 or 1.

Proposition 3.1. *Assume that Π is positively homogeneous*

$$\Pi(\lambda X) = \lambda \Pi(X), \quad \forall \lambda \geq 0,$$

then the optimal insurance coverage α^ only takes the values 0 or 1.*

Proof. By our assumptions, we get that the objective function L_1 is linear in α , since

$$L_1(\alpha, e) = (1 - \alpha)\rho_1(X_e) + \alpha\Pi(X_e) + c(e).$$

So the optimal insurance demand only depends on the sign of $\Pi(X_e) - \rho_1(X_e)$, and is given by

$$\alpha^*(e) = \begin{cases} 1 & \text{if } e \in \mathcal{I} \\ 0 & \text{if } e \in \mathcal{N} := \mathcal{I}^c, \end{cases}$$

where $\mathcal{I} := \{e \in \mathbb{R}^+ \mid \Pi(X_e) \leq \rho_1(X_e)\}$. □

In other words, the optimal coverage will be the full insurance coverage $\alpha^* = 1$ when the price asked by the insurer is lower than the risk associated to the claims distribution, and it will be 0 otherwise.

In the actuarial science literature, various premium functionals are analyzed (see [36] or [19]). One standard example is the safety loading:

$$\Pi(X) := (1 + \theta)\mathbb{E}[X], \tag{3.1}$$

where θ is a non-negative parameter. In this case, $\mathcal{I} := \left\{ e \in \mathbb{R}^+ \mid 1 + \theta \leq \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \right\}$. For $\theta = 0$, this is the so-called pure premium principle.

Another example is the standard deviation principle:

$$\Pi(X) := \mathbb{E}[X] + \gamma\sigma_X,$$

where σ_X denotes the standard deviation of X and γ is a non-negative parameter. In this case we have $\mathcal{I} := \left\{ e \in \mathbb{R}^+ \mid \gamma \leq \frac{\rho_1(X_e) - \mathbb{E}[X_e]}{\sigma_{X_e}} \right\}$.

Remark 3.1. *The safety loading and the standard deviation principles satisfy the positive homogeneity property given in Proposition 3.1. If this property is not satisfied, then one can find an interior solution $\alpha^* \in (0, 1)$ for the optimal insurance coverage. In the paper [20], the authors avoid corner solutions in two ways: by considering a non linear contract with a deductible, or by introducing a source of background risk. In our case, we could consider a premium functional Π which is not homogeneous, and introduce interior solutions this way, even in the case of a single insurable source of risk. An example of such functional Π is the variance principle $\Pi(X) := \mathbb{E}[X] + \beta V(X)$.*

Remark 3.2. *Note that for Proposition 3.1 to hold true, we only need ρ_1 to satisfy the properties 2 and 4 of cash-additivity and positive homogeneity of Definition 2.1.*

Throughout the rest of the paper, we assume for simplicity that $\Pi(X) = (1 + \theta)\mathbb{E}[X]$, with $\theta \geq 0$, but our results can be generalized to other insurance pricing functionals. Recall that IB demands (full) insurance, that is $\alpha^*(e) = 1$, if $e \in \mathcal{I}$ and IB demands no insurance, that is $\alpha^*(e) = 0$ if $e \in \mathcal{N}$, where

$$\mathcal{I} := \left\{ e \in \mathbb{R}^+ \mid 1 + \theta \leq \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \right\}.$$

Let us define then the following function G and state our main assumption on it, that says how the effort impacts the risk and the price of insurance (represented by the expectation, given the safety loading premium).

$$G(e) := \frac{\rho_1(X_e)}{\mathbb{E}[X_e]}. \quad (3.2)$$

In this paper we assume the function G is monotonic, which gives two different cases that we study separately. We first assume that G is non-increasing, which is a natural economic assumption, since it means that an increased effort has a bigger impact on the risk than it has on the price: we analyze the self-insurance problem under this assumption. Then, in Subsection 3.3 we assume that G is non-decreasing: we prove that this assumption is always satisfied in the self-protection case. In both cases, we will solve completely the problem of IB and discuss Statements (A) and (B) mentioned in the Introduction.

3.2 Non-increasing G

As discussed in Section 2, suppose that the agent can exert an effort to reduce the risk that he is exposed to. In this subsection, we assume that an increased effort has a bigger impact on the risk than it has on the price. We will prove next that for any model satisfying Assumption 3.1 below, Statement (A) of the Introduction holds true, i.e. less market insurance translates into more prevention effort.

Assumption 3.1. *The function G defined in (3.2) is non-increasing, or equivalently*

$$e_1 \leq e_2 \implies \frac{\rho_1(X_{e_1}) - \rho_1(X_{e_2})}{\rho_1(X_{e_1})} \geq \frac{\mathbb{E}[X_{e_1}] - \mathbb{E}[X_{e_2}]}{\mathbb{E}[X_{e_1}]}.$$

We already know that the functions $e \mapsto \mathbb{E}[X_e]$ and $e \mapsto \rho(X_e)$ are non-increasing and convex. The assumption above asserts that the effort impact is bigger on the risk measure ρ_1 than it is on the expectation, meaning that riskiness is more sensitive to effort than the expected loss. This assumption is satisfied in various models of self-insurance, which is the main example we have in mind, although the analysis we develop and our results are more general.

Example 3.1. Consider the distribution of the loss random variable X_e given by

$$P_{X_e} := (1-p)\delta_{\{0\}} + p P_{Y_e}, \quad (3.3)$$

where $0 < p < 1$, P_{Y_e} denotes the distribution (controlled by IB) of a positive random variable, and $\delta_{\{0\}}$ is the Dirac mass at 0. As shown in Section 5.1, when the random variables Y_e are Pareto distributed, Assumption 3.1 is satisfied. We also mention the cases where the random variables Y_e have a Fréchet, Weibull or Log-Normal distribution, as other standard examples in which Assumption 3.1 is satisfied.

We proceed now to solve the problem (2.5) of IB, consisting in the minimization of the functional L_1 . For fixed e , we already solved the problem of minimizing L_1 in α . For fixed α , we know from Lemma 2.1 and Remark 2.2 that

$$L_1(\alpha, e) = (1 - \alpha)\rho_1(X_e) + \alpha(1 + \theta)\mathbb{E}[X] + c(e),$$

is convex, as the sum of three convex functions. Note now that the function $e \mapsto L_1(\alpha^*(e), e)$ is not convex, so its minimization is not straightforward. However, we can take advantage of the fact that α^* only takes the values 0 and 1 to study $L_1(\alpha^*(e), e)$ separately on the sets \mathcal{N} and \mathcal{I} , on which it is a convex function, and then compare the local minima on \mathcal{N} and \mathcal{I} to get a global minimum.

Define $L_{\mathcal{N}}$ and $L_{\mathcal{I}}^\theta$ as follows, for $e \in \mathbb{R}^+$,

$$\begin{aligned} L_{\mathcal{N}}(e) &:= \rho_1(X_e) + c(e), \\ L_{\mathcal{I}}^\theta(e) &:= (1 + \theta)\mathbb{E}[X_e] + c(e). \end{aligned}$$

$L_{\mathcal{N}}$ and $L_{\mathcal{I}}^\theta$ are convex and such that

$$L_1(\alpha^*(e), e) = \begin{cases} L_{\mathcal{N}}(e), & \text{if } e \in \mathcal{N}, \\ L_{\mathcal{I}}^\theta(e), & \text{if } e \in \mathcal{I}. \end{cases} \quad (3.4)$$

Lemma 3.1. There exists $e_\theta \in \mathbb{R}^+$ such that \mathcal{N} and \mathcal{I} are two intervals of the form

$$\begin{aligned} \mathcal{N} &= (e_\theta, +\infty), \\ \mathcal{I} &= [0, e_\theta]. \end{aligned}$$

Proof. We already know from Lemma 2.1 that both functions $e \mapsto \rho_1(X_e)$ and $e \mapsto \mathbb{E}[X_e]$ are continuous. Also, since X_e has finite expectation, $G(e) = \frac{\rho_1(X_e)}{\mathbb{E}[X_e]}$ is continuous and non-increasing. We can define its right continuous inverse $G^{-1}(y) := \inf\{e \in \mathbb{R}^+ \mid G(e) < y\}$ and

$$e_\theta := G^{-1}(1 + \theta). \quad (3.5)$$

Note that $G(e_\theta) = 1 + \theta$, since G is continuous. Then by definition, for every $e \leq e_\theta$, $G(e) \geq 1 + \theta$, which means that $e \in \mathcal{I}$ and for every $e > e_\theta$, $G(e) < 1 + \theta$, which means that $e \in \mathcal{N}$. \square

Let us now study the function L_I^θ . It is a continuous and strictly convex function, that goes to $+\infty$ as e tends to $+\infty$.³ So it has a unique minimizer on \mathbb{R}^+ that we denote $e_I(\theta)$. The following proposition identifies the local minimizer of L_I^θ on the set \mathcal{I} . Its proof is postponed to the Appendix A.

Proposition 3.2. *Suppose that $e \mapsto \mathbb{E}[X_e]$ and $e \mapsto c(e)$ are differentiable. There exist constants $\theta_0 < \theta_I$ such that*

- (a) $\theta \leq \theta_I \implies \operatorname{argmin}_{\mathcal{I}} L_I^\theta = \{e_I(\theta)\}$.
- (b) $\theta_I < \theta \implies \operatorname{argmin}_{\mathcal{I}} L_I^\theta = \{e_\theta\}$.

Moreover, the function $\theta \rightarrow e_I(\theta)$ is non-decreasing and $\theta \leq \theta_0 \implies e_I(\theta) = 0$.

Let us now study the local minimizer of L_N on the set \mathcal{N} . The function $L_N(e) = \rho(X_e) + c(e)$ is non-negative, strictly convex and coercive. Therefore it admits a unique minimizer on \mathbb{R}^+ , that we denote e_N . In the next proposition, we identify the local minimizer of L_N over the set \mathcal{N} . Its proof can be found in the Appendix A .

Proposition 3.3. *There exists a positive constant θ_N such that*

- (a) $\theta \leq \theta_N \implies \operatorname{argmin}_{\mathcal{N}} L_N = \{e_\theta\}$.
- (b) $\theta > \theta_N \implies \operatorname{argmin}_{\mathcal{N}} L_N = \{e_N\}$.

There remains to determine the global minimum of the function $e \mapsto L_1(\alpha^*(e), e)$ defined in (3.4). This can be done by carefully analyzing $e_I(\theta)$ and e_θ as functions of θ . By doing so, we obtain the main result of this subsection, whose proof can be found in Appendix 2.

Theorem 3.1. *Let $e^* := \operatorname{argmin}_e L_1(\alpha^*(e), e)$ be the global minimum of the function $e \mapsto L_1(\alpha^*(e), e)$. There exists a positive constant θ_M such that*

- (a) $\theta \leq \theta_M \implies e^* = e_I(\theta)$.

3. Since $L_I^\theta(e) \geq c(e)$, and $c(e)$ being non-decreasing and strictly convex, we have $\lim_{e \rightarrow +\infty} c(e) = +\infty$.

(b) $\theta > \theta_M \implies e^* = e_N$.

θ_M represents the maximum value of the safety loading such that the insurance buyer accepts to enter the contract and pay the price given by $(1 + \theta_M)\mathbb{E}[X_{e_I(\theta_M)}]$. For $\theta > \theta_M$, he chooses not to buy any insurance and exerts the optimal effort e_N . Note that the optimal effort $e^*(\theta)$ is not a continuous function of θ : it has a jump at the point $\theta = \theta_M$, and this function becomes constant (equal to e_N) after θ_M .

To conclude this section, let us state the following corollary of Theorem 3.1 which says that in self-insurance, Statement (A) from the Introduction is always true. An increase in the premium leads to an increase in the effort, as long as the price is attractive for IB.

Corollary 3.1. *The optimal effort $e^*(\theta)$ is a non-decreasing function of θ .*

3.3 Self-Protection

In the previous subsection, we made the Assumption 3.1 stating that the impact of an increased effort is bigger on the risk measure than it is on the expectation. This assumption implies that the conclusion (A) of the seminal paper of Ehrlich and Becker [22] is satisfied. We study now the remaining case where the function G is non-decreasing, and prove that in that case, the conclusion (B) holds true. This will be the case for any model of self-protection, in the form (3.6) described below.

Assumption 3.2. *The function G defined in (3.2) is non-decreasing, or equivalently*

$$e_1 \leq e_2 \implies \frac{\rho_1(X_{e_1}) - \rho_1(X_{e_2})}{\rho_1(X_{e_1})} \leq \frac{\mathbb{E}[X_{e_1}] - \mathbb{E}[X_{e_2}]}{\mathbb{E}[X_{e_1}]}.$$

An example of such a situation, is the one in which the insurance buyer controls his loss probability by exerting an effort e on it. More precisely, let the probability distribution P_e of the random variable X_e take the following form

$$P_e := (1 - p(e))\delta_{\{0\}} + p(e)P_Y, \quad (3.6)$$

where $e \mapsto p(e)$ is a decreasing function, P_Y denotes the distribution of a positive random variable Y , and $\delta_{\{0\}}$ is the Dirac mass at 0. In other words, the loss random variable X_e is strictly positive with probability $p(e)$, in which case its value is given by Y , and takes the value 0 with probability $1 - p(e)$.

Lemma 3.2. *Assume that the loss distribution has the form (3.6) and ρ_1 is a law invariant and coherent risk measure. Then Assumption 3.2 is satisfied.*

This result says that when only the loss probability is controlled, then the impact of the effort is bigger on the price (the expectation) than it is on the risk. As it will become clear from the

next results, this implies that there are situations in which an increase in the insurance price leads to less effort to reduce the loss probability, that is the well known conclusion (B) of [22].

Remark 3.3. *Given the particular form (3.6) of probability distributions corresponding to self-protection, Assumption 2.1, which ensures that $e \mapsto \rho_1(X_e)$ is convex, says in this setting that $e \mapsto \bar{q}_Y\left(\frac{u}{p(e)}\right), 0 \leq u \leq p(e)$, is convex. This is the case, for example, if $u \mapsto \bar{q}_Y(u)$ and $e \mapsto p(e)$ are both concave.*

In the case of distortion risk measures, we can provide a sufficient condition for the convexity of $e \mapsto \rho_1(X_e)$, which is independent of the distribution of the random variable Y . This is the purpose of the next lemma.

Lemma 3.3. *Assume that ρ_1 is a distortion risk measure (see (2.1)) with a distortion function ψ such that $e \mapsto \psi[p(e)]$ is convex. Then both functions $e \mapsto \rho_1(X_e)$ and $e \mapsto \mathbb{E}[X_e]$ are convex.*

Remark 3.4. *A standard assumption in insurance economics is that $e \mapsto p(e)$ is convex (see [14] for instance), which means that an increased effort leads to a decrease in the loss probability $p(e)$, with a decreasing marginal impact of the effort. In particular, this implies that the objective probabilities $\mathbb{P}(X_e > x) = p(e)P_Y((x, +\infty))$ are also decreasing and convex. The assumption in the previous lemma entails that the subjective probabilities $\psi(\mathbb{P}(X_e > x))$ are also decreasing in a convex way. Note that the assumption that $\psi \circ p$ is convex is also needed in [12]. In Subsection 5.1.3, we provide classes of distortion functions and of maps $e \mapsto p(e)$ satisfying this assumption.*

We proceed now to solve the problem of IB. By Lemma 3.2, the sets \mathcal{N} and \mathcal{I} take the form

$$\mathcal{N} := \left\{ e \in (0, +\infty) \text{ such that } (1 + \theta) > \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \right\} = (0, e_\theta), \quad (3.7)$$

$$\mathcal{I} := \left\{ e \in (0, +\infty) \text{ such that } (1 + \theta) \leq \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \right\} = (e_\theta, +\infty). \quad (3.8)$$

where $e_\theta := \inf\{e \in \mathbb{R}^+ \mid \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \geq 1 + \theta\}$.

We know that for all $e < +\infty$, $\mathbb{E}[X_e] = p(e)\mathbb{E}[Y] \neq 0$ (since Y is positive a.s.) and from Remark 2.2 (or from Lemma 3.3), the functions $e \mapsto \rho_1(X_e)$ and $e \mapsto \mathbb{E}[X_e]$ are continuous. We have that $e \mapsto \frac{\rho_1(X_e)}{\mathbb{E}[X_e]}$ is continuous, hence e_θ is the solution to the equation $\frac{\rho_1(X_e)}{\mathbb{E}[X_e]} = 1 + \theta$. So e_θ is defined as the generalized inverse of $e \mapsto \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} - 1$, which is non-decreasing, therefore e_θ is a non-decreasing function of θ .

We proceed now with a similar analysis to the one developed in Section 3.2, based on finding the minima of the functions $L_{\mathcal{N}}$ and $L_{\mathcal{I}}^\theta$ on their respective domains and compare them properly. Recall that $L_{\mathcal{N}}(e) = \rho_1(X_e) + c(e)$ and $L_{\mathcal{I}}^\theta(e) = (1 + \theta)\mathbb{E}[X_e] + c(e)$.

On the set \mathcal{N}

Since $L_{\mathcal{N}}$ is convex, non-negative and coercive, it admits an unique minimizer on \mathbb{R}^+ that we denote e_N . Since the random variables $(X_e)_{e \in \mathbb{R}^+}$ have finite risk measure, the function $L_{\mathcal{N}}$ is finite at $e = 0$, the right hand derivative $\partial_e^+ L_{\mathcal{N}}(0)$ is well defined and we can suppose without loss of generality that it is non-positive.

Proposition 3.4. *There exists a positive constant θ_N such that*

- (a) $\theta \leq \theta_N \implies \operatorname{argmin}_{\mathcal{N}} L_{\mathcal{N}} = \{e_\theta\}$,
- (b) $\theta > \theta_N \implies \operatorname{argmin}_{\mathcal{N}} L_{\mathcal{N}} = \{e_N\}$.

Proof. We define $\theta_N := \inf\{\theta \geq 0 \mid e_\theta > e_N\}$. Since $\theta \mapsto e_\theta$ is a non-decreasing function, $\theta > \theta_N$ implies that $e_N \in \mathcal{N}$ therefore e_N is the local minimizer of $L_{\mathcal{N}}$ on \mathcal{N} . If $\theta \leq \theta_N$ then the concave map $L_{\mathcal{N}}$ is non-decreasing on \mathcal{N} which means that its minimizer is given by the boundary point e_θ . \square

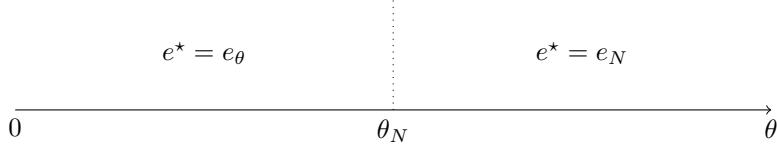


Figure 1.1 – Local minimizer of $L_{\mathcal{N}}$ on the set \mathcal{N} , as a function of θ .

On the set \mathcal{I}

Since $L_{\mathcal{I}}^\theta$ is a convex, coercive function it admits an unique minimizer on \mathbb{R}^+ that we call $e_I(\theta)$. We will denote it e_I when no confusion can arise. The same argument as in the proof of Proposition 3.2 gives that e_I is a non-decreasing function of θ . Different from the previous section, in this setting there is no obvious way to compare e_I and e_θ with respect to the values of θ because both efforts are non-decreasing functions of θ .

Global minimum

We recall that on \mathcal{N} we have $\rho_1(X_e) < (1 + \theta)\mathbb{E}[X_e]$ and on \mathcal{I} we have the reverse inequality. Therefore we can distinguish six possible cases concerning the order of e_I , e_N and e_θ depending on the value of θ

- | | | |
|------------------------------------|------------------------------------|------------------------------------|
| (a) $e_\theta \leq e_I \leq e_N$, | (c) $e_N \leq e_I \leq e_\theta$, | (e) $e_\theta \leq e_N \leq e_I$, |
| (b) $e_I \leq e_\theta \leq e_N$, | (d) $e_I \leq e_N \leq e_\theta$, | (f) $e_N \leq e_\theta \leq e_I$. |

In the next proposition, we prove that case (b) is not possible and we solve case (a). As a corollary, we solve also case (e). We postpone the proofs to Appendix 2

Proposition 3.5. If $e_I \leq e_N$ and $e_\theta \leq e_N$ then $e_\theta \leq e_I$ and the global minimum of L_1 is e_I .

Corollary 3.2. If $e_\theta \leq e_N$ and $e_\theta \leq e_I$ then the global minimum of L_1 is e_I .

The cases (c) and (d) are addressed in the proposition that follows.

Proposition 3.6. If $e_N \leq e_\theta$ and $e_I \leq e_\theta$ then the global minimum of L_1 is e_N .

Proof. If $e_N \leq e_\theta$ and $e_I \leq e_\theta$, then e_N and e_I both belong to the set \mathcal{N} . Moreover, for all $e \in \mathcal{N}$ we have $L_I^\theta(e) > L_N(e)$, which implies $L_I^\theta(e_I) > L_N(e_I) \geq L_N(e_N)$, since e_N is the global minimum of L_N . So we can conclude that e_N is the global minimum of L_1 . \square

The Theorem that follows addresses the last case (f). Its proof is postponed to Appendix 2.

Theorem 3.2. There exists a positive constant θ_M such that

- (i) If $e_N \leq e_\theta \leq e_I$ and $\theta \leq \theta_M$, the global minimum of L_1 is e_I .
- (ii) If $e_N \leq e_\theta \leq e_I$ and $\theta > \theta_M$, the global minimum of L_1 is e_N .

Moreover we have the following inequality

$$\theta_N \leq \theta_M,$$

where θ_N is defined in Proposition 3.4.

By combining the results of all the cases, we obtain the graph that follows.

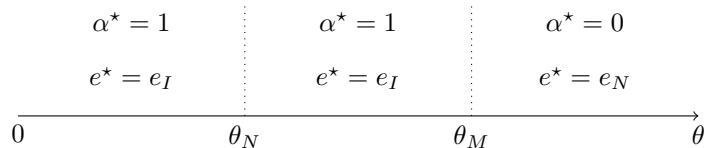


Figure 1.2 – Comparison of optimal insurance demand and effort levels for different safety loadings.

Let us conclude this section with the following corollary stating that Statement (B) holds true in the case of self-protection. The behavior of the optimal effort $e^*(\theta)$ is very similar to the one described in [35] and [12].

Corollary 3.3. The map $\theta \mapsto e^*(\theta)$ is non-decreasing for $\theta \leq \theta_M$. It becomes constant, equal to e_N , for $\theta > \theta_M$. It has a jump at $\theta = \theta_M$ equal to $e_N - e_I(\theta_M)$. The cases (c), (e) and (f) described above imply that this jump can be negative, in which case we deduce that Statement (B) holds true.

4 The insurance seller's optimization problem

We now move to the problem of IS, which consists of finding the best premium for the insurance coverage that IB is willing to accept. Recall that IS uses a law-invariant and coherent risk measure ρ_2 and her goal is to solve problem (2.3). Since we are considering safety loading premiums, the problem of IS reduces to

$$\underset{\theta \in \mathbb{R}_+}{\text{minimize}} \quad \alpha^*(\theta) (\rho_2(X_{e^*(\theta)}) - (1 + \theta)\mathbb{E}[X_{e^*(\theta)}]).$$

Remark 4.1. *In a standard Principal-Agent relationship, it is assumed that the agent accepts only contracts for which his utility (loss) is greater (less) than some reservation value R_0 , representing the utility (loss) he can guarantee under some outside option. However, this feature is endogenous in our model, since IB is free to choose $\alpha^* = 0$ when the contract does not seem convenient for him.*

4.1 Solving the problem of IS

In this section we describe the general methodology to solve the problem of IS. This analysis applies to both the self-insurance and self-protection type of IB, described in Subsections 3.2 and 3.3.

We know that under Assumption 3.1 or 3.2, there exists a positive constant θ_M such that $\alpha^*(\theta) = \mathbf{1}_{\theta \in [0, \theta_M]}$. Given the form of the objective function of IS, her problem is reduced to a minimization over a compact set

$$\underset{\theta \in [0, \theta_M]}{\text{minimize}} \quad \rho_2(X_{e_I(\theta)}) - (1 + \theta)\mathbb{E}[X_{e_I(\theta)}]. \quad (4.1)$$

Remark 4.2. *Here we assume of course, that the value of problem (4.1) is negative. We consider uninteresting the case in which the value of (4.1) is positive, since IS would be better without entering into the contract and should simply offer any $\theta > \theta_M$. Note that a necessary condition is then $\rho_2(X_{e_I(\theta^*)}) < \rho_1(X_{e_I(\theta^*)})$, because of the definition of the set \mathcal{I} of efforts where full insurance is demanded. In the next sections, we will assume directly that $\rho_2 < \rho_1$, that is the risk affects IS less than IB.*

By continuity, there exists an optimal loading factor θ^* which minimizes IS's risk. Moreover, if the functions involved are smooth enough (which is the case in most of the models studied in the literature), we can characterize θ^* as the solution to the following system

$$\begin{aligned} \left(\frac{\partial}{\partial e} \rho_2(X_{e_I(\theta)}) - (1 + \theta) \frac{\partial}{\partial e} \mathbb{E}[X_{e_I(\theta)}] \right) \frac{\partial e_I(\theta)}{\partial \theta} &= \mathbb{E}[X_{e_I(\theta)}] - \lambda, \\ \lambda(\theta - \theta_M) &= 0, \end{aligned}$$

where $\lambda \geq 0$ is a KKT multiplier associated to the constraint $\theta \leq \theta_M$.

Not much more can be said about IS's problem, without specifying a particular model. In Section 5.1, as an application, we solve explicitly the problem of IS in the Pareto case.

4.2 Adverse Selection

As an extension of the problem of IS, suppose now a situation in which there are two types of IB in the population. Let us assume directly that the losses have the form (3.3), and the difference between the types of IB is the probability p of having a loss.

More precisely, let $p_b > p_g$ and let us refer to the two types of IB as the *good* type and the *bad* type. Then the distribution of the losses for IB of good and bad type are respectively

$$P_e^g := (1 - p_g)\delta_{\{0\}} + p_g P_{Y_e},$$

$$P_e^b := (1 - p_b)\delta_{\{0\}} + p_b P_{Y_e}.$$

We assume the type of a particular IB is private information. When asking for insurance, IS does not know what is the type of IB but only the distribution of types in the population: with probability π_g , the IB that she is interacting with is of good type and with probability π_b , he is of bad type.

In this setting, we will consider two different problems that may be of interest for IS, the *shutdown* problem and the *screening* problem. As explained for instance in [16, 28], the difference between these two problems lies on whether IS wants to sign contracts with the IB of bad type, so they are defined by different optimization programs.

Shutdown problem

In the shutdown problem, IS wants to provide insurance only to the good type of IB, because of the low risk probability p_g of suffering a loss compared to p_b . In other words, IS wants to exclude the bad type of IB and sign a contract only with IB of good type. IS chooses then a premium that will be rejected by IB of bad type and will only be accepted by IB of good type.

Given any loading factor $\theta \geq 0$, denote by $(\alpha_g^*(\theta), e_g^*(\theta))$ and $(\alpha_b^*(\theta), e_b^*(\theta))$ the optimal level of insurance and effort chosen by IB of good and bad type respectively. The shutdown problem is given as follows

$$\begin{aligned} & \underset{\theta \geq 0}{\text{minimize}} && \rho_2(X_{e_g^*(\theta)}) - (1 + \theta)\mathbb{E}[X_{e_g^*(\theta)}] \\ & \text{s.t.} && \alpha_b^*(\theta) = 0, \\ & && \alpha_g^*(\theta) = 1. \end{aligned}$$

In Section 5.1, we approach the shutdown problem of IS in the Pareto case.

Screening problem

In the screening problem, IS is willing to contract both types of IB and differentiate them by designing simultaneously two premiums, one for each type. Since IB's type is private and he could claim being of any type, IS will design a menu of premiums satisfying the so-called *truth-revealing* property, that is to say, providing incentives to IB for choosing the contract made for his respective type.

Denote by L_1^g and L_1^b respectively the risk measures associated to the total loss of IB of type good and bad. The screening problem faced by IS is given as follows

$$\inf_{\theta_g, \theta_b \geq 0} \pi_g \left(\rho_2(X_{e_g^*(\theta_g)}) - (1 + \theta_g) \mathbb{E}[X_{e_g^*(\theta_g)}] \right) + \pi_b \left(\rho_2(X_{e_b^*(\theta_b)}) - (1 + \theta_b) \mathbb{E}[X_{e_b^*(\theta_b)}] \right), \quad (4.2)$$

$$\text{such that } \begin{cases} L_1^g(\alpha_g^*(\theta_g), e_g^*(\theta_g)) \leq L_1^g(\alpha_g^*(\theta_b), e_g^*(\theta_b)), \\ L_1^b(\alpha_b^*(\theta_b), e_b^*(\theta_b)) \leq L_1^b(\alpha_b^*(\theta_g), e_b^*(\theta_g)), \end{cases} \quad (4.3)$$

$$\text{and such that } \begin{cases} \alpha_g^*(\theta_g) = 1, \\ \alpha_b^*(\theta_b) = 1. \end{cases} \quad (4.4)$$

Constraint (4.3) states that whatever the type of IB is, he does not have any incentive to lie and take the premium designed for the other type, since the loss when doing so is greater than the one obtained when revealing the truth. Constraint (4.4) makes sure that the premiums are sufficiently attractive to both types of IB in the population.

In Section 5.1, we solve explicitly the screening problem of IS in the Pareto case.

5 Case studies

5.1 The Pareto Case

In the following case study, we consider that the distribution of the loss random variable X_e is given by

$$P_{X_e} = (1 - p)\delta_{\{0\}} + pP_{Y_e},$$

where $0 < p < 1$, $\delta_{\{0\}}$ is the Dirac mass at 0 and P_{Y_e} denotes the distribution of a Pareto random variable of parameter $\hat{x} > 0$ and $k > 0$. The Pareto distribution is often used in standard actuarial models (estimation of risk premiums or quotation in non-proportional reinsurance for example).

For tractability purposes, we define

$$\Pi(X_e) = (1 + \theta) \mathbb{E}[X_e], \quad c(e) = \frac{e^2}{2}, \quad \psi_1(u) = u^{r_1}, \quad r_1 \in (0, 1), \quad \psi_2(u) = u^{r_2}, \quad r_2 \in (0, 1), \quad r_2 > r_1.$$

5.1.1 Self-insurance

To study self-insurance in this context, we suppose that IB's effort to reduce the size of the loss expresses itself in the control of the parameter $e \mapsto k(e)$. We assume this function to be smooth, non-decreasing, concave and such that $k(0) = 1$. We then have the following

$$\mathbb{P}[X_e > x] = p \left(\frac{\hat{x}}{x} \right)^{k(e)}, \quad \forall x > \hat{x}, \quad \mathbb{E}[X_e] = \frac{p\hat{x}k(e)}{k(e)-1}, \quad \bar{q}_{X_e}(u) = \begin{cases} \hat{x} \left(\frac{p}{u} \right)^{1/k(e)}, & \text{if } u \in [0, p], \\ 0, & \text{if } p < u \leq 1. \end{cases}$$

Since k is a concave function, the quantile function \bar{q}_{X_e} is convex. We can calculate the risk measure $\rho_1(X_e)$ which is given by Equation (2.1)

$$\rho_1(X_e) = r_1 \hat{x} \int_0^p u^{r_1-1-1/k(e)} p^{1/k(e)} du = \begin{cases} \frac{r_1 p^{r_1} \hat{x} k(e)}{r_1 k(e) - 1}, & \text{if } k(e) > \frac{1}{r_1}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore, IB's problem is

$$\inf_{(\alpha, e) \in [0, 1] \times (0, \infty)} \left\{ (1 - \alpha) \frac{p^{r_1} r_1 \hat{x} k(e)}{r_1 k(e) - 1} + \alpha(1 + \theta) \frac{p \hat{x} k(e)}{k(e) - 1} + \frac{e^2}{2} \right\}. \quad (5.1)$$

Remark 5.1. *The explicit value of $\rho_1(X_e)$ given above entails that IB will never choose an effort below the level $k^{-1}(\frac{1}{r_1})$. This also implies that $k(e) > \frac{1}{r_1} > 1$.*

Moreover we can compute the value of the function G defined by Equation (3.2)

$$G(e) = \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} = \frac{r_1 p^{r_1-1} (k(e) - 1)}{r_1 k(e) - 1},$$

so G is decreasing and we can apply all the results of Subsection 3.2 to this case.

To solve problem (5.1), we start by characterizing the main constants introduced in Section 3. The point e_θ is given by

$$e_\theta = k^{-1} \left(\frac{1 + \theta - p^{r_1-1} r_1}{r_1 (1 + \theta - p^{r_1-1})} \right), \quad (5.2)$$

and the corresponding sets \mathcal{N} and \mathcal{I}

$$\begin{aligned} \mathcal{N} &= \left\{ e \in \mathbb{R}^+ \mid (1 + \theta) > \frac{p^{r_1-1} r_1 (k(e) - 1)}{(r_1 k(e) - 1)} \right\} = (e_\theta, +\infty), \\ \mathcal{I} &= \left\{ e \in \mathbb{R}^+ \mid (1 + \theta) \leq \frac{p^{r_1-1} r_1 (k(e) - 1)}{(r_1 k(e) - 1)} \right\} = [0, e_\theta]. \end{aligned}$$

As the global minimum of $L_{\mathcal{N}}$, the point e_N is characterized by the optimality condition

$$e_N = \frac{p^{r_1} r_1 \hat{x} k'(e_N)}{(r_1 k(e_N) - 1)^2}, \quad (5.3)$$

which shows that e_N is a non-decreasing function of the loss probability p .

As the global minimum of L_I^θ , $e_I(\theta)$ is characterized by the equation $(L_I^\theta)'(e_I(\theta)) = 0$ which rewrites as

$$e_I(\theta) = \frac{(1+\theta)\hat{x}pk'(e_I(\theta))}{(k(e_I(\theta))-1)^2}. \quad (5.4)$$

We observe then that $e_I(\theta)$ is non-decreasing with p and with θ .

Finally, we have from Theorem 3.1 the existence of a constant θ_M that defines the global minimum of the objective function L_1 . The point θ_M is the solution to the equation $L_I^{\theta_M}(e_I(\theta_M)) = L_N(e_N)$ and we have

- If $\theta \leq \theta_M$ then the solution to (5.1) is $(\alpha^*, e^*) = (1, e_I(\theta))$.
- If $\theta > \theta_M$ then the solution to (5.1) is $(\alpha^*, e^*) = (0, e_N)$.

Concerning the problem of IS, we can calculate the distortion risk measure $\rho_2(X_e)$ ⁴

$$\rho_2(X_e) = r_2\hat{x} \int_0^p u^{r_2-1-1/k(e)} p^{1/k(e)} du = \begin{cases} \frac{p^{r_2} r_2 \hat{x} k(e)}{r_2 k(e)-1}, & \text{if } k(e) > \frac{1}{r_2}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore, problem (4.1) writes

$$\inf_{\theta \in \Theta_M} \left\{ \frac{p^{r_2} r_2 \hat{x} k(e_I(\theta))}{r_2 k(e_I(\theta))-1} - (1+\theta) \frac{p \hat{x} k(e_I(\theta))}{k(e_I(\theta))-1} \right\}, \quad (5.5)$$

where $\Theta_M := \{\theta \in [0, \theta_M] : k(e_I(\theta)) > \frac{1}{r_2}\}$. In the Pareto case, IS's problem is easier to solve than in the general case, because its objective function turns out to be monotone. In fact, we provide now the solution to the problem of IS, which is given by the maximum premium IB is willing to pay.

Theorem 5.1. *The solution to problem (5.5) is $\theta^* = \theta_M$.*

Remark 5.2. *Since k' is continuous, we see from (5.4) that the map $\theta \rightarrow e_I(\theta)$ is continuous. It follows from the definition⁵ of θ_M that when IS offers such a loading factor, IB will be indifferent between having insurance or not.*

5.1.2 Adverse selection in the case of self-insurance

Recall from Subsection 4.2, the shutdown problem of IS

$$\begin{aligned} \underset{\theta \geq 0}{\text{minimize}} \quad & \rho_2(X_{e_g^*(\theta)}) - (1+\theta)\mathbb{E}[X_{e_g^*(\theta)}] \\ \text{s.t.} \quad & \alpha_b^*(\theta) = 0, \\ & \alpha_g^*(\theta) = 1. \end{aligned}$$

4. Since $r_2 > r_1$, the risk measures satisfy $\rho_2 < \rho_1$, which matches the assumption in Remark 4.2.

5. Indeed, in the proof of Proposition 3.1, θ_M is explicitly defined by $\theta_M := \inf\{\theta \geq 0 \mid L_I^\theta(e_I(\theta)) \geq L_N(e_N)\}$, see the complete proof in the Appendix A for more details.

IS needs to optimize over the prices θ such that $\theta \leq \theta_M(p_g)$ and $\theta > \theta_M(p_b)$ where $\theta_M(p_g)$ is the optimal insurance price for the IB of good type and $\theta_M(p_b)$ is the optimal insurance price for IB of bad type.

Recall that θ_M is characterized by $L_I^{\theta_M}(e_I(\theta_M)) = L_N(e_N)$. From the dependances of $e_I(\theta)$ and e_N on p , we can notice that the higher the probability of risk is for an IB the more he is willing to pay to insure this risk, this leads to $\theta_M(p_g) \leq \theta_M(p_b)$. Therefore IS cannot choose a price θ respecting the constraints of the shut-down contract which means that such contract does not exist in our context.

Recall also the screening problem of IS

$$\inf_{\theta_g, \theta_b \geq 0} \pi_g \left(\rho_2(X_{e_g^*(\theta_g)}) - (1 + \theta_g) \mathbb{E}[X_{e_g^*(\theta_g)}] \right) + \pi_b \left(\rho_2(X_{e_b^*(\theta_b)}) - (1 + \theta_b) \mathbb{E}[X_{e_b^*(\theta_b)}] \right), \quad (5.6)$$

$$\text{such that } \begin{cases} L_1^g(\alpha_g^*(\theta_g), e_g^*(\theta_g)) \leq L_1^g(\alpha_g^*(\theta_b), e_g^*(\theta_b)), \\ L_1^b(\alpha_b^*(\theta_b), e_b^*(\theta_b)) \leq L_1^b(\alpha_b^*(\theta_g), e_b^*(\theta_g)), \end{cases} \quad (5.7)$$

$$\text{and such that } \begin{cases} \alpha_g^*(\theta_g) = 1, \\ \alpha_b^*(\theta_b) = 1. \end{cases} \quad (5.8)$$

From Equation (5.8), IS chooses a price θ_g such that $\theta_g \leq \theta_M(p_g)$ and a price θ_b such that $\theta_b \leq \theta_M(p_b)$. From Equation (5.7), IS needs to select the prices such that $\theta_g \leq \theta_b$ and $\theta_b \leq \theta_g$. Therefore the only way in which IS can contract both types of IB is by offering the same price to them $\theta_g = \theta_b$.

One can observe that adverse selection problems in our context of linear contracts, and coherent risk measures are rather degenerate. The insurance seller cannot use the standard tools to deal with adverse selection: she cannot offer a menu of contracts since the optimal insurance coverage is either 0 or 1, she cannot exclude the bad type of agent since shut-down contracts do not exist, and she cannot use differentiated pricing since the constraints in the screening problem imply that the price is unique.

5.1.3 Self-protection

To study self-protection, we suppose that IB's effort reduces the probability of loss $e \mapsto p(e)$ which we assume non-increasing and convex. We then have the following

$$\mathbb{P}[X_e > x] = p(e) \left(\frac{\hat{x}}{x} \right)^k, \quad \forall x > \hat{x}, \quad \mathbb{E}[X_e] = \frac{\hat{x}k}{k-1} p(e), \quad \bar{q}_{X_e}(u) = \begin{cases} \hat{x} \left(\frac{p(e)}{u} \right)^{1/k}, & \text{if } u \in [0, p(e)], \\ 0, & \text{if } p(e) < u \leq 1. \end{cases}$$

Note that since p is a convex function, the quantile function \bar{q}_{X_e} is also a convex function of e . From Equation (2.1) the risk measure $\rho_1(X_e)$ is

$$\rho_1(X_e) = r_1 \hat{x} \int_0^{p(e)} u^{r_1-1-1/k} p(e)^{1/k} du = \begin{cases} \frac{r_1 p(e)^{r_1} \hat{x} k}{r_1 k - 1}, & \text{if } k > \frac{1}{r_1}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore, IB's problem is

$$\inf_{(\alpha, e) \in [0, 1] \times (0, \infty)} \left\{ (1 - \alpha) \frac{r_1 \hat{x} k}{r_1 k - 1} p(e)^{r_1} + \alpha(1 + \theta) \frac{\hat{x} k}{k - 1} p(e) + \frac{e^2}{2} \right\}. \quad (5.9)$$

We can also compute the value of the function G from Equation (3.2)

$$G(e) = \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} = \frac{r_1(k-1)}{r_1 k - 1} p(e)^{r_1-1},$$

which is an increasing function of e since p is non-increasing and $r_1 < 1$. Therefore we can apply the results of Subsection 3.3. In this setting, e_θ is defined by

$$e_\theta = p^{-1} \left(\left(\frac{(1 + \theta)(r_1 k - 1)}{r_1(k-1)} \right)^{1/(r_1-1)} \right), \quad (5.10)$$

which is an increasing function of θ and the sets \mathcal{N} and \mathcal{I} are defined as follows

$$\begin{aligned} \mathcal{N} &= \left\{ e \in (0, +\infty) \text{ such that } (1 + \theta) > \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \right\} = (0, e_\theta), \\ \mathcal{I} &= \left\{ e \in (0, +\infty) \text{ such that } (1 + \theta) < \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \right\} = (e_\theta, +\infty). \end{aligned}$$

We have

$$L_{\mathcal{N}}(e) = \frac{r_1 \hat{x} k}{r_1 k - 1} p(e)^{r_1} + \frac{e^2}{2}, \quad L_{\mathcal{I}}^\theta(e) = \frac{(1 + \theta) \hat{x} k}{k - 1} p(e) + \frac{e^2}{2}.$$

If the map $u \mapsto \bar{q}_Y(u)$ is convex and if $-2\frac{p'}{p} \leq -\frac{p''}{p'}$, then⁶ Assumption 2.1 is satisfied, and by Theorem 2.1, $L_{\mathcal{N}}$ is a convex function of e . Alternatively, since the expression for $L_{\mathcal{N}}$ is more explicit here, we can give another condition: $L_{\mathcal{N}}$ is convex if and only if $e \mapsto p(e)^{r_1}$ is convex, which is the case if $\frac{p''}{(p')^2} \geq \frac{1-r_1}{p}$. An example of function satisfying this last condition is given by $p(e) = \frac{1}{a+e}$, with $a > 1$.

Let us now study the minimum of the function $L_{\mathcal{N}}$. Its first derivative is given by:

$$L'_{\mathcal{N}}(e) = e + \frac{r_1^2 \hat{x} k}{r_1 k - 1} p'(e) p(e)^{r_1-1}.$$

Since

$$L'_{\mathcal{N}}(0) = \frac{r_1^2 \hat{x} k}{r_1 k - 1} p'(0) p(0)^{r_1-1} < 0, \quad \text{and} \quad \lim_{e \rightarrow +\infty} L'_{\mathcal{N}}(e) = +\infty,$$

6. This assumption means that we have non-increasing returns on self-protection which implies the convexity of the function $e \mapsto p(e)$. It is a standard assumption in the literature, see [13] and [14].

the global minimum e_N of the function $L_{\mathcal{N}}$ is thus the solution to

$$e_N = -\frac{r_1^2 \hat{x} k}{r_1 k - 1} p'(e_N) p(e_N)^{r_1 - 1}.$$

The global minimum of $L_{\mathcal{I}}$ is characterized by

$$e_I(\theta) = -\frac{(1 + \theta) \hat{x} k}{k - 1} p'(e_I(\theta)). \quad (5.11)$$

This equation allows us to conclude that $e_I(\theta)$ is an increasing function of θ . We have from Theorem 3.2 the existence of a constant θ_M that defines the global minimum of the objective function L_1 . The point θ_M is the solution to the equation $L_{\mathcal{I}}^{\theta_M}(e_I(\theta_M)) = L_{\mathcal{N}}(e_N)$ and we have

- If $\theta < \theta_M$ then the solution to (5.9) is $(\alpha^*, e^*) = (1, e_I(\theta))$.
- If $\theta \geq \theta_M$ then the solution to (5.9) is $(\alpha^*, e^*) = (0, e_N)$.

To study IS's problem in the self-protection case, we will make the following assumption:

Assumption 5.1.

$$p(e) = \frac{1}{a + e}, \quad a > 1.$$

The distortion risk measure is given by

$$\rho_2(X_e) = r_2 \hat{x} \int_0^{p(e)} u^{r_2 - 1 - 1/k} p(e)^{1/k} du = \begin{cases} \frac{p(e)^{r_2} r_2 \hat{x} k}{r_2 k - 1}, & \text{if } k > \frac{1}{r_2}, \\ +\infty, & \text{otherwise.} \end{cases}$$

From Section 3, we know that the equilibrium for IB's problem in the self-protection case is $(\alpha^*, e^*) = (1, e_I(\theta))$, therefore, IS's problem is as follows

$$\inf_{\theta \in [0, \theta_M]} \left\{ \frac{r_2 \hat{x} k p(e_I(\theta))^{r_2}}{r_2 k - 1} - (1 + \theta) \frac{\hat{x} k p(e_I(\theta))}{k - 1} \right\}. \quad (5.12)$$

Given Assumption 5.1, the characterization (5.11) of $e_I(\theta)$ now writes

$$e_I(\theta) = c \frac{1 + \theta}{(a + e_I(\theta))^2},$$

with $c := \hat{x} \frac{k}{k - 1} > 0$. After some cumbersome, but easy calculations (see Appendix 3), we get:

$$e_I(\theta) = t_1(\theta) + t_2(\theta) - \frac{2a}{3}, \quad (5.13)$$

where

$$\begin{aligned} t_1(\theta) &:= \left[\frac{1}{2} \left(c(1 + \theta) + \frac{2}{27} a^3 - \sqrt{\frac{4}{27} c(1 + \theta) a^3 + c^2 (1 + \theta)^2} \right) \right]^{1/3} \quad \text{and} \\ t_2(\theta) &:= \left[\frac{1}{2} \left(c(1 + \theta) + \frac{2}{27} a^3 + \sqrt{\frac{4}{27} c(1 + \theta) a^3 + c^2 (1 + \theta)^2} \right) \right]^{1/3} \end{aligned}$$

In this setting of self-protection, the solution to the problem of IS is also to charge the maximum premium θ_M that IB is willing to pay. We have then the following result, whose proof is postponed to Appendix 3.

Theorem 5.2. *Under Assumption 5.1, the solution to Problem (5.12) is $\theta^* = \theta_M$.*

Remark 5.3. *Analogously to Remark 5.2, in this case since p' is continuous, we see from (5.13) that the map $\theta \rightarrow e_I(\theta)$ is continuous. It follows that when IS offers the loading factor θ_M , IB will be indifferent between having insurance or not.*

5.2 Discrete distributions

As already mentioned in the Introduction, most of the literature on self-insurance and self-protection concentrates on the case of discrete distributions, with only two possible loss values. We show in this subsection that our results also apply to the case of discrete distributions, with an arbitrary number of possible losses. We prove that, contrary to the Bernoulli case, more conditions are needed for Statements (A) and (B) to hold true.

Let $\{p_i, i = 0, \dots, n-1\}$ be positive numbers such that $\sum_{i=0}^{n-1} p_i = 1$, and let $0 < x_1 < \dots < x_{n-1}$ be positive numbers. We consider a discrete loss with $n \in \mathbb{N}$ values, given by

$$X := \begin{cases} x_i & \text{with probability } p_i, i \in \{1, \dots, n-1\} \\ 0 & \text{with probability } p_0. \end{cases}$$

To simplify the calculations, we suppose here that the risk measure ρ_1 is a distortion risk measure, whose definition is given in Section 2, Equation (2.1)

$$\rho_1(X) := \int_0^1 \bar{q}_X(u) d\psi(u) = \int_0^{+\infty} \psi(\mathbb{P}[X > u]) du.$$

5.2.1 Self-insurance: the discrete case

To study self-insurance in this case, we now suppose that the agent exerts an effort e , that has an impact on every possible loss value x_i , which will be denoted $x_i(e)$ and the loss X will be denoted X_e . Then we can write

$$\begin{aligned} \rho_1(X_e) &= \int_0^{x_1(e)} \psi \left(\sum_{i=1}^{n-1} p_i \right) du + \int_{x_1(e)}^{x_2(e)} \psi \left(\sum_{i=2}^{n-1} p_i \right) du + \dots + \int_{x_{n-2}(e)}^{x_{n-1}(e)} \psi(p_{n-1}) du, \\ &= \left(\psi \left(\sum_{i=1}^{n-1} p_i \right) - \psi \left(\sum_{i=2}^{n-1} p_i \right) \right) x_1(e) + \dots + \psi(p_{n-1}) x_{n-1}(e), \\ &= \sum_{j=1}^{n-1} \left(\psi \left(\sum_{i=j}^{n-1} p_i \right) - \psi \left(\sum_{i=j+1}^{n-1} p_i \right) \right) x_j(e). \end{aligned}$$

To simplify the formulas, we will denote

$$\kappa_j := \psi \left(\sum_{k=j}^{n-1} p_k \right) - \psi \left(\sum_{k=j+1}^{n-1} p_k \right) \geq 0, \quad \forall j = 1, \dots, n-1. \quad (5.14)$$

Moreover we have

$$\mathbb{E}[X_e] = \sum_{j=1}^{n-1} p_j x_j(e).$$

Hence the function G defined in Lemma 3.1 is given by

$$G(e) = \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} = \frac{\sum_{j=1}^{n-1} \kappa_j x_j(e)}{\sum_{j=1}^{n-1} p_j x_j(e)}. \quad (5.15)$$

Assumption 5.2. We suppose that for all j and for all $i > j$, the function $e \mapsto \frac{x_j(e)}{x_i(e)}$ is a non-decreasing function (recall that $x_j(e) < x_i(e)$).

This assumption means that the marginal impact of the effort is higher on small losses than it is on catastrophic losses. It is referred as *increasing difference* in the literature (see [30] for definitions, [39] for a recent application to self-insurance and self-protection, [18] and [7] for illustrations).

We need to study the monotony of G to determine the optimal prevention effort of self-insurance.

$$\begin{aligned} G'(e) &= \frac{1}{\left(\sum_{j=1}^{n-1} p_j x_j(e) \right)^2} \left(\sum_{j=1}^{n-1} \kappa_j x'_j(e) \sum_{j=1}^{n-1} p_j x_j(e) - \sum_{j=1}^{n-1} \kappa_j x_j(e) \sum_{j=1}^{n-1} p_j x'_j(e) \right), \\ &= \frac{1}{\left(\sum_{j=1}^{n-1} p_j x_j(e) \right)^2} \sum_{i,j=1}^{n-1} \kappa_j p_i \left(x_i(e) x'_j(e) - x'_i(e) x_j(e) \right), \\ &= \frac{1}{\left(\sum_{j=1}^{n-1} p_j x_j(e) \right)^2} \sum_{i \neq j} \kappa_j p_i a_{i,j}, \end{aligned}$$

where $a_{j,i} := x_i(e) x'_j(e) - x'_i(e) x_j(e)$, which implies that $a_{j,i} = -a_{i,j}$. We have that

$$G'(e) = \frac{1}{\left(\sum_{j=1}^{n-1} p_j x_j(e) \right)^2} \left(\sum_{i>j} \kappa_j p_i a_{i,j} - \sum_{i<j} \kappa_j p_i a_{j,i} \right) = \frac{1}{\left(\sum_{j=1}^{n-1} p_j x_j(e) \right)^2} \sum_{i>j} (\kappa_j p_i - \kappa_i p_j) a_{i,j}.$$

Thus, the sign of the derivative $G'(e)$ depends on the sign of $\kappa_j p_i - \kappa_i p_j$ for all $i > j$. Let us prove that for a concave distortion function, for all $i > j$, we have the inequality $\kappa_j p_i - \kappa_i p_j \leq 0$. First we recall the following standard result.

Lemma 5.1. *If f is a concave function on an interval I , for all points x, y, z in I such that $x < y < z$ we have*

$$\frac{f(z) - f(y)}{z - y} \leq \frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x}. \quad (5.16)$$

Proposition 5.1. *For all $i > j$, $\kappa_j p_i - \kappa_i p_j \leq 0$.*

Proof. Proving that $\kappa_j p_i - \kappa_i p_j \leq 0$ is equivalent to proving that $\frac{\kappa_j}{p_j} - \frac{\kappa_i}{p_i} \leq 0$.

From Equation (5.14) for all $i > j$

$$\frac{\kappa_j}{p_j} - \frac{\kappa_i}{p_i} = \frac{1}{p_j} \left(\psi \left(\sum_{k=j}^{n-1} p_k \right) - \psi \left(\sum_{k=j+1}^{n-1} p_k \right) \right) - \frac{1}{p_i} \left(\psi \left(\sum_{k=i+1}^{n-1} p_k \right) - \psi \left(\sum_{k=i}^{n-1} p_k \right) \right).$$

Applying Lemma 5.1 to ψ and with

$$x := \sum_{k=i+1}^{n-1} p_k, \quad y := \sum_{k=j+1}^{n-1} p_k \text{ and } z := \sum_{k=j}^{n-1} p_k,$$

we obtain

$$\frac{\kappa_j}{p_j} - \frac{\kappa_i}{p_i} \leq \frac{1}{\sum_{k=j}^{n-1} p_k} \left(\psi \left(\sum_{k=j}^{n-1} p_k \right) - \psi \left(\sum_{k=i+1}^{n-1} p_k \right) \right) - \frac{1}{\sum_{k=i+1}^{n-1} p_k} \left(\psi \left(\sum_{k=i+1}^{n-1} p_k \right) - \psi \left(\sum_{k=i}^{n-1} p_k \right) \right).$$

Applying again Lemma 5.1 to ψ and with

$$x := \sum_{k=i+1}^{n-1} p_k, \quad y := \sum_{k=i}^{n-1} p_k \text{ and } z := \sum_{k=j}^{n-1} p_k,$$

we get

$$\frac{\kappa_j}{p_j} - \frac{\kappa_i}{p_i} \leq \frac{1}{\sum_{k=i}^{n-1} p_k} \left(\psi \left(\sum_{k=i}^{n-1} p_k \right) - \psi \left(\sum_{k=i+1}^{n-1} p_k \right) \right) - \frac{1}{\sum_{k=j}^{n-1} p_k} \left(\psi \left(\sum_{k=j}^{n-1} p_k \right) - \psi \left(\sum_{k=i}^{n-1} p_k \right) \right) = 0,$$

which ends the proof. \square

In conclusion, from Assumption 5.2 and Proposition 5.1 we obtain $G'(e) \leq 0$, therefore G is a non-increasing function of e and Statement (A) holds true.

5.2.2 Self-protection: the discrete case

Now, let us focus on the study of self-protection: we assume here that the effort only impacts the loss probabilities, which we denote $p_i(e)$ for all $i = 0, \dots, n-1$. Using the same calculations as above, we get

$$G(e) := \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} = \frac{\sum_{j=1}^{n-1} \kappa_j(e)x_j}{\sum_{j=1}^{n-1} p_j(e)x_j},$$

with

$$\kappa_j(e) := \psi \left(\sum_{k=j}^{n-1} p_k(e) \right) - \psi \left(\sum_{k=j+1}^{n-1} p_k(e) \right) \geq 0, \quad \forall j = 1, \dots, n-1.$$

As usual, we further assume that for each $i = 0, \dots, n-1$, p_i is a decreasing and convex function of e . The first derivative of G writes

$$\begin{aligned} G'(e) &= \frac{1}{\left(\sum_{j=1}^{n-1} p_j(e)x_j \right)^2} \left(\sum_{j=1}^{n-1} \kappa'_j(e)x_j \sum_{i=1}^{n-1} p_i(e)x_i - \sum_{j=1}^{n-1} \kappa_j(e)x_j \sum_{i=1}^{n-1} p'_i(e)x_i \right), \\ &= \frac{1}{\left(\sum_{j=1}^{n-1} p_j(e)x_j \right)^2} \sum_{i,j=1}^{n-1} x_j x_i (\kappa'_j(e)p_i(e) - p'_i(e)\kappa_j(e)). \end{aligned}$$

The monotonicity of the function G depends on the sign of $(\kappa'_j(e)p_i(e) - p'_i(e)\kappa_j(e))$ for all i, j . It is thus possible to give rather strong conditions for G to be monotonic and therefore for Statements (A) or (B) to hold true. For instance, if for all $i, j \in \{1, \dots, n-1\}$ $e \mapsto \frac{\kappa_j(e)}{p_i(e)}$ is a non-decreasing function, we can conclude that $G'(e) > 0$ therefore G is a non-decreasing function of e and Statement (B) holds true. This last assumption means that the effort impact is bigger on the probabilities than it is on the distorted probabilities, whatever the risk level (small or catastrophic losses). As already noted in [12], for high probability values, there is an underestimation of the impact of the effort in reducing the loss probability. So paradoxically, for extremely risk averse agents who perceive probabilities as strongly distorted towards 1, effort can have a bigger impact on probabilities than on distorted probabilities. But note that this assumption is not always satisfied in practice: for the distortion function associated to the tail Value-at-Risk (TVaR), used in actuarial practice in the Swiss Solvency Test, straightforward calculations show that the ratio $\frac{\kappa_j(e)}{p_i(e)}$ need not be increasing, nor even monotonic. When this ratio is decreasing, our analysis shows that in this simple model of self-protection, market insurance and self-protection are substitutes.

6 Numerical analysis

6.1 The Pareto case

In Subsection 5.1, the distribution of the loss random variable X_e is given by

$$P_{X_e} = (1-p)\delta_{\{0\}} + pP_{Y_e},$$

where $0 < p < 1$, $\delta_{\{0\}}$ is the Dirac mass at 0 and P_{Y_e} denotes the distribution of a Pareto random variable of parameter $\hat{x} > 0$ and $k > 0$.

In this subsection, we present the numerical solution to an example of the self-insurance problem studied in 5.1.1. For the purpose of this numerical analysis, we set

$$k(e) = \sqrt{e} + 1, \quad \hat{x} = 1, \quad r_1 = 0.5, \quad r_2 = 0.6.$$

Figure 1.3 shows the evolution of the optimal efforts $e_I(\theta)$ (defined by Equation (5.4)), e_N (defined by Equation (5.3)) and the boundary effort e_θ (defined by Equation (5.2)) as functions of θ . Recall that e_N is a constant, equal here to 2.16, which is independent of the safety loading. IB's optimal effort while being insured is non-decreasing with θ and the threshold e_θ is non-increasing with θ , both of them being equal at θ_I which is equal here to 8.19. e_θ also crosses e_N at $\theta_N = 3.41$. We can read from the graph the optimal effort e^* , equal to $e_I(\theta)$, for $\theta \leq \theta_M = 4.57$ and e^* jumps at this value from 1.14 to $e_N = 2.16$.

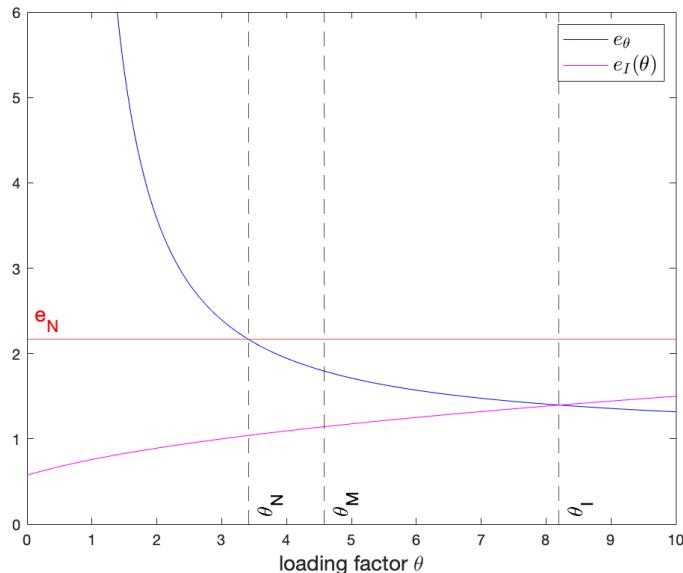


Figure 1.3 – Optimal efforts $e_I(\theta)$, e_N and boundary effort e_θ for the Pareto distribution, as functions of θ .

Figure 1.4 illustrates the changes with respect to the safety loading θ , of the objective functions of IB L_I^θ and L_N evaluated in their respective minimum $e_I(\theta)$ and e_N . As $L_N(e_N)$ is the

objective function defined on the set where IB does not take any insurance, it is independent of θ and is equal to 6.05. As expected, $L_I^\theta(e_I(\theta))$ is increasing with θ and it matches $L_N(e_N)$ at θ_M which is estimated to 4.57. This is the maximum insurance price IS can offer for IB to subscribe full insurance. By offering this loading factor, the value that IS obtains is equal to -2.00.

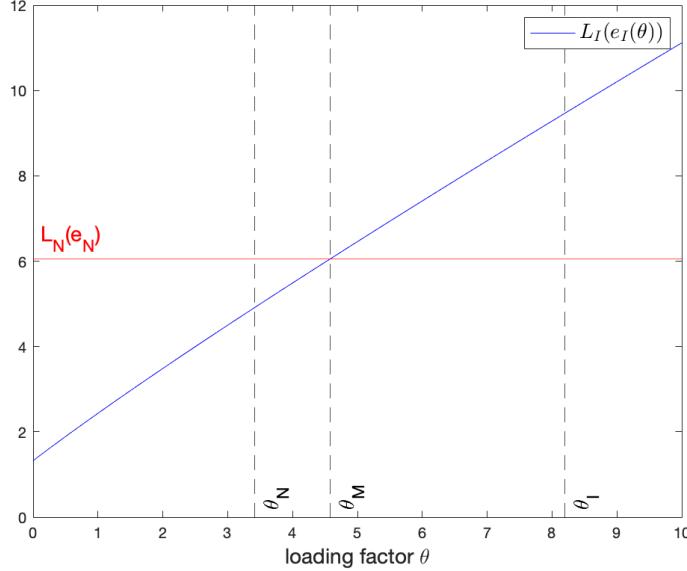


Figure 1.4 – Objective functions $L_I^\theta(e_I(\theta))$ and $L_N(e_N)$ for the Pareto distribution, as functions of θ .

6.2 Approximation of the compound Poisson distribution

In Subsection 3.3, as in Section 5, the distribution of the loss random variable X (we temporarily skip the subscript e for simplicity) is given by

$$P_X := (1-p)\delta_{\{0\}} + p P_Y, \quad (6.1)$$

where $0 < p < 1$, P_Y denotes the distribution of a positive random variable Y , and $\delta_{\{0\}}$ is the Dirac mass at 0. Let $\lambda > 0$ be a real parameter and P be a probability measure on \mathbb{R}^+ . A random variable \tilde{X} which has the compound Poisson distribution $\mathcal{P}(\lambda, P)$ takes the form

$$\tilde{X} = \sum_{j=1}^N Z_j,$$

where $(Z_j)_{j \geq 1}$ is a sequence of independent and identically distributed (i.i.d.) random variables with distribution P and N is a Poisson random variable with parameter λ , independent of the sequence $(Z_j)_{j \geq 1}$. This model is commonly referred to as the *frequency-severity* model: the random variable N represents the frequency, i.e. the number of claims in a given time interval

(typically one year), and P is the claim severity distribution. Most practical insurance and reinsurance models are given in the form of a compound Poisson distribution ([1]).

In this section, we show that the distribution P_X is close to the distribution $\mathcal{P}(\lambda, P)$, for the Kolmogorov distance, with $\lambda = -\log(1-p)$ and $P = P_Y$. The approximation is good in the regime where p is close to 0, which is the case for various practical insurance claims, as natural catastrophes, third party liability, or certain health insurance market segments (see [1], Chapters 2 and 3 and the references therein).

The Kolmogorov distance between the probability distributions of the variables X and \tilde{X} is defined by

$$d_{KOL}(X, \tilde{X}) = \sup_{x \in \mathbb{R}} |\mathbb{P}(X > x) - \mathbb{P}(\tilde{X} > x)|.$$

By computing the characteristic functions, it is not hard to show that both X_p and \tilde{X}_p converge in distribution to $\delta_{\{0\}}$ when p goes to 0. The following lemma shows how close the distributions of X and \tilde{X} are with respect to the Kolmogorov distance by providing an explicit upper bound. We write X_p and \tilde{X}_p to emphasize the dependence of these variables on the parameter p .

Lemma 6.1. *We have*

$$d_{KOL}(X_p, \tilde{X}_p) \leq 2 \{p + (1-p) \log(1-p)\}.$$

The right hand side of the above inequality is approximately equal to 10^{-4} for $p = 1\%$, and to 10^{-6} when $p = 0.1\%$. Figure 1.5 below shows the error bound function of the previous lemma, together with an approximation of the Kolmogorov distance between \tilde{X}_p and X_p , in the case where X_p has a distribution given by (6.1), where P_Y is the Pareto law with scale parameter $\hat{x} = 20$ and shape parameter $k = 5$ (see Section 5.1 for details on the Pareto distribution). As mentioned above, the approximation of \tilde{X}_p by X_p is good for relatively small values of p : Figure 1.5 provides the distance and error values for p varying between 0.1% and 10%. We subdivided the interval $[0.1\%, 10\%]$ with a uniform step equal to 0.001, and for each value of p , the cumulative distribution function of \tilde{X}_p has been approximated by the empirical CDF constructed from a set of 10×10^6 simulations of the compound Poisson distribution. This explains the variations in the black curve, coming from the variance of the simulations.

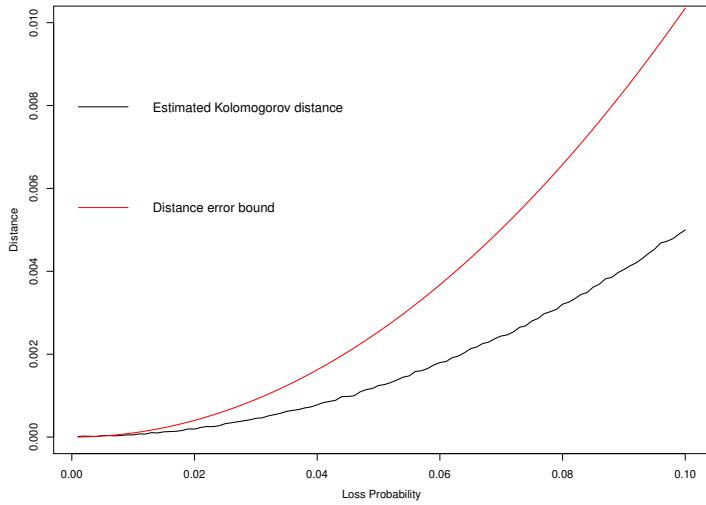


Figure 1.5 – Estimated Kolmogorov distance and error bound from Lemma 6.1, as functions of p .

Figure 1.6 below shows both CDFs of X_p and \tilde{X}_p in the case of the Pareto distribution described above, for $p = 1.98\%$, i.e. $\lambda = -\log(1 - p) = 2\%$. The jump at 0 of the CDF represents the mass at 0, which is high here ($1 - p$ being close to 1). Since the scale parameter $\hat{x} = 20$ of the Pareto distribution is also a minimal value, one can observe that the CDFs below are constant in the interval $(0, 20)$.

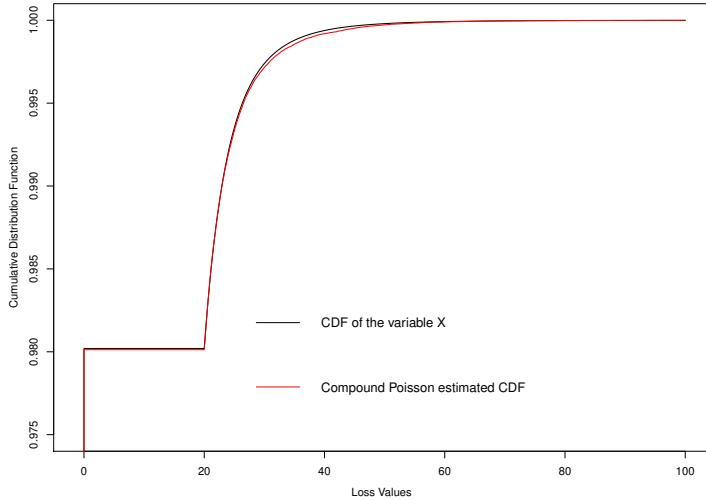


Figure 1.6 – Cumulative distribution functions of X (black curve) and X_p (red curve).

Proof. [of Lemma 6.1] On the one hand, by conditioning on N , we get for $x \geq 0$,

$$\begin{aligned}\mathbb{P}(\tilde{X}_p > x) &= \sum_{n=0}^{+\infty} \mathbb{P}\left(\sum_{j=1}^N Z_j > x | N = n\right) \mathbb{P}(N = n) \\ &= \lambda \exp(-\lambda) \mathbb{P}(Z_1 > x) + \sum_{n=2}^{+\infty} \mathbb{P}\left(\sum_{j=1}^N Z_j > x | N = n\right) \exp(-\lambda) \frac{\lambda^n}{n!}.\end{aligned}$$

On the other hand, still for $x \geq 0$,

$$\mathbb{P}(X_p > x) = p \mathbb{P}(Y > x) = (1 - \exp(-\lambda)) \mathbb{P}(Z_1 > x),$$

which implies that

$$\left| \mathbb{P}(X_p > x) - \mathbb{P}(\tilde{X}_p > x) \right| \leq |\lambda \exp(-\lambda) + \exp(-\lambda) - 1| + \sum_{n=2}^{+\infty} \mathbb{P}\left(\sum_{j=1}^N Z_j > x | N = n\right) \exp(-\lambda) \frac{\lambda^n}{n!},$$

and, since both X_p and \tilde{X}_p are a.s. non-negative,

$$\begin{aligned}d_{KOL}(X_p, \tilde{X}_p) &\leq |\lambda \exp(-\lambda) + \exp(-\lambda) - 1| + \mathbb{P}(N \geq 2) = 2\mathbb{P}(N \geq 2) \\ &= 2(p + (1 - p) \log(1 - p)).\end{aligned}$$

□

Remark 6.1. If N does not follow the Poisson distribution, then one can still use the same type of approximation as above and prove, using the exact same arguments that

$$d_{KOL}(X_p, \tilde{X}_p) \leq 2\mathbb{P}(N \geq 2),$$

where $p = 1 - \mathbb{P}(N = 0)$.

7 Conclusion

The approach we adopted in this paper allowed us to study the standard problems of self-insurance and self-protection in a unified way, and in the form of a sequential game between a buyer and a seller. We provided conditions for market insurance and self-insurance to be substitutes and for market insurance and self-protection to be complements. We proved that in both cases the formulated game admits a unique equilibrium. On the demand side, either the insurance buyer subscribes to full insurance and makes the corresponding optimal effort or he will take no insurance while still providing an optimal prevention effort. On the supply side, we obtained the existence of an optimal loading factor allowing the insurance seller to know the maximum price for which she will contract the insurance buyer.

Since we are using coherent risk measures, which are in particular cash-additive, the initial wealth of the agent do not play any role in our analysis. Initial wealth is however known to

impact insurance choices, in particular in health insurance [8]. Our model could be extended to the case of cash-subadditive risk measures, or analyzed using expected utility, aspects that we leave for further research.

Bibliographie

- [1] ALBRECHER, H., TEUGELS, J., AND BEIRLANT, J. *Reinsurance : Actuarial and Statistical Aspects*. Statistics in Practice. Wiley, 2017.
- [2] ARTZNER, P., DELBAEN, F., EBER, J.-M., AND HEATH, D. Coherent measures of risk. *Math. Finance* 9, 3 (1999), 203–228.
- [3] ATHEY, S. Monotone comparative statics under uncertainty. *The Quarterly Journal of Economics* 117, 1 (2002), 187–223.
- [4] BARRIEU, P., AND EL KAROUI, N. Inf-convolution of risk measures and optimal risk transfer. *Finance Stoch.* 9, 2 (2005), 269–298.
- [5] BLEICHRODT, H., AND EECKHOUDT, L. Willingness to pay for reductions in health risks when probabilities are distorted. *Health Economics* 15, 2 (2006), 211–214.
- [6] BOWERS, N., GERBER, H., HICKMAN, J., JONES, D., AND NESBITT, C. *Actuarial Mathematics*. Society of Actuaries, 1997.
- [7] BRIYS, E., SCHLESINGER, H., AND V. D. SCHULENBURG, J.-M. G. Reliability of Risk Management : Market Insurance, Self-Insurance and Self-Protection Reconsidered. *The Geneva Risk and Insurance Review* 16, 1 (June 1991), 45–58.
- [8] CAWLEY, J., AND PHILIPSON, T. An empirical examination of information barriers to trade in insurance. *American Economic Review* 89, 4 (1999), 827–846.
- [9] CHAN, F.-Y., AND GERBER, H. U. The reinsurer’s monopoly and the bowley solution. *ASTIN Bulletin : The Journal of the IAA* 15, 2 (1985), 141–148.
- [10] CHEN, L., AND SHEN, Y. On a new paradigm of optimal reinsurance : A stochastic stackelberg differential game between an insurer and a reinsurer. *ASTIN Bulletin : The Journal of the IAA* 48, 2 (2018), 905–960.
- [11] CHEUNG, K. C., YAM, S. C. P., AND ZHANG, Y. Risk-adjusted bowley reinsurance under distorted probabilities. *Insurance : Mathematics and Economics* 86 (2019), 64–72.
- [12] COURBAGE, C. Self-insurance, self-protection and market insurance within the dual theory of choice. *The Geneva Risk and Insurance Review* 26, 1 (2001), 43–56.

- [13] COURBAGE, C., LOUBERGÉ, H., AND PETER, R. Optimal prevention for multiple risks. *Journal of Risk and Insurance* 84, 3 (2017), 899–922.
- [14] COURBAGE, C., REY, B., AND TREICH, N. Prevention and precaution. In *Handbook of Insurance*, G. Dionne, Ed. Huebner International Series on Risk, Insurance, and Economic Security (New York : Springer), New York, 2013, ch. 8, pp. 185–204.
- [15] CVITANIĆ, J., POSSAMAÏ, D., AND TOUZI, N. Dynamic programming approach to principal-agent problems. *Finance and Stochastics* 22, 1 (2018), 1–37.
- [16] CVITANIĆ, J., WAN, X., AND YANG, H. Dynamics of contract design with screening. *Management Science* 59, 5 (2013), 1229–1244.
- [17] CVITANIĆ, J., AND ZHANG, J. *Contract theory in continuous-time models*. Springer, 2012.
- [18] DEAN HIEBERT, L. Optimal loss reduction and increases in risk aversion. *The Journal of Risk and Insurance* 56 (06 1989), 300.
- [19] DEELSTRA, G., AND PLANTIN, G. Risk theory and reinsurance. Sciences po publications, Sciences Po, 2014.
- [20] DOHERTY, N. A., AND EECKHOUDT, L. Optimal insurance without expected utility : The dual theory and the linearity of insurance contracts. *Journal of Risk and Uncertainty* 10, 2 (Mar 1995), 157–179.
- [21] EECKHOUDT, L., GOLLIER, C., AND SCHLESINGER, H. *Economic and Financial Decisions Under Risk*. Economic and Financial Decisions Under Risk. Princeton University Press, 2005.
- [22] EHRLICH, I., AND BECKER, G. S. Market insurance, self-insurance, and self-protection. *Journal of political Economy* 80, 4 (1972), 623–648.
- [23] EMBRECHTS, P., AND HOFERT, M. Risk measures and dependence modeling. In *Handbook of Insurance*, G. Dionne, Ed. Huebner International Series on Risk, Insurance, and Economic Security (New York : Springer), New York, 2013, ch. 6, pp. 135–166.
- [24] ETNER, J., AND JELEVA, M. Risk perception, prevention and diagnostic tests. *Health Economics* 22, 2 (2013), 144–156.
- [25] FÖLLMER, H., AND SCHIED, A. *Stochastic Finance : An Introduction in Discrete Time*. De Gruyter Textbook. De Gruyter, 2016.
- [26] FRITTELLI, M., AND ROSAZZA GIANIN, E. Putting order in risk measures. *Journal of Banking & Finance* 26, 7 (July 2002), 1473–1486.
- [27] GOLLIER, C., HAMMITT, J. K., AND TREICH, N. Risk and choice : A research saga. *Journal of Risk and Uncertainty* 47, 2 (Oct 2013), 129–145.

- [28] HERNÁNDEZ SANTIBÁÑEZ, N., POSSAMAÏ, D., AND ZHOU, C. Bank monitoring incentives under moral hazard and adverse selection. *arXiv preprint arXiv :1701.05864* (2017).
- [29] HÖLMSTROM, B. Moral hazard and observability. *The Bell Journal of Economics* 10, 1 (1979), 74–91.
- [30] HOY, M. The value of screening mechanisms under alternative insurance possibilities. *Journal of Public Economics* 39, 2 (July 1989), 177–206.
- [31] JEWITT, I. Justifying the first-order approach to principal–agent problems. *Econometrica* 56, 5 (1988), 1177–1190.
- [32] JEWITT, I. Choosing between risky prospects : The characterization of comparative statics results, and location independent risk. *Management Science* 35, 1 (1989), 60–70.
- [33] JOUINI, E., SCHACHERMAYER, W., AND TOUZI, N. Optimal risk sharing for law invariant monetary utility functions. *Math. Finance* 18, 2 (2008), 269–292.
- [34] JULLIEN, B., SALANIÉ, B., AND SALANIÉ, F. Should more risk-averse agents exert more effort ? *The Geneva Papers on Risk and Insurance Theory* 24, 1 (Jun 1999), 19–28.
- [35] KONRAD, K., AND SKAPERDAS, S. Self-insurance and self-protection : A nonexpected utility analysis. *The Geneva Risk and Insurance Review* 18, 2 (1993), 131–146.
- [36] LAEVEN, R. J., AND GOOVAERTS, M. J. *Premium Calculation and Insurance Pricing*. American Cancer Society, 2014.
- [37] LANGLAIS, E. Willingness to pay for risk reduction and risk aversion without the expected utility assumption. *Theory and Decision* 59, 1 (Aug 2005), 43–50.
- [38] LEE, K. Risk aversion and self-insurance-cum-protection. *Journal of Risk and Uncertainty* 17, 2 (Nov 1998), 139–151.
- [39] LI, L., AND PETER, R. Should we do more when we know less? optimal risk reduction under technological uncertainty. *University of Iowa (Working Paper)* (2018).
- [40] MACHINA, M. Non-expected utility and the robustness of the classical insurance paradigm. *The Geneva Risk and Insurance Review* 20, 1 (1995), 9–50.
- [41] MASTROGIACOMO, E., AND ROSAZZA GIANIN, E. Pareto optimal allocations and optimal risk sharing for quasiconvex risk measures. *Mathematics and Financial Economics* 9, 2 (Mar 2015), 149–167.
- [42] MEYER, D. J., AND MEYER, J. A diamond-stiglitz approach to the demand for self-protection. *Journal of Risk and Uncertainty* 42, 1 (Feb 2011), 45–60.
- [43] MIRRLEES, J. The theory of moral hazard and unobservable behaviour : part i (reprint of the unpublished 1975 version). *The Review of Economic Studies* 66, 1 (1999), 3–21.

- [44] ROGERSON, W. P. The first-order approach to principal-agent problems. *Econometrica* 53, 6 (1985), 1357–1367.
- [45] SANNIKOV, Y. A continuous-time version of the principal–agent problem. *The Review of Economic Studies* 75, 3 (2008), 957–984.
- [46] SHOGREN, J. F., AND CROCKER, T. D. Risk, self-protection, and ex ante economic value. *Journal of Environmental Economics and Management* 20, 1 (1991), 1 – 15.
- [47] TOUZI, N. Optimal insurance demand under marked point processes shocks. *Ann. Appl. Probab.* 10, 1 (2000), 283–312.
- [48] WINTER, R. Optimal insurance contracting under moral hazard. In *Handbook of Insurance*, G. Dionne, Ed. Huebner International Series on Risk, Insurance, and Economic Security (New York : Springer), New York, 2013, ch. 9, pp. 205–230.

Chapitre 2

Self-insurance and non-concave distortion risk measures

Abstract

This article considers an optimization problem for an insurance buyer in the context of proportional insurance and furnishing effort to reduce the size of a potential loss. The buyer's risk preference is given by a distortion risk measure, with risk probabilities being evaluated via a non-concave distortion function. This kind of distortion function reflects potential cognitive biases in the way in which individuals perceive risk probabilities. The buyer will select the optimal level of both insurance coverage and the prevention effort that he/she will furnish to reduce the amount that he/she stands to lose. The distribution of losses is given by a family of stochastically-ordered probability measures, indexed by the prevention effort. Contrary to what is found in the standard economic literature, introducing a non-concave distortion function leads to indeterminacy in the relationship between market insurance and self-insurance. Self-insurance and market insurance may be either substitutes, with a rise in one producing a fall in the other, or complements, so that the two rise or fall together, depending on the price elasticity.

Key words: Prevention, Self-insurance, Distortion risk measures, Distortion function, Non-concave distortion, Cognitif bias.

1 Introduction

In [17], Ehrlich and Becker developed one of the first theoretical models relating prevention to market insurance, considering a rational agent who maximizes expected utility and who takes out an insurance policy against a potential loss. In addition to the insurance contract, the agent can actively reduce the potential loss via preventive effort. Ehrlich and Becker define two types of prevention that affect different aspects of the risk: the size of the claim can

be reduced via self-insurance or the probability of loss via self-protection activities. Their comparative-statics analysis yields the following two conclusions:

- (A) In the general case, market insurance and self-insurance are substitutes, in the sense that a higher insurance price implies more prevention efforts in order to reduce the size of the loss.
- (B) In the general case, market insurance and self-protection are either substitutes or complements, so that there is an indeterminacy. Complementarity is found at the actuarial premium, in the sense that more expensive market insurance brings about less effort to reduce the probability of loss.

Although this model is the foundation of prevention in insurance, its results only hold under the Von Neuman-Morgenstern hypothesis, defining Expected Utility (EU) theory, which states that individuals have a completely unbiased view of any risk they face. However, advances in decision-making theory have shown that individuals in risky situations who need to evaluate their risk exposure use rapid mental processes that may produce inaccurate judgments. These fast mental processes yielding less-good outcomes are called cognitive biases. In [23], Kahneman describes the functioning of the human brain in decision-making. Most, if not all, individuals are subject to numerous cognitive biases. The human brain uses two opposing systems. One is a fast and automatic intuition-based system, which operates by associating ideas together via cause and effect to create a flow of thinking. These associations are consistent, but do not represent logic; they are also affected by the subject's past personal experiences. It is this first system that creates biases. On the contrary, the second system is based on assessment: it is slow, calculative and controls the first system. This second system requires concentration and effort. As such, it is inclined to accept the conclusions of the first system, which are biased as they rely on causal rather than statistical judgments.

This functioning of the human brain produces in particular an over- or under-estimation of the actual risk the individual faces, depending on their personal experience and their appetite for risk (see [24] and [34]). This results in a distortion of risk probabilities that can be translated mathematically into a probability-distortion function. This latter is a better description of how individuals evaluate risk probabilities. Different forms of distortion functions have been described in the literature. In [25], the authors design a series of experiments to determine how individuals evaluate probabilities, concluding that there is a tendency to over-estimate the probability of rare events and under-estimate that of common events. [39] describes a survey of different experiments producing evidence in support of this theory. The resulting distortion functions are inverse S-shaped. These functions vary in their concavity and convexity across agents, as individuals may have different attitudes towards risk (see [32], [21] and [1]).

Prevention activities to reduce risk exposure and their economic implications have been analyzed under the assumption that individuals overestimate their risk probabilities (see [37], [11] and, more recently, [3]). The existence of cognitive biases modifies the perception of loss

probabilities and challenges the global validity of the EU model as the reference for decision theory. It is therefore natural to test the robustness of the results in [17] with different theories of choice under risk including hypotheses that better reflect the way in which individuals estimate risks.

As EU theory cannot take into account the inverse S-shaped distortion of probabilities, [33] proposes Rank Dependent Expected Utility (RDEU) theory, which allows for the non-linear distortion of probabilities. First, events are ranked by an agent, and then weights are associated to each event, allowing for the over- or under-weighting of good or bad events. For example, the preferences of more risk-averse individuals are defined through the shapes of both the utility and probability-transformation functions. Under RDEU, [27] investigates the relationships between market insurance, self-insurance and self-protection. It is shown that most of the results in the reference model carry over to RDEU. In particular, self-insurance demand increases for risk-averse agents. The ambiguous result for self-protection is also found in this alternate theory. More recently, [19] used the RDEU model to consider the impact of risk perceptions on self-protection in the context of health risks. Their main result underlines the key role of the shape of the probability-transformation function in understanding prevention choices.

The particular RDEU model with a linear utility function is called the Dual Theory (DT), and is presented in [40]. EU theory transforms the value of a prospect by an expectation that is linear in probabilities but non-linear in wealth. DT theory is the opposite: it is linear in wealth but transforms probabilities via a weighting function defined by the cumulative distribution function over wealth. [9] analyzes the relationships between first market insurance and second self-insurance and self-protection under DT. Results (A) and (B) in [17] hold under DT theory: the substitutability of market insurance and self-insurance and the complementarity of self-protection. However in the special case where insurance companies may not price the premium according to effective prevention activities, market insurance and self-protection are substitutes.

More recently, in [3] the robustness of the results in [17] is tested under coherent risk measures. The interaction between an insurance buyer, referred to as IB in the remainder of the paper, and an insurance seller is modeled as a Stackelberg-type game, where both parties' risk preferences are given by convex risk measures. The insurance seller first proposes prices, in the form of safety loadings, and then the insurance buyer chooses both the optimal proportional insurance share and prevention effort to minimize his/her risk measure. In this context, optimal self-insurance and self-protection both exist and are unique. They are also related to the insurance share chosen by the agent as, in both cases, the agent exerts an optimal effort according to the insurance share chosen.

Summary of the approach

The problem. Consider an economic agent who faces a risk and who wishes to take out an insurance policy and who also the opportunity of taking prevention actions. His/her action is given by a vector (α, e) , where α is the insurance cover he/she chooses and e is the prevention effort to reduce risk. The IB pays a premium in exchange for the chosen α . In this premium, the actual insurance price is given by the loading factor θ . We here assume moral hazard, as the loading factor θ in the insurance seller's contract with the IB does not depend on the effort e . The prevention effort exerted by the agent is therefore not observable by the insurance company. The problem presented here is of a particular type: the optimization over (α, e) can be separated into two separate problems, the first covering α and the second e . The first shows that the optimal α only takes on two values, 0 or 1. The second is more technical, but can still be solved by finding local solutions and then comparing them carefully in order to have a global view of the results. As in [3] and [16], the positive homogeneity of the risk measure (see Definition 2.1) and the linearity of the insurance contract imply a corner-type solution for optimal coverage. Linear insurance contracts are found in practice, especially in health insurance and in reinsurance with "quota-share" contracts. Note that this analysis readily extends to reinsurance, where the effort of the agent consists for instance in underwriting policy control, or in prevention campaigns for clients.

Distortion Risk Measures. The standard assumption in behavioral economics with prevention is that the agent maximizes his/her expected utility (see [17], [15], [29]). In this paper, the criterion of the agent is given by a distortion risk measure (see Equation (2.3)). Risk measures are a common tool in actuarial practice, as the European regulatory texts, Solvency II and the Swiss Solvency Test, require insurance and reinsurance companies to calculate the Value-at-Risk or the Expected Shortfall (see [20] for the definitions, and the differences between these measures). These two risk measures are examples of the class of distortion risk measures. This class is a subfamily of law-invariant risk measures. If ρ is a distortion risk measure, then a monetary utility function U can be defined as $U(X) := -\rho(-X)$. The results presented here can be reformulated in terms of law-invariant monetary utility functions. Monetary utility functions are typically assumed to be concave, but in the current framework they can be taken as non-concave. See [18] and [20] for more details on risk measures and insurance applications. The use of a non-concave distortion risk measure takes into account cognitive biases that interfere with the evaluation of risk probabilities. As underlined in [23], various thought mechanisms influence our perception of events, leading in particular to the modification of risk perception. This phenomenon has been studied intensively in the economic literature (see [25], [39] and [1]). It has been shown that the probability of an event differs from the subjective probability that people attach to it in a non-concave way that depends on risk aversion. The risk-averse tend to overestimate small-probability events and underweight medium- and large-probability events; risk lovers behave in the opposite

way. Moreover, under the cash-additivity property (see Definition 2.1), the diversification principle is equivalent to the convexity property of a risk measure (see [20] for more details). Using non-concave distortion risk measures allows us to consider situations where there is no diversification of risk sources.

The choice of distributions. In the standard literature, the distributions of losses are discrete, and are concentrated at two points: the IB faces a loss L , defined as $L = \ell B$ with B being a Bernoulli random variable. In the framework of this paper (see Section 2 for more details), a family of random variables $(X_e)_{e \in (0, +\infty)}$ is considered, indexed by a prevention-effort parameter e , and such that the distributions of $(X_e)_{e \in (0, +\infty)}$ form a family of probability measures which is decreasing under first-order stochastic dominance (FSD) (see Assumption 2.1 for more details).

Main contributions

Our work here extends the problem of self-insurance to the case where the objective function of the IB is characterized by a law-invariant distortion risk measure, where the distortion is non-concave. We prove that there exists a prevention effort that reduces the IB's risk optimally.

The main contribution regarding self-insurance is that either statement (A) or (B) will hold. This last conclusion results from considering how higher prices affect optimal insurance demand and prevention effort, in particular the self-insurance effort, and by focusing on the relative impact of effort on the loss distribution through the risk measure and on prices.

The structure of the paper

In Section 2 the problem and the main assumptions are presented. Section 3 sets out and solves the buyer's problem. Section 4 presents an illustration of the method developed in this paper, in the particular case of the Weibull law and with a specified inverse S-shaped distortion function. Last, Section 5 concludes.

2 Model and assumptions

Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a fixed probability space. For a given random variable X , \bar{F}_X denotes the survival function defined by $\bar{F}_X(x) := \mathbb{P}(X > x)$ and \bar{q}_X denotes the tail quantile function, defined as the inverse of \bar{F}_X

$$\bar{q}_X(u) := \inf\{x \in \mathbb{R} \mid \bar{F}_X(x) \leq u\}.$$

Assume that the insured can make some effort to reduce their loss, which is modeled by a parameter $e \in [0, \infty)$ that we can consider as prevention. As our analysis is based on the

definition of self-insurance in [17], the effort furnished by the insured affects the size of losses by directly affecting their distribution.

In what follows, we introduce the insurance-claims distribution used in this paper and the specific risk measure that is analyzed.

Let $(X_e)_{e \in [0, \infty)}$ be a family of non-negative random variables, representing losses, such that the distributions of X_e are defined by the following equation:

$$P_{X_e} = (1 - p)\delta_{\{0\}} + pP_{Y_e}, \quad (2.1)$$

where $0 < p < 1$, $\delta_{\{0\}}$ is the Dirac mass at 0 and P_{Y_e} denotes the distribution of a positive random variable Y_e , which IB can influence through her choice of preventive effort.

Example 2.1. One example of the distributions of the random variable Y_e is the Pareto distribution, which is often used in standard actuarial models (see [5]). Other well-known possibilities are the Fréchet, Weibull and Log-Normal distributions.

Assumption 2.1. The family $(X_e)_{e \in [0, +\infty)}$ is such that the distribution of each X_e is decreasing under first-order stochastic dominance (FSD)

$$e_1 < e_2 \implies \mathbb{E}[f(X_{e_1})] > \mathbb{E}[f(X_{e_2})], \quad (2.2)$$

for every non-decreasing function $f : \mathbb{R} \rightarrow \mathbb{R}$.

$X \underset{\text{(mon)}}{\leq} Y$ denotes that X is dominated by Y for FSD, also called the *monotone* order. The following property of FSD (see Theorem 2.68 in [20]) will also be used

$$X \underset{\text{(mon)}}{\leq} Y \iff \bar{q}_X(u) \leq \bar{q}_Y(u), \quad \forall u \in (0, 1).$$

In (2.2), no convexity or concavity assumption about f is made. This implies that the situation in which the distributions of $(X_e)_{e \in (0, +\infty)}$ are decreasing for the monotone convex order is not excluded. Moreover, Assumption 2.1 implies that, for each $u \in (0, 1)$, the function $e \mapsto \bar{q}_{X_e}(u)$ is non-increasing.

Since the family of distributions used is indexed by some prevention effort, it is assumed that the effort affects the tail quantile of the loss distribution in such a way that the marginal impact decreases. This classic assumption in economics translates into the following in mathematics.

Assumption 2.2. For every $u \in (0, 1)$, the function $e \mapsto \bar{q}_{X_e}(u)$ is convex.

The problem of the IB is presented as an optimization problem. When the IB searches for an insurance contract, he/she has to choose optimal insurance coverage and the optimal effort parameter. It is assumed that he/she minimizes a distortion risk measure, where the distortion function is non-concave. Definition 2.1 states the properties of all risk measures.

Definition 2.1. $\rho : L^\infty \rightarrow \mathbb{R}$ is a risk measure if

1. ρ is monotone: $X \geq Y \Rightarrow \rho(X) \geq \rho(Y)$ almost surely.
2. ρ is cash-additive: $\forall m \in \mathbb{R}, \rho(X + m) = \rho(X) + m$.
3. ρ is positively homogeneous: $\forall \lambda \in \mathbb{R}^+, \rho(\lambda X) = \lambda \rho(X)$.
4. ρ is law-invariant: $X \stackrel{d}{=} Y \Rightarrow \rho(X) = \rho(Y)$, where $\stackrel{d}{=}$ denotes equality in distribution.

The first property states that if a loss Y is larger than a loss X in all states of nature, then Y is riskier than X . The second property is a strong mathematical assumption that translates in economics as the risk measure ρ having a unit. In the context of this paper this corresponds to a measure in capital (Euros, Pounds, Dollars etc.), although the fourth property suggests that this risk measure should be independent of this currency. The last property can be interpreted as the dependence of the risk measurement on only the distribution of the losses and not on the underlying assets.

For more details and properties of risk measures, refer to [20] (Chapter 4).

Distortion risk measures are an important class of risk measures defined by

$$\rho(X) := \int_0^1 \bar{q}_X(u) d\psi(u) = \int_0^1 \bar{q}_X(u) d\psi(u) du, \quad (2.3)$$

where ψ is a distortion function, i.e. a non-decreasing function from $[0, 1]$ to $[0, 1]$ such that $\psi(0) = 0$ and $\psi(1) = 1$.

A distortion risk measure is not convex in general. If the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ has no atoms, then we can prove that a risk measure of the form given by Equation (2.3) is convex if and only if the distortion function is concave (see [20] for the distribution without atoms case, and [4] for the atom-distribution case). The convexity of the risk measure incites economic agents to use diversification. The risk measure is then called coherent. Note that non-coherent risk measures are used in insurance and reinsurance, as insurance and reinsurance companies need to calculate a risk measure called Value-at-Risk according to the Solvency II rules.

One important example of a non-concave distortion function is the inverse S-shaped distortion function. Figure 2.1 depicts that given by $\psi(u) = \exp(-(-\log(u))^\alpha)$, with $0 < \alpha < 1$.

1. Recall that L^∞ is the space of measurable functions, which take values in \mathbb{R} and are \mathbb{P} -essentially bounded, where the essential supremum of the absolute value of the functions serves as an appropriate norm.

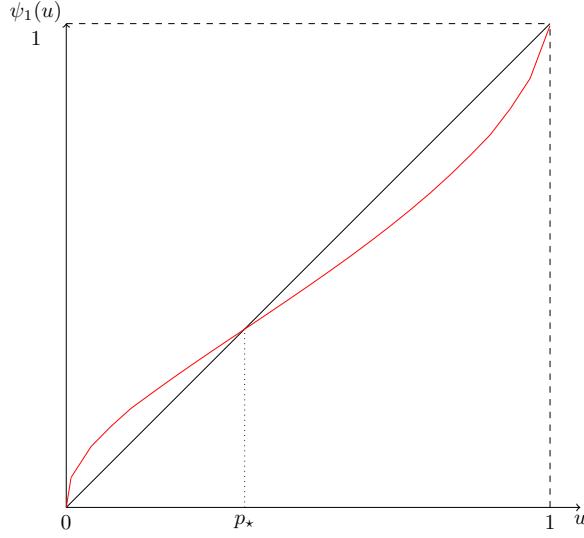


Figure 2.1 – An example of an inverse S-shaped distortion function: $\psi(u) = \exp(-(-\log(u))^\alpha)$, with $0 < \alpha < 1$.

As noted in the Introduction, choosing a non-concave, and in particular an inverse S-shaped, probability-distortion more accurately represents how individuals evaluate their risk probabilities (see [21], [25], [35] and [39]). This function divides the probability interval into two: a first part of small probabilities, where the function is concave, and a second part of medium and high probabilities, where the distortion function is convex. The value of the boundary between those two intervals is denoted by p_* . As described in [35], [38] and [32], the point p_* is an invariant fixed point of the distortion function ψ . Moreover, [1] and [36] underline the relationship between the form of this type of distortion function and individual optimism regarding the best outcomes and pessimism regarding the worst outcomes. The concavity below p_* reflects the overweighting of low probabilities, and the convexity after p_* the underweighting of high probabilities. This change in the shape of the curve reveals economic agents' sensitivity to extreme outcomes and the change in sensitivity moving towards medium probabilities. For an optimistic agent, the distortion function will be quite steep away from 0 and quite flat away from 1, meaning that a small increase in the risk probability has a greater impact on low than on high probabilities. On the contrary, for a pessimistic agent, the inverse S-shaped distortion curve is relatively flat away from 0 and relatively steep away from 1: high probabilities are more distorted than are small probabilities. Last, as noted in [35], the inverse S-shaped curve implies that people are insensitive to probabilities in the middle of the range. For more examples of this type of distortion function, see [1] and [32].

Proposition 2.1. *Let ρ be a distortion risk measure defined by Equation (2.3). If the distributions of $(X_e)_{e \in [0, \infty)}$ satisfy Assumption 2.2 then*

- i) *The function $e \mapsto \rho(X_e)$ is convex.*

ii) The function $e \mapsto \rho(X_e)$ is continuous and non-increasing.

Proof. i) Since ψ_1 is non-decreasing and non-negative, the distortion risk measure ρ_1 depends on e only through the function $e \mapsto \bar{q}_{X_e}$, which is convex by Assumption 2.2. Hence $e \mapsto \rho_1(X_e)$ is convex.

ii) Convexity entails that $e \mapsto \rho_1(X_e)$ is continuous. The assumption that the distribution of the family $(X_e)_{e \in [0, \infty)}$ is decreasing by FSD implies that the tail quantile function \bar{q}_{X_e} is a non-increasing function of effort. Equation (2.3) then implies that $e \mapsto \rho_1(X_e)$ is non-increasing. \square

Remark 2.1. As the convexity of the risk measure in e comes entirely from the convexity of the tail quantile function, an example for ρ non convex in e would be a distortion function ψ that is sometimes non-increasing. This means that the cumulative distribution function of X_e would be non-increasing for some intervals of $[0, \infty)$ and the density of X_e would allow negative probabilities in the sense that the economic agent who distorts objective probabilities using ψ perceives some probabilities as being negative.

Remark 2.2. The last point of the proof states that for $\psi(u) = u$, $\rho(X_e) = \mathbb{E}[X_e] \geq 0$, as X_e is non-negative. Hence, by Proposition 2.1, $e \mapsto \mathbb{E}[X_e]$ is a continuous, non-increasing, non-negative and convex function.

From Equation (2.3), the risk measure $\rho_1(X_e)$ can be written as

$$\rho_1(X_e) = \int_0^1 \bar{q}_{X_e}(u) d\psi_1(u) du$$

which, from Equation (2.1), is equal to

$$\rho_1(X_e) = \int_0^p \bar{q}_{X_e}(u) d\psi_1(u) du.$$

Remark 2.3. Note that since ψ_1 is non-decreasing and non-negative, and as the random variable X_e is non-negative, the risk measure ρ_1 is non-negative.

Remark 2.4. In the case where ψ is differentiable, $d\psi(u) = \psi'(u)$ for all $u \in (0, 1)$, then the risk measure ρ_1 can be written as $\rho_1(X_e) = \int_0^p \bar{q}_{X_e}(u) \psi'_1(u) du$.

Assume that the IB has initial wealth denoted by ω_0 and faces a risk with loss X_e , against which he/she takes out a proportional insurance contract with an insurance company. In this insurance contract, he/she chooses her level of insurance, denoted by $\alpha \in [0, 1]$, and pays a premium denoted by Π . The prevention effort $e \in [0, \infty)$ made by the IB has a monetary cost $c(e)$, where c is a non-decreasing and strictly convex function.² Common examples of the

2. If c is convex but not strictly convex, all of our analysis holds by considering the smallest effort that minimizes the IB's objective function.

functions for the cost of the effort can be found in [12]. Suppose that the IB uses a distortion risk measure ρ_1 . Her goal is to minimize the risk measure associated with her total loss, which is given by

$$L_1(\alpha, e) := \rho_1\left((1 - \alpha)X_e + \Pi(\alpha X_e) + c(e) - \omega_0\right).$$

Using the positive homogeneity and cash-additivity properties of ρ , the IB's objective function then simplifies to

$$L_1(\alpha, e) = (1 - \alpha)\rho_1(X_e) + \Pi(\alpha X_e) + c(e) - \omega_0. \quad (2.4)$$

The IB therefore solves the following optimization problem

$$\inf_{(\alpha, e) \in [0, 1] \times [0, \infty)} L_1(\alpha, e). \quad (2.5)$$

Remark 2.5. As ρ_1 is a cash-additive risk measure, without loss of generality initial wealth ω_0 is considered to be 0 throughout the remainder of the paper: the initial wealth of the agent, ω_0 , plays no role here. However, in Economics (see [7] for health insurance) initial wealth is known to affect insurance choices. This economic consideration could be included by considering cash-subadditive risk measures, or analyzed using expected utility. This point is left for further research.

For tractability reasons, and to facilitate reading, suppose that the insurance premium Π is defined as follows:

$$\Pi(X_e) := (1 + \theta)\mathbb{E}[X_e]. \quad (2.6)$$

This form of premium with a safety loading is a standard choice in the actuarial-science literature (see [14] and [28]).

3 The optimization problem

This section solves the IB's problem. First, optimal insurance coverage is determined, and due to the special form of the IB's objective function, the optimal insurance coverage can be written as a function of the IB's prevention effort. Optimal self-insurance effort is then outlined, depending on both insurance coverage and the safety loading of the insurance premium, to obtain the solution to the IB's problem.

3.1 Optimal insurance coverage

The resolution of the IB's problem begins by the calculation of her optimal insurance coverage.

Proposition 3.1. *Optimal insurance coverage, α^* , takes only two values, 0 or 1, depending on the value of effort e .*

Proof. In Equation (2.6), note that the pricing functional is positively homogeneous as it is defined by an expectation, so that for all $\alpha \in [0, 1]$

$$(1 + \theta)\mathbb{E}[\alpha X_e] = \alpha(1 + \theta)\mathbb{E}[X_e].$$

Equation (2.4) can be rewritten as

$$\begin{aligned} L_1(\alpha, e) &= (1 - \alpha)\rho_1(X_e) + \alpha(1 + \theta)\mathbb{E}[X_e] + c(e), \\ &= \rho_1(X_e) + \alpha[(1 + \theta)\mathbb{E}[X_e] - \rho_1(X_e)] + c(e). \end{aligned}$$

The value of α^* hence depends only on the sign of $(1 + \theta)\mathbb{E}[X_e] - \rho_1(X_e)$. Optimal insurance coverage can then be written as a function of e as

$$\alpha^*(e) = \begin{cases} 1 & \text{if } e \in \mathcal{I} \\ 0 & \text{if } e \in \mathcal{N} := \mathcal{I}^c, \end{cases} \quad (3.1)$$

$$\text{where } \mathcal{I} := \left\{ e \in [0, \infty) \mid 1 + \theta \leq \frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \right\}.$$

□

This proposition means that the insurance buyer will either take out full insurance or no insurance. This type of choice is found in other models of prevention with insurance (see [9]), and also empirically (see [16]).

3.2 Optimal self-insurance effort

It is important to recall that the IB makes an effort to reduce the size of the potential loss that is faced. This translates into the choice of the distribution of this loss.

As was the case for fixed e , the Problem (2.5) is already solved in α ; it remains to minimize the functional L_1 in effort for fixed α . Suppose α is fixed, the function $e \mapsto L_1(\alpha^*(e), e)$ is not convex so the minimization is not straightforward. From Equation (3.1), the optimal insurance coverage only takes on values 0 or 1 in separate sets, so $e \mapsto L_1(\alpha^*(e), e)$ can be analyzed in these respective sets where the function is now convex. We can then compare the local minima in each domain to identify the global minimum.

For $e \in [0, \infty)$, define

$$L_{\mathcal{N}}(e) := \rho_1(X_e) + c(e), \text{ and} \quad (3.2)$$

$$L_{\mathcal{I}}^\theta(e) := (1 + \theta)\mathbb{E}[X_e] + c(e), \quad (3.3)$$

such that

$$L_1(\alpha^*(e), e) = \begin{cases} L_{\mathcal{N}}(e), & \text{if } e \in \mathcal{N}, \\ L_{\mathcal{I}}^\theta(e), & \text{if } e \in \mathcal{I}. \end{cases}$$

Define also

$$G(e) := \frac{\rho_1(X_e)}{\mathbb{E}[X_e]}, \quad (3.4)$$

which reveals how effort affects the risk and the price of insurance.

The sets \mathcal{I} and \mathcal{N} can be rewritten as

$$\begin{aligned} \mathcal{I} &:= \{e \in [0, \infty) \mid 1 + \theta \leq G(e)\}, \\ \mathcal{N} &:= \{e \in [0, \infty) \mid 1 + \theta > G(e)\}. \end{aligned}$$

As the two intervals \mathcal{I} and \mathcal{N} are defined according to the function G , establishing the monotonicity of G will help to solve the IB's problem.

Assumption 3.1. *The function G is non-monotonic.*

Assumption 3.1 implies that, for a fixed θ , G can potentially cross the horizontal line at $1 + \theta$ multiple times if G is not monotonic, as illustrated in Figure 2.2.

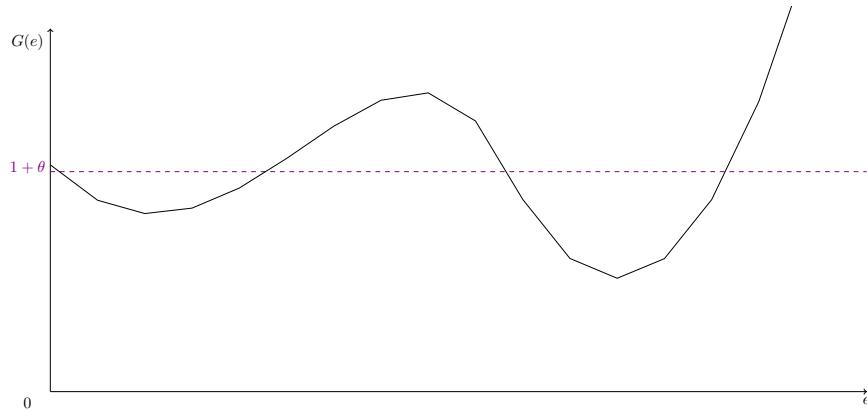


Figure 2.2 – Illustration of a non-monotonic function G crossing $1 + \theta$ multiple times for a fixed θ .

As the sets \mathcal{I} and \mathcal{N} depend on θ by definition, their shapes will change as θ rises. As θ rises so does $1 + \theta$, so that the set \mathcal{N} that satisfies the inequality $1 + \theta > G(e)$ increases in size and its complement, the set \mathcal{I} , shrinks. Figure 2.3 shows how these two sets change as θ rises.

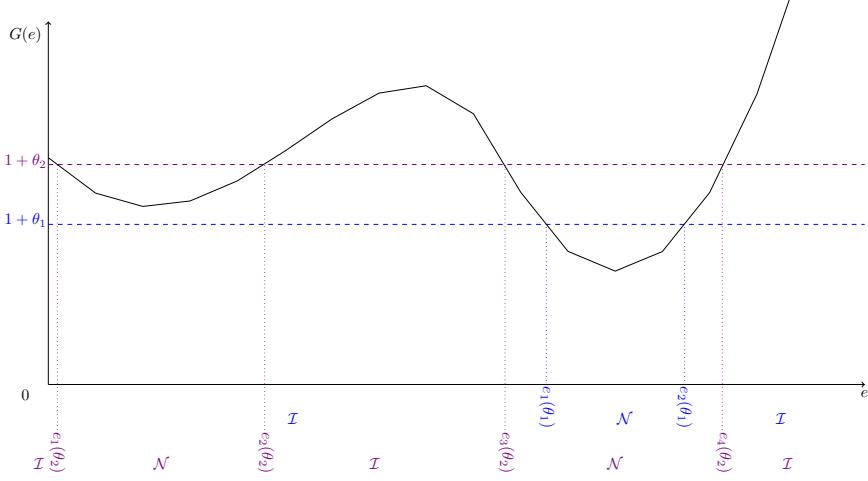


Figure 2.3 – Illustration of the change in the \mathcal{I} and \mathcal{N} sets as θ rises.

The finite boundary points of the sets \mathcal{I} and \mathcal{N} are denoted by $e_i(\theta)$, with $i \in \mathbb{N}$ satisfying $e_i(\theta) \leq e_j(\theta)$ for $i < j$. Note that for all $i \in \mathbb{N}$, $L_{\mathcal{I}}^\theta(e_i(\theta)) = L_{\mathcal{N}}(e_i(\theta))$.

Assumption 3.2. *If Assumption 3.1 does not hold then the function G is non-decreasing.*

In the remainder of the paper, the function G satisfies Assumption 3.1. The case of Assumption 3.2 will be discussed in Remark 3.1.

The next steps focus on each set \mathcal{I} and \mathcal{N} and establish the local minima of each for fixed θ . In the set \mathcal{I} , the analysis of the function $e \mapsto L_1(\alpha^*(e), e)$ reduces to that of

$$L_{\mathcal{I}}^\theta(e) = (1 + \theta)\mathbb{E}[X_e] + c(e).$$

Recall that $e \mapsto \mathbb{E}[X_e]$ and $e \mapsto c(e)$ are differentiable almost everywhere. In the following proposition, it is assumed that they are differentiable everywhere.

Proposition 3.2. *$L_{\mathcal{I}}^\theta$ has a unique minimizer, denoted by $e_I(\theta)$ on $[0, \infty)$. Moreover, the function $\theta \mapsto e_I(\theta)$ is non-decreasing and there exists a positive constant θ_0 such that if $\theta \leq \theta_0$ then $e_I(\theta) = 0$.*

Proof. From Equation (3.3), $L_{\mathcal{I}}^\theta$ is continuous, strictly convex and goes to $+\infty$ as e goes to $+\infty$; it then has a unique minimizer on $[0, \infty)$ denoted by $e_I(\theta)$, defined as the solution to the equation

$$(1 + \theta) = \frac{-c'(e)}{\partial_e \mathbb{E}[X_e]}. \quad (3.5)$$

The remainder of this proof follows that in Proposition 3.2 in [3]; for completeness, we show the arguments below.

As the random variables $(X_e)_{e \in \mathbb{R}^+}$ have finite expected values, the function $L_{\mathcal{I}}^\theta$ is finite at $e = 0$. Its right-hand derivative $\frac{d}{de} L_{\mathcal{I}}^\theta(0)$ is well-defined and

$$\frac{d}{de} L_{\mathcal{I}}^\theta(0) \geq 0 \iff \theta \leq -\frac{c'(0)}{E'(0)} - 1 =: \theta_0. \quad (3.6)$$

In this case, $L_{\mathcal{I}}^\theta$ is non-decreasing at 0, and by convexity it is non-decreasing everywhere, so its minimum is attained at $e = 0$. Note that for the constant θ_0 to be well-defined, the expectation needs to satisfy $E'(0) \neq 0$. If this is not the case, then $\frac{d}{de} L_{\mathcal{I}}^\theta(0) = c'(0) \geq 0$ and the function $L_{\mathcal{I}}^\theta$ is non-decreasing for any $\theta \in \mathbb{R}$.

Now suppose that $\theta > \theta_0$. This means that the function $L_{\mathcal{I}}^\theta$ is decreasing at 0. Therefore, the minimizer $e_I(\theta)$ of $L_{\mathcal{I}}^\theta$ on \mathbb{R}^+ is at an interior point characterized by

$$1 + \theta = -\frac{c'(e_I(\theta))}{E'(e_I(\theta))}. \quad (3.7)$$

$\theta \mapsto e_I(\theta)$ is non-decreasing, as the function $e \mapsto c'(e) \frac{-1}{E'(e)}$ is non-decreasing (as it is the product of two non-negative non-decreasing functions). As $1 + \theta$ obviously rises with θ , (2.2) implies that $e_I(\theta)$ is non-decreasing in θ . \square

Proposition 3.3. *Suppose $\theta > \theta_0$. If $e_I(\theta)$ belongs to \mathcal{I} then it is the local minimizer of $L_{\mathcal{I}}^\theta$ on this set.*

Proof. If $e_I(\theta)$ is in the set \mathcal{I} , then by Proposition 3.2 it is the local minimizer of $L_{\mathcal{I}}^\theta$ on \mathcal{I} . Otherwise, the restriction of $L_{\mathcal{I}}^\theta$ to \mathcal{I} is first non-increasing and then non-decreasing, and so the local minimizer is given by a boundary point of the set \mathcal{I} . \square

Recall that on the set \mathcal{N} , $e \mapsto L_{\mathcal{N}}(e)$ is the function to be analyzed.

Proposition 3.4. *$L_{\mathcal{N}}$ has a unique minimizer, denoted by e_N on $[0, \infty)$.*

Proof. Note that $e \mapsto L_{\mathcal{N}}(e)$ is continuous and non-negative, as the risk measure ρ_1 and the cost function c are continuous and non-negative functions of e . In addition, $L_{\mathcal{N}}$ is strictly convex as ρ_1 is non-increasing and convex, and $c(e)$ is non-decreasing and strictly convex.

We have to consider two cases to prove this proposition. First, assume that the risk measure ρ_1 is finite at 0. Then the right-hand derivative of $L_{\mathcal{N}}$ at 0 is well-defined. $L_{\mathcal{N}}$ is therefore non-decreasing at 0 if and only if $-\partial_e \rho_1(X_0) \leq c'(0)$. The convexity of $L_{\mathcal{N}}$ leads to its monotonicity on $[0, \infty)$ and there exists a unique minimizer $e_N = 0$. In the case where $-\partial_e \rho_1(X_0) > c'(0)$, $L_{\mathcal{N}}$ starts by being decreasing at 0 and its strict convexity establishes the existence of a unique minimizer $e_N > 0$ on $[0, \infty)$.

Now assume that $\rho_1(X_e) \rightarrow \infty$ as $e \rightarrow 0^+$. Since $c(e) \rightarrow \infty$ as $e \rightarrow \infty$, $L_{\mathcal{N}}$ is a strictly convex and coercive function. Hence, it has a unique minimizer on $[0, \infty)$. \square

Proposition 3.5. *There exists a unique positive constant θ_N such that*

If $\theta \leq \theta_N$, then e_N belongs to the set \mathcal{I} .

If $\theta > \theta_N$, then e_N belongs to the set \mathcal{N} .

Moreover, when e_N is in \mathcal{N} then it is the local minimizer of $L_{\mathcal{N}}$ on \mathcal{N} .

Proof. Recall the definition of $\mathcal{N} := \{e \in [0, \infty) \mid 1 + \theta > G(e)\}$. Since e_N is independent of θ , as θ rises there exists a constant denoted by θ_N such that for $\theta \geq \theta_N$ e_N belongs to the set \mathcal{N} . Hence the definition $\theta_N := \inf\{\theta \geq 0 \mid 1 + \theta > G(e_N)\}$. This constant is unique by definition, as e_N is independent of θ . Moreover when e_N is on \mathcal{I} , the concavity of $e \mapsto L_{\mathcal{N}}$ in e_N proves that its restriction to the set \mathcal{N} is non-decreasing. Hence the local minimizer on \mathcal{N} is given by a boundary point of \mathcal{N} . \square

The remainder of our analysis of self-insurance effort aims to find the global minimum by comparing the local minima on each set \mathcal{N} and \mathcal{I} , depending on the value of θ . In particular, $e_I(\theta)$ can switch from one set to another as θ rises. To do so, the values $L_{\mathcal{I}}^\theta(e_I(\theta))$ and $L_{\mathcal{N}}(e_N)$ have to be compared to each other. As the bounds of \mathcal{I} and \mathcal{N} , and $e_I(\theta)$ cannot be clearly compared, all possible orders of $e_I(\theta)$ and e_N need to be considered. For a fixed θ , four cases can be distinguished

- | | |
|---|---|
| (a) $e_I(\theta) \in \mathcal{I}$ and $e_N \in \mathcal{I}$, | (c) $e_I(\theta) \in \mathcal{I}$ and $e_N \in \mathcal{N}$, |
| (b) $e_I(\theta) \in \mathcal{N}$ and $e_N \in \mathcal{N}$, | (d) $e_I(\theta) \in \mathcal{N}$ and $e_N \in \mathcal{I}$ |

The next Lemma proves that case (d) is not possible.

Lemma 3.1. *There is no positive θ such that both $e_I(\theta) \in \mathcal{N}$ and $e_N \in \mathcal{I}$.*

Proof. By contradiction, suppose there exists $\theta \geq 0$ such that both $e_I(\theta) \in \mathcal{N}$ and $e_N \in \mathcal{I}$. Recall that

$$\begin{aligned} \forall e \in \mathcal{I}, L_{\mathcal{I}}^\theta(e) &\leq L_{\mathcal{N}}(e), \\ \forall e \in \mathcal{N}, L_{\mathcal{N}}(e) &< L_{\mathcal{I}}^\theta(e), \end{aligned}$$

then $L_{\mathcal{I}}^\theta(e_N) \leq L_{\mathcal{N}}(e_N)$ and $L_{\mathcal{N}}(e_I(\theta)) \leq L_{\mathcal{I}}^\theta(e_I(\theta))$. As $e_I(\theta)$ is the global minimum of $L_{\mathcal{I}}^\theta$, $L_{\mathcal{N}}(e_I(\theta)) < L_{\mathcal{N}}(e_N)$, which contradicts the global minimality of e_N for $L_{\mathcal{N}}$. \square

The next Proposition deals with case (a).

Proposition 3.6. *If e_N belongs to \mathcal{I} , then the global minimum of L_1 is $e_I(\theta)$.*

Proof. If e_N belongs to \mathcal{I} , then by Lemma 3.1 $e_I(\theta)$ also belongs to the set \mathcal{I} , and from the definition of this set and the global optimality of $e_I(\theta)$ for $e \mapsto L_{\mathcal{I}}^\theta(e)$, $L_{\mathcal{I}}^\theta(e_I(\theta)) \leq L_{\mathcal{I}}^\theta(e_N) < L_{\mathcal{N}}(e_N)$.

In the following Proposition, we consider case (b).

Proposition 3.7. *If $e_I(\theta)$ belongs to \mathcal{N} , then the global minimum of L_1 is e_N .*

Proof. By Lemma 3.1, $e_I(\theta) \in \mathcal{N}$ implies that e_N is in \mathcal{N} . From the definition of \mathcal{N} and the global optimality of e_N for $e \mapsto L_{\mathcal{N}}(e)$, $L_{\mathcal{N}}(e_N) < L_{\mathcal{N}}(e_I(\theta)) \leq L_{\mathcal{I}}^\theta(e_I(\theta))$. \square

The Theorem that follows addresses case (c).

Theorem 3.1. *There exists a positive constant θ_M such that*

If $\theta \leq \theta_M$, then $e_I(\theta)$ is the global minimum of the objective function L_1 .

If $\theta > \theta_M$, then e_N is the global minimum of the objective function L_1 .

Moreover θ_M is larger than θ_N .

Proof. The proof of the Theorem follows the argument of the proof of Theorem 3.2 in [3]. For completeness, it is reproduced in the following.

First note that, for all effort e , the objective function $\theta \mapsto L_{\mathcal{I}}^\theta$ is non-decreasing. For all e , if $\theta < \theta'$ then

$$(1 + \theta)\mathbb{E}[X_e] + c(e) < (1 + \theta')\mathbb{E}[X_e] + c(e) \iff L_{\mathcal{I}}^\theta(e) < L_{\mathcal{I}}^{\theta'}(e).$$

From Proposition 3.2, $\theta \mapsto e_I(\theta)$ is non-decreasing, so that $\theta \mapsto L_{\mathcal{I}}^\theta(e_I(\theta))$ is also non-decreasing.

Since $L_{\mathcal{N}}(e_N)$ is independent of θ , and since at the actuarial premium it is better to have full insurance coverage than no insurance at all, meaning $L_{\mathcal{I}}^0(e_I(0)) < L_{\mathcal{N}}(e_N)$, the non-decreasing mapping $\theta \mapsto L_{\mathcal{I}}^\theta(e_I(\theta))$ leads to the definition of a constant $\theta_M \geq 0$ such that if $\theta \leq \theta_M$ then $L_{\mathcal{I}}^\theta(e_I(\theta)) < L_{\mathcal{N}}(e_N)$, and if $\theta > \theta_M$ then $L_{\mathcal{I}}^\theta(e_I(\theta)) > L_{\mathcal{N}}(e_N)$.

We can prove by contradiction that $\theta_M \geq \theta_N$. Suppose that $\theta_N > \theta_M$, and take a θ such that $\theta_N > \theta > \theta_M$. From the definition of θ_M , $L_{\mathcal{I}}^\theta(e_I(\theta)) < L_{\mathcal{N}}(e_N)$. From the definition of θ_N , e_N belongs to \mathcal{I} and so does $e_I(\theta)$, so that $L_{\mathcal{I}}^\theta(e_I(\theta)) > L_{\mathcal{N}}(e_N)$, which is impossible. \square

Remark 3.1. *Note that the case of monotonic non-decreasing G is simpler. Here the function G will cross $1 + \theta$ only once, dividing the sets \mathcal{I} and \mathcal{N} into only two intervals of \mathbb{R}^+ at a single boundary effort denoted by e_θ . The examination of the global minimum of the value function of the agent remains similar to that developed above, as effort e_I and boundary effort e_θ are both non-decreasing functions of the insurance price θ . As there is no obvious way to compare these two effort levels with respect to the values of θ , all of the possibilities need to be evaluated.*

To conclude this Section, the following Corollary characterizes the behavior of the optimal effort $e^*(\theta)$, which is very similar to that described in [30].

Corollary 3.1. *The map $\theta \mapsto e^*(\theta)$ is non-decreasing for $\theta \leq \theta_M$. It becomes constant, and equal to e_N , for $\theta > \theta_M$. It exhibits a jump at $\theta = \theta_M$ equal to $e_N - e_I(\theta_M)$. As $e_I(\theta)$ can be either larger or smaller than e_N , this jump can be negative.*

Corollary 3.1 can be interpreted in economic terms as follows. An increase in the price of insurance can either imply greater prevention effort to reduce risk or less prevention effort to reduce his/her loss. Hence this result reflects an indeterminacy in the relationship between market insurance and self-insurance. Substitutability between market insurance and self-insurance may result: a higher insurance price leads to less insurance demand and increases the demand for prevention. This is the most-commonly discussed case in the literature, as it is the main conclusion of the reference model [17], as well as in the two extensions under RDEU [9] and coherent risk measures [3]. This result is also found in the experimental analysis in [31].

However, market insurance and self-insurance may also be complementary: a higher insurance price implies less insurance demand and a lower demand for prevention. To the best of the author's knowledge, this result is new in the prevention literature and only appears in the theoretical model in [6], which considers market insurance and self-insurance in the presence of nonreliability risk with mixed distributions (continuous/discrete). The risk averse economic agent can use three main tools of risk management that may not be effective and he/she assigns a probability distribution to their effectiveness. In addition to these prevention tools, he/she may purchase a market insurance. This complementarity could be explained by individual risk aversion, as in advantageous-selection (as opposed to adverse-selection) models. Advantageous selection is based on individual preferences and risk aversion: less risk-averse agents purchase less insurance and carry out less prevention, with the opposite holding for the more risk-averse (see [13] and [22]). These results have been confirmed empirically (see [10] and [26]).

4 Case study

Consider the loss given by a random variable X_e , with distribution as in Equation (2.1), with Y_e following a Weibull distribution. The Weibull distribution is chosen as it is often used in insurance and reinsurance. Depending on how the shape parameter of the Weibull distribution is defined, it is possible to have either light tail claims (e.g. small claims) or heavy tail claims (e.g. large claims) (see [2] for further discussion).

As we are analyzing self-insurance, the IB's prevention action translates into the shape parameter denoted by τ . $e \mapsto \tau(e)$ is assumed to be a positive, non-decreasing and concave function. The inverse S-shaped distortion function of the agent is chosen as $\psi_1(u) = \exp(-(-\log(u))^\eta)$, where $\eta \in (0, 1)$.

The survival function of X_e is

$$\mathbb{P}[X_e > u] = p \exp\left(-\lambda u^{\tau(e)}\right). \quad (4.1)$$

where $\lambda > 0$ is the scale parameter of the distribution.

One useful representation of the risk measure ρ for this case study will be the following

$$\rho(X_e) = \int_0^\infty \psi(\mathbb{P}[X_e > u]) du. \quad (4.2)$$

First, note that the Weibull distribution is decreasing by FSD, as for $e_1 < e_2$,

$$pe^{-\lambda u^{\tau(e_1)}} > pe^{-\lambda u^{\tau(e_2)}},$$

and hence $\mathbb{P}[X_{e_1} > u] > \mathbb{P}[X_{e_2} > u]$. Using the representation for any function g such that $g(0) = 0$

$$\mathbb{E}[g(X_e)] = \int_0^\infty g'(u) \mathbb{P}[X_e > u] du,$$

we have

$$\mathbb{E}[g(X_{e_1})] \geq \mathbb{E}[g(X_{e_2})].$$

From Equations (4.1) and (4.2), the risk measure ρ is

$$\begin{aligned} \rho(X_e) &= \int_0^\infty \exp\left(-\left(-\log\left(p \exp\left(-\lambda u^{\tau(e)}\right)\right)\right)^\eta\right) du, \\ &= \int_0^\infty \exp\left(-\left(-\log(p) + \lambda u^{\tau(e)}\right)^\eta\right) du, \end{aligned}$$

by taking $v := -\log(p) + \lambda u^{\tau(e)}$, $du = \frac{(v+\log(p))^{1/\tau(e)-1}}{\tau(e)\lambda^{1/\tau(e)}} dv$

$$\rho(X_e) = \frac{1}{\tau(e)\lambda^{1/\tau(e)}} \int_{-\log(p)}^\infty (v + \log(p))^{1/\tau(e)-1} \exp(-v^\eta) dv. \quad (4.3)$$

Moreover the expectation of X_e is given by

$$\mathbb{E}[X_e] = \frac{p}{\lambda} \Gamma\left(1 + \frac{1}{\tau(e)}\right), \quad (4.4)$$

where Γ denotes the Gamma function defined as $\Gamma(u) = \int_0^\infty t^{u-1} \exp(-t) dt$ for $u > 0$.

The function G can be calculated from Equations (3.4), (4.3) and (4.4):

$$\begin{aligned} G(e) &= \frac{p\lambda^{1-1/\tau(e)}}{\tau(e)\Gamma\left(1 + \frac{1}{\tau(e)}\right)} \int_{-\log(p)}^\infty (v + \log(p))^{1/\tau(e)-1} \exp(-v^\eta) dv, \\ &= \frac{p\lambda^{1-1/\tau(e)}}{\Gamma\left(\frac{1}{\tau(e)}\right)} \int_{-\log(p)}^\infty (v + \log(p))^{1/\tau(e)-1} \exp(-v^\eta) dv. \end{aligned} \quad (4.5)$$

In Equation (4.5), the integral is not easy to calculate, so we cannot further describe the function G without specifying some parameters. In the following, four graphical possibilities are presented with particular parameter choices, to provide an idea of what may result.

First consider small claims, where the shape parameter $\tau(e)$ is larger than 1. For the set of parameters $p = 0.06$, $\eta = 0.4$ and $\lambda = 10$, and taking $\tau(e) = \sqrt{e + \frac{1}{100}}$, the function G is graphically non-increasing convex at first then non-decreasing concave as show in Figure 2.4. It can be seen that the function G is rapidly decreasing then increases at a slower pace in a concave way.

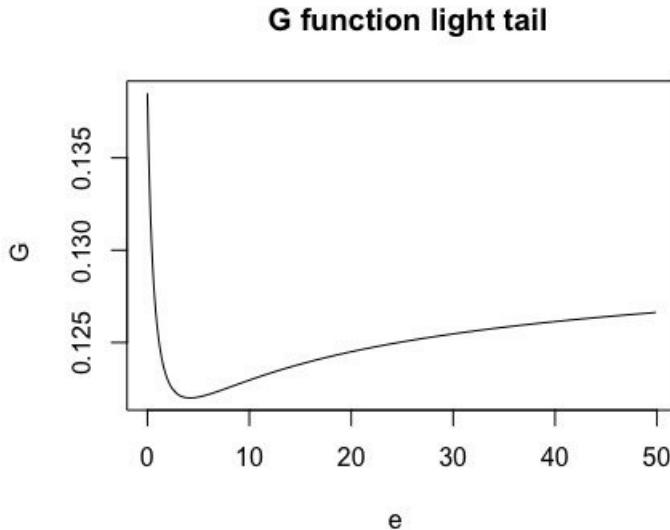


Figure 2.4 – Illustration of G for small claims with $p = 0.06$, $\eta = 0.4$, $\lambda = 10$ and $\tau(e) = \sqrt{e + \frac{1}{100}}$.

Figure 2.5, shows the function G for the set of parameters $p = 0.08$, $\eta = 0.2$, $\lambda = 150$ and for a shape parameter function $\tau(e) = \sqrt{e}$. G decreases almost instantly and then increases only slowly.

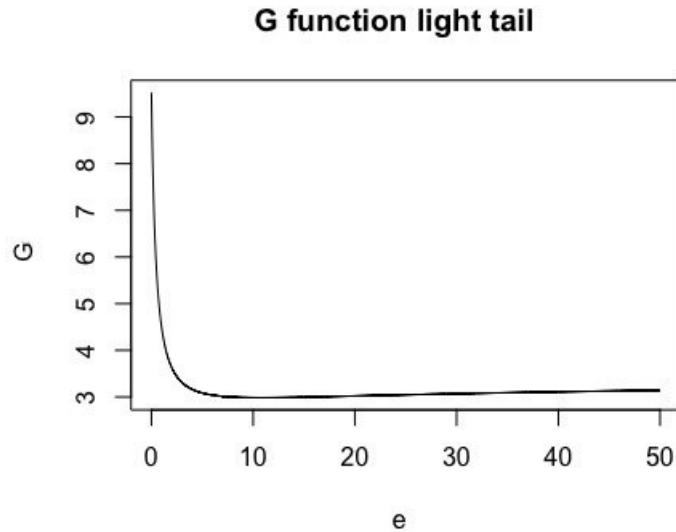


Figure 2.5 – Illustration of G for small claims with $p = 0.08$, $\eta = 0.2$, $\lambda = 150$ and $\tau(e) = \sqrt{e}$.

Second, consider large claims, where the shape parameter $\tau(e)$ is such that $1 > \tau(e) > 0$. By taking $p = 0.05$, $\eta = 0.5$, $\lambda = 8$ and $\tau(e) = \sqrt{e}$, the G function is non-decreasing and convex, as shown in Figure 2.6.

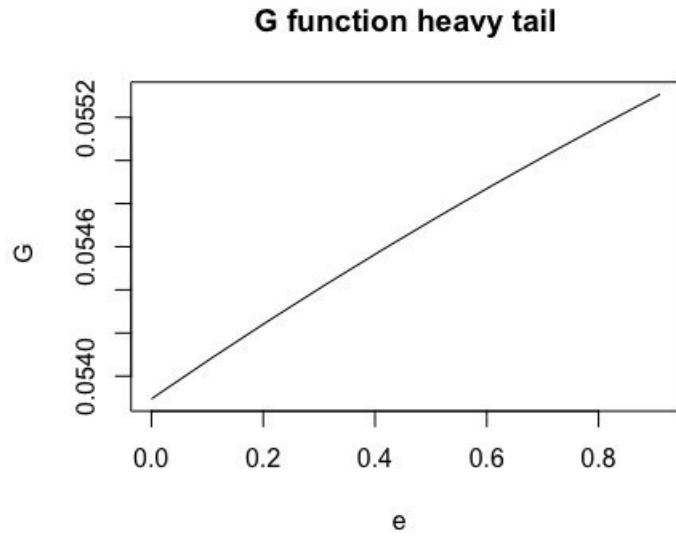


Figure 2.6 – Illustration of G for small claims with $p = 0.05$, $\eta = 0.4$, $\lambda = 8$ and $\tau(e) = \sqrt{e}$.

By playing with the parameters used in the calculation of G , different results can be obtained for small- and large-claims insurance. The function G can be non-decreasing or not monotonic,

as assumed in Section 3.2.

With these specified parameters, it is also possible to calculate $e_I(\theta)$ and e_N . The value of $e_I(\theta)$ is given by Proposition 3.3 as the minimizer of L_T^θ . From Equation (4.4), and assuming that the cost function is given by $c(e) := \frac{e^2}{2}$, $e_I(\theta)$ is the solution of the following equation

$$-\frac{(1+\theta)p\tau'(e)}{\tau(e)^2}\partial_e\Gamma\left(1+\frac{1}{\tau(e)}\right)+e=0.$$

The value of e_N is given by Proposition 3.5, which is numerically solvable. Figure 2.7 shows the evolution of $e_I(\theta)$ and e_N , as functions of θ , for the set of parameters $p = 0.06$, $\eta = 0.4$ and $\lambda = 10$, and taking $\tau(e) = \sqrt{e + \frac{1}{100}}$. The distortion used for this illustration is given by $\psi_1(u) = \exp(-(-\log(u))^\eta)$. The effort $e_I(\theta)$, as a function of θ , evolves in a concave way, and e_N is independent of θ and equal to 1. This can be interpreted as the IB making less effort to reduce risk as the insurance price rises, while being insured.

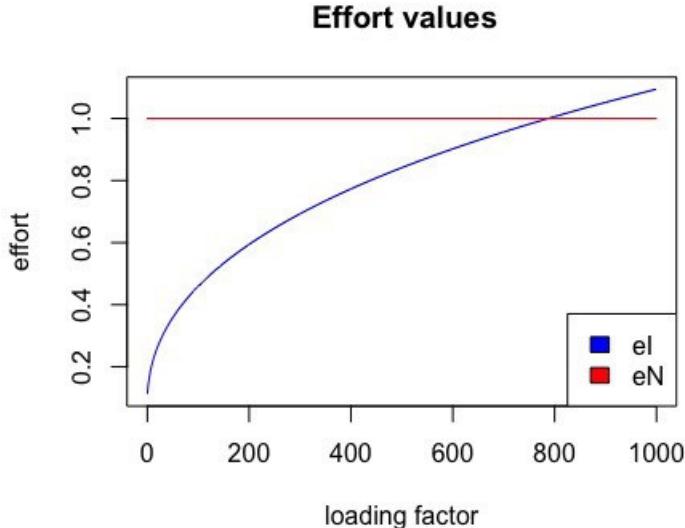


Figure 2.7 – Illustration of the values of $e_I(\theta)$ and e_N for small claims with $p = 0.06$, $\eta = 0.4$, $\lambda = 10$ and $\tau(e) = \sqrt{e + \frac{1}{100}}$.

$L_T^\theta(e_I(\theta))$ and $L_N(e_N)$ can also be compared graphically, to find the maximum value of the insurance price acceptable to the IB, θ_M . Figure 2.8 illustrates the changes with respect to θ of the minimum value of the objective functions L_T^θ and L_N , with the set of parameters $p = 0.06$, $\eta = 0.4$ and $\lambda = 10$, taking $\tau(e) = \sqrt{e + \frac{1}{100}}$, and with the distortion function $\psi_1(u) = \exp(-(-\log(u))^\eta)$.

As $L_N(e_N)$ is the objective function of the IB while being uninsured, it is independent of θ and equals 2.31. $L_T^\theta(e_I(\theta))$ is increasing in θ and equals $L_N(e_N)$ at θ_M , which latter is estimated

to be 313.

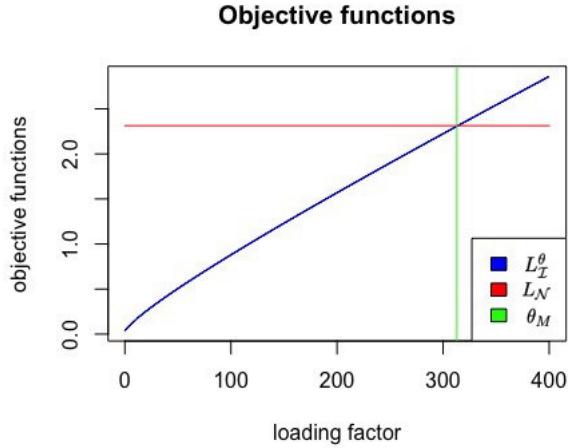


Figure 2.8 – Illustration of the values of $e_I(\theta)$ and e_N for small claims with $p = 0.06$, $\eta = 0.4$, $\lambda = 10$ and $\tau(e) = \sqrt{e + \frac{1}{100}}$.

5 Conclusion

This paper has modeled the standard problem of self-insurance as an optimization problem for an insurance buyer. The criterion of the insurance buyer comes from a distortion risk measure, where the distortion is non-concave. Non-concave distortion functions yield new conclusions for market insurance and self-insurance: these are either substitutes or complements depending on the elasticity of the price of insurance. Either the insurance buyer enters the insurance contract, and by doing so takes out full insurance, or does not. In either case the IB chooses the corresponding optimal effort.

An important limitation of this study is that the initial wealth of the agent is not taken into account in the presented model. Indeed, as the considered risk measure is cash-additive, the initial wealth, being a constant, can be assumed equal to zero. As underlined in [8] for health insurance and [29] for self-insurance, initial wealth can influence insurance choices. A possible extension of this model is to consider cash-subadditive risk measures. This aspect is left for future research.

Bibliographie

- [1] ABDELLAOUI, M., L'HARIDON, O., AND ZANK, H. Separating curvature and elevation : A parametric probability weighting function. *Journal of Risk and Uncertainty* 41, 1 (2010), 39–65.
- [2] ALBRECHER, H., TEUGELS, J., AND BEIRLANT, J. *Reinsurance : Actuarial and Statistical Aspects*. Statistics in Practice. Wiley, 2017.
- [3] BENSALEM, S., SANTIBÁÑEZ, N. H., AND KAZI-TANI, N. Prevention efforts, insurance demand and price incentives under coherent risk measures. *Insurance : Mathematics and Economics* 93 (2020), 369 – 386.
- [4] BERTSIMAS, D., AND BROWN, D. B. Constructing uncertainty sets for robust linear optimization. *Oper. Res.* 57, 6 (Nov. 2009), 1483–1495.
- [5] BOWERS, N., GERBER, H., HICKMAN, J., JONES, D., AND NESBITT, C. *Actuarial Mathematics*. Society of Actuaries, 1997.
- [6] BRIYS, E., SCHLESINGER, H., AND V. D. SCHULENBURG, J.-M. G. Reliability of Risk Management : Market Insurance, Self-Insurance and Self-Protection Reconsidered. *The Geneva Risk and Insurance Review* 16, 1 (June 1991), 45–58.
- [7] CAWLEY, J., AND PHILIPSON, T. An Empirical Examination of Information Barriers to Trade in Insurance. NBER Working Papers 5669, National Bureau of Economic Research, Inc, July 1996.
- [8] CAWLEY, J., AND PHILIPSON, T. An empirical examination of information barriers to trade in insurance. *American Economic Review* 89, 4 (1999), 827–846.
- [9] COURBAGE, C. Self-insurance, self-protection and market insurance within the dual theory of choice. *The Geneva Risk and Insurance Review* 26, 1 (2001), 43–56.
- [10] COURBAGE, C., AND DE COULON, A. Prevention and private health insurance in the u.k. *The Geneva Papers on Risk and Insurance. Issues and Practice* 29, 4 (2004), 719–727.
- [11] COURBAGE, C., LOUBERGÉ, H., AND PETER, R. Optimal prevention for multiple risks. *Journal of Risk and Insurance* 84, 3 (2017), 899–922.

- [12] CVITANIĆ, J., AND ZHANG, J. *Contract theory in continuous-time models*. Springer, 2012.
- [13] DE MEZA, D., AND WEBB, D. C. Advantageous selection in insurance markets. *The RAND Journal of Economics* 32, 2 (2001), 249–262.
- [14] DEELSTRA, G., AND PLANTIN, G. Risk theory and reinsurance. Sciences po publications, Sciences Po, 2014.
- [15] DIONNE, G., AND EECKHOUDT, L. Self-insurance, self-protection and increased risk aversion. *Economics Letters* 17, 1-2 (1985), 39–42.
- [16] DOHERTY, N. A., AND EECKHOUDT, L. Optimal insurance without expected utility : The dual theory and the linearity of insurance contracts. *Journal of Risk and Uncertainty* 10, 2 (1995), 157–79.
- [17] EHRLICH, I., AND BECKER, G. S. Market insurance, self-insurance, and self-protection. *Journal of political Economy* 80, 4 (1972), 623–648.
- [18] EMBRECHTS, P., AND HOFERT, M. Risk measures and dependence modeling. In *Handbook of Insurance*, G. Dionne, Ed. Huebner International Series on Risk, Insurance, and Economic Security (New York : Springer), New York, 2013, ch. 6, pp. 135–166.
- [19] ETNER, J., AND JELEVA, M. Risk perception, prevention and diagnostic tests. *Health Economics* 22, 2 (2013), 144–156.
- [20] FÖLLMER, H., AND SCHIED, A. *Stochastic Finance : An Introduction in Discrete Time*. De Gruyter Textbook. De Gruyter, 2016.
- [21] GONZALEZ, R., AND WU, G. On the shape of the probability weighting function. *Cognitive Psychology* 38, 1 (1999), 129–166.
- [22] HEMENWAY, D. Propitious selection. *The Quarterly Journal of Economics* 105, 4 (1990), 1063–1069.
- [23] KAHNEMAN, D. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.
- [24] KAHNEMAN, D., AND TVERSKY, A. Subjective probability : A judgment of representativeness. *Cognitive Psychology* 3 (1972), 430 – 454.
- [25] KAHNEMAN, D., AND TVERSKY, A. Prospect theory : An analysis of decision under risk. *Econometrica* 47, 2 (1979), 263–91.
- [26] KENKEL, D. Prevention. In *Handbook of Health Economics*, A. J. Culyer and J. P. Newhouse, Eds. Elsevier, 2000, ch. 31, pp. 1675–1720.
- [27] KONRAD, K., AND SKAPERDAS, S. Self-insurance and self-protection : A nonexpected utility analysis. *The Geneva Risk and Insurance Review* 18, 2 (1993), 131–146.

- [28] LAEVEN, R. J., AND GOOVAERTS, M. J. *Premium Calculation and Insurance Pricing*. American Cancer Society, 2014.
- [29] LEE, K. Wealth effects on self-insurance. *The Geneva Risk and Insurance Review* 35, 2 (Dec 2010), 160–171.
- [30] LESUEUR, J.-Y. Health insurance coverage vs self-prevention : What impact of the insurer’s healthcare prevention programs ? working paper or preprint, Feb. 2018.
- [31] PANNEQUIN, F., CORCOS, A., AND MONTMARQUETTE, C. Behavioral foundations of the substitutability between insurance and self-insurance : An experimental study. CIRANO Working Papers 2016s-12, CIRANO, Mar. 2016.
- [32] PRELEC, D. The probability weighting function. *Econometrica* 66, 3 (1998), 497–528.
- [33] QUIGGIN, J. A theory of anticipated utility. *Journal of Economic Behavior & Organization* 3, 4 (December 1982), 323–343.
- [34] ROGERS, P. The cognitive psychology of lottery gambling : A theoretical review. *Journal of Gambling Studies* 14 (1998), 111–134.
- [35] TVERSKY, A., AND KAHNEMAN, D. Advances in prospect theory : Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 4 (Oct 1992), 297–323.
- [36] WAKKER, P. P. *Prospect Theory : For Risk and Ambiguity*. Cambridge University Press, 2010.
- [37] WINTER, R. A. *Optimal Insurance under Moral Hazard*. Springer Netherlands, Dordrecht, 2000, pp. 155–183.
- [38] WU, G., AND GONZALEZ, R. Curvature of the probability weighting function. *Management Science* 42, 12 (1996), 1676–1690.
- [39] WU, G., AND GONZALEZ, R. Nonlinear decision weights in choice under uncertainty. *Management Science* 45, 1 (1999), 74–85.
- [40] YAARI, M. E. The dual theory of choice under risk. *Econometrica* 55, 1 (1987), 95–115.

Chapitre 3

A Continuous-Time Model of Self-Protection

Ce chapitre reprend un article co-écrit avec Nicolás Hernández Santibáñez et Nabil Kazi-Tani.

Abstract

This paper deals with an optimal linear insurance demand model, where the protection buyer can also exert time-dynamic costly prevention effort to reduce her risk exposure. This is expressed as a stochastic control problem, in which the agent maximizes an exponential utility of her terminal wealth. We assume that the effort reduces the intensity of the jump arrival process, and we interpret this as dynamic self-protection. We solve the problem using a dynamic programming principle approach, and we provide a representation of the value function using a particular backward SDE. This allows us to solve the problem explicitly: we identify the dynamic certainty equivalent of the agent, and prove that the dynamic effort is actually constant, for a large class of loss processes. This shows in particular that the Lévy property is preserved under exponential utility maximization. We also characterize the constant effort as the unique minimizer of an explicit Hamiltonian, from which we can determine the optimal effort in particular cases. Finally, after studying the dependence of the BSDE on the linear insurance contract parameter, we prove the existence of an optimal linear cover, that is not necessarily zero or full insurance.

Key words: Self-Protection, Prevention effort, Dynamic programming, Continuation utility, Backward SDEs

1 Introduction

Self-protection activities, also called *primary prevention* measures, are actions taken by an agent facing a loss risk, to reduce the probability that a loss happens, but not the severity of the loss. The losses can be monetary or not, so examples include actions taken by individuals to enhance their health condition, i.e. to reduce the probability of getting an illness. The

situation where the size of the loss is impacted by the effort is referred to as self-insurance, or *secondary prevention*.

In the seminal paper [10], the arbitrage between market insurance and prevention measures is studied, in both cases of self-insurance and self-protection. In this paper, we analyze a self-protection model, in a time-dynamic framework, for a general class of losses distribution, containing in particular all compound Poisson processes, which are close to actuarial models and practice. The processes we consider are described by a counting process N , modeling the claim arrivals and a one-dimensional probability distribution, modeling the cost, or severity of each claim. We formulate and solve an optimization problem in which an agent chooses both the purchased quantity of market insurance and the exerted prevention effort by maximizing an expected utility. Since we focus here on self-protection, we assume that only the counting process is impacted by the effort. Our main result is that the optimal effort is constant, i.e. it is fixed by the agent at time 0, and then it does not change with time or with loss occurrences. This means that an exponential utility maximizer has never any incentive to modify her effort dynamically. We formulate the control problem in a weak sense: the agent only controls the probability distribution of the loss process, and her effort is not observable by the insurance seller. Our main result says that the agent, who can choose among a large class of semimartingales distributions, should maximize her utility only in the subclass of Lévy processes, hence making the loss process homogeneous and Markovian.

Let us describe more precisely the modeling approach that we retain.

Summary of the approach

A Stochastic Control Problem. The utility maximization problem that we consider is formulated as a stochastic control problem, in which the agent, through dynamic prevention effort, controls the jump intensity of the claim counting process. In a continuous-time setting, the distinction between self-protection and self-insurance is made clear by the decomposition of the compensator of the considered jump measures as a claim counting part and a claim cost part. More precisely, for a pure jump càdlàg process X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let μ_X be its associated random jump measure given by $\mu_X(\omega, [0, t], A) = \sum_{s \leq t} \mathbf{1}_{\{\Delta X_s \in A\}}$, for all $A \in \mathcal{B}(\mathbb{R})$, where $\Delta X_s := X_s - X_{s-}$. Provided it exists, let $\nu(\omega, dt, dx)$ be the compensator of μ_X . We assume in all the paper that ν is absolutely continuous with respect to the Lebesgue measure and takes the following form

$$\nu(\omega, dt, dx) = \lambda_t(\omega) dG(x) dt,$$

where G is a given cumulative distribution function that represents the cost of the claims. We emphasize that we allow the compensator of the jump measure to be random, since in our control setting, λ_t will be impacted by the effort process, which will naturally depend on the driving randomness. Hence if the effort only impacts λ_t , it is a self-protection framework,

whereas if the effort only impacts the distribution G , it is a self-insurance setting. In this paper, we focus on the control of the jump intensity $\lambda_t(\omega)$.

The controller optimally determines her linear insurance contract parameter and her dynamic effort process in order to maximize an expected exponential utility of her wealth at a terminal time $T > 0$.

Dynamic Programming Principle. In order to solve the formulated stochastic control problem, we first prove that it satisfies a dynamic programming principle (DPP). It says that the problem of determining the optimal effort between times 0 and T can be simplified into determining the optimal effort first in $[0, \tau]$, secondly in $[\tau, T]$, and then pasting the solutions together, where τ is a stopping time for the natural filtration of the loss process. Because of the exponential utility framework, the DPP takes here a multiplicative form (see Lemma 1.1) and this allows us to use martingale techniques to characterize the value function of the problem. In particular, using the martingale representation property, we are able to link explicitly the value function with the solution (Y, H) of a well-posed backward stochastic differential equation (BSDE). The component Y is interpreted as a monetary certainty equivalent for the agent, while the component H gives the level of decrease of dynamic utility when a claim arrives. This representation also allows to characterize the optimal effort at each time $t \leq T$ as the minimizer of an explicit Hamiltonian. The BSDE takes a particular homogeneous form: the terminal condition is null, while the generator only depends on H . This implies that $H \equiv 0$ and Y is continuous, which means in turn that the optimal effort is constant in time and in ω .

The effort process of the protection buyer is unobservable by the insurance seller: our model thus includes moral hazard. Indeed, only the wealth process distribution is impacted by the effort and not the wealth values directly, since we consider a weak formulation of the control problem. The wealth distributions that we consider, seen as probability measures on the space of càdlàg trajectories from $[0, T]$ to \mathbb{R} , are all equivalent. In particular, all these measures have the same negligible sets and this considerably simplifies the analysis. The initial distribution of the loss process, when there is no effort, is given by a particular Lévy process: a compound Poisson process. The fact that the optimal effort is constant says that the Lévy property is preserved under exponential utility maximization, which is the main result of this paper.

Related literature

The first systematic study of prevention effort, in the form of self-protection or self-insurance, was made in the economic literature in the seminal paper of Ehrlich and Becker [10]. Their main result states that market insurance and self-insurance are substitutes, while surprisingly, market insurance and self-protection can be complements. These results triggered a large literature about prevention in insurance economics, in which the robustness of Ehrlich and

Becker's results are tested. We refer the reader to [14], [9] (chapter 9) and to the survey paper [6] for an account of the large body of results around prevention. In the case of observable effort, prevention is well understood, both for self-protection and self-insurance. However, in the case of moral hazard, little is known on optimal choices under these two forms of risk reduction. [17] and [23] both determine the form of the optimal insurance contract, under expected utility, respectively in case of self-protection, where a deductible is optimal, and in case of self-insurance, where this time low losses are fully covered while large losses are partially covered. When the criteria of both the protection buyer and seller are given by coherent risk measures, [2] shows under moral hazard, that if the buyer's risk measure decreases faster in effort than his expected loss, optimal effort is non-decreasing in the insurance price with a potential discontinuity when optimal coverage switches from full to zero. On the contrary, if the decrease of the buyer's risk measure is slower than the expected loss, optimal effort may or may not be non-decreasing with the insurance price. Our results differ from these papers in various aspects. First, we fix the class of possible contracts, to consider only linear insurance covers. Note that this is not restrictive, since many contracts in practice take this form, in particular in health insurance. These contracts are also standard in a reinsurance context, under the name of quota-share. Secondly, our most important contribution is that we consider a continuous-time version of the problem of self-protection, using the full power and generality of semimartingale processes.

Risk reduction effort, in a time dynamic context is known as *precaution* in the economic literature (see [6]). The main issue there is how the knowledge of arrival of future "better" information impacts decisions today [15], [11]. The definition of information taken in this literature dates back to [4] and is based on the notion of convex order between one-dimensional probability distributions. Our model is again very different as it relies on a filtration generated by a jump process to define a notion of information. The papers mentioned above deal with the precaution problem in discrete time, in a one or two period models. In a continuous-time setting, [3], [5], [20] and [16] all consider costly effort to control the jump frequency of a stochastic process in a Principal-Agent context. The main difference between a Principal-Agent problem and our framework is the payment time: in optimal contract theory, a contract value is paid by the principal to the agent at the end of the period (or when the Principal decides to end the contract), while on the contrary, an insurance premium is paid at the start of the cover. If the insurance cover itself is interpreted as the contract, then again the timing of the contract payments is not controlled by the principal, but is determined by the random claim arrival times. Still in a Principal-Agent framework, [13] determines the optimal compensation scheme of a financial market maker, as well as the optimal quotes that he should display. The control problem introduced in [13] consists among other things in controlling the arrival intensity of a jump process, so it has strong methodological similarities with the control problem under study in our paper. Finally, let us mention that the problem of entropy minimization, under a martingale constraint is solved in [19] in dimension 1, and in [12] and [1]

in dimension $d \geq 1$. All these papers show, using different arguments, which are also different from the arguments in this paper, that entropy minimization preserves the Lévy property: in particular, the optimal jump measure compensator is constant. This is reminiscent of our result, where the Lévy property is preserved under exponential utility maximization.

Structure of the paper

Section 2 discusses the stochastic control problem at hand and our main assumptions. In Section 3.1, we determine the optimal effort of the agent and show that it is constant, by solving an associated BSDE. Section 3.2 is dedicated to the optimal insurance cover: we prove that it exists by studying the dependence of the BSDE in the linear insurance contract. Section 4 illustrates the proposed methodology in a particular case.

2 Model and Assumptions

Let $(\Omega, \mathcal{F}, \mathbb{P}^0)$ be a fixed probability space and N be a Poisson process of fixed intensity $\lambda_0 > 0$ under \mathbb{P}^0 . Consider a risk averse economic agent with initial wealth ω_0 , who is facing potential losses, whose distribution is described below, in the form of a Compound Poisson process. The agent has the possibility to enter a proportional insurance contract for a period of time $[0, T]$, where she chooses her level of insurance denoted $\alpha \in [0, 1]$, for which she pays a premium $\Pi(\alpha)$. The insurance coverage is assumed to be fixed over $[0, T]$ and the premium $\Pi(\alpha)$ is paid at time 0. In addition to the contract, she may undertake time-dynamic self-protection actions to reduce the occurrence of the losses. The prevention effort is assumed to be a predictable process e with values in $[0, \infty)$. Prevention efforts come with a monetary cost function c , that is assumed to be increasing, twice continuously differentiable and strictly convex¹. We also define Σ , the set of prevention efforts verifying $\mathbb{E} \left[e^{\gamma p \int_0^T c(e_s) ds} \right]^{1/p} < \infty$ for some $p > 1$, such processes are said admissible. Moreover, we assume that the agent determines her level of effort and of purchased market insurance by maximizing her expected utility at terminal time T .

Wealth process dynamics. Let X^e denote the wealth process of the agent, after buying proportional insurance, which is given by

$$X_t^e = \omega_0 - \Pi(\alpha) - \int_0^t c(e_s) ds - (1 - \alpha) J_t, \quad \text{for } 0 \leq t \leq T, \quad (2.1)$$

where J_t is a compound Poisson process, written as

$$J_t = \sum_{i=1}^{N_t} Z_i, \quad 0 \leq t \leq T,$$

1. Notice that since c is non-decreasing and strictly convex, we have $\lim_{e \rightarrow \infty} c(e) = \infty$.

where $(Z_i)_{i \geq 1}$ is a sequence of non-negative independent and identically distributed random variables with cumulative distribution function denoted G and independent of N . Recall that under \mathbb{P}^0 , N is a Poisson process with intensity λ_0 . At a jump time of J , we say that an accident occurred and it affects negatively the wealth process X^e . The random variable Z_i quantifies the cost of the i -th accident, and under the proportional insurance contract, the agent only pays the fraction $(1 - \alpha)Z_i$ of each accident's cost. We denote by $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ the \mathbb{P} -completion of the filtration generated by the process J .

Let μ_J be the jump measure associated to J , so that J can be written as

$$J_t = \int_0^t \int_0^\infty x \mu_J(ds, dx).$$

Furthermore, we assume that the jump sizes are bounded by a constant $C \in \mathbb{R}$. As a consequence, the support of the jump size distribution is contained in $[0, M_J]$ for some $M_J > 0$. Note that Equation (2.1) can be rewritten using the random measure μ_J as

$$X_t^e = \omega_0 - \Pi(\alpha) - \int_0^t c(e_s) ds - (1 - \alpha) \int_0^t \int_0^\infty x \mu_J(ds, dx), \quad 0 \leq t \leq T.$$

Effort. The agent undertakes self-protection activities, which means that her actions reduce the probability of a claim arrival, but not the severity of the claims: this is included in our model in the fact that the insurance buyer controls the intensity $e \mapsto \lambda(e)$ of the compound Poisson process. We assume that the function λ is non-increasing, and satisfies $\lambda(0) =: \lambda_0 > 0$ and $\lim_{e \rightarrow \infty} \lambda(e) =: \lambda_\infty > 0$. For a given effort process $(e_t)_{0 \leq t \leq T}$, we introduce the probability measure \mathbb{P}^e , defined by the following change of measure

$$D_T^e := \frac{d\mathbb{P}^e}{d\mathbb{P}^0} = e^{-\int_0^T (\lambda(e_s) - \lambda_0) ds + \int_0^T (\log(\lambda(e_s)) - \log(\lambda_0)) dN_s}, \quad (2.2)$$

so that D^e is an \mathbb{F} -adapted martingale. The compensated measure associated to μ_J under \mathbb{P}^e , denoted $\tilde{\mu}_J^e$, is defined by

$$\tilde{\mu}_J^e(dt, dx) = \mu_J(dt, dx) - \lambda(e_t) dt dG(x).$$

For $\beta \geq 0$, we define the following spaces, that will be useful in Theorem 3.1, to characterize existence and uniqueness of a solution to the BSDE related to the optimization problem of the insurance buyer:

$$\begin{aligned} \mathcal{S}_\beta^2 &:= \left\{ \phi \text{ is an } \mathbb{R}\text{-valued } \mathcal{O}\text{-semimartingale with càdlàg paths and } \|\phi\|_{\mathcal{S}_\beta^2}^2 := \mathbb{E} \left[\sup_{t \in [0, T]} e^{\beta t} |\phi_t|^2 \right] < \infty \right\}, \\ \mathbb{H}_\beta^{2,\natural} &:= \left\{ H : (\Omega \times [0, T] \times \mathbb{R}, \mathcal{P} \otimes \mathcal{B}(\mathbb{R})) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})), \quad \|H\|_{\beta}^{2,\natural} := \mathbb{E} \left[\int_0^T e^{\beta t} d \langle \int_0^t \int_0^\infty H_s(x) \tilde{\mu}_J^0(ds, dx) \rangle_t \right] < \infty \right\}, \\ \mathbb{H}_\beta^2 &:= \left\{ \phi \text{ is an } \mathbb{R}\text{-valued } \mathcal{O}\text{-semimartingale with càdlàg paths and } \|\phi\|_{\mathbb{H}_\beta^2}^2 := \mathbb{E} \left[\int_0^T e^{\beta t} |\phi_t|^2 dt \right] < \infty \right\}, \end{aligned}$$

where \mathcal{O} denotes the optional sigma-algebra and \mathcal{P} denotes the predictable sigma-algebra on $\Omega \times [0, T]$. We also define the two following spaces for $dt \otimes d\mathbb{P}$ -almost every $(t, \omega) \in \mathbb{R}_+ \times \Omega$

$$\begin{aligned}\mathfrak{H}_{t,\omega} &:= \overline{\{H : (\mathbb{R}, \mathcal{B}(R)) \rightarrow (\mathbb{R}, \mathcal{B}(R)), |||H(\cdot)|||_t(\omega) < \infty\}}, \\ \mathfrak{H} &:= \{H : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}, H_t(\omega, \cdot) \in \mathfrak{H}_{t,\omega} \text{ for } dt \otimes d\mathbb{P}\text{-almost every } (t, \omega) \in \mathbb{R}_+ \times \Omega\},\end{aligned}$$

where $|||H_t(\omega, \cdot)|||_t(\omega) := \int_{\mathbb{R}} H_t(\omega, x) \lambda_0 dG(x)$, $0 \leq t \leq T$. From [21], Lemma 2.14, the space $(\mathfrak{H}_{t,\omega}, |||\cdot|||_t(\omega))$ is Polish for $dt \otimes d\mathbb{P}$ -almost every $(t, \omega) \in \mathbb{R}_+ \times \Omega$.

The optimization problem. Let u be the following exponential utility function

$$u(x) = -\frac{1}{\gamma} \exp(-\gamma x), \quad x \in \mathbb{R},$$

where $\gamma > 0$ is a constant relative risk aversion coefficient. We assume that the agent maximizes her expected utility at terminal time T , $\mathbb{E}^e[u(X_T^e)]$, where \mathbb{E}^e denotes the expectation with respect to the measure \mathbb{P}^e . As ω_0 is a constant independent of α and e , the utility maximization problem of the insurance buyer reduces to

$$V := \sup_{(\alpha, e) \in [0, 1] \times \Sigma} \mathbb{E}^e \left[-\frac{1}{\gamma} \exp \left(-\gamma \left(-\Pi(\alpha) - (1 - \alpha) \int_0^T \int_0^\infty x \mu_J(ds, dx) - \int_0^T c(e_s) ds \right) \right) \right] \quad (2.3)$$

where V denotes the value function of the agent. We can write V as

$$V = \sup_{\alpha \in [0, 1]} e^{\gamma \Pi(\alpha)} V^\alpha, \quad (2.4)$$

where

$$V^\alpha := \sup_{e \in \Sigma} \mathbb{E}^e \left[-\frac{1}{\gamma} e^{-\gamma \left(-(1 - \alpha) \int_0^T \int_0^\infty x \mu_J(ds, dx) - \int_0^T c(e_s) ds \right)} \right], \quad (2.5)$$

i.e. we first solve the stochastic control problem for fixed α , and then we maximize utility over the real variable α . In the next section, we obtain a representation of V^α using a BSDE: so in order to solve the problem in α , we have to understand the dependence of the BSDE solution on α , which is the purpose of Section 3.2.

Remark 2.1. *We can assume in our problem that the agent can reject the insurance contract if her value is not high enough, that is the agent will take the cover only if $V \geq R_0$ for some reservation value $R_0 \geq 0$. In this case rejection of the contract simply means choosing $\alpha = 0$.*

3 Solving the insurance problem

3.1 Optimal prevention effort

In this section, we fix $\alpha \in [0, 1]$ and we find the value of V^α . To do so, we follow the approach of [22] and [7] and introduce the dynamic continuation utility process of the agent. We write the

problem of self-protection as a standard control problem that we solve by using the dynamic programming principle and a particular backward stochastic differential equation.

The next theorem is the main result of this paper: by solving explicitly BSDE (3.3) below, it identifies the agent's value and the optimal constant effort she needs to make in order to maximize her utility.

Theorem 3.1. *Let e^* be the unique minimizer of the following one-dimensional problem*

$$\inf_{e \in [0, \infty)} \left\{ c(e) + \frac{1}{\gamma} \lambda(e) [L_G(\gamma(1 - \alpha)) - 1] \right\},$$

where $L_G(y) := \int_0^\infty e^{yx} dG(x)$ denotes the Laplace transform of G . Then the constant control $e_t(\omega) := e^*$ is the unique optimal control of problem (2.5). Moreover, the value function admits the following explicit expression

$$V^\alpha = -\frac{1}{\gamma} \exp \left(\gamma T \left\{ c(e^*) + \frac{1}{\gamma} \lambda(e^*) [L_G(\gamma(1 - \alpha)) - 1] \right\} \right).$$

The proof of Theorem 3.1 relies on the following result, that links the value function (2.5) with a particular homogeneous, explicitly solvable BSDE. First let us define the BSDE. For $e \in [0, \infty)$ and $h \in \mathfrak{H}$, we define the function $f^\alpha : [0, \infty) \times \mathfrak{H} \rightarrow \mathbb{R}$ by

$$f^\alpha(e, h) := c(e) + \frac{\lambda(e)}{\gamma} \int_0^\infty (\exp [\gamma(h(x) + (1 - \alpha)x)] - 1) dG(x) - \int_0^\infty \lambda_0 h(x) dG(x), \quad (3.1)$$

and

$$F^\alpha(h) := \inf_{e \in [0, \infty)} f^\alpha(e, h). \quad (3.2)$$

Consider the following BSDE with jumps

$$\begin{aligned} Y_t^\alpha &= \int_t^T \int_0^\infty H_s^\alpha(x) \tilde{\mu}_J^0(ds, dx) - \int_t^T F^\alpha(H_s^\alpha) ds, \quad 0 \leq t \leq T, \\ Y_T^\alpha &= 0. \end{aligned} \quad (3.3)$$

The next proposition shows that solving the problem of the agent is equivalent to solving this BSDE. Its proof is postponed to Appendix 2.

Proposition 3.1. (a) *There exists a unique solution $(Y^\alpha, H^\alpha) \in \mathcal{S}_\beta^2 \times \mathbb{H}_\beta^{2,\natural}$ to the BSDE (3.3).*

(b) *The optimal prevention effort for the utility maximisation problem (2.5) is given by*

$$e_t^* = e^*(H_t^\alpha),$$

where $e^*(h)$ denotes the minimizer in (3.2). The value function of the agent is given by

$$V^\alpha = -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha}. \quad (3.4)$$

Proof. [of Theorem 3.1]

The unique solution of the BSDE (3.3) is given by (Y^α, H^α) such that

$$\begin{cases} Y_t^\alpha = - \left(c(e^*) + \frac{\lambda(e^*)}{\gamma} (L_G(\gamma(1-\alpha)) - 1) \right) (T-t), & 0 \leq t \leq T, \\ H_t^\alpha \equiv 0, & 0 \leq t \leq T, \end{cases}$$

where e^* is the minimizer in (3.2), that is $F^\alpha(0) = f^\alpha(e^*, 0)$. Indeed, note that the optimal response to $H^\alpha \equiv 0$ is finite and constant, therefore we have the admissibility condition

$$\int_0^T (|H_s^\alpha|^2 + |F^\alpha(H_s^\alpha)|) ds = \int_0^T \left| c(e^*) + \frac{\lambda(e^*)}{\gamma} (L_G(\gamma(1-\alpha)) - 1) \right| ds < \infty.$$

Since $Y_T^\alpha = 0$, the pair (Y^α, H^α) is a solution to the BSDE. We conclude from the uniqueness result. \square

The value Y_t^α can be interpreted as the deterministic amount of cash that the agent would pay at time t rather than taking a chance on an uncertain higher loss after time t . This value Y_t^α is known in the economic literature as the *certainty equivalent* of the agent at time t .

The process H_t^α tells by how much the value function V^α decreases when there is a jump. Having $H^\alpha \equiv 0$ means that the accidents that happen between time 0 and time T do not affect the continuation value of the agent.

Proposition 3.1 means that by making the optimal prevention effort, the agent gets her certainty equivalent before entering the insurance contract at time 0. It is then equivalent for the agent to lose the amount in cash Y_0^α or running the project with a coverage α and perform the optimal effort over the period $[0, T]$.

3.2 Optimal insurance cover

Recall Problem (2.4)

$$V = \sup_{\alpha \in [0,1]} e^{\gamma \Pi(\alpha)} V^\alpha,$$

where the value of V^α is given by Theorem 3.1 and Π is the insurance premium the agent pays for a linear insurance coverage α . We study now this optimization problem.

Assumption 3.1. *The insurance premium Π is a lower semicontinuous, non-decreasing function of α .*

In the actuarial science literature, various premium functionals are analyzed (see [18] or [8]), such as the safety loading, the standard deviation principle, the zero utility principle, the Escher principle etc. For all these cases, the premium Π would be continuous and non decreasing, and satisfies in particular our mild Assumption 3.1.

Theorem 3.2. *Under Assumption 3.1, there exists an optimal insurance cover, denoted α^* . The value function of the agent is then given by $V = V^{\alpha^*} e^{\gamma \Pi(\alpha^*)}$.*

Proof. The proof of this theorem will be in two steps. First, we prove that $\alpha \mapsto Y_0^\alpha$ is continuous. And then we look at the continuity of $\alpha \mapsto V^\alpha e^{\gamma \Pi(\alpha)}$.

Step 1. To prove that $\alpha \mapsto Y_0^\alpha$ is continuous, we will show that the generator F^α of the BSDE (3.3) verifies the following two properties:

- (a) The family $(F^\alpha)_{\alpha \in [0,1]}$ is such that for any $h \in \mathfrak{H}$, the map $(t, \omega) \mapsto F^\alpha(h_t(\omega, .))$ is $\mathcal{F}_t \otimes \mathcal{B}([0, t])$ -measurable. Moreover the family $(F^\alpha)_{\alpha \in [0,1]}$ is equi-Lipschitz, i.e., there exists $K_1 > 0$ such that, for $dt \times d\mathbb{P}$ -almost every $(t, \omega) \in \mathbb{R}_+ \times \Omega$,

$$\forall \alpha \in [0, 1], |F^\alpha(h_t^1(\omega)) - F^\alpha(h_t^2(\omega))| \leq K_1(\omega) ||| h_t^1(\omega, .) - h_t^2(\omega, .) |||_t(\omega).$$

- (b) The function $\alpha \mapsto F^\alpha(h)$ is equi-Lipschitz, i.e., there exists $K_2(\omega) > 0$ such that, for $dt \times d\mathbb{P}$ -almost every $(t, \omega) \in \mathbb{R}_+ \times \Omega$,

$$\forall h \in \mathfrak{H}, |F^{\alpha_1}(h) - F^{\alpha_2}(h)| \leq K_2(\omega) |\alpha_1 - \alpha_2|.$$

Note that from Theorem 3.1, we can assume without loss of generality that h is bounded by a constant M_h . It follows then that the effort e is bounded from above by a constant that we denote e_{max} . Indeed, note that the term $\frac{1}{\gamma} \int_0^\infty (e^{\gamma(h(x)+(1-\alpha)x)} - 1) dG(x) - \int_0^\infty \lambda_0 h(x) dG(x)$ is bounded so then $f^\alpha(e, h) \rightarrow \infty$ as $e \rightarrow \infty$. It follows that for $\alpha \in [0, 1]$, the function f^α defined by Equation (3.1) is a bounded function in e . Then, we have that the optimization of f^α over e is actually over the set $[0, e_{max}]$. The lower bound is trivial and we also have $f^\alpha(e, h) \leq c(e_{max}) + \frac{\lambda_0}{\gamma} \int_0^{M_J} (e^{\gamma(M_h+(1-\alpha)x)} - 1) dG(x)$.

From the known identities for infimum and supremum we get for $\alpha \in [0, 1]$, $h^1, h^2 \in \mathfrak{H}$ and for $dt \times d\mathbb{P}$ -almost every $(t, \omega) \in \mathbb{R}_+ \times \Omega$

$$\begin{aligned} & |F^\alpha(h_t^1(\omega, x)) - F^\alpha(h_t^2(\omega, x))| \\ & \leq \sup_{e \in [0, e_{max}]} \left| \frac{\lambda(e)}{\gamma} \int_0^{M_J} (e^{\gamma(h_t^1(\omega, x)+(1-\alpha)x)} - 1) dG(x) - \frac{\lambda(e)}{\gamma} \int_0^{M_J} (e^{\gamma(h_t^2(\omega, x)+(1-\alpha)x)} - 1) dG(x) \right. \\ & \quad \left. + \lambda_0 \int_0^{M_J} h_t^1(\omega, x) dG(x) - \lambda_0 \int_0^{M_J} h_t^2(\omega, x) dG(x) \right|, \\ & = \sup_{e \in [0, e_{max}]} \left| \frac{\lambda(e)}{\gamma} \int_0^{M_J} e^{\gamma(1-\alpha)x} (e^{\gamma h_t^1(\omega, x)} - e^{\gamma h_t^2(\omega, x)}) dG(x) + \lambda_0 \int_0^{M_J} (h_t^1(\omega, x) - h_t^2(\omega, x)) dG(x) \right|, \\ & \leq \sup_{e \in [0, e_{max}]} \left| \frac{\lambda(e)}{\gamma} e^{\gamma M_J} \int_0^{M_J} (e^{\gamma h_t^1(\omega, x)} - e^{\gamma h_t^2(\omega, x)}) dG(x) \right| + ||| h_t^1(\omega, \cdot) - h_t^2(\omega, \cdot) |||_t(\omega). \end{aligned}$$

Since h is assumed bounded, we have that for every $x \in [0, M_J]$ and for $dt \times d\mathbb{P}$ -almost every $(t, \omega) \in \mathbb{R}_+ \times \Omega$, $h \mapsto F^\alpha(h)$ takes its values in a compact set. Hence from the mean-value

theorem, we have the existence of $h_t^3(\omega, x) \in [h_t^1(\omega, x), h_t^2(\omega, x)]$ such that

$$\begin{aligned} |F^\alpha(h_t^1(\omega, x)) - F^\alpha(h_t^2(\omega, x))| &\leq \lambda_0 e^{\gamma M_J} \int_0^{M_J} e^{\gamma h_t^3(\omega, x)} |h_t^1(\omega, x) - h_t^2(\omega, x)| dG(x) + |||h_t^1(\omega, \cdot) - h_t^2(\omega, \cdot)|||_t(\omega), \\ &\leq K_1(\omega) |||h_t^1(\omega, \cdot) - h_t^2(\omega, \cdot)|||_t(\omega), \end{aligned}$$

with $K_1(\omega) := e^{\gamma M_J} e^{\gamma M_h} + 1$.

Assumption (b) is also verified as for $h \in \mathfrak{H}$ and $\alpha_1, \alpha_2 \in [0, 1]$

$$\begin{aligned} |F^{\alpha_1}(h) - F^{\alpha_2}(h)| &\leq \sup_{e \in [0, e_{max}]} \left| \frac{\lambda(e)}{\gamma} \int_0^{M_J} e^{\gamma(h_t(\omega, x) + (1-\alpha_1)x)} dG(x) - \frac{\lambda(e)}{\gamma} \int_0^{M_J} e^{\gamma(h_t(\omega, x) + (1-\alpha_2)x)} dG(x) \right|, \\ &= \sup_{e \in [0, e_{max}]} \left| \frac{\lambda(e)}{\gamma} \int_0^{M_J} e^{\gamma h_t(\omega, x)} (e^{\gamma(1-\alpha_1)x} - e^{\gamma(1-\alpha_2)x}) dG(x) \right|. \end{aligned}$$

From the mean-value theorem, there exists $\alpha_3 \in [\alpha_2, \alpha_3]$ such that

$$\begin{aligned} |F^{\alpha_1}(h) - F^{\alpha_2}(h)| &\leq \sup_{e \in [0, e_{max}]} \frac{\lambda(e)}{\gamma} \int_0^{M_J} e^{\gamma h_t(\omega, x)} |\gamma x e^{\gamma(1-\alpha_3)x}| |\alpha_1 - \alpha_2| dG(x), \\ &\leq \lambda_0 \int_0^{M_J} e^{\gamma M_h} x e^{\gamma x} |\alpha_1 - \alpha_2| dG(x), \\ &\leq K_2(\omega) |\alpha_1 - \alpha_2|, \end{aligned}$$

with $K_2(\omega) := \lambda_0 e^{\gamma M_h} \int_0^{M_J} x e^{\gamma x} dG(x)$.

We prove now that $\alpha \mapsto Y_0^\alpha$ is continuous. For any $\alpha_1, \alpha_2 \in [0, 1]$, property (a) allows us to use the following estimate from [21], Proposition 3.13

$$\|Y^{\alpha_1} - Y^{\alpha_2}\|_{\mathcal{S}_\beta^2}^2 + \|H^{\alpha_1} - H^{\alpha_2}\|_{\mathbb{H}_\beta^2}^2 \leq K_0 \|F^{\alpha_1}(H_s^{\alpha_2}) - F^{\alpha_2}(H_s^{\alpha_2})\|_{\mathbb{H}_\beta^2}^2,$$

with $K_0 > 0$. In particular, we have, using property (b)

$$\begin{aligned} \mathbb{E} [|Y_0^{\alpha_1} - Y_0^{\alpha_2}|^2] &\leq K_0 \|F^{\alpha_1}(H_s^{\alpha_2}) - F^{\alpha_2}(H_s^{\alpha_2})\|_{\mathbb{H}_\beta^2}^2, \\ &\leq K_0 (K_2)^2 \frac{e^{\beta T}}{\beta} |\alpha_1 - \alpha_2|^2, \end{aligned}$$

and by taking $\alpha_1 \rightarrow \alpha_2$ we then have $\mathbb{E} [|Y_0^{\alpha_1} - Y_0^{\alpha_2}|^2] \rightarrow 0$ leading to the continuity of Y_0^α in α . We can conclude from Theorem 3.1, that V^α is continuous in α .

Step 2. As the premium Π is assumed to be lower semicontinuous and V^α is negative, we obtain that $\alpha \mapsto V^\alpha e^{\gamma \Pi(\alpha)}$ is upper semicontinuous and therefore attains a maximum over the compact $[0, 1]$. This leads to the existence of an optimal insurance cover α^* defined as the solution of Problem (2.4). \square

4 Example: Linear cost and linear effect on the frequency

In this section, we present a simple example in which problem (2.5), the choice of the optimal effort, can be solved explicitly. Assume that the intensity of the Poisson process N is linear in e , equal to

$$\lambda(e) = \begin{cases} 1 - e & \text{if } e \leq 1 - \varepsilon, \\ \varepsilon & \text{if } e > 1 - \varepsilon, \end{cases}$$

with ε small. This type of function allows to reduce the intensity until a minimum value of ε . Exerting an effort e cost to the agent the value $c(e) = ce$ with $c > 0$. The agent will not exert an effort bigger than $1 - \varepsilon$ as the intensity cannot further be reduced by making a bigger effort. Note that $\lambda_0 = 1$.

In this particular case, the wealth process X^e is given by

$$X_t^e = \omega_0 - \Pi(\alpha) - (1 - \alpha) \int_0^t \int_0^\infty x \mu_J(ds, dx) - c \int_0^t e_s ds, \quad 0 \leq t \leq T.$$

The maximization problem (2.3) becomes then

$$V = \sup_{(\alpha, e) \in [0, 1] \times \Sigma} \mathbb{E}^e \left[-\frac{1}{\gamma} e^{-\gamma(-\Pi(\alpha) - (1 - \alpha) \int_0^t \int_0^\infty x \mu_J(ds, dx) - c \int_0^t e_s ds)} \right] = \sup_{\alpha \in [0, 1]} e^{\gamma \Pi(\alpha)} V^\alpha.$$

We follow the study in Section 3.1, we first find the value of V^α for each $\alpha \in [0, 1]$ and then we find the optimal insurance cover for a particular type of premium. We have therefore

Proposition 4.1. *The value $V^\alpha = -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha}$ is given by*

- (1) *If $L_G(\gamma(1 - \alpha)) \leq c\gamma + 1$ then $V^\alpha = -\frac{1}{\gamma} e^{(L_G(\gamma(1 - \alpha)) - 1)T}$.*
- (2) *If $L_G(\gamma(1 - \alpha)) > c\gamma + 1$ then $V^\alpha = -\frac{1}{\gamma} e^{(\varepsilon(L_G(\gamma(1 - \alpha)) - 1) + \gamma c(1 - \varepsilon))T}$.*

Proof. From the proof of Theorem 3.1, we see that the unique solution of the BSDE (3.3) is given by (Y^α, H^α) such that

$$\begin{cases} Y_t^\alpha = -\left(ce^* + \frac{(1-e^*)}{\gamma}(L_G(\gamma(1 - \alpha)) - 1)\right)(T - t), & 0 \leq t \leq T, \\ H_t^\alpha \equiv 0, & 0 \leq t \leq T, \end{cases}$$

where the optimal effort $e^* \equiv 0$ if $L_G(\gamma(1 - \alpha)) \leq c\gamma + 1$ and $e^* \equiv 1 - \varepsilon$ otherwise². As the optimal effort e^* is constant we can rewrite V^α as

$$\begin{aligned} V^\alpha &= \mathbb{E}^{e^*} \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha)J_T + ce^*T)} \right], \\ &= -\frac{1}{\gamma} e^{\gamma ce^*T} \mathbb{E}^{e^*} \left[e^{\gamma(1-\alpha) \sum_{i=1}^{N_T} Z_i} \right], \\ &= -\frac{1}{\gamma} e^{\gamma ce^*T} \mathbb{E}^{e^*} [L_Z(\alpha)^{N_T}], \\ &= -\frac{1}{\gamma} e^{\gamma ce^*T} e^{(1-e^*)T(L_G(\gamma(1-\alpha))-1)}. \end{aligned}$$

2. We assume the constants satisfy $e^\gamma < 1 + c\gamma$ to simplify the exposition of the results in Proposition 4.2.

□

Remark 4.1. We can see in the proof of the previous proposition, that in this example the optimal effort of the agent can decrease with the insurance coverage and therefore be complements. In case a zero coverage enforces maximal effort, there will exist a value α_0 such that for every coverage greater than α_0 the agent will feel secure enough to make no effort.

We assume next that the insurance premium is given by

$$\Pi(\alpha) := \alpha(1 + \theta)\mathbb{E}^0[J_T] = \alpha(1 + \theta)T \mathbb{E}^0[Z_i],$$

where $\theta \geq 0$ is the safety loading that reflects the insurance price. For this particular premium, we can find explicitly the optimal coverage of the agent. To simplify the proof of the next proposition, suppose that $Z_i = 1$ for all i so then J is equal to the Poisson process N of intensity $\lambda(e)$ under \mathbb{P}^e .

Proposition 4.2. There exist values $\theta_1 \leq \theta_2$ such that the optimal coverage is given by

$$\alpha^*(\theta) = \left(1 - \frac{\log(1 + \theta)}{\gamma}\right) 1_{\{1+\theta < \theta_1\}} + \left(1 - \frac{\log(\frac{1+\theta}{\varepsilon})}{\gamma}\right) 1_{\{1+\theta < \theta_2\}}.$$

Moreover $\varepsilon(1 + c\gamma) \leq 1 + \theta_i \leq e^\gamma$, for $i = 1, 2$.

Proof. From Equation (2.3) we have

$$V = \sup_{\alpha \in [0,1]} V^\alpha e^{\gamma\alpha(1+\theta)T},$$

so from Proposition 4.1, we have the objective function

$$V^\alpha e^{\gamma\alpha(1+\theta)T} = \begin{cases} -\frac{1}{\gamma} e^{T(e^{\gamma(1-\alpha)} - 1 + \gamma\alpha(1+\theta))}, & \alpha \geq 1 - \frac{1}{\gamma} \ln(c\gamma + 1), \\ -\frac{1}{\gamma} e^{T(\varepsilon(e^{\gamma(1-\alpha)} - 1) + \gamma c + \gamma\alpha(1+\theta))}, & \alpha < 1 - \frac{1}{\gamma} \ln(c\gamma + 1). \end{cases}$$

Define

$$f_1^\theta(\alpha) := -\frac{1}{\gamma} e^{T(e^{\gamma(1-\alpha)} - 1 + \gamma\alpha(1+\theta))}, \quad f_2^\theta(\alpha) := -\frac{1}{\gamma} e^{T(\varepsilon(e^{\gamma(1-\alpha)} - 1) + \gamma c + \gamma\alpha(1+\theta))},$$

and the point $\bar{\alpha} := 1 - \frac{1}{\gamma} \ln(c\gamma + 1)$. Notice that f_1^θ and f_2^θ are concave functions, with a global maximum over \mathbb{R} given by $\alpha_1^\theta := 1 - \frac{\log(1+\theta)}{\gamma}$ and $\alpha_2^\theta := 1 - \frac{\log(\frac{1+\theta}{\varepsilon})}{\gamma}$ respectively. There are two main cases that we analyze separately

Case 1: If $e^\gamma < 1 + \gamma$, then for any $\alpha \in [0, 1]$ we have $e^*(\alpha) = 0$. The problem reduces to maximize f_1^θ over the interval $[0, 1]$. If $(1 + \theta) \leq e^\gamma$ then the global maximum of the function

belongs to $[0, 1]$ and we have $\alpha^*(\theta) = \alpha_1^\theta$. If $(1 + \theta) > e^\gamma$ then $\alpha_1^\theta < 0$ and we have $\alpha^*(\theta) = 0$.

We have then

$$\alpha^*(\theta) = \left(1 - \frac{\log(1 + \theta)}{\gamma}\right) 1_{\{1+\theta < e^\gamma\}}.$$

Case 2: If $e^\gamma > 1 + c\gamma$, then $e^*(\alpha) = (1 - \varepsilon)1_{\{\alpha \in [0, \bar{\alpha}]\}}$. The problem reduces to maximize f_1^θ and f_2^θ over the intervals $[0, \bar{\alpha}]$ and $[\bar{\alpha}, 1]$ respectively and compare their values.

If $(1 + \theta) \leq \varepsilon(1 + c\gamma)$, then we have $\bar{\alpha} \leq \alpha_2^\theta < \alpha_1^\theta$. Therefore, the map f_2^θ is increasing over $[0, \bar{\alpha}]$, and the map f_1^θ is increasing over $[\bar{\alpha}, \alpha_1^\theta]$ and decreasing over $[\alpha_1^\theta, 1]$. It follows that in this case $\alpha^*(\theta) = \alpha_1^\theta$.

If $\varepsilon(1 + c\gamma) \leq (1 + \theta) \leq (1 + c\gamma)$, then $\alpha_2^\theta \leq \bar{\alpha} \leq \alpha_1^\theta$. The map f_1^θ is increasing over $[\bar{\alpha}, \alpha_1^\theta]$ and decreasing over $[\alpha_1^\theta, 1]$. Depending on whether or not $(1 + \theta) \leq \varepsilon e^\gamma$, the map f_2^θ will be increasing over $[0, \alpha_2^\theta]$ and decreasing over $[\alpha_2^\theta, \bar{\alpha}]$ or just decreasing over $[0, \bar{\alpha}]$. We compare the values $f_1^\theta(\alpha_1^\theta)$ and $f_2^\theta(\alpha_2^\theta)$ or $f_2^\theta(0)$ to determine $\alpha^*(\theta)$. This will give us a smooth transition of $\alpha^*(\theta)$ from the value α_1^θ to either α_2^θ or 0 and will also dictate what happens for the next values of θ .

If $(1 + c\gamma) \leq (1 + \theta) \leq e^\gamma$, then $\alpha_2^\theta \leq \alpha_1^\theta \leq \bar{\alpha}$. Both maps f_1^θ and f_2^θ are decreasing over $[\bar{\alpha}, 1]$, so $\alpha^*(\theta)$ corresponds to the maximum of f_2^θ over $[0, \bar{\alpha}]$. Depending on whether or not $(1 + \theta) \leq \varepsilon e^\gamma$, the maximum will be attained either at α_2^θ or 0. The solution $\alpha^*(\theta)$ in this interval pastes smoothly with the one in the previous paragraph.

If $e^\gamma \leq (1 + \theta)$, then f_2^θ and f_1^θ are decreasing over $[0, 1]$ so $\alpha^*(\theta) = 0$. \square

Let us comment the results in the example of this section. Depending on the value of θ , the agent could decide to take insurance or not. When it is the case, the agent can elect an intermediate cover, meaning a cover that may not be full cover. As a function of the price θ , the coverage choice is decreasing: it can decrease continuously from 1 to 0 if $(1 + \theta_0) = e^\gamma$, or it will jump suddenly to no-coverage in the other case. This differs from some results in the literature, for instance [2], where the optimal proportion of coverage α is either 0 or 1.

Conclusion

We considered in this paper an optimal insurance demand model, where the protection buyer has the possibility to exert a costly, non observable effort to reduce the probability that a loss occurs. This corresponds to a self-protection model in continuous time, formulated as a stochastic control problem in the weak sense. After proving that this problem satisfies a

dynamic programming principle, we use it to connect the value function with an explicitly solvable backward SDE. Our main result is that, surprisingly, the optimal effort process of the agent is constant. The optimal effort is determined at time 0, and there is no incentive for the agent to modify it afterwards. These results heavily depend on the fact that the considered process distributions are all equivalent. If the effort also impacted the jumps size distribution, the process laws would not necessarily be absolutely equivalent to a reference probability measure, and much more arguments would be needed to solve the optimization problem, that we leave for further research.

Bibliographie

- [1] ANDRUSIV, A., AND ENGELBERT, H.-J. On the minimal entropy martingale measure for lévy processes. *Stochastics* (2019), 1–21.
- [2] BENSELEM, S., SANTIBÁÑEZ, N. H., AND KAZI-TANI, N. Prevention efforts, insurance demand and price incentives under coherent risk measures. *Insurance : Mathematics and Economics* 93 (2020), 369 – 386.
- [3] BIAIS, B., MARIOTTI, T., ROCHE, J.-C., AND VILLENEUVE, S. Large risks, limited liability, and dynamic moral hazard. *Econometrica* 78, 1 (2010), 73–118.
- [4] BOHNENBLUST, H. F., SHAPLEY, L. S., AND SHERMAN, S. Reconnaissance in game theory. *RAND Corporation research memorandum series* (1949).
- [5] CAPONI, A., AND FREI, C. Dynamic contracting : accidents lead to nonlinear contracts. *SIAM Journal on Financial Mathematics* 6, 1 (2015), 959–983.
- [6] COURBAGE, C., REY, B., AND TREICH, N. Prevention and precaution. In *Handbook of Insurance*, G. Dionne, Ed. Huebner International Series on Risk, Insurance, and Economic Security (New York : Springer), New York, 2013, ch. 8, pp. 185–204.
- [7] CVITANIĆ, J., POSSAMAÏ, D., AND TOUZI, N. Dynamic programming approach to principal-agent problems. *Finance and Stochastics* 22, 1 (2018), 1–37.
- [8] DEELSTRA, G., AND PLANTIN, G. Risk theory and reinsurance. Sciences po publications, Sciences Po, 2014.
- [9] EECKHOUDT, L., GOLLIER, C., AND SCHLESINGER, H. *Economic and Financial Decisions Under Risk*. Economic and Financial Decisions Under Risk. Princeton University Press, 2005.
- [10] EHRLICH, I., AND BECKER, G. S. Market insurance, self-insurance, and self-protection. *Journal of political Economy* 80, 4 (1972), 623–648.
- [11] EPSTEIN, L. G. Decision making and the temporal resolution of uncertainty. *International economic review* (1980), 269–283.
- [12] ESCHE, F., AND SCHWEIZER, M. Minimal entropy preserves the lévy property : how and why. *Stochastic processes and their applications* 115, 2 (2005), 299–327.

- [13] EUCH, O. E., MASTROLIA, T., ROSENBAUM, M., AND TOUZI, N. Optimal make-take fees for market making regulation. *Papers* 1805.02741, arXiv.org, May 2018.
- [14] GOLLIER, C., HAMMITT, J. K., AND TREICH, N. Risk and choice : A research saga. *Journal of Risk and Uncertainty* 47, 2 (Oct 2013), 129–145.
- [15] HENRY, C. Investment decisions under uncertainty : the " irreversibility effect". *The American Economic Review* 64, 6 (1974), 1006–1012.
- [16] HERNÁNDEZ SANTIBÁÑEZ, N., POSSAMAÏ, D., AND ZHOU, C. Bank monitoring incentives under moral hazard and adverse selection. *arXiv preprint arXiv :1701.05864* (2017).
- [17] HÖLMSTROM, B. Moral hazard and observability. *The Bell Journal of Economics* 10, 1 (1979), 74–91.
- [18] LAEVEN, R. J., AND GOOVAERTS, M. J. *Premium Calculation and Insurance Pricing*. American Cancer Society, 2014.
- [19] MIYAHARA, Y. Minimal entropy martingale measures of jump type price processes in incomplete assets markets. *Asia-Pacific Financial Markets* 6, 2 (1999), 97–113.
- [20] PAGÈS, H., AND POSSAMAÏ, D. A mathematical treatment of bank monitoring incentives. *Finance and Stochastics* 18, 1 (2014), 39–73.
- [21] PAPAPANTOLEON, A., POSSAMAÏ, D., AND SAPLAOURAS, A. Existence and uniqueness results for bsde with jumps : the whole nine yards. *Electron. J. Probab.* 23 (2018), 68 pp.
- [22] SANNIKOV, Y. A continuous-time version of the principal–agent problem. *The Review of Economic Studies* 75, 3 (2008), 957–984.
- [23] WINTER, R. Optimal insurance contracting under moral hazard. In *Handbook of Insurance*, G. Dionne, Ed. Huebner International Series on Risk, Insurance, and Economic Security (New York : Springer), New York, 2013, ch. 9, pp. 205–230.

Conclusion générale

Ce travail de thèse a étudié trois modélisations de la prévention. Le chapitre 1 s'intéresse à la relation entre la demande en assurance et l'effort de prévention sous ses deux pendants, auto-assurance et auto-protection. Il étudie également le côté de l'offre en assurance en considérant la relation assureur-assuré comme un jeu de Stackelberg. Le chapitre 2 se concentre sur la relation entre auto-assurance et demande en assurance de marché. Ces deux chapitres se placent dans un cadre statique au travers de la minimisation de mesures de risque, un outil largement utilisé dans les pratiques actuarielles. Le chapitre 3 s'intéresse à la relation entre demande en assurance et auto-protection dans un cadre dynamique au travers d'un problème de contrôle stochastique.

Le chapitre 1 se place dans le cadre d'une minimisation de mesures de risque cohérentes pour l'assuré et pour l'assureur. Il permet en plus d'étudier le problème du côté de l'offre. L'approche proposée, qui consiste à voir la relation entre un assureur et un assuré comme un jeu séquentiel, permet d'étudier les deux types de prévention, auto-assurance et auto-protection, de manière uniforme. L'identification d'un unique équilibre à ce jeu implique, du côté de la demande, que l'individu choisira soit de s'assurer entièrement et fournira un effort de prévention correspondant à ce choix, soit il ne s'assurera pas du tout mais fournira tout de même un effort de prévention lui permettant de réduire son risque de manière optimale. De plus, les deux résultats identifiés dans la littérature économique de la prévention sont retrouvés : l'auto-assurance est substituable à l'assurance de marché et l'auto-protection peut être substituable ou complémentaire à l'assurance de marché suivant le prix de cette dernière. Du côté de l'offre, il existe un prix que l'assureur peut proposer pour réduire de manière optimale son risque et pour lequel l'assuré choisira de souscrire au contrat, permettant ainsi à l'assureur de connaître le prix maximal pour lequel l'assuré choisira de souscrire au contrat.

Le chapitre 2 se place dans le cadre d'une minimisation de mesures de risque de distorsion, dans le cas particulier où elles sont définies par des fonctions de distorsion non concaves. Les fonctions de distorsion non-concave permettent de prendre en compte les travaux sur les biais comportementaux initiés par Kahneman et Tversky (voir [18], [19] et [17]). Il a été montré que les individus placent un filtre face aux probabilités de risque. Ce filtre, subjectif, est soumis à un certain nombre de biais cognitifs et dépend de l'expérience personnelle de chaque individu. Ce chapitre se concentre sur la relation entre auto-assurance et contrat d'assurance.

Dans la littérature standard, cette relation est définie comme substituable, c'est-à-dire que la demande en auto-assurance diminue la demande en assurance et inversement. Etudier une mesure de risque de distorsion avec distorsion non concave comme critère de préférence du risque permet d'apporter une nouvelle conclusion sur cette relation. Elle présente dans ce cas une indétermination, suivant la sensibilité de l'agent au prix de l'assurance, l'auto-assurance peut être substituable ou complémentaire à l'assurance.

Le chapitre 3 se place dans le cadre d'une maximisation d'utilité espérée. Sur une période de temps donnée $[0, T]$, $T > 0$ qui représente la durée d'un contrat d'assurance, l'agent cherche à maximiser sa satisfaction au temps T . Il veut donc déterminer de manière optimale son niveau de couverture assurantielle et son niveau d'effort d'auto-protection dynamique, pour maximiser l'utilité espérée exponentielle de sa richesse finale. Les sinistres de l'agent sont modélisés par les saut d'un processus de Poisson composé. L'effort de prévention se fait donc sur l'intensité de ces sauts. Cette richesse finale est vue comme la solution d'une équation différentielle stochastique rétrograde à saut. Grâce à une approche basée sur les articles [8] et [11], l'unicité et la caractérisation de la solution de cette équation différentielle est obtenue ainsi que l'existence d'un effort de prévention optimal. En analysant le problème en la couverture assurantielle, on montre qu'il existe un niveau optimal. L'individu pourra alors souscrire à une assurance en fournissant un effort de prévention qui lui permettra de maximiser sa satisfaction ou bien choisir de ne pas souscrire au contrat mais tout de même prendre part à des actions d'auto-protection.

Les travaux exposés dans ce travail de thèse présentent cependant des limites qui représentent également des futures pistes de recherche. Une première limite à souligner est la non-prise en compte de la richesse initiale de l'agent. En effet, dans les chapitres 1 et 2, les mesures de risque choisies sont invariantes par translation et dans le chapitre 3, la fonction d'utilité de l'agent étant exponentielle, la richesse initiale de l'agent (qui est constante) ne joue aucun rôle dans les analyses proposées. Bien que ces modèles se veulent généraux et s'appliquent à un large éventail de risques, il semble que leurs applications dans le cadre de l'assurance santé soit incomplète. En effet, Cawley (voir [3]) ont montré que le niveau de richesse initiale influence les décisions d'assurance. Plus récemment, Lee (voir [21]) a montré que la richesse initiale de l'individu a également un impact sur la demande en auto-assurance. Ceci pourrait être pris en compte en considérant une mesure de risque sous-additive par translation par exemple.

De plus, les trois chapitres ne considèrent qu'un contrat d'assurance proportionnelle (ce qu'on retrouve dans la littérature avec [10], [6] ou encore [9] qui propose également une comparaison entre assurance proportionnelle avec assurance avec franchise) qui combiné à l'homogénéité des mesures de risque proposées, donne un problème linéaire en la couverture assurantielle. Il serait opportun de considérer des contrats d'assurance non-proportionnelle, comme on en retrouve notamment en assurance santé aux États-Unis ou encore en réassurance (voir [1] pour plus de détails à ce sujet). En effet, Holmstrom (voir [14]) montre que pour les actions

d'auto-protection, il est optimal d'avoir un contrat d'assurance avec franchise et couverture totale au-delà du niveau de cette franchise.

Il est également à souligner que seuls les risques unidimensionnels sont étudiés dans les trois chapitres de ce travail de thèse. Courbage, Loubergé et Peter (voir [7]) ont étudié le cas des risques multi-dimensionnels sous la théorie de l'utilité espérée. Une extension de ce travail de thèse pourrait être proposée dans le cadre des mesures de risque définies sur des vecteurs aléatoires avec une distribution de probabilités générale et dont les dépendances seraient traduites par des copules.

Les trois chapitres exposés ne tiennent pas compte de paramètres individuels en dehors de l'aversion au risque. Il serait intéressant de prendre en compte d'autres paramètres individuels qui pourraient influencer les décisions d'assurance. On pourrait imaginer une modélisation en terme de mesures de risque, qui tiendrait compte d'autres facteurs comme dans le modèle de Grossman (voir [12]). Ce modèle adapte la notion de capital humain introduit par Becker dans [2], à la santé, et souligne les effets de facteurs tels que le niveau d'études, le salaire ou encore la classe socio-professionnelle et l'âge.

De plus, il serait pertinent d'étendre les différents travaux proposés à d'autres types de relation assureur-assuré. Par exemple, une relation d'agence avec un Principal faisant face à plusieurs agents. Ceci se modéliseraient selon une approche basée sur la théorie des jeux à champ moyen (voir [20] et [13] pour la théorie et [5] pour une extension) où les individus sont connectés par des graphes. Une application possible serait la propagation d'une épidémie dans le réseau d'agents. On pourrait également étudier un problème à trois étages avec un assuré, un assureur et un réassureur. Ceci pourrait se modéliser comme un problème d'agence dans lequel l'assuré serait un Agent, l'assureur aurait à la fois le statut de Principal et d'Agent et le réassureur serait un Principal. Une modélisation générale a été proposée par Hubert (voir [16]).

Une dernière limite à souligner dans ce travail de thèse est que l'assureur est supposé être en situation de monopole de marché³. Il est le seul à pouvoir proposer un contrat à l'agent. Dans le monde assurantiel, cette hypothèse pourrait être questionnée dans de futurs travaux, au regard de la concurrence observée sur certains secteurs. Il serait alors possible de regarder le cas d'une compétition de Cournot ou de Bertrand (voir [4]) entre deux assureurs. Il serait également possible de modéliser ceci sous la forme d'une relation d'agence avec plusieurs principaux et un agent dans laquelle les principaux sont en compétition (voir [15], [22]).

3. Cette hypothèse est retrouvé dans la littérature voir [10], [6]

Bibliographie

- [1] ALBRECHER, H., TEUGELS, J., AND BEIRLANT, J. *Reinsurance : Actuarial and Statistical Aspects*. Statistics in Practice. Wiley, 2017.
- [2] BECKER, G. *Human Capital : A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*. National Bureau of Economic Research, Inc, 1964.
- [3] CAWLEY, J., AND PHILIPSON, T. An empirical examination of information barriers to trade in insurance. *American Economic Review* 89, 4 (1999), 827–846.
- [4] CHAN, P., AND SIRCAR, R. Bertrand and cournot mean field games. *Appl. Math. Optim.* 71, 3 (June 2015), 533–569.
- [5] CLAISSE, J., REN, Z., AND TAN, X. Mean field games with branching, 2019.
- [6] COURBAGE, C. Self-insurance, self-protection and market insurance within the dual theory of choice. *The Geneva Risk and Insurance Review* 26, 1 (2001), 43–56.
- [7] COURBAGE, C., LOUBERGÉ, H., AND PETER, R. Optimal prevention for multiple risks. *Journal of Risk and Insurance* 84, 3 (2017), 899–922.
- [8] CVITANIĆ, J., POSSAMAÏ, D., AND TOUZI, N. Dynamic programming approach to principal-agent problems. *Finance and Stochastics* 22, 1 (2018), 1–37.
- [9] DOHERTY, N. A., AND EECKHOUDT, L. Optimal insurance without expected utility : The dual theory and the linearity of insurance contracts. *Journal of Risk and Uncertainty* 10, 2 (Mar 1995), 157–179.
- [10] EHRLICH, I., AND BECKER, G. S. Market insurance, self-insurance, and self-protection. *Journal of political Economy* 80, 4 (1972), 623–648.
- [11] EUCH, O. E., MASTROLIA, T., ROSENBAUM, M., AND TOUZI, N. Optimal make-take fees for market making regulation. Papers 1805.02741, arXiv.org, May 2018.
- [12] GROSSMAN, M. On the concept of health capital and the demand for health. *Journal of Political Economy* 80, 2 (1972), 223–255.
- [13] GUÉANT, O., LASRY, J.-M., AND LIONS, P.-L. Mean field games and applications. In *Paris–Princeton lectures on mathematical finance 2010*, J.-M. Morel, F. Takens, and

B. Teissier, Eds., vol. 2003 of *Lecture notes in mathematics*. Springer Berlin Heidelberg, 2011, pp. 205–266.

- [14] HOLMSTROM, B. Moral hazard and observability. *Bell Journal of Economics* 10, 1 (1979), 74–91.
- [15] HU, K., REN, Z., AND YANG, J. Principal-agent problem with multiple principals, 2019.
- [16] HUBERT, E. Continuous-time incentives in hierarchies, 2020.
- [17] KAHNEMAN, D. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.
- [18] KAHNEMAN, D., AND TVERSKY, A. Subjective probability : A judgment of representativeness. *Cognitive Psychology* 3 (1972), 430 – 454.
- [19] KAHNEMAN, D., AND TVERSKY, A. Prospect theory : An analysis of decision under risk. *Econometrica* 47, 2 (1979), 263–91.
- [20] LASRY, J.-M., AND LIONS, P.-L. Mean field games. *Japanese Journal of Mathematics* 2, 1 (2007), 229–260.
- [21] LEE, K. Wealth effects on self-insurance. *The Geneva Risk and Insurance Review* 35, 2 (Dec 2010), 160–171.
- [22] MASTROLIA, T., AND REN, Z. Principal-agent problem with common agency without communication. *SIAM Journal on Financial Mathematics* 9, 2 (2018), 775–799.

Annexe A

Preuves du Chapitre 1

1 Proofs of section 2

Proof of Lemma 2.1

1. Assume first that ρ takes the form given in (2.1), which we can write as

$$\rho(X_e) = \int_0^1 Q(u, e) d\psi(u).$$

Since $e \mapsto Q(u, e)$ is convex, then $e \mapsto \int_0^1 Q(u, e) d\psi(u)$ is also convex. Now, if ρ is a general risk measure satisfying 1,2,3 and 5 in Definition 2.1, by Theorem 4.62 and Corollary 4.78 in [1], we have the following Legendre-Fenchel type representation :

$$\rho(X_e) = \sup_{\psi} \left\{ \int_0^1 \bar{q}_{X_e}(u) d\psi(u) - \gamma(\psi) \right\}, \quad (1.1)$$

where the supremum is taken over all concave distortion functions ψ and where

$$\gamma(\psi) = \sup_{Y \in A_\rho} \left\{ \int_0^1 \bar{q}_Y(u) d\psi(u) \right\}, \quad (1.2)$$

with $A_\rho := \{Y \in L^1 \mid \rho(Y) \leq 0\}$. Note that to be able to apply Corollary 4.78 in [2], we need a continuity property of ρ (the so-called Fatou property), which is automatically satisfied by Theorem 2.1 in [3]. By (1.1), the function $e \mapsto \rho(X_e)$ is a supremum of convex functionals, and hence it is convex.

2. Convexity entails that $e \mapsto \rho(X_e)$ is continuous. From representation (1.1), the map $e \mapsto \rho(X_e)$ is also non-increasing and by taking $\psi(u) = u$ we have $\rho(X_e) \geq \mathbb{E}[X_e] \geq 0$.

□

2 Proofs of section 3

2.1 Proofs of subsection 3.2, G non-increasing

Proof of Proposition 3.2

For simplicity, we will write $E(e) := \mathbb{E}[X_e]$. Since the random variables $(X_e)_{e \in \mathbb{R}^+}$ have finite expectation, the function $L_{\mathcal{I}}^\theta$ is finite at $e = 0$. Its right hand derivative $\frac{d}{de}L_{\mathcal{I}}^\theta(0)$ is well defined and

$$\frac{d}{de}L_{\mathcal{I}}^\theta(0) \geq 0 \iff \theta \leq -\frac{c'(0)}{E'(0)} - 1 =: \theta_0. \quad (2.1)$$

In that case, $L_{\mathcal{I}}^\theta$ is non-decreasing at 0, and by convexity it is non-decreasing everywhere, so its minimum is attained at $e = 0$. Note that for the constant θ_0 to be well defined, we need that $E'(0) \neq 0$. If this is not the case, then $\frac{d}{de}L_{\mathcal{I}}^\theta(0) = c'(0) \geq 0$ and the function $L_{\mathcal{I}}^\theta$ is non-decreasing for any $\theta \in \mathbb{R}$.

We analyze now the case where $\theta > \theta_0$. This entails that the function $L_{\mathcal{I}}^\theta$ is decreasing at 0. Therefore, the minimizer $e_I(\theta)$ of $L_{\mathcal{I}}^\theta$ on \mathbb{R}^+ is an interior point characterized by

$$1 + \theta = -\frac{c'(e_I(\theta))}{E'(e_I(\theta))}. \quad (2.2)$$

Let us prove that $\theta \mapsto e_I(\theta)$ is non-decreasing. Indeed, the function $e \mapsto c'(e)\frac{-1}{E'(e)}$ is non-decreasing, being the product of two non-negative non-decreasing functions. Since $1 + \theta$ is obviously increasing with θ , (2.2) implies that $e_I(\theta)$ is non-decreasing with θ .

Note that since the function $e \mapsto \frac{\rho_1(X_e)}{\mathbb{E}[X_e]}$ is non-increasing, $\theta \mapsto e_\theta$ is non-increasing. Note also that since $\frac{\rho_1(X_e)}{\mathbb{E}[X_e]} \geq 1$ for every e in \mathbb{R}^+ , when $\theta = 0$, $\mathcal{I} = \mathbb{R}^+$ and e_θ goes to $+\infty$ when θ tends to 0. Thus the following constant θ_I is well defined :

$$\theta_I := \inf\{\theta > 0 \mid e_\theta \leq e_I(\theta)\}. \quad (2.3)$$

Said otherwise, if θ is greater than θ_I , then the global minimizer $e_I(\theta)$ lies outside of the set \mathcal{I} , and if θ is lower than θ_I , $e_I(\theta) \in \mathcal{I}$. In particular, $\theta \leq \theta_I$ implies that the local minimizer of $L_{\mathcal{I}}^\theta$ on \mathcal{I} is given by $e_I(\theta)$. On the other hand, when $\theta > \theta_I$, then the restriction of $L_{\mathcal{I}}^\theta$ on \mathcal{I} is non-increasing, and its local minimizer is given by the bound e_θ . To conclude the proof, Note that $\theta_0 < \theta_I$, since for $e_I(\theta)$ to be greater than e_θ (or even to be greater than 0), the function $L_{\mathcal{I}}^\theta$ has to be decreasing at 0. \square

Proof of Proposition 3.3

Step 1 : We first prove that there exists $\theta > 0$ such that

$$e_\theta \leq e_N. \quad (2.4)$$

By assumption, the expectation $\mathbb{E}[X_e]$ is finite for every $e \geq 0$. This does not imply the finiteness of the risk measure values $\rho(X_e)$. Thus, if $e \mapsto \rho(X_e)$ is not finite everywhere, we define the minimal effort e_{\min} such that the agent will never exert an effort less or equal to that value, in other words :

$$e_{\min} := \inf\{e > 0 \mid \rho(X_{\tilde{e}}) < +\infty, \forall \tilde{e} > e\}.$$

In particular, $\rho(X_{e_{\min}}) = +\infty$ and $\rho(X_e) < +\infty$ for each $e > e_{\min}$. If $e \mapsto \rho(X_e)$ is finite everywhere, we define $e_{\min} := 0$. Let us show that $e_{\min} \leq e_N$. Indeed, if $\rho(X_{e_{\min}}) = +\infty$, since e_N is the global minimizer of $e \mapsto \rho(X_e) + c(e)$, we have $e_N > e_{\min}$. If $e_{\min} = 0$, the inequality is also satisfied since $e_N \geq 0$ by definition. Since, also by definition, e_θ converges to e_{\min} as θ goes to $+\infty$, we obtain that for a large enough θ , the inequality (2.4) is satisfied.

Step 2 : We know that $\theta \mapsto e_\theta$ is non-increasing. Define then

$$\theta_N := \inf\{\theta \geq 0 \mid e_\theta < e_N\}. \quad (2.5)$$

Step 1 guarantees that θ_N is finite. Then $\theta > \theta_N$ implies that $e_N \in \mathcal{N}$. If $\theta \leq \theta_N$, then L_N is non-decreasing on \mathcal{N} , and its minimizer is given by the boundary point e_θ . \square

We present now a lemma which is needed for proving Theorem 3.1. Its economic interpretation is quite natural : it says that for prices not too high (safety loadings not too high), the agent cannot be fully insured and exert more effort than when he has no insurance.

Lemma 2.1. *There does not exist any $\theta \geq 0$ such that*

$$e_N < e_I(\theta) < e_\theta. \quad (2.6)$$

Proof. The proof is by contradiction. Let us assume the existence of $\tilde{\theta} \geq 0$ satisfying (2.6).

Since the function $\theta \mapsto e_I(\theta)$ is non-decreasing and $\theta \mapsto e_\theta$ is non-increasing, from (2.4), we have the existence of θ larger than $\tilde{\theta}$ such that

$$e_N < e_\theta < e_I(\theta). \quad (2.7)$$

However, remember that e_θ defines the sets \mathcal{I} and \mathcal{N} over which it is optimal to buy insurance and not to buy insurance respectively. That is

$$\begin{aligned} \forall e \leq e_\theta, \quad e \in \mathcal{I} \text{ and } L_{\mathcal{I}}^\theta(e) \leq L_N(e), \\ \forall e > e_\theta, \quad e \in \mathcal{N} \text{ and } L_N(e) < L_{\mathcal{I}}^\theta(e). \end{aligned}$$

In particular, using (2.7) we get $L_N(e_N) \geq L_{\mathcal{I}}^\theta(e_N)$ and $L_{\mathcal{I}}^\theta(e_I(\theta)) > L_N(e_I(\theta))$. Since $e_I(\theta)$ is the global minimizer of $L_{\mathcal{I}}^\theta$, this contradicts the global optimality of e_N for L_N . This proves the claim. \square

We proceed now to the proof of the theorem.

Proof of Theorem 3.1

Step 1 : We prove that the constants θ_I and θ_N defined in Propositions 3.2 and 3.3 satisfy $\theta_N \leq \theta_I$.

Since $\rho(X_e) \geq \mathbb{E}[X_e]$ for every $e \geq 0$, e_θ converges to $G^{-1}(1) = +\infty$ as θ approaches 0^+ . Note then that $e_I(0) \leq e_N$, since the reverse inequality is impossible by Lemma 2.1. Recall the definition (2.5) of the constant θ_N . So the situation for θ small enough ($\theta \leq \theta_N$) is :

$$e_I(\theta) \leq e_N < e_\theta. \quad (2.8)$$

It follows from the definitions of θ_I and θ_N that $\theta_N \leq \theta_I$.

Step 2 : Let $\theta \leq \theta_N$ so (2.8) holds. In this case both $e_N, e_I(\theta) \in \mathcal{I}$ so to find the global minimum of L_1 we can ignore¹ the domain \mathcal{N} . Since in \mathcal{I} , the set over which it is optimal to buy insurance, $e_I(\theta)$ dominates we conclude that $e_I(\theta)$ is the global minimizer of L_1 .

Consider now $\theta \geq \theta_I$. Since $\theta_I \geq \theta_N$ in this case we have both $e_N, e_I(\theta) \in \mathcal{N}$ and similarly to the previous case, we can ignore the domain \mathcal{I} . Since in \mathcal{N} , the set where not to buy insurance is optimal, e_N dominates we conclude that the global minimizer of L_1 is e_N .

Finally, take $\theta \in (\theta_N, \theta_I)$. In this case we have

$$e_I(\theta) \leq e_\theta \leq e_N, \quad (2.9)$$

and we need to compare $L_{\mathcal{I}}^\theta(e_I(\theta))$ and $L_{\mathcal{N}}(e_N)$. Let us prove that $\theta \mapsto L_{\mathcal{I}}^\theta(e_I(\theta))$ is non-decreasing. To this end, let $0 < \theta_1 < \theta_2$, the fact that $e_I(\theta_1)$ is the global minimizer of $L_{\mathcal{I}}^{\theta_1}$ gives (recall the notation $E(e) := \mathbb{E}[X_e]$) :

$$\begin{aligned} L_{\mathcal{I}}^{\theta_1}(e_I(\theta_1)) &= (1 + \theta_1)E(e_I(\theta_1)) + c(e_I(\theta_1)) \leq (1 + \theta_1)E(e_I(\theta_2)) + c(e_I(\theta_2)) \\ &\leq (1 + \theta_2)E(e_I(\theta_2)) + c(e_I(\theta_2)) = L_{\mathcal{I}}^{\theta_2}(e_I(\theta_2)). \end{aligned}$$

Thus we can define θ_M by

$$\theta_M := \inf\{\theta \geq \theta_N \mid L_{\mathcal{I}}^\theta(e_I(\theta)) \geq L_{\mathcal{N}}(e_N)\}.$$

Note that $\theta_M \in [\theta_N, \theta_I]$. Then the conclusions of the theorem hold true by definition. \square

Proof of Corollary 3.1

Since $e_I(\theta)$ is a non-decreasing function of θ , we just need to show that the jump at θ_M is non-negative. Indeed, since $\theta_M \leq \theta_I$ we know that $e_I(\theta_M) < e_{\theta_M}$. Then, necessarily $e_N \geq e_I(\theta_M)$ otherwise Lemma 2.1 is contradicted. \square

2.2 Proofs of subsection 3.3, self-protection

Proof of Lemma 3.2

1. Indeed, $e_N, e_I(\theta) \in \mathcal{I}$ implies that the maps $L_{\mathcal{I}}^\theta$ and $L_{\mathcal{N}}$ are necessarily non-decreasing over \mathcal{N} .

Let us first suppose that ρ_1 is a distortion risk measure in the form of (2.1), with a concave distortion ψ . We get then :

$$\begin{aligned}\rho_1(X_e) &= \int_0^1 \bar{q}_{X_e}(u) d\psi(u), \\ &= \int_0^{p(e)} \bar{q}_Y\left(\frac{u}{p(e)}\right) d\psi(u), \\ &= p(e) \int_0^1 \bar{q}_Y(v) d\psi(p(e)v) dv.\end{aligned}$$

Since $\mathbb{E}[X_e] = p(e)\mathbb{E}[Y]$, we have

$$\frac{\rho_1(X_e)}{\mathbb{E}[X_e]} = \frac{\int_0^1 \bar{q}_Y(v) d\psi(p(e)v) dv}{\mathbb{E}[Y]}.$$

The concavity of ψ implies that the function $e \mapsto \frac{\rho_1(X_e)}{\mathbb{E}[X_e]}$ is non-decreasing, which in turn implies that Assumption 3.2 is satisfied. In the case where ρ_1 is a general law invariant coherent risk measure, using the representation (1.1) and the same computations as above, we get :

$$\frac{\rho_1(X_e)}{\mathbb{E}[X_e]} = \frac{\sup_{\psi} \left\{ \int_0^1 \bar{q}_Y(v) d\psi(p(e)v) dv - \gamma(\psi) \right\}}{\mathbb{E}[Y]},$$

where the supremum is taken over all concave distortion functions ψ , and where $\gamma(\psi)$ is defined in (1.2). Since for any fixed concave ψ , $e \mapsto \int_0^1 \bar{q}_Y(v) d\psi(p(e)v) dv$ is non-decreasing, we can conclude that $e \mapsto \frac{\rho_1(X_e)}{\mathbb{E}[X_e]}$ is also non-decreasing. \square

Proof of Lemma 3.3.

To prove the convexity of $e \mapsto \rho_1(X_e)$, we write

$$\rho_1(X_e) = \int_0^{+\infty} \psi(\mathbb{P}(X_e > x)) dx = \int_0^{+\infty} \psi[p(e)P_Y((x, +\infty))] dx,$$

which is convex, since by assumption $e \mapsto \psi[p(e)P_Y((x, +\infty))]$ is convex for any $x \geq 0$. The particular case $\psi(x) = x$ gives that $e \mapsto \mathbb{E}[X_e]$ is convex. \square

Proof of Proposition 3.5.

The proof is by contradiction. Suppose $e_I < e_\theta$, since e_I is the minimum of the convex function hence $L_{\mathcal{I}}^\theta$ is non-decreasing on $[e_\theta, +\infty)$ so $\partial_e^+ L_{\mathcal{I}}^\theta(e_\theta) > 0$. Moreover from the definition of the sets \mathcal{N} and \mathcal{I} we have that $\partial_e^+ L_{\mathcal{N}}(e_\theta) > \partial_e^+ L_{\mathcal{I}}^\theta(e_\theta) > 0$ and $\partial_e^- L_{\mathcal{N}}(e_\theta) < \partial_e^- L_{\mathcal{I}}^\theta(e_\theta)$. Since $L_{\mathcal{N}}$ is convex in e_N this means that $e_N < e_\theta$.

We proved that if $e_I \leq e_N$ and $e_\theta \leq e_N$ then $e_\theta \leq e_I$ so e_I and e_N belong to the set \mathcal{I} . Moreover we know for all $e \in \mathcal{I}$, $L_{\mathcal{I}}^\theta(e) < L_{\mathcal{N}}(e)$ therefore $L_{\mathcal{I}}^\theta(e_I) < L_{\mathcal{N}}(e_N)$. This means that e_I is the global minimum of $L_{\mathcal{I}}$. \square

Proof of Corollary 3.2.

The last argument in the proof of Proposition 3.5 let us conclude. Indeed, in this case e_N and e_I belong to the set \mathcal{I} therefore the same inequality holds and e_I will be the global minimum of L_1 .

Proof of Theorem 3.2.

Note that for all effort e , the objective function $L_{\mathcal{I}}^{\theta}$ as a function of θ is non-decreasing. Indeed,

$$\theta_1 < \theta_2 \iff (1 + \theta_1)\mathbb{E}[X_e] + c(e) < (1 + \theta_2)\mathbb{E}[X_e] + c(e) \iff L_{\mathcal{I}}^{\theta_1}(e) < L_{\mathcal{I}}^{\theta_2}(e) \quad \forall e. \quad (2.10)$$

In particular for $e = e_I(\theta)$, $L_{\mathcal{I}}^{\theta}(e_I(\theta))$ is non-decreasing with θ .

Since $L_{\mathcal{N}}(e_N)$ is independent of θ and since in the absence of loading factor it is better to have full insurance coverage than no coverage, that is $L_{\mathcal{I}}^{\theta}(e_I(0)) < L_{\mathcal{N}}(e_N)$, we define $\theta_M := \sup\{\theta \geq 0 \mid L_{\mathcal{I}}^{\theta}(e_I(\theta)) \leq L_{\mathcal{N}}(e_N)\}$. By the definition of θ_M , (i) and (ii) hold true.

Let us prove that $\theta_M \geq \theta_N$ by contradiction. Suppose that $\theta_N > \theta_M$ and let θ be such that $\theta_N > \theta > \theta_M$. By the definition of these two constants we have that $e_{\theta} \leq e_N$ and $L_{\mathcal{I}}^{\theta}(e_I) > L_{\mathcal{N}}(e_N)$. This is not possible, since the global optimality of e_I and the point e_N belonging to \mathcal{I} imply $L_{\mathcal{I}}^{\theta}(e_I) \leq L_{\mathcal{I}}^{\theta}(e_N) \leq L_{\mathcal{N}}(e_N)$. \square

3 Proofs of section 5

Proof of Theorem 5.1.

We prove that the objective function of IS is non-increasing, from which the conclusion follows. Write it as the sum of two functions of θ for which we study the monotonicity :

$$L_2(\theta) := f_1(\theta) - f_2(\theta), \quad (3.1)$$

with

$$f_1 := \frac{p^{r_2} r_2 \hat{x} k(e_I(\theta))}{r_2 k(e_I(\theta)) - 1}, \quad f_2 := (1 + \theta) \frac{p \hat{x} k(e_I(\theta))}{k(e_I(\theta)) - 1}.$$

Recall that $e_I(\theta)$ is increasing in θ and then so is $k(e_I(\theta))$. Note that f_1 is a non-increasing function as

$$\frac{\partial f_1}{\partial \theta} = -\frac{p^{r_2} \hat{x} r_2 k'(e_I(\theta)) e'_I(\theta)}{(r_2 k(e_I(\theta)) - 1)^2} < 0.$$

We now study the derivative of f_2 . Rewrite the characterization of $e_I(\theta)$ as follows

$$\frac{e_I(\theta)(k(e_I(\theta)) - 1)^2}{p \hat{x} k'(e_I(\theta))} - (1 + \theta) = 0. \quad (3.2)$$

Consider the following function

$$F(\theta, e) := \frac{e(k(e) - 1)^2}{p\hat{x}k'(e)} - (1 + \theta).$$

By the Implicit Function Theorem, we have

$$\frac{\partial e_I(\theta)}{\partial \theta} = \frac{-\partial F(\theta, e_I(\theta))/\partial \theta}{\partial F(\theta, e_I(\theta))/\partial e}.$$

Remark that

$$\begin{aligned} \frac{\partial F(\theta, e)}{\partial e} &= \frac{((k(e) - 1)^2 + 2ek'(e)(k(e) - 1))k'(e)p\hat{x} - p\hat{x}k''(e)e(k(e) - 1)^2}{(p\hat{x}k'(e))^2}, \\ &= \frac{(k(e) - 1)^2}{p\hat{x}k'(e)} + \frac{2ek'(e)(k(e) - 1)}{p\hat{x}k'(e)} - \frac{k''(e)e}{k'(e)} \frac{(k(e) - 1)^2}{p\hat{x}k'(e)}. \end{aligned}$$

In the next few lines, we will write e_I instead of $e_I(\theta)$ for simplicity of notation. By rewriting Equation (3.2), we notice that

$$\frac{(k(e_I) - 1)^2}{p\hat{x}k'(e_I)} = \frac{(1 + \theta)}{e_I},$$

therefore

$$\frac{\partial F(\theta, e_I)}{\partial e} = \frac{1 + \theta}{e_I} \left(1 + \frac{2e_I k'(e_I)}{k(e_I) - 1} - \frac{k''(e_I)e_I}{k'(e_I)} \right),$$

and

$$\frac{\partial F(\theta, e_I)}{\partial \theta} = -1.$$

This leads to

$$\frac{\partial e_I}{\partial \theta} = \frac{e_I}{1 + \theta} \left(\frac{1}{1 + \frac{2e_I k'(e_I)}{k(e_I) - 1} - \frac{k''(e_I)e_I}{k'(e_I)}} \right).$$

Since

$$1 + \frac{2e_I k'(e_I)}{k(e_I) - 1} - \frac{k''(e_I)e_I}{k'(e_I)} > 1, \quad \frac{e_I}{1 + \theta} > 0,$$

we obtain the following inequality

$$\frac{\partial e_I}{\partial \theta} < \frac{e_I}{1 + \theta}. \tag{3.3}$$

By differentiating f_2 and from inequality (3.3) we have

$$\begin{aligned}\frac{\partial f_2}{\partial \theta} &= p\hat{x} \frac{(k(e_I(\theta)) - 1)k(e_I(\theta)) - (1 + \theta)k'(e_I(\theta))e'_I(\theta)}{(k(e_I(\theta)) - 1)^2} \\ &> \frac{p\hat{x}}{(k(e_I(\theta)) - 1)^2} ((k(e_I(\theta)) - 1)k(e_I(\theta)) - k'(e_I(\theta))e_I(\theta)).\end{aligned}\quad (3.4)$$

To conclude, define the following function

$$h(x) := (k(x) - 1)k(x) - xk'(x),$$

its derivative $h'(x) = 2k'(x)[k(x) - 1] - xk''(x)$ is positive since k is non-decreasing, concave and bounded from below by 1 (see Remark 5.1). This entails that $h(x) > h(0) = 0$, for any $x > 0$. Therefore from Inequality (3.4), f_2 is non-decreasing in θ .

From Equation (3.1), this leads to L_2 being a non-increasing function of θ . In conclusion the optimal insurance price is the upper bound of the interval where we minimize this function, that is to say $\theta^* = \theta_M$. \square

Self-protection in the Pareto case : Equality (5.13)

We now prove the explicit value of $e_I(\theta)$ given in (5.13). Under Assumption 5.1,

$$p(e) = \frac{1}{a+e} \text{ with } a > 1.$$

From the characterization (5.11) of $e_I(\theta)$, we have

$$e_I(\theta) = -\frac{(1+\theta)\hat{x}k}{k-1} p'(e_I(\theta)) = \frac{(1+\theta)\hat{x}k}{k-1} \frac{1}{(a+e_I(\theta))^2} = \frac{c(1+\theta)}{(a+e_I(\theta))^2}, \quad (3.5)$$

where $c = \frac{\hat{x}k}{k-1} > 0$. By rewriting Equation (3.5), we obtain the following equality

$$e_I(\theta)(a+e_I(\theta))^2 - c(1+\theta) = 0.$$

We define the following polynomial function of third degree with coefficients in \mathbb{R}

$$P(X) = X^3 + 2aX^2 + a^2X - c(1+\theta).$$

We can write P as

$$P(X) = Q\left(X + \frac{2a}{3}\right),$$

where $Q(X) = X^3 - \frac{a^3}{3}X - c(1+\theta) - \frac{2a^3}{27}$. Therefore to find the roots of P we need to determine the roots of Q .

Let $X = x + y$, we have

$$Q(X) = x^3 + y^3 + \left(3xy - \frac{a^3}{3}\right)(x + y) - c(1 + \theta) - \frac{2a^3}{27},$$

where x^3 and y^3 are roots of the following polynomial function

$$R(T) = T^2 - \left(c(1 + \theta) + \frac{2a^3}{27}\right)T + \frac{a^6}{27^2}.$$

We denote Δ the discriminant of R :

$$D := 27\Delta = 4a^6 + 27(c(1 + \theta))^2 - 32a^3c(1 + \theta) - 4a^6 + 36a^3c(1 + \theta) = c(1 + \theta)(27c(1 + \theta) + 4a^3) > 0.$$

Therefore R has two distinct solutions in \mathbb{R} given by

$$\begin{aligned} u &:= \frac{1}{2} \left(c(1 + \theta) + \frac{2a^3}{27} - \sqrt{\frac{4}{27}c(1 + \theta)a^3 + (c(1 + \theta))^2} \right), \\ v &:= \frac{1}{2} \left(c(1 + \theta) + \frac{2a^3}{27} + \sqrt{\frac{4}{27}c(1 + \theta)a^3 + (c(1 + \theta))^2} \right). \end{aligned}$$

Q has one solution in \mathbb{R} given by $u^{1/3} + v^{1/3}$. Therefore the solution $e_I(\theta) \in \mathbb{R}$ of $P(e_I(\theta)) = 0$ is given by

$$e_I(\theta) = t_1(\theta) + t_2(\theta) - \frac{2a}{3},$$

where $t_1(\theta) = u^{1/3}$ and $t_2(\theta) = v^{1/3}$.

Proof of Theorem 5.2.

We can rewrite (5.12) as

$$\inf_{\theta \in [0, \theta_M]} \left\{ c_2 \frac{1}{(t_1(\theta) + t_2(\theta) + \frac{a}{3})^{r_2}} - c \frac{1 + \theta}{t_1(\theta) + t_2(\theta) + \frac{a}{3}} \right\},$$

where $c_2 := \frac{r_2 \hat{x} k}{r_2 k - 1} > 0$. We already know that $\theta \mapsto e_I(\theta)$ is non decreasing, this implies that $\theta \mapsto \frac{1}{(t_1(\theta) + t_2(\theta) + \frac{a}{3})^{r_2}}$ is non increasing. The conclusion of the Theorem will follow from the fact that the function m given by

$$m(\theta) := \frac{1 + \theta}{t_1(\theta) + t_2(\theta) + \frac{a}{3}},$$

is an increasing function. By computing the derivative of m , we see that it is increasing if and only if

$$t_1(\theta) + t_2(\theta) + \frac{a}{3} > (1 + \theta)(t'_1(\theta) + t'_2(\theta)) = (1 + \theta) \frac{d}{d\theta} e_I(\theta). \quad (3.6)$$

To compute $\frac{d}{d\theta}e_I(\theta)$, Note that the characterization (5.11) of $e_I(\theta)$ can be written as $F(\theta, e_I(\theta)) = 0$ with $F(\theta, e) := \frac{1}{c}e(a+e)^2 - (1+\theta)$. Using the Implicit Function Theorem we get

$$\begin{aligned}\frac{d}{d\theta}e_I(\theta) &= \frac{c}{3e_I(\theta)^2 + 4ae_I(\theta) + a^2} = \frac{c}{3(e_I(\theta) + a)(e_I(\theta) + \frac{a}{3})} \\ &= \frac{c}{3(t_1(\theta) + t_2(\theta) + \frac{a}{3})(t_1(\theta) + t_2(\theta) - \frac{a}{3})}.\end{aligned}$$

After noticing that $(t_1(\theta) + t_2(\theta) - \frac{a}{3}) > e_I(\theta) \geq 0$, we obtain that (3.6) is equivalent to

$$\begin{aligned}&3\left(t_1(\theta) + t_2(\theta) + \frac{a}{3}\right)^2 \left(t_1(\theta) + t_2(\theta) - \frac{a}{3}\right) > (1+\theta)c \\ \iff &3(x+y)^3 + a(x+y)^2 - \frac{a^2}{3}(x+y) - \frac{a^3}{3^2} - c(1+\theta) > 0 \\ \iff &2(x+y)^3 + a(x+y)^2 - \left(\frac{a}{3}\right)^3 + Q(x+y) > 0,\end{aligned}$$

where for simplicity, we write $x = t_1(\theta)$, $y = t_2(\theta)$ and where $Q(X) := X^3 - \frac{a^2}{3}X - c(1+\theta) - 2\left(\frac{a}{3}\right)^3$. Recall that $Q(x+y) = 0$, which implies that $2(x+y)^3 = \frac{2}{3}a^2(x+y) + 2c(1+\theta) + 4\left(\frac{a}{3}\right)^3$. From these observations, we get that the last inequality above is equivalent to

$$\frac{2}{3}a^2(x+y) + 2c(1+\theta) + 3\left(\frac{a}{3}\right)^3 + a(x+y)^2 > 0,$$

which is satisfied since all the terms are positive. This ends the proof. \square

Bibliographie

- [1] FÖLLMER, H., AND SCHIED, A. *Stochastic finance : an introduction in discrete time.* Walter de Gruyter, 2011.
- [2] FÖLLMER, H., AND SCHIED, A. *Stochastic Finance : An Introduction in Discrete Time.* De Gruyter Textbook. De Gruyter, 2016.
- [3] JOUINI, E., SCHACHERMAYER, W., AND TOUZI, N. Law invariant risk measures have the Fatou property. In *Advances in mathematical economics. Vol. 9*, vol. 9 of *Adv. Math. Econ.* Springer, Tokyo, 2006, pp. 49–71.

Annexe B

Preuves du Chapitre 3

1 Dynamic programming principle

For all \mathbb{F} -stopping time τ with values in $[t, T]$ and for any $e \in \Sigma_\tau$, we define

$$L_T(\tau, e) = \mathbb{E}_\tau^e \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(e_s) ds)} \right],$$

where \mathbb{E}_τ^e denotes the conditional expectation with respect to \mathcal{F}_τ under the probability measure \mathbb{P}^e and Σ_τ denotes the restriction of Σ to the controls on $[\tau, T]$. The continuation utility of the agent is defined for all \mathbb{F} -stopping time τ by

$$V_\tau^\alpha = \underset{e \in \Sigma_\tau}{\text{ess sup}} L_T(\tau, e). \quad (1.1)$$

Lemma 1.1. *Let $t \in [0, T]$ and τ be a \mathbb{F} -stopping time with values in $[t, T]$. Then*

$$V_t^\alpha = \underset{\varepsilon \in \Sigma}{\text{ess sup}} \mathbb{E}_t^\varepsilon \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} V_\tau^\alpha \right].$$

Proof. Denote $\tilde{V}_t^\alpha := \underset{\varepsilon \in \Sigma}{\text{ess sup}} \mathbb{E}_t^\varepsilon \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} V_\tau^\alpha \right]$. First we prove that $V_t^\alpha \leq \tilde{V}_t^\alpha$.

From Equation (1.1) and by the tower property we have

$$V_t^\alpha = \underset{\varepsilon \in \Sigma}{\text{ess sup}} \mathbb{E}_t^\varepsilon \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} \mathbb{E}_\tau^\varepsilon \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(\varepsilon_s) ds)} \right] \right].$$

For $\varepsilon \in \Sigma$, recall the definition of D^ε given in (2.2). From Bayes' formula and by noticing that for all $\varepsilon \in \Sigma$, $\frac{D_T^\varepsilon}{D_\tau^\varepsilon}$ does not depend on the values of ε before time τ , we have

$$\begin{aligned} \mathbb{E}_\tau^\varepsilon \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(\varepsilon_s) ds)} \right] &= \mathbb{E}_\tau \left[-\frac{D_T^\varepsilon}{D_\tau^\varepsilon} \frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(\varepsilon_s) ds)} \right], \\ &\leq \underset{e \in \Sigma_\tau}{\text{ess sup}} \mathbb{E}_\tau^e \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(e_s) ds)} \right], \\ &= V_\tau^\alpha. \end{aligned}$$

Hence $V_t^\alpha \leq \tilde{V}_t^\alpha$.

Now we prove $V_t^\alpha \geq \tilde{V}_t^\alpha$. Let $\varepsilon \in \Sigma$ and $e \in \Sigma_\tau$ and define $\nu_u = \varepsilon_u \mathbb{1}_{0 \leq u \leq \tau} + e_u \mathbb{1}_{\tau \leq u \leq T}$, notice that ν belongs to Σ . By the definition of V_t^α and by the tower property we have

$$\begin{aligned} V_t^\alpha &\geq \mathbb{E}_t^\nu \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha) \int_t^T \int_0^\infty x \mu_J(ds, dx) + \int_t^T c(\nu_s) ds)} \right], \\ &= \mathbb{E}_t^\nu \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds + (1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(e_s) ds)} \right], \\ &= \mathbb{E}_t^\nu \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} \mathbb{E}_\tau^\nu \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(e_s) ds)} \right] \right]. \end{aligned}$$

From Bayes' formula

$$\begin{aligned} \mathbb{E}_\tau^\nu \left[-\frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(e_s) ds)} \right] &= \mathbb{E}_\tau \left[-\frac{D_T^\nu}{D_\tau^\nu} \frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(e_s) ds)} \right], \\ &= \mathbb{E}_\tau \left[-\frac{D_T^e}{D_\tau^e} \frac{1}{\gamma} e^{\gamma((1-\alpha) \int_\tau^T \int_0^\infty x \mu_J(ds, dx) + \int_\tau^T c(e_s) ds)} \right], \\ &= L_T(\tau, e), \end{aligned}$$

$$\text{and } V_t^\alpha \geq \mathbb{E}_t^\nu \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} L_T(\tau, e) \right].$$

From Bayes' formula and by the tower property

$$\mathbb{E}_t \left[\frac{D_T^\nu}{D_t^\nu} e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} L_T(\tau, e) \right] = \mathbb{E}_t \left[\frac{D_\tau^\nu}{D_t^\nu} e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} L_T(\tau, e) \right].$$

Notice that $\frac{D_T^\nu}{D_t^\nu} = \frac{D_\tau^\varepsilon}{D_t^\varepsilon}$. Therefore $V_t^\alpha \geq \mathbb{E}_t^\varepsilon \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} L_T(\tau, e) \right]$ and this inequality holds for all $e \in \Sigma_\tau$.

To conclude, note that the family $(L_T(\tau, e))_{e \in \Sigma_\tau}$ is directed upwards. Indeed, for any e_1 and e_2 in Σ_τ , define

$$e_3 := e_1 \mathbb{1}_{L_T(\tau, e_1) \geq L_T(\tau, e_2)} + e_2 \mathbb{1}_{L_T(\tau, e_1) < L_T(\tau, e_2)}.$$

By definition, e_3 belongs to Σ_τ and we also have the following inequality

$$L_T(\tau, e_3) \geq \max \{L_T(\tau, e_1), L_T(\tau, e_2)\}.$$

From [4], Proposition VI.1.1, there exists a sequence $(e_n)_{n \in \mathbb{N}}$ in Σ_τ such that $(L_T(\tau, e_n))_{n \in \mathbb{N}}$ is non-decreasing almost surely and $\operatorname{ess\,sup}_{e \in \Sigma_\tau} L_T(\tau, e) = \lim_{n \rightarrow \infty} L_T(\tau, e_n)$ almost surely. We have finally from the monotone convergence theorem

$$\begin{aligned} V_t^\alpha &\geq \lim_{n \rightarrow \infty} \mathbb{E}_t^\varepsilon \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} L_T(\tau, e_n) \right], \\ &= \mathbb{E}_t^\varepsilon \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} \lim_{n \rightarrow \infty} L_T(\tau, e_n) \right], \\ &= \mathbb{E}_t^\varepsilon \left[e^{\gamma((1-\alpha) \int_t^\tau \int_0^\infty x \mu_J(ds, dx) + \int_t^\tau c(\varepsilon_s) ds)} V_\tau^\alpha \right]. \end{aligned}$$

□

2 Proof of Proposition 3.1.

Proof. The proof follows the ideas of the proof of Theorem 3.1 in [1], and is divided in five steps.

Step 1. For $e \in \Sigma$, let us define the process

$$U_t^e := V_t^\alpha e^{\gamma((1-\alpha) \int_0^t \int_0^\infty x \mu_J(ds, dx) + \int_0^t c(e_s) ds)}.$$

Lemma 1.1 implies that the process U^e is a \mathbb{P}^e -supermartingale for all e in Σ . From the Doob regularization theorem (see [3] Theorem 3.13), we can consider the càdlàg version of U^e . Let us prove that

$$\sup_{0 \leq t \leq T} \mathbb{E}^e[|U_t^e|^p] < \infty,$$

for a certain $p > 1$ which will entail that U^e is uniformly integrable for every admissible effort e . Note that from Hölder's inequality, by the definition of $e \in \Sigma$ and since the intensity λ and the jump sizes are bounded, there exists a $p > 1$ such that for r and q defined such that $\frac{1}{q} + \frac{1}{r} = \frac{1}{p}$

$$\sup_{0 \leq t \leq T} \mathbb{E}^e[|U_t^e|^p] \leq \sup_{0 \leq t \leq T} \mathbb{E}^e [|V_t^\alpha|^q]^{\frac{p}{q}} K^{\frac{p}{r}} \mathbb{E} \left[e^{\gamma r(1-\alpha)\lambda_0 M_J T} \right]^{\frac{p}{r}}.$$

We then need to prove that

$$\sup_{0 \leq t \leq T} \mathbb{E}^e[|V_t^\alpha|^q] < \infty,$$

which is the case since $|L_T(t, e)|^q$ is bounded by the definition of $e \in \Sigma$ and as the intensity λ and the jump sizes are bounded

$$\begin{aligned} |L_T(t, e)|^q &= \left| -\frac{1}{\gamma} \mathbb{E}_t^e \left[e^{\gamma((1-\alpha) \int_t^T \int_0^\infty x \mu_J(ds, dx) + \int_t^T c(e_s) ds)} \right] \right|^q, \\ &= \left| \frac{1}{\gamma} \mathbb{E}_t \left[\frac{D_T^e}{D_t^e} e^{\gamma((1-\alpha) \int_t^T \int_0^\infty x \mu_J(ds, dx) + \int_t^T c(e_s) ds)} \right] \right|^q. \end{aligned}$$

Note that we have the following inequality

$$\begin{aligned} \frac{D_T^e}{D_t^e} &= \exp \left(- \int_t^T (\lambda(e_s) - \lambda_0) ds + \int_t^T \log \left(\frac{\lambda(e_s)}{\lambda_0} \right) dN_s \right), \\ &\leq e^{(\lambda_0 - \lambda_\infty)T} =: \varsigma_T > 0. \end{aligned} \tag{2.1}$$

Hence

$$|L_T(t, e)|^p \leq \left| \frac{1}{\gamma} \mathbb{E} \left[K \varsigma_T e^{\gamma(1-\alpha)\lambda_0 M_J T} \right] \right|^p.$$

By applying the Doob-Meyer decomposition theorem, we obtain

$$U_t^e = M_t^e - A_t^e, \tag{2.2}$$

where M^e is a \mathbb{P}^e -martingale and A^e an integrable non-decreasing predictable process with $A_0^e = 0$.

As A^e is a càdlàg non-decreasing process, there exists a continuous non-decreasing process $A^{e,c}$ and a purely discontinuous process $A^{e,d}$, equal to the sum of the jumps of A^e , i.e. $A_t^{e,d} = \sum_{0 \leq s \leq t} \Delta A_s^e$, such that

$$A^e = A^{e,c} + A^{e,d}$$

We take $A_0^{e,c} = A_0^{e,d} = 0$. The martingale representation theorem¹ gives that under \mathbb{P}^e there exists a predictable function H^e such that

$$M_t^e = V_0^\alpha + \int_0^t \int_0^\infty H_s^e(x) \mu_J(ds, dx) - \int_0^t \int_0^\infty H_s^e(x) \lambda(e_s) ds dG(x). \quad (2.3)$$

Step 2. From (1.1), V^α is a negative process. Let Y^α be the process defined by

$$V_t^\alpha = -\frac{1}{\gamma} e^{-\gamma Y_t^\alpha}, \quad \forall t \in [0, T]. \quad (2.4)$$

Note that $Y_T^\alpha = 0$. The jump times of J are equal to the jump times of the Poisson process N which are totally inaccessible stopping times². It follows that the jump times of J are totally inaccessible stopping times under \mathbb{P}^e . From [2] Proposition I.2.24, since $A^{e,d}$ is a predictable process, $\Delta A_\tau^{e,d} = 0$ a.s. on $\{\tau < \infty\}$ for all totally inaccessible stopping time τ . Hence, we have $[A^{e,d}, J] = 0$ a.s.

From the definition of U_t^e

$$\begin{aligned} Y_t^\alpha &= -\frac{1}{\gamma} \log \left(-e^{-\gamma((1-\alpha) \int_0^t \int_0^\infty x \mu_J(ds, dx) + \int_0^t c(e_s) ds)} \gamma U_t^e \right) \\ &= -\frac{1}{\gamma} \log (-\gamma U_t^e) + \int_0^t \int_0^\infty (1-\alpha)x \mu_J(ds, dx) + \int_0^t c(e_s) ds. \end{aligned} \quad (2.5)$$

From (2.2) and (2.3),

$$dU_t^e = \int_0^\infty H_t^e(x) \mu_J(dt, dx) - \int_0^\infty \lambda(e_t) H_t^e(x) dt dG(x) - dA_t^{e,c} - \Delta A_t^{e,d}.$$

By applying Ito's formula

$$d \log (-\gamma U_t^e) = - \int_0^\infty \lambda(e_t) \frac{H_t^e(x)}{U_{t-}^e} dt dG(x) + \int_0^\infty \log \left(1 + \frac{H_t^e(x)}{U_{t-}^e} \right) \mu_J(dt, dx) - \frac{dA_t^{e,c}}{U_{t-}^e} + \log \left(1 - \frac{\Delta A_t^{e,d}}{U_{t-}^e} \right).$$

1. From [2], Theorem III.1.26, \mathbb{P}^0 is the unique weak solution to the martingale problem associated to X^0 . From Theorem III.4.29 it follows that \mathbb{P}^0 satisfies the martingale representation theorem, as so does any measure \mathbb{P}^e due to Theorem III.5.24.

2. From [2] Definition I.2.20 : "A stopping time T is called totally inaccessible if $\mathbb{P}[T = S < \infty] = 0$ for all predictable times S ".

Hence from Equation (2.5), we get

$$\begin{aligned} dY_t^\alpha &= \left(c(e_t) + \frac{\lambda(e_t)}{\gamma} \int_0^\infty \frac{H_t^e(x)}{U_{t-}^e} dG(x) \right) dt + \frac{dA_t^{e,c}}{\gamma U_{t-}^e} - \frac{1}{\gamma} \log \left(1 - \frac{\Delta A_t^{e,d}}{U_{t-}^e} \right) \\ &\quad - \frac{1}{\gamma} \int_0^\infty \left(\log \left(1 + \frac{H_t^e(x)}{U_{t-}^e} \right) - \gamma(1-\alpha)x \right) \mu_J(dt, dx). \end{aligned} \quad (2.6)$$

By defining

$$\hat{H}_t^\alpha(x) := \frac{1}{\gamma} \left(\log \left(1 + \frac{H_t^e(x)}{U_{t-}^e} \right) - \gamma(1-\alpha)x \right), \quad (2.7)$$

$$I_t := \int_0^t \left(f^\alpha(e_s, \hat{H}_s^\alpha(x)) ds + \frac{dA_s^{e,c}}{\gamma U_{s-}^e} \right), \quad (2.8)$$

$$\tilde{A}_t^d := \frac{1}{\gamma} \sum_{s \leq t} \log \left(1 - \frac{\Delta A_s^{e,d}}{U_{s-}^e} \right),$$

Equation (2.6) becomes

$$dY_t^\alpha = dI_t - d\tilde{A}_t^d - \int_0^\infty \hat{H}_t^\alpha(x) \tilde{\mu}_J^0(dt, dx), \quad (2.9)$$

with $Y_T^\alpha = 0$ and where Y^α is independent of e , by the definition (2.4). By the equality $d[Y^\alpha, J]_t = \int_0^\infty \hat{H}_t^\alpha(x) \mu_J(dt, dx)$ and since the bracket $[Y^\alpha, J]_t$ can be defined pathwise, independently of any measure \mathbb{P}^e , we get that

$$\int_0^\infty \hat{H}_t^\alpha(x) \tilde{\mu}_J^0(dt, dx) = \int_0^\infty \hat{H}_t^\alpha(x) (\mu_J(dt, dx) - \lambda_0 dG(x) dt)$$

is independent of e .

Next, Y^α being independent of e , its predictable jump part \tilde{A}^d is independent of e and in particular

$$\hat{A}_t := \frac{\Delta A_t^{e,d}}{U_{t-}^e} \quad \text{is also independent of } e.$$

Equation (2.9) then leads to I being independent of e .

Equation (2.8) leads to $0 \leq -\frac{dA_t^{e,c}}{\gamma U_{t-}^e} = f^\alpha(e_t, \hat{H}_t^\alpha(x)) dt - dI_t$ since U^e is a negative process and $A^{e,c}$ is non-decreasing.

Step 3. We prove in this step that Y^α and H^α defined respectively in (2.4) and (2.7) solve the BSDE (3.3).

Note that we have $V_T^\alpha = -\frac{1}{\gamma}$, hence

$$\mathbb{E}^e[U_T^e] = \mathbb{E}^e[V_T^\alpha e^{\gamma(\int_0^T \int_0^\infty (1-\alpha)\mu_J(ds, dx) + \int_0^T c(e_s) ds)}] = \mathbb{E}^e \left[-\frac{1}{\gamma} e^{\gamma(\int_0^T \int_0^\infty (1-\alpha)\mu_J(ds, dx) + \int_0^T c(e_s) ds)} \right] = L_T(0, e).$$

We obtain

$$0 = \sup_{e \in \Sigma} \mathbb{E}^e[U_T^e] - V_0^\alpha = \sup_{e \in \Sigma} \mathbb{E}^e[U_T^e - M_T^e], \quad (2.10)$$

using the inequality for the difference of suprema and $U^e - M^e$ being negative.

From Equation (2.2), we have the following equality $U_T^e - M_T^e = -A_T^{e,c} - A_T^{e,d}$. Note that $A^{e,c}$ can be rewritten from Equations (2.8) for all $t \in [0, T]$ as

$$-dA_t^{e,c} = \gamma U_{t-}^e \left(f^\alpha(e_t, \hat{H}_t^\alpha(x)) dt - dI_t \right).$$

Hence Equation (2.10) becomes

$$0 = \gamma \sup_{e \in \Sigma} \mathbb{E} \left[D_T^e \int_0^T U_{s-}^e \left(f^\alpha(e_s, \hat{H}_s^\alpha) ds - dI_s - \frac{\hat{A}_s}{\gamma} \right) \right]. \quad (2.11)$$

Note that from the boundedness of the controls and of the intensity λ , we have

$$\begin{aligned} U_t^e &\leq V_t^\alpha e^{\gamma \left(\int_0^t \int_0^\infty (1-\alpha) \mu_J(ds, dx) + \int_0^t c(0) ds \right)}, \\ &= V_t^\alpha e^{\gamma((1-\alpha)J_t + c(0)t)}, \\ &=: \zeta_t < 0, \quad \text{and} \\ D_T^e &\geq \exp \left(\int_0^T \log \left(\frac{\lambda_\infty}{\lambda_0} \right) dN_s \right) =: \eta_T > 0. \end{aligned}$$

Since $A^{e,d} \geq 0$ and $f^\alpha(e_t, \hat{H}_t(x)) dt - dI_t \geq 0$, we obtain from Equation (2.11) and the monotone convergence theorem

$$\begin{aligned} 0 &\leq \gamma \sup_{e \in \Sigma} \mathbb{E} \left[\eta_T \int_0^T \zeta_{s-} \left(f^\alpha(e_s, \hat{H}_s^\alpha) ds - dI_s - \frac{\hat{A}_s}{\gamma} \right) \right], \\ &= -\gamma \inf_{e \in \Sigma} \mathbb{E} \left[\eta_T \int_0^T (-\zeta_{s-}) \left(f^\alpha(e_s, \hat{H}_s^\alpha) ds - dI_s - \frac{\hat{A}_s}{\gamma} \right) \right], \\ &= -\gamma \mathbb{E} \left[\eta_T \int_0^T (-\zeta_{s-}) \left(F^\alpha(\hat{H}_s^\alpha) ds - dI_s - \frac{\hat{A}_s}{\gamma} \right) \right], \\ &\leq 0, \end{aligned}$$

leading to

$$\begin{aligned} \hat{A} &\equiv 0, \\ I_t &= \int_0^t F^\alpha(\hat{H}_s^\alpha) ds. \end{aligned}$$

Equation (2.9) becomes

$$dY_t^\alpha = F^\alpha(\hat{H}_t^\alpha) dt - \int_0^\infty \hat{H}_t^\alpha(x) \tilde{\mu}_J^0(dt, dx).$$

Step 4. The aim of the following step is to prove the representation (3.4) of the value function and to find the optimal prevention effort. To do so, consider from Equation (3.3)

$$Y_0^\alpha = \int_0^T \int_0^\infty H_s^\alpha(x) \tilde{\mu}_J^0(ds, dx) - \int_0^T F^\alpha(H_s^\alpha) ds,$$

Hence for all $e \in \Sigma$

$$\begin{aligned} & \mathbb{E}^e \left[-\frac{1}{\gamma} e^{-\gamma(-\int_0^T \int_0^\infty (1-\alpha)x \mu_J(ds, dx) - \int_0^T c(e_s) ds)} \right] \\ &= \mathbb{E}^e \left[-\frac{1}{\gamma} e^{-\gamma(-\int_0^T \int_0^\infty (1-\alpha)x \mu_J(ds, dx) - \int_0^T c(e_s) ds + Y_0^\alpha - \int_0^T \int_0^\infty H_s^\alpha(x) \tilde{\mu}_J^0(ds, dx) + \int_0^T F^\alpha(H_s^\alpha) ds)} \right], \\ &= -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha} \mathbb{E}^e \left[e^{\gamma(\int_0^T \int_0^\infty ((1-\alpha)x + H_s^\alpha(x)) \mu_J(ds, dx) + \int_0^T (c(e_s) - \lambda_0 \int_0^\infty H_s^\alpha(x) dG(x) - F^\alpha(H_s^\alpha)) ds)} \right]. \end{aligned}$$

Denote for $t \in [0, T]$

$$\hat{Y}_t^\alpha := \int_0^t \int_0^\infty ((1-\alpha)x + H_s^\alpha(x)) \mu_J(ds, dx) + \int_0^t \left(c(e_s) - \lambda_0 \int_0^\infty H_s^\alpha(x) dG(x) - F^\alpha(H_s^\alpha) \right) ds.$$

By applying Ito's formula, we obtain

$$\begin{aligned} & \mathbb{E}^e \left[-\frac{1}{\gamma} e^{-\gamma(-\int_0^T \int_0^\infty (1-\alpha)x \mu_J(ds, dx) - \int_0^T c(e_s) ds)} \right] \\ &= -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha} \mathbb{E}^e \left[1 + \int_0^T \gamma e^{\gamma \hat{Y}_{s-}^\alpha} \left(c(e_s) - \lambda_0 \int_0^\infty H_s^\alpha(x) dG(x) - F^\alpha(H_s^\alpha) \right) ds \right. \\ &\quad \left. + \int_0^T \int_0^\infty e^{\gamma \hat{Y}_{s-}^\alpha} \left(e^{\gamma((1-\alpha)x + H_s^\alpha(x))} - 1 \right) \mu_J(ds, dx) \right], \\ &= -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha} \mathbb{E}^e \left[1 + \int_0^T \gamma e^{\gamma \hat{Y}_{s-}^\alpha} (f^\alpha(e_s, H_s^\alpha) - F^\alpha(H_s^\alpha)) ds + \int_0^T \int_0^\infty e^{\gamma \hat{Y}_{s-}^\alpha} \left(e^{\gamma((1-\alpha)x + H_s^\alpha(x))} - 1 \right) \tilde{\mu}_J^e(ds, dx) \right], \\ &= -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha} \mathbb{E}^e \left[1 + \int_0^T \gamma e^{\gamma \hat{Y}_{s-}^\alpha} (f^\alpha(e_s, H_s^\alpha) - F^\alpha(H_s^\alpha)) ds \right], \\ &\leq -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha}. \end{aligned}$$

By taking the supremum over $e \in \Sigma$, we then obtain

$$V^\alpha \leq -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha}.$$

It follows that the optimal effort is given by the minimizer $e^*(H^\alpha)$ and $V^\alpha = -\frac{1}{\gamma} e^{-\gamma Y_0^\alpha}$.

Step 5. In this last step, we want to prove the uniqueness of the solution to Equation (3.3). By repeating the same arguments from step 4, we have that for any two solutions (Y^α, H^α) , $(\bar{Y}^\alpha, \bar{H}^\alpha)$ of BSDE (3.3) and for any $t \geq 0$ we have

$$V_t^\alpha = -\frac{1}{\gamma} e^{-\gamma Y_t^\alpha} = -\frac{1}{\gamma} e^{-\gamma \bar{Y}_t^\alpha},$$

leading to $Y_t^\alpha = \bar{Y}_t^\alpha$ and $H_t^\alpha = \bar{H}_t^\alpha$, for $t \in [0, T]$. \square

Bibliographie

- [1] EUCH, O. E., MASTROLIA, T., ROSENBAUM, M., AND TOUZI, N. Optimal make-take fees for market making regulation. *Papers* 1805.02741, arXiv.org, May 2018.
- [2] JACOD, J., AND SHIRYAEV, A. *Limit theorems for stochastic processes*, vol. 288. Springer, 1987.
- [3] KARATZAS, I., AND SHREVE, S. E. Brownian motion. In *Brownian Motion and Stochastic Calculus*. Springer, 1998, pp. 47–127.
- [4] NEVEU, J. *Martingales à temps discret*. 1972.