Économie du risque et de l'assurance

Partie "Assurance"

Renaud Bourlès

3ème année 2015-2016

Ce cours s'adresse en particulier aux élèves de troisième année à l'École Centrale Marseille (option "Mathématiques, Management, Économie et Finance") et aux étudiants des Masters 2 "Aix-Marseille School of Economics" (spécialité Finance-Assurance) et ISMA (Ingénierie Statistique et Mathématiques Actuarielles).

Il s'inscrit dans le module "Économie du risque et de l'assurance" et complète la partie "Risque" enseignée par Dominique Henriet (cf. http://www.dominique.henriet-mrs.fr/). Le but de cet enseignement est de fournir les bases de l'analyse théorique de la tarification de l'assurance. Dans un premier temps, nous analyserons le cas simple dans lequel tous les individus sont identiques (Chapitre 1) puis nous étudierons les implications d'une possible hétérogénéité des assurés en termes d'exposition au risque, que celle-ci soit observable (Chapitre 2) ou non (Chapitre 3). Cette analyse sera complétée par l'étude des comportements des assurés en présence d'assurance (Chapitre 4) avant la mise en pratique des concepts via quelques exercices (Chapitre 5).

Sommaire

Bibliographie

- Henriet, D. and Rochet, J.-C., "Microéconomie de l'assurance", Economica, 1990.
- Picard, P., "Economic Analysis of Insurance Fraud", Handbook of Insurance, 2nd edition, G. Dionne. (ed), Springer, 2013.
- Rees, R. and Wambach, A., "The Microeconomics of Insurance", Foundations and Trends in Microeconomics, 4(1-2), 2008.
- Schlessinger, H., "The Theory of Insurance Demand", Handbook of Insurance, G. Dionne. (ed), Kluwer, 2000.

Chapitre 1

Le modèle à risque unique

Afin d'appréhender simplement la tarification des contrats d'assurance, nous considèrerons dans un premier temps que tous les assurés (potentiels) sont identiques et notamment qu'ils font face au même risque. Ce modèle, appelé modèle de Mossin, nous permettra d'introduire les notions d'assurance complète, d'assurance partielle et de franchise d'assurance. On en déduira ensuite comment la demande d'assurance évolue en fonction de son prix et du revenu de l'assuré.

1.1 Le modèle de Mossin

On considère un ménage, au revenu R qui est confronté à un risque d'accident – correspondant à une perte monétaire ou dommage D – pouvant survenir avec une probabilité p. Nous désignons par $u(.)$ la fonction d'utilité au sens de Von Neumann Morgenstern de ce ménage, et nous supposerons que u est strictement croissante, concave et de classe C^2. En l'absence d'assurance, l'espérance d'utilité d'un consommateur s'écrit donc :

$$V_0 = pu(R - D) + (1 - p)u(R)$$

Dans ce contexte, un contrat d'assurance correspond à un couple $z = (\Pi, q)$ s'interprétant comme une couverture q en cas de sinistre, en échange du versement d'une prime Π.

On fait l'hypothèse que $q \in [0, D]$. On parlera de "couverture complète" si $q = D$ et de franchise si $q < D$ (la franchise sera alors égale à $D - q$). L'utilité espérée d'un ménage achetant un tel contrat d'assurance vaut donc :

$$V(\Pi, q) = pu(R - D + q - \Pi) + (1 - p)u(R - \Pi)$$

Avec $R - D + q - \Pi \equiv A$ le revenu si le sinistre (l'accident) est survenu et $R - \Pi \equiv N$ le revenu si le sinistre n'est pas survenu.

L'hypothèse $q \leq D$ implique que le revenu en cas de sinistre est inférieur au revenu sans sinistre ($A \leq N$) et permet de s'assurer que les ménages n'ont pas intérêt à provoquer volontairement des sinistres ou à en déclarer exagérément (on reviendra sur ce type de phénomènes – dits de risque moral – lors du chapitre 4).

Afin de représenter ces contrats et l'espérance d'utilité correspondante (dans le plan (q, Π)), étudions dans un premier temps l'effet des termes du contrat sur $V(.)$:

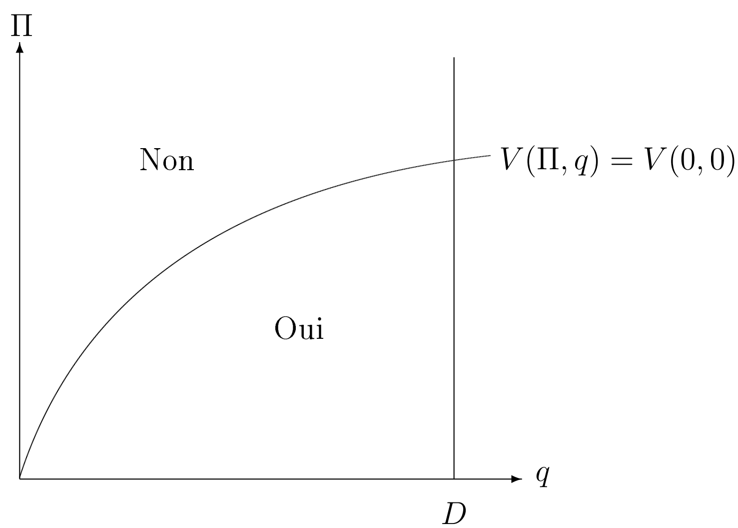
$$\begin{aligned} \frac{\partial V(\Pi, q)}{\partial \Pi} &= -\left[pu'(R - D + q - \Pi) + (1-p)u'(R - \Pi)\right] < 0, \text{ et} \\ \frac{\partial V(\Pi, q)}{\partial q} &= pu'(R - D + q - \Pi) > 0 \end{aligned}$$

Dans le plan (couverture, prime) les utilités espérées sont donc croissantes vers le Sud-Est. L'espérance d'utilité croit avec la couverture et décroit avec la prime. Par ailleurs, on peut remarquer que les courbes d'indifférences (c'est-à-dire l'ensemble des contrats procurant la même espérance d'utilité) sont croissantes et concaves dans le plan (q, Π). En effet, en utilisant le théorème des fonctions implicites $\left(\frac{\partial \Pi}{\partial q} = -\frac{\partial V / \partial q}{\partial V / \partial \Pi}\right)$, on obtient :

$$\begin{aligned} \frac{\partial \Pi}{\partial q} &= \frac{pu'(R - D + q - \Pi)}{pu'(R - D + q - \Pi) + (1-p)u'(R - \Pi)} > 0 \qquad (*) \\ \text{et } \frac{\partial^2 \Pi}{\partial q^2} &= \frac{p(1-p)u''(R - D + q - \Pi)u'(R - \Pi)}{\left[pu'(R - D + q - \Pi) + (1-p)u'(R - \Pi)\right]^2} < 0 \end{aligned}$$

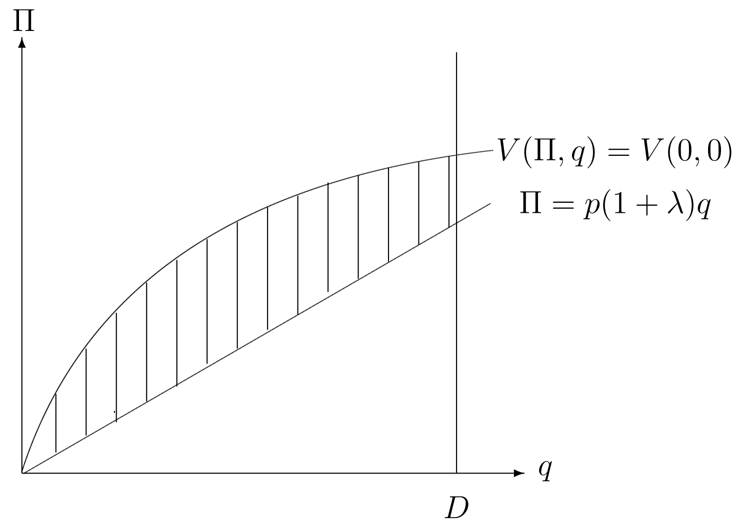
Un ménage acceptera alors un contrat $z = (\Pi, q)$ si il obtient une utilité espérée supérieure à celle qu'il a sans assurance. En traçant la courbe d'indifférence $V(\Pi, q) = V(0, 0)$ dans le plan (q, Π), on obtient donc l'ensemble des contrats acceptables pour les ménages.

Graphique 1



Étudions maintenant le comportement de l'assureur. Pour chaque police d'assurance vendue, le bénéfice est dans tous les cas (qu'il y ait un sinistre ou non) la prime Π, alors que le coût est la couverture q qu'il ne doit rembourser qu'en cas de sinistre. On supposera par ailleurs qu'il peut exister un coût de transaction λ (appelé aussi taux de chargement) de telle sorte que le coût effectif pour l'assureur s'écrit $(1+\lambda)q$ en cas de sinistre. Pour chaque prime vendue, le profit espéré de l'assureur s'écrit donc $\Pi - p(1+\lambda)q$. La zone dans laquelle la transaction d'assurance est bénéfique (à la fois pour l'assureur et pour l'assuré) peut être représentée dans le plan (q, Π) comme suit :

Graphique 2



Considérons d'abord le cas où $\lambda = 0$. On remarque alors que la pente de la courbe d'isoprofit de l'assureur (c'est-à-dire l'ensemble de contrats procurant le même niveau de profit espéré) est égale à p. Chaque euro de couverture offerte coûte p€ à l'assureur. Or, on remarque que le consentement à payer des ménages (la prime supplémentaire que le ménage est prêt à payer pour 1€ de couverture supplémentaire) est supérieur à p sauf en $q = D$ (où elle est exactement égale à p). En effet,

- En $q = D$,

$$u(A) = u(N) \Rightarrow \frac{\partial \Pi}{\partial q} = -\frac{\partial V/\partial q}{\partial V/\partial \Pi} = p$$

- alors que si $q < D$,

$$\begin{aligned} A < N \quad &\Rightarrow \quad u'(A) > u'(N) \text{ car } u \text{ est concave.} \\ &\Rightarrow \quad pu'(A) + (1-p)u'(N) < u'(A) \\ &\Rightarrow \quad \frac{\partial \Pi}{\partial q} = -\frac{\partial V/\partial q}{\partial V/\partial \Pi} > p \end{aligned}$$

Il y a donc, du point de vue de l'assureur, un gain à augmenter sa couverture jusqu'à $q = D$ puisque, pour $q < D$, l'assuré est prêt à payer plus que cela ne coûte à l'assureur.
On obtient ainsi, le résultat central du modèle de Mossin.

Théorème. *En l'absence de taux de chargement ($\lambda = 0$) l'assurance complète est optimale*

Remarque. *Ceci est vrai, autant en monopole (profit maximal) qu'en concurrence parfaite (profit nul)*

Dans le cas du monopole, l'assuré a ex-post le même revenu dans les deux états de la nature (sinistre / non sinistre) et a la même utilité espérée que sans assurance. On parle alors d'**équivalent certain**.

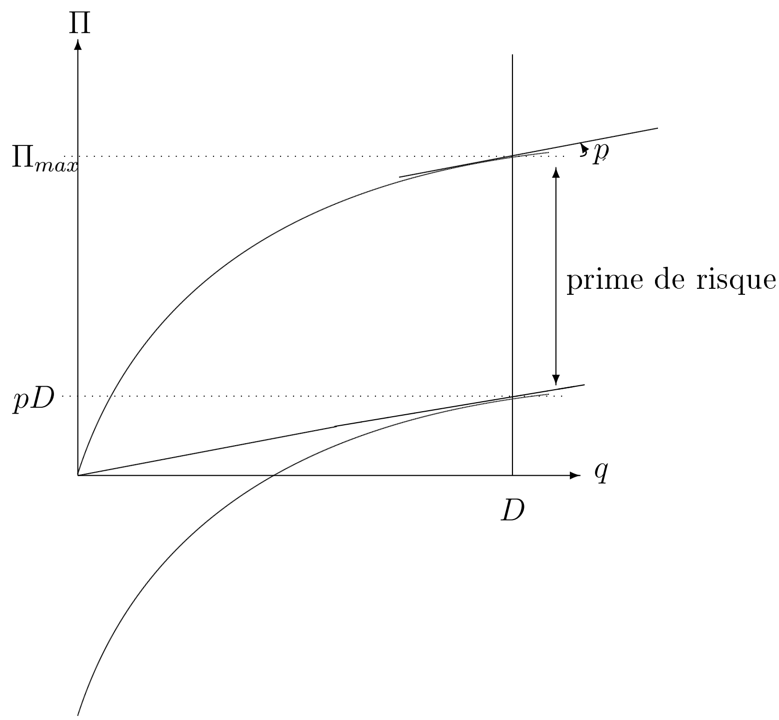
Définition. *L'équivalent certain d'un risque $\widetilde{x}$ est la constante :*

$$e(\widetilde{x}) \quad / \quad \mathbb{E}(u(\widetilde{x})) = u(e(\widetilde{x}))$$

Ici, on cherche donc $e/pu(R - D) + (1 - p)u(R) = u(e)$. On a alors $e = R - \Pi_{\max}$.

La **prime de risque** est alors définie comme la différence entre l'espérance de revenu et l'équivalent certain (c'est le montant que l'assuré est prêt à payer pour être assuré complètement). Ici l'espérance de revenu est égale à $\mathbb{E} = p(R - D) + (1 - p)R = R - pD$ et la prime de risque vaut $R - pD - R + \Pi_{\max} = \Pi_{\max} - pD$.

Graphique 3



Si on considère maintenant que le taux de chargement λ est strictement positif, c'est-à-dire qu'il existe des coûts de transaction, la pente de la courbe d'isoprofit (le coût d'1€ de couverture supplémentaire pour l'assureur) est supérieure à p (égale à $(1 + \lambda)p$) et l'optimum est atteint avec $q < D$.

Théorème. *On peut compléter le théorème précédent :*
- *En l'absence de taux de chargement ($\lambda = 0$) l'assurance complète est optimale ($q = D$)*
- *Si le taux de chargement est strictement positif ($\lambda > 0$), seule une couverture partielle est optimale ($q < D$)*

Remarque. *Le modèle peut être simplifié en considérant que la prime est une fonction linéaire de la couverture : $\Pi = \pi q$, ce qui est notamment le cas en concurrence ($\pi = (1 + \lambda)p$). Sous cette simplification, le choix de l'assuré revient à choisir le montant de couverture qui maximise son utilité espérée :*

$$V(q) = pu(R - D + (1 - \pi)q) + (1 - p)u(R - \pi q)$$

c'est-à-dire, qu'il choisit q^* tel que :

$$V'(q^*) = p(1 - \pi)u'(R - D + (1 - \pi)q^*) - (1 - p)\pi u'(R - \pi q^*) = 0 \tag{CPO}$$

(la concavité de $u(\cdot)$ assurant la condition du second ordre). Ainsi, on obtient

$$u'(R - D + (1 - \pi)q^*) = \frac{\pi}{p}\frac{1 - p}{1 - \pi}u'(R - \pi q^*)$$

Et on retrouve le théorème précédent :
- *si $\pi = p$ (c'est-à-dire $\lambda = 0$ dans le cas de la concurrence) alors $q^* = D$*
- *si $\pi > p$ (c'est-à-dire $\lambda > 0$ dans le cas de la concurrence) alors $q^* < D$ (puisque $u(\cdot)$ est concave)*

Cette simplification nous permet d'étudier le lien entre la demande optimale d'assurance (q^*) et les variables exogènes du modèle telles que le revenu R, la probabilité d'accident p ou la taille de la perte D, voire même le prix de l'assurance (π, rendu exogène par la simplication). En effet, la condition d'optimalité (CPO) peut s'écrire :

$$F(q; p, D, R, \pi) \equiv p(1 - \pi)u'(R - D + (1 - \pi)q^*) - (1 - p)\pi u'(R - \pi q^*) = 0$$

avec $\frac{\partial F}{\partial q} < 0$.

Ainsi, en utilisant le théorème des fonctions implicites, on est capable d'étudier l'effet de p, D, R et π sur q^*.

1.2 Effet richesse

Analysons d'abord comment la couverture optimale varie en fonction du revenu R. On a :

$$\frac{\partial q^*}{\partial R} = -\frac{\partial F/\partial R}{\partial F/\partial q^*}$$

et

$$\frac{\partial F}{\partial R} = p(1-\pi)u''(A) - \pi(1-p)u''(N)$$

où $A \equiv R - D + (1-\pi)q$ et $N = R - \pi q$. Ainsi, $\frac{\partial F}{\partial R}$, et donc $\frac{\partial q^*}{\partial R}$, ne peuvent être signés en toute généralité. Cependant, en utilisant la condition d'optimalité (CPO), on obtient :

$$p(1-\pi) = \pi(1-p)\frac{u'(N)}{u'(A)}$$

on peut alors écrire :

$$\frac{\partial F}{\partial R} = \pi(1-p)u'(N)\left(\frac{u''(A)}{u'(A)} - \frac{u''(N)}{u'(N)}\right)$$

On a donc $\frac{\partial F}{\partial R} \geq 0$ si et seulement si $-\frac{u''(N)}{u'(N)} \geq -\frac{u''(A)}{u'(A)}$. Comme $N \geq A$ $\forall q \leq D$, on obtient le résultat suivant :

Proposition. *La couverture optimale est une fonction croissante du revenu (et l'assurance est un bien normal) si l'aversion pour le risque est croissante ou constante avec le revenu. Au contraire, si l'aversion au risque est strictement décroissante, la couverture optimale est une fonction strictement décroissante du revenu et l'assurance est un bien inférieur.*

Étant donné que la dernière hypothèse est généralement celle choisie pour modéliser les comportements en univers risqué, il semblerait que l'assurance soit un bien inférieur. L'intuition est assez simple : si la propension à prendre du risque augmente avec le revenu, alors toutes choses égales par ailleurs, la demande d'assurance devrait diminuer. Cependant il semble que ce ne soit pas le cas dans la réalité où la plupart des études statistiques montrent que la demande d'assurance croit avec le revenu. Cette observation n'est toutefois pas nécessairement en contradiction avec le résultat théorique précédent puisqu'il n'est pas établi “toutes choses égales par ailleurs”. En effet, il est fort probable que dans le monde réel, la taille de la perte D soit elle aussi croissante avec le revenu. Comme, toutes choses égales par ailleurs, la demande d'assurance est croissante avec la taille du sinistre ($\partial F/\partial D = -p(1-\pi)u''(A) > 0$ et $\partial q^*/\partial D = -\frac{\partial F/\partial D}{\partial F/\partial q} > 0$), l'observation empirique précédente peut être rationalisée dans le présent modèle.

1.3 Effet prix

Étudions maintenant comment varie la demande d'assurance en fonction du prix par unité de couverture : π.

$$\frac{\partial F}{\partial \pi} = -pu'(A) - (1-p)u'(N) + q\left[(1-p)\pi u''(N) - p(1-\pi)u''(A)\right]$$

Cette fois encore, l'effet est ambigu. Ainsi, on ne peut pas être certain que la demande d'assurance décroisse quand la prime par une unité de couverture augmente. Cependant, on peut remarquer que $\partial q^*/\partial \pi$ peut s'écrire :

$$\frac{\partial q^*}{\partial \pi} = -\frac{\partial F/\partial \pi}{\partial F/\partial q} = \frac{pu'(A)+(1-p)u'(N)}{\partial F/\partial q} + q^*\frac{\partial F/\partial R}{\partial F/\partial q} = \frac{pu'(A)+(1-p)u'(N)}{\partial F/\partial q} - q\frac{\partial q^*}{\partial R}$$

c'est-à-dire qu'on peut isoler un "effet substitution" (négatif) et un "effet richesse". Ainsi, si $\frac{\partial q^*}{\partial R} > 0$, alors $\frac{\partial q^*}{\partial \pi} < 0$. En utilisant la proposition précédente, on a donc :

Proposition. *La demande d'assurance décroit avec la prime par unité de couverture si l'aversion (absolue) au risque est constante ou croissante. Cependant, si l'aversion au risque est (suffisamment) décroissante, la demande peut augmenter avec le prix (l'assurance est alors un bien de Giffen).*

Encore une fois, l'intuition de ce résultat est relativement simple. En effet, l'augmentation de la prime par unité de couverture a pour effet d'augmenter le prix relatif de la richesse dans l'état A ("accident") par rapport à celui de l'état N ("pas d'accident"), ce qui – à utilité constante – a tendance à diminuer la demande de couverture. Cependant, la hausse de la prime à aussi pour effet d'appauvrir les assurés et donc d'augmenter leur demande d'assurance si l'aversion au risque est décroissante avec la richesse.

Il est par ailleurs possible de trouver une autre condition suffisante à la décroissance de q^* en π en écrivant $\frac{\partial F}{\partial \pi}$ comme :

$$\begin{aligned}
\frac{\partial F}{\partial \pi} &= pu'(A)\left[-1 - \frac{(1-\pi)qu''(A)}{u'(A)}\right] + (1-p)u'(N)\left[-1 + \frac{q\pi u''(N)}{u'(N)}\right] \\
&= pu'(A)\left[-1 - \frac{Au''(A)}{u'(A)}\right] + (1-p)u'(N)\left[-1 - \frac{Nu''(N)}{u'(N)}\right] \\
&\quad + pu''(A)(R-D) + (1-p)u''(N)R
\end{aligned}$$

Ainsi, si le coefficient relatif d'aversion au risque $\left(-\frac{cu''(c)}{u'(c)}\right)$ est partout inférieur à 1 (ce qui est plutôt infirmé par les études récentes[1]) alors la demande d'assurance décroit avec le "taux de prime".

1. cf. Meyer, D. and Meyer, J. (2005), "Relative Risk Aversion : What Do We Know ?", *The Journal of Risk and Uncertainty*, 31(3), 243-262 ou Chetty, R. (2006), "A New Method of Estimating Risk Aversion", *The American Economic Review*, 96(5), 1821-1834.

Encore une fois, les exercices de statique comparée menés ici supposent que les autres variables exogènes du modèle restent constantes. Notamment, les variations de "primes" considérées précédemment sont supposées ne pas être liées à des variations de la probabilité d'accident. Toutefois, via la théorie de l'offre d'assurance exposée en début de chapitre, les deux variables ne sont certainement pas indépendantes. Par ailleurs, on montre aisément qu'à taux de prime constant, une augmentation de la probabilité de sinistre augmente la demande d'assurance ($\frac{\partial F}{\partial p} > 0$). Ainsi, en prenant en compte la réponse de l'offre (via π), l'augmentation de la probabilité d'assurance aura un effet généralement ambigu sur le montant de couverture achetée.

Chapitre 2

La différenciation des produits

La réalité du marché de l'assurance est bien entendu plus complexe que celle décrite par le modèle Mossin, avec notamment un grand nombre de contrats et des tarifications différentes. Par la suite, on s'attachera donc à comprendre les raisons de cette diversité en introduisant tour à tour de l'hétérogénéité parmi les agents et de l'asymétrie d'information. Nous étudierons ainsi comment une compagnie d'assurance prend en compte le fait que ses assurés soient différents, mais également comment elle s'y prend pour repérer ces différences.

2.1 L'introduction d'hétérogénéité dans le modèle de Mossin

Considérons le modèle suivant avec une population de ménages soumis à un même risque de dommage, représenté par une perte monétaire D. Les ménages diffèrent uniquement par leur probabilité d'accident p_i ; ils sont par ailleurs identiques, avec la même fonction d'utilité VNM u et la même richesse initiale R. Cette hétérogénéité sur la probabilité de sinistre affecte à la fois le coût d'un contrat d'assurance pour l'assureur et les préférences des assurés :

$$V(p_i, z) = p_i u(R - D + q - \Pi) + (1 - p_i)u(R - \Pi)$$

Si on suppose que la compagnie et l'assuré possèdent une information complète sur ces probabilités de sinistre et que le taux de chargement est nul, alors on peut aisément étendre les résultats de la section précédente :

- En concurrence parfaite, ($\max_z V(p_i, z)$ sous contrainte de profit nul), pour une couverture q le consommateur de type i devra payer $\Pi = p_i q$ et chaque consommateur se couvrira complètement $q = D$
- En situation de monopole, la compagnie propose au ménage de type i le contrat qui maximise son profit parmi ceux que le ménage a intérêt à accepter ($\max_{q,\Pi} \Pi - p_i q \;/\; V(p_i, z) \geq V_0$). On montre facilement que la solution correspond à l'assurance complète : $q = D$ à un prix tel que la contrainte est saturée.

Dans les deux cas, il y a autant de tarifs que de type de clients.

Le raisonnement précédent présuppose toutefois que la compagnie connaisse parfaitement la probabilité de sinistre de tous ses assurés. Ce n'est bien sûr pas le cas en réalité. Les assureurs ont toutefois développé des méthodes dites de "scoring" permettant d'évaluer ces probabilités de sinistres en fonction des caractéristiques observables des assurés (âge, sexe, revenu, CSP, puissance de la voiture assurée dans le cas de l'assurance automobile...).

2.2 Mesurer les probabilités de sinistre individuelles : le "scoring"

Le "scoring" est une méthode statistique permettant d'assigner une probabilité de sinistre ex-ante à un assuré se présentant pour la première fois. Cette méthode permet donc d'estimer (du côté de l'assureur) les p_i du modèle précédent. En se basant sur l'historique des contrats passés, la compagnie d'assurance évalue l'incidence de chacune des variables observables sur la probabilité de sinistre.

Pour cela, on définit Y_i la variable indicatrice de sinistre dans l'année pour l'individu i ($Y_i = 1$ si l'assuré i a eu un sinistre, $Y_i = 0$ sinon) et X_i le vecteur de ses caractéristiques. On souhaite alors connaître l'effet de chacune des caractéristiques X_{it} sur Y_i. Le caractère binaire de Y_i nous empêche cependant d'appliquer ici les méthodes de régressions linéaires classiques, c'est-à-dire de chercher le vecteur de coefficients β, tel que

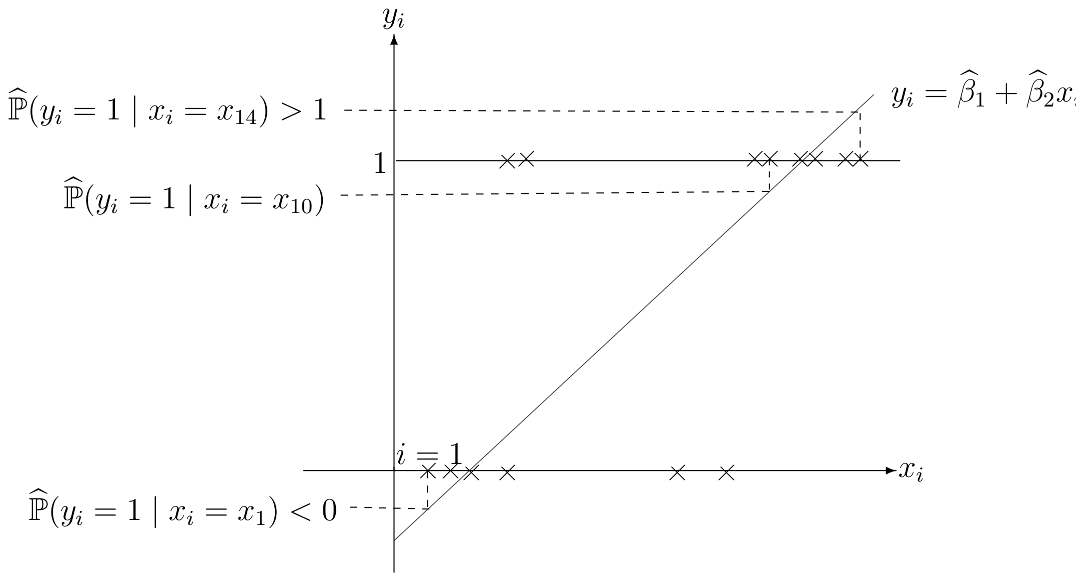
$$Y_i = X_i'\beta + u_i, \qquad \text{avec } u_i \sim_{iid} \mathcal{N}(0,1)$$

En effet même si ce modèle peut s'interpréter en termes de probabilités :

$$\left.\begin{array}{ll} \mathbb{E}(Y_i \mid X_i) & = \mathbb{E}(X_i'\beta \mid X_i) + \underbrace{\mathbb{E}(u_i \mid X_i)}_{=0} = X_i'\beta, \text{ et} \\ \mathbb{E}(Y_i \mid X_i) & = 1.\mathbb{P}(Y_i = 1 \mid X_i) + 0.\mathbb{P}(Y_i = 0 \mid X_i) \\ & = \mathbb{P}(Y_i = 1 \mid X_i) \end{array}\right\} \Rightarrow X_i'\beta = \mathbb{P}(Y_i = 1 \mid X_i)$$

rien ne contraint ici la probabilité estimée (c'est-à-dire le score) $\widehat{\mathbb{P}}_i = \mathbb{P}(\widehat{Y_i = 1 \mid} X_i) = X_i'\widehat{\beta}$ à être comprise entre 0 et 1 (cf. Graphique 4, dans le cas où X_{it} est unidimentionnel).

Graphique 4



C'est pourquoi on ne spécifie pas, dans les modèle de scoring, une relation linéaire (entre Y_i et les X_{it}). Pour répondre à cette problématique (et contraindre la probabilité estimée à être comprise entre 0 et 1) on cherche donc une fonction $F(X_i'\beta)$ telle que :

- $\lim_{v \to +\infty} F(v) = 1$
- $\lim_{v \to -\infty} F(v) = 0$
- F est continuement différentiable et $\frac{dF(v)}{dv} > 0$

On spécifie alors $\mathbb{E}(Y_i \mid X_i) = F(X_i'\beta)$ et $\frac{\partial P(Y_i=1|X_i)}{\partial X_{it}} = \frac{\partial F(X_i'\beta)}{\partial X_{it}} = f(X_i'\beta)\beta_t$

Deux solutions pour F ont été particulièrement développées :

- Le modèle Probit, qui utilise la fonction de répartition de la loi normale centrée réduite :

$$\begin{aligned}\mathbb{P}(Y_i = 1 \mid X_i) = F(X_i'\beta) &= \Phi(X_i'\beta) \\ &= \int_{-\infty}^{X_i'\beta} \frac{1}{\sqrt{2\pi}} e^{-t/2} dt\end{aligned}$$

- Le modèle Logit, qui utilise la fonction de répartition de la loi logistique :

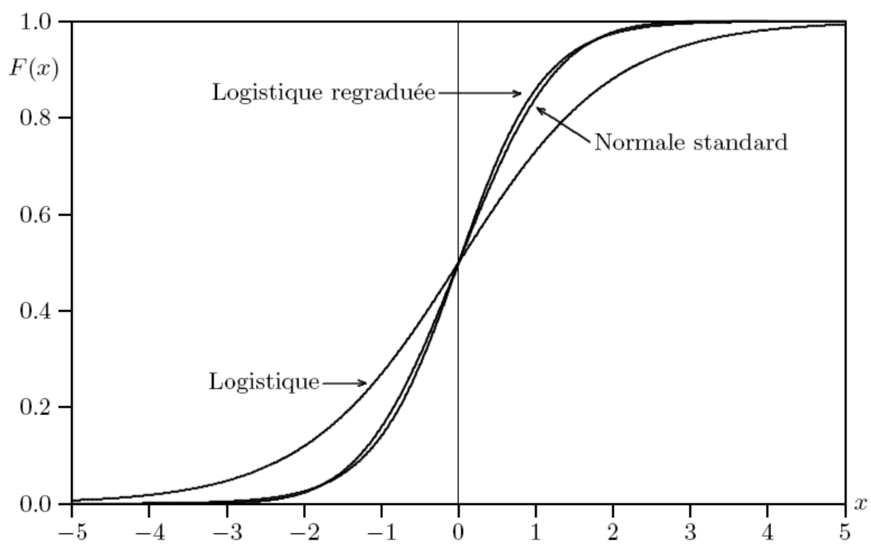
$$\begin{aligned}\mathbb{P}(Y_i = 1 \mid X_i) = F(X_i'\beta) &= \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)} \\ &= \frac{1}{1 + (\exp(X_i'\beta))^{-1}} = \frac{1}{1 + (\exp(-X_i'\beta))}\end{aligned}$$

Afin d'analyser les caractéristiques de ses modèles et de comprendre leurs similitudes, il est utile de rappeler quelques caractéristiques des lois normale et logistique.

- Si $Z \sim \mathcal{N}(0,1)$, sa fonction de répartition est notée $\Rightarrow F_Z(z) = \Phi(z)$ et sa densité vaut $f_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-t/2}$. On obtient alors $E(Z) = 0$ et $\text{Var}(Z) = 1$. Si la loi normale centrée réduite a l'avantage d'être symétrique et d'avoir été très largement étudiée, elle présente l'inconvénient d'avoir une fonction de répartition ne possédant pas d'expression explicite (mais uniquement une écriture sous forme d'intégrale).

- Au contraire, si $Z \sim$ Logistique, sa fonction de répartition est parfaitement déterminée et s'écrit $F_Z(z) = \frac{e^z}{1+e^z}$. On a par ailleurs $f_Z(z) = \frac{e^z}{(1+e^z)^2}$ avec $f_z(z) = F_Z(z)(1 - F_Z(z))$ (propriété qui va se révéler très utile), $E(Z) = 0$ et $\text{Var}(X) = \frac{\pi^2}{3}$. La loi logistique possède donc l'avantage d'être très proche de la loi normale centrée réduite tout en ayant une fonction de répartition possèdant une expression analytique simple (cf. Graphique 5, où la loi logistique a également été réduite).

Graphique 5



En plus d'être statistiquement simple, les modèles Logit et Probit peuvent être justifiés théoriquement.

Le modèle Probit peut ainsi être obtenu à partir d'un modèle comprenant une variable Y_i^* non observée, ou latente. Supposons que $Y_i^* = X_i\beta + u_i$ où $u_i \sim iid\mathcal{N}(0,1)$ mais que nous n'observons que le signe de Y_i^* , qui détermine la valeur de la variable binaire Y_i selon la relation $Y_i = 1$ si $Y_i^* > 0$ et $Y_i = 0$ si $Y_i^* \leq 0$. Alors

$$\begin{aligned}\mathbb{P}(Y_i = 1) &= \mathbb{P}(Y_i^* > 0) = \mathbb{P}(X_i\beta + u_i > 0) = 1 - \mathbb{P}(u_i \leq -X_i'\beta) \\ &= 1 - \Phi(-X_i'\beta) = \Phi(X_i'\beta)\end{aligned}$$

Le modèle Logit peut également être obtenu avec une variable latente, cette fois avec des erreurs suivant une loi logistique. Il peut toutefois être encore plus facilement dérivé en supposant que les caractéristiques observées (X_i) ne determinent pas p_i mais $\ln(p_i/(1-p_i))$:

$$\ln\left(\frac{p_i}{1-p_i}\right) = X_i'\beta$$

On obtient bien alors :

$$p_i = \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}$$

2.3 Estimation des modèles de scoring

Une fois le modèle choisi, afin de calculer les "scores" il convient d'estimer le vecteur β par la méthode du maximum de vraisemblance.

Comme

$$Y_i = \begin{cases} 1 \\ 0 \end{cases}$$

il existe deux types de contributions à la vraisemblance :

- $i/Y_i = 1 : L_i(\beta) = \mathbb{P}(y_i = 1 \mid X_i) = F(X_i'\beta)$
- $i/Y_i = 0 : L_i(\beta) = \mathbb{P}(y_i = 0 \mid X_i) = 1 - F(X_i'\beta)$

La vraisemblance s'écrit donc

$$\forall i, L_i(\beta) = \mathbb{P}(y_i = 1 \mid X_i)^{Y_i}.\left[1 - \mathbb{P}(y_i = 1 \mid X_i)\right]^{1-Y_i} = F(X_i'\beta)^{Y_i}.\left[1 - F(X_i'\beta)\right]^{1-Y_i}$$

et

$$L(\beta) = \prod_{i=1}^{n} F(X_i'\beta)^{Y_i}.\prod_{i=1}^{n}\left[1 - F(X_i'\beta)\right]^{1-Y_i}$$

ou plus simplement (en se ramenant à la log-vraissemblance)

$$l(\beta) = \sum_{i=1}^{n}\left[Y_i \ln(F(X_i'\beta)) + (1-Y_i)\ln\left(1 - F(X_i'\beta)\right)\right]$$

Ainsi, le maximum de vraisemblance est atteint en β tel que :

$$\begin{aligned}
\frac{\partial l(\beta)}{\partial \beta} &= \sum_{i=1}^{n}\left[Y_i \frac{1}{F(X_i'\beta)} f(X_i'\beta)X_i - (1-Y_i)\frac{1}{1-F(X_i'\beta)} f(X_i'\beta)X_i\right] = 0 \\
&\Leftrightarrow \sum_{i=1}^{n}\left[\frac{Y_i - F(X_i'\beta)}{F(X_i'\beta)(1-F(X_i'\beta))}\right] f(X_i'\beta)X_i = 0
\end{aligned}$$

Ce qui se simplifie dans le cas du modèle Logit (pour lequel on $f_z(z) = F_Z(z)(1-F_Z(z))$) en :

$$\begin{aligned}
&\sum_{i=1}^{n}\left[Y_i - F(X_i'\beta)\right] X_i = 0 \\
\Leftrightarrow &\sum_{i=1}^{n}\left[Y_i - \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right] X_i = 0
\end{aligned}$$

Remarque. *Plutôt que d'estimer la probabilité qu'un sinistre (au moins) survienne ou non dans l'année, il est également possible de prendre en compte le fait que plusieurs sinistres puissent survenir la même année pour le même assuré. On estime alors une probabilité pour chaque nombre de sinistres possible ($Y_i = 0, 1, ..., J$), en utilisant :*

- *soit un modèle Logit multivarié :*
  On veut alors

  $$\begin{cases} \ln\left(\frac{\mathbb{P}(Y_i=j|X_i)}{\mathbb{P}(Y_i=l|X_i)}\right) = X_i'(\beta_j - \beta_l)\forall j \forall l \\ \sum_{j=0}^{L}\mathbb{P}(Y_i = j \mid X_i) = 1 \end{cases}$$

  ce qui donne

  $$\begin{cases} \mathbb{P}(Y_i = J \mid X_i) = \frac{1}{1+\sum_{j=0}^{J-1}\exp\left[X_i'(\beta_j-\beta_J)\right]} \\ \mathbb{P}(Y_i = j \mid X_i) = \frac{\exp\left[X_i'(\beta_j-\beta_J)\right]}{1+\sum_{l=0}^{J-1}\exp\left[X_i'(\beta_l-\beta_J)\right]} \forall j = 0, ..., J-1 \end{cases}$$

- *soit un modèle de Poisson : $\mathbb{P}(Y_i = k) = \frac{\lambda_i^k}{k!}e^{-\lambda_i}$ avec $\lambda_i = X_i'\beta$ espérance et variance du nombre d'accidents sur la période pour l'individu i*

Chapitre 3

Les critères inobservables

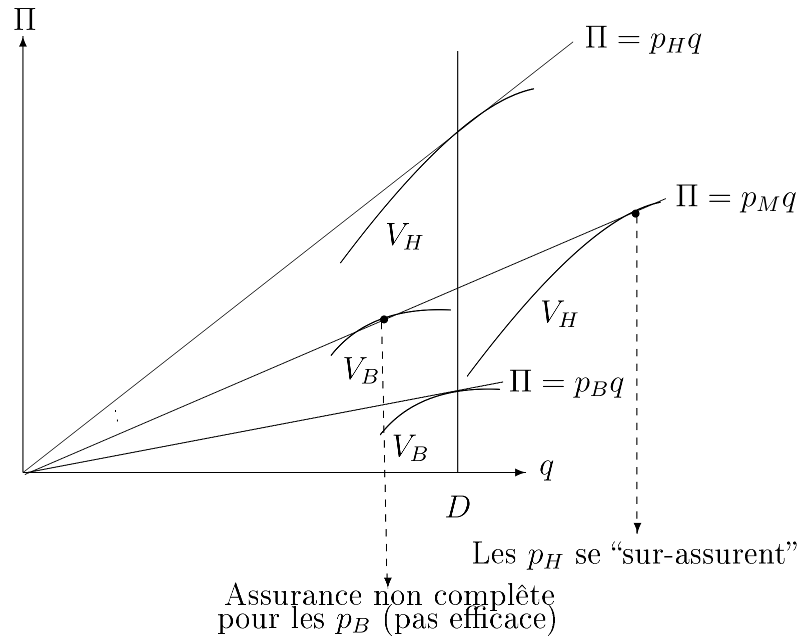
Malgrè les techniques de scoring, il restera toujours de l'hétérogénéité inobservée c'est-à-dire des critères non observés par l'assureur qui impactent la probabilité de dommage. Dans ce chapitre on cherche à analyser les conséquences de cette hétérogénéité inobservée et les méthodes que peuvent utiliser les assureurs pour la prendre en compte. On supposera plus précisément que les assurés peuvent être exposés (de manière exogène) différemment au risque – c'est-à-dire avoir une probabilité de dommage différente – sans que l'assureur ne puisse le savoir ex-ante. On se placera toutefois toujours dans le cas où l'assuré connait parfaitement cette probabilité de sinistre.

3.1 Le problème d'anti-sélection

Afin d'analyser les conséquences du fait que l'assureur n'arrive pas à discerner les assurés, plaçons nous dans une situation où l'assureur ne connait que le risque moyen, noté p_M. En situation de concurrence (et en l'absence de taux de chargement), l'assureur propose alors une prime $\Pi = p_M q$. Cette prime est trop élevée pour les assurés moins exposés au risque que la moyenne $p_i < p_M$: elle est supérieure à leur concentement à payer pour l'assurance complète. Ceux-ci décident donc de ne pas s'assurer ou de s'assurer partiellement (en fonction des contrats disponibles). Au contraire, cette prime est inférieure à ce que les individus les plus exposés au risque ($p_i > p_M$) seraient prêts à payer pour l'assurance complète. Ceux-ci aimeraient même s'assurer plus que complètement ($q > D$).

Dans le cas simple où il y a deux types de risque (ou d'assuré) : p_H et p_B (avec $p_B < p_H$) en proportion λ_H et $\lambda_B = 1 - \lambda_H$ respectivement ($p_M = \lambda_H p_H + \lambda_B p_B$), cette situation peut être représentée comme suit :

Graphique 6



Du fait de l'hétérogénéité inobservée, la compagnie d'assurance se retrouve donc avec un profit espéré négatif (la prime actuarielle étant calculée sur la base d'une situation où tous les assurés achètent la même quantité d'assurance). Dans le cas où seuls des contrats d'assurance complète sont proposés, celle-ci se retrouve même à n'assurer que les individus les plus risqués. C'est pourquoi on parle dans ce cas d'**anti-sélection**, les contrats proposés ne "sélectionnant" que les individus les moins "rentables" pour l'assureur.

Remarque. *Il est à noter qu'une telle situation ne se produit que très rarement dans la réalité, où on observe généralement l'inverse, c'est-à-dire des situations dans lesquelles les individus les moins exposés au risque sont "correctement" assurés alors que les plus exposés sont plutôt mal assurés (voire pas assurés du tout). Ceci peut être expliqué soit par le fait que l'asymétrie d'information est en réalité dans l'autre sens (l'assureur connaît mieux le risque que l'assuré*[2]*) ; soit par le fait que les différences d'exposition au risque sont dues à des différences de comportements et que les individus les plus tolérants face au risque sont à la fois ceux qui achètent le moins d'assurance et qui prennent le plus de risque*[3]*. On étudiera plus précisément cette notion de comportement réduisant le risque dans le chapitre suivant.*

2. voir à ce sujet B. Villeneuve (2000), "The Consequences for a Monopolistic Insurer of Evaluating Risk Better than Customers : The Adverse Selection Hypothesis Reversed", *Geneva Papers on Risk and Insurance Theory*, 25, 65-79

3. voir à ce sujet D. de Meza et D. Webb (2001), "Advantageous selection in insurance markets", *The RAND Journal of Economics*, 32(2), 249-262

3.2 L'auto-sélection : le modèle de Rothschild et Stiglitz

Afin de lutter contre ces problèmes d'anti-sélection, il est nécessaire que l'assureur puisse différencier les contrats offerts aux individus "haut-risque" (dont la probabilité de sinistre est p_H) et "bas-risque" (dont la probabilité de défaut est p_B). Comme elle ne peut pas différencier entre les (deux) types d'individus, elle va construire les contrats tels que les assurés se différencient d'eux même. On parlera alors d'**auto-sélection**. Chaque type d'assuré / de risque choisira de manière optimale le contrat qui lui est destiné.

Définition. *On appelle tarif auto-sélectif un couple de contrat* (z_B, z_H) *tel que*

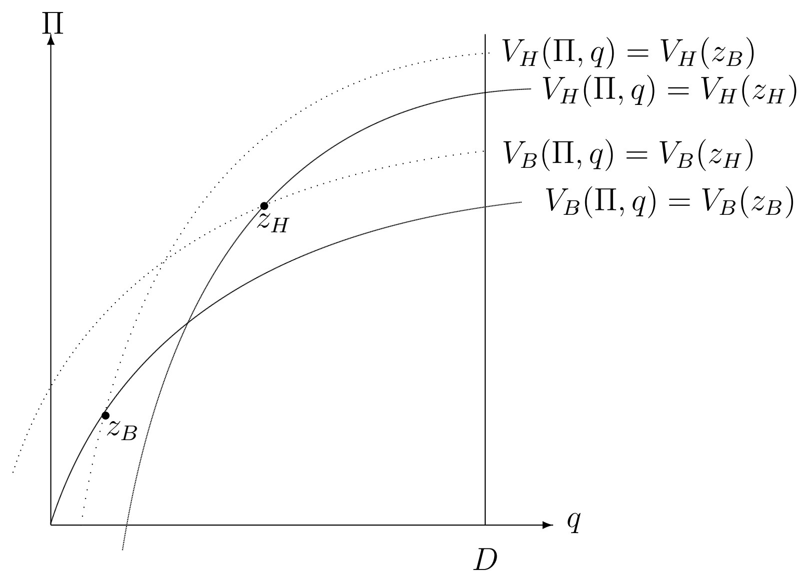
$$\begin{cases} V_H(z_H) \geq V_H(z_B) \\ V_B(z_B) \geq V_B(z_H) \end{cases}$$

c'est-à-dire (Π_H, q_H, Π_B, q_B) *tel que*

$$\begin{cases} p_H u(R - D + q_H - \Pi_H) + (1 - p_H) u(R - \Pi_H) \geq p_H u(R - D + q_B - \Pi_B) + (1 - p_H) u(R - \Pi_B) \\ p_B u(R - D + q_B - \Pi_B) + (1 - p_B) u(R - \Pi_B) \geq p_B u(R - D + q_H - \Pi_H) + (1 - p_B) u(R - \Pi_H) \end{cases}$$

Les individus de type H, les "haut risque" préfèreront alors le contrat z_H, alors que les "bas risque" (type B) préfèreront le contrat z_B.

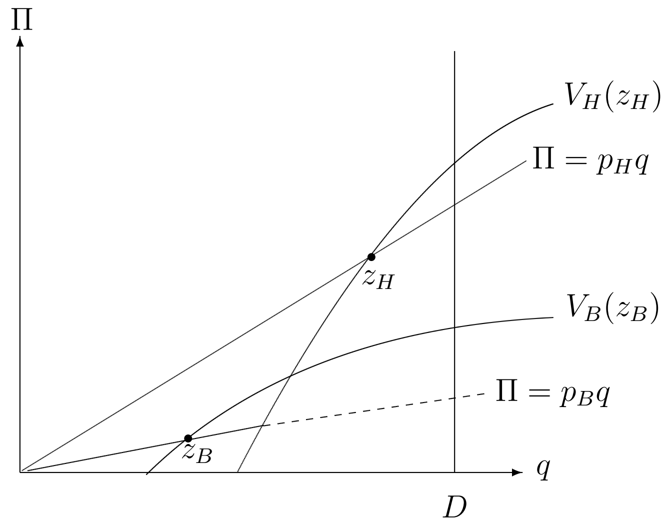
Graphique 7



Dans le cadre d'une concurrence parfaite, à l'équilibre, le profit espéré doit être nul **pour chaque contrat**. On parlera alors de tarif neutre.

Définition. *Un tarif neutre est un tarif auto-sélectif pour lequel le profit espéré est nul pour chaque contrat :* $\Pi_i = p_i q_i \; \forall i = B, H$

Graphique 8



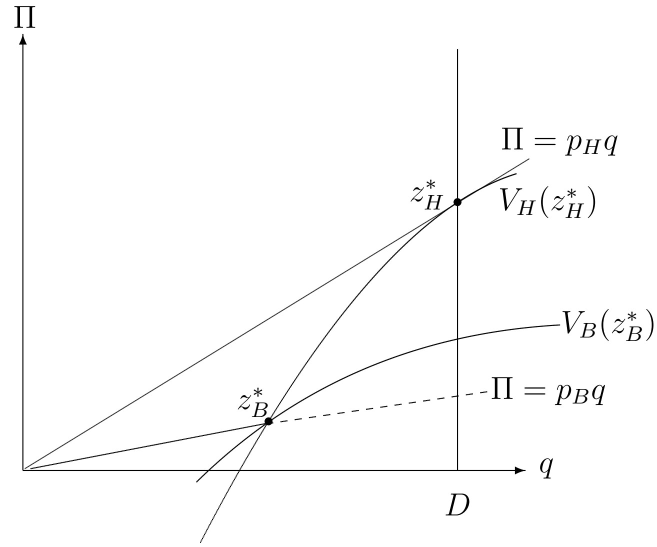
Dans la représentation graphique ci-dessus, (z_B, z_H) définit bien un tarif neutre. Cependant, ce ne serait plus le cas si z_B se trouvait dans la zone pointillée (dans ce cas les "haut risque" préfèreraient z_B à z_H) ou si il existait une intersection entre la courbe $V_B(z_B)$ et la droite $\Pi = p_H q$ (alors, une firme pourrait attirer à la fois les "haut risque" et les "bas risque" tout en faisant un profit strictement positif).

Il existe une infinité de tarifs neutres, cependant le théorème de Rothschild et Stiglitz met en évidence qu'un seul domine tous les autre au sens de Pareto (aucun autre tarif neutre n'est mutuellement avantageux pour les deux types).

Théorème. ***(Rothschild et Stiglitz, 1976)*** *Il existe un tarif neutre unique qui domine au sens de Pareto tous les tarifs neutres. Il est défini par :*

$$\forall i \quad \Pi_i^* = p_i q_i^*$$
$$q_H^* = D \qquad et \qquad V_H(z_H^*) = V_H(z_B^*)$$

Graphique 9



On peut donc parler du "meilleur tarif neutre" qui est ainsi obtenu en offrant une assurance parfaite aux "haut risque", et en proposant aux "bas rique" une franchise juste suffisante pour éviter que les "haut risque" ne soient attirés par leur contrat : z_B^*. Il y a donc un effet externe supporté par les bons risques, puisque contrairement aux "bas risque", les "haut risque" se voient proposer le contrat qu'ils préfèrent parmi les contats neutres.

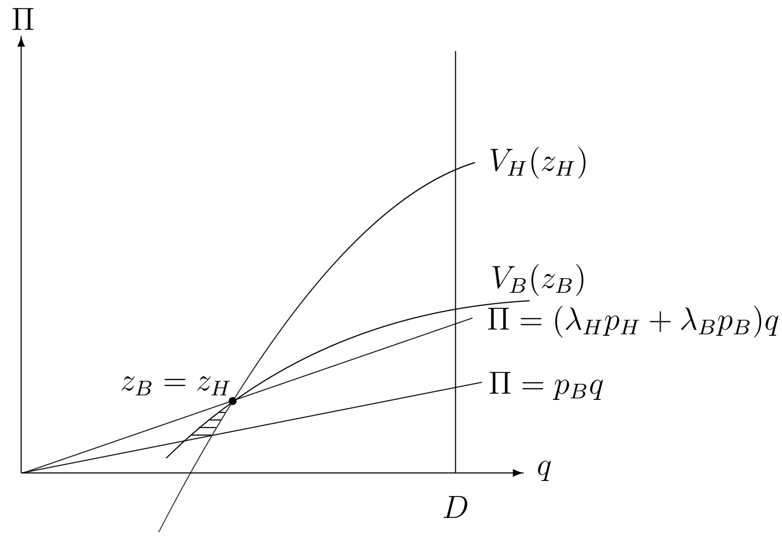
Ce tarif, cette allocation, semble être un bon candidat à l'équilibre, dans le sens où aucun assuré n'a alors intérêt à changer de contrat et aucun assureur ne peut proposer de contrat mutuellement avantageux (d'après le théorème précédent). Afin de vérifier que ce "meilleur tarif neutre" est bien l'équilibre, il est cependant nécessaire de définir la notion d'équilibre étudiée, c'est-à-dire l'évolution du "jeu" en dehors de l'équilibre. Le concept d'équilibre le plus simple est celui définit par Rothschild et Stiglitz en 1976 (on étudiera par la suite des concepts alternatifs d'équilibre) :

Définition. *Un équilibre au sens de Rothschild et Stiglitz est un ensemble de contrats* $z = (z_B, z_H)$ *non déficitaires, tel qu'il n'existe pas de contrat qui, s'il était proposé conjointement à* z *ferait un profit positif.*

Théorème. *Si un équilibre au sens de Rothschild et Stiglitz existe, alors c'est* $z^* = (z_B^*, z_H^*)$

On a vu que z^* était le "meilleur" tarif neutre, c'est-à-dire le meilleur tarif auto-sélectif (ou séparateur). z^* attirerait donc tous les assurés (et ferait un profit nul) si il était proposé conjointement à n'importe quel autre tarif séparateur. Aucun autre tarif séparateur ne peut donc être équilibre au sens de Rothschild et Stiglitz. Afin de démontrer le théorème précédent il suffit donc de montrer qu'aucun contrat rassemblant (où $z_B = z_H$) ne peut être un équilibre ; c'est-à-dire que pour tout contract rassemblant, il existe un contrat qui ferait un profit positif, si proposé conjointement.

Graphique 10



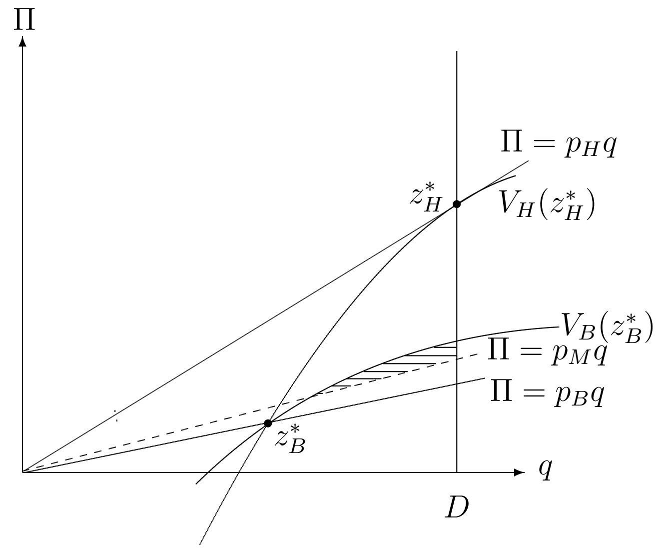
Partons d'un contrat "rassemblant" (pooling). Un assureur peut dévier vers un contrat profitable avec une couverture et une prime plus faible (dans la zone hachurée du graphique 10) en attirant uniquement les "bas risque". On parlera alors d'écrémage. Tous les contrats rassemblant étant sujet à l'écrémage par un contrat séparateur, seul un tarif séparateur peut être équilibre. On a donc z^* comme seul candidat à l'équilibre.

Cependant, il peut exister des situations dans lesquelles aucun équilibre n'existe, c'est-à-dire dans lesquelles un contrat peut faire un profit positif tout en étant proposé conjointement à z^*.

Théorème. *Aucun équilibre au sens de Rothschild et Stiglitz n'existe si la proportion de "bas risque" est trop forte ($\lambda_B > k$).*

On a montré que seul le contrat (z_B^*, z_H^*) pouvait être un équilibre au sens de Rothschild et Stiglitz. Pour qu'un tel équilibre existe, il faut toutefois qu'aucun contrat proposé en plus de (z_B^*, z_H^*) ne fasse un profit positif. Comme z^* est le "meilleur" des tarifs séparateurs cela ne peut être le cas que pour un équilibre rassemblant. On peut montrer – comme illustré dans le graphique 11 – que dans les cas où la proportion de "bas risque" est (trop) grande, un contrat rassemblant peut attirer les assurés et faire un profit positif. Ce sera notamment le cas quand, graphiquement, la droite de profit nul d'un assureur proposant un contrat rassemblant ($\Pi = (\lambda_H p_H + \lambda_B p_B)q = p_M q$) sera très proche de celle d'un assureur proposant un contrat aux "bas risque" uniquement ($\Pi = p_B q$).

Graphique 11



Ainsi, en se plaçant dans la zone hachurée (située entre la droite $\Pi = p_M q$ et la courbe d'indifférence $V_B(\Pi, q) = V_B(z_B^*)$) un assureur pourra faire un profit positif (on est au dessus de $\Pi = p_M q$) et attirer les deux types de consommateurs (on est au Sud-Est de $V_B(\Pi, q) = V_B(z_B^*)$ et de $V_H(\Pi, q) = V_H(z_H^*)$). Cependant, on a vu que n'importe quel contrat rassemblant (même dans cette zone) était sujet à l'écrémage. Ainsi, un équilibre rassemblant n'existera pas à cause d'un écrémage profitable et un équilibre séparateur n'existera pas non plus quand le poids des "haut risque" est trop faible. Cela vient du fait que, comme on l'a vu, le meilleur contrat séparateur entraine une perte de bien être pour les "bas risque" (avec une faible couverture) afin de faire payer aux "haut risque" la prime maximale par unité de couverture (obtenue avec l'assurance complète). Cette stratégie ne sera donc profitable que si la proportion de "haut risque" est élevée.

3.3 L'existence de l'équilibre

L'absence d'équilibre au sens de Rothschild-Stiglitz dans le cas $\lambda_B > k$ a beaucoup intrigué les économistes et plusieurs solutions ont été proposées afin de rétablir l'existence d'un équilibre. Ces solutions concernent notamment (i) la définition de l'équilibre (Wilson 1977 [4]), (ii) la nature des contrats d'assurance (Picard 2014 [5]) ou (iii) la prise en compte de l'engagement limité, c'est-à-dire de la possibilité de faillite des assureurs (Mimra et Wambach 2015 [6]).

4. C. Wilson, C. (1977), "A model of insurance markets with incomplete information", Journal of Economic Theory, 16(2), 167-207.

5. P. Picard (2014), "Participating Insurance Contracts and the Rothschild-Stiglitz Equilibrium Puzzle", Geneva Risk and Insurance Review, 39, 153-175.

6. W. Mimra and A. Wambach (2015), "Endogenous Capital in the Rothschild-Stiglitz Model", mimeo

- Le premier développement, proposé par Wilson (1977), porte sur une modification du concept d'équilibre et la définition d'un équilibre dit "anticipatif". Dans le modèle proposé, tout entrant potentiel anticipe l'éventualité du retrait de contrats par l'entreprise en place.

  Définition. *Un équilibre au sens de Wilson est un ensemble de contrats non déficitaires, tel qu'il n'existe pas de contrat qui, s'il était proposé conjointement à z ferait un profit positif, après retrait des contrats devenus déficitaires (s'il y en a).*

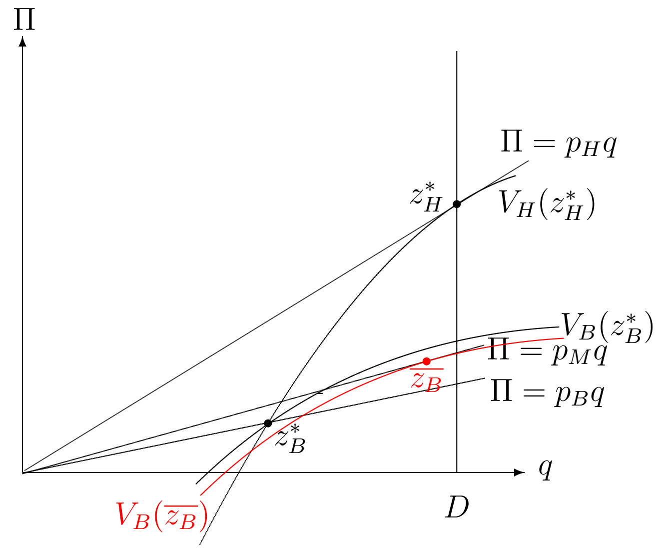
  Cette notion "adaptative" de l'équilibre permet de rétablir l'existence quelque soit la proportion de bas risque. En effet, dans ce cas, lorsque (z_B^*, z_H^*) est dominé par un contrat de pooling, l'écrémage (de ce dernier contrat) n'est plus profitable. Un contrat de pooling est alors équilibre. En effet, l'écrémage du contrat rassemblant (c'est-à-dire le fait d'attirer uniquement les "bas risque") conduit à la "faillite" des assureurs proposant ce contrat (puisque celui-ci se trouve en dessous de la droite $\Pi = p_H q$). Après l'élimination de ces contrats devenus déficitaires, l'assureur qui a "dévié" se retrouve donc avec les deux types d'assurés. Comme son contrat se trouve nécessairement en dessous de la droite $\Pi = p_M q$ (afin d'écrémer les "bas risque", cf. Graphique 10), le fait de proposer un tel contrat n'est plus profitable "après retrait des contrats devenus déficitaires". Selon la définition de Wilson, un équilibre rassemblant existe donc quand $\lambda_B < k$

  Théorème. *Il existe toujours un équilibre au sens de Wilson, le tarif correspondant est unique :*
  - *si la proportion de "bas risque" est faible ($\lambda_B \leq k$) c'est le meilleur tarif neutre z^**
  - *si la proportion de "bas risque" est trop élevé ($\lambda_B > k$) c'est le tarif rassemblant avec profit espéré nul qui satisfait le plus les "bas risque", c'est-à-dire $\{\overline{z_B}, \overline{z_B}\}$, où $\overline{z_B}$ est défini par :*
  $$V_B(\overline{z_B}) = \max_q V_B(p_M q, q) \text{ avec } p_M = \lambda_H p_H + \lambda_B p_B$$

- Plus récemment, Picard (2009) a proposé une solution n'impliquant non pas un changement du concept d'équilibre, mais plus de liberté dans la construction des contrats d'assurance. Il a notamment montré qu'en acceptant que les contrats soient participatifs, le problème d'existence disparaissait. Alors, le fait que les assurés reçoivent une partie des bénéfices ou payent une partie des pertes agit comme une menace implicite contre l'écrémage. En effet, supposons que les "haut risque" aient choisi une police participative (ce qui revient à une subvention croisée par les "bas risque"). Alors, quand les "bas risque" changent d'assureur, la situation des "haut risque" se détèriore, ce qui rend difficile pour un assureur déviant (de l'équilibre) d'attirer les "bas risque" sans les "haut risque". Alors, un équilibre (au sens de Rothschild et Stiglitz) existe de nouveau lorsque la proportion de "bas risque" est forte : il s'agit d'un équilibre totalement participatif (profit totalement redistribué), type mutuel, correspondant en terme de richesses espérées à $(\overline{z_B}, \overline{z_B})$. Lorsque la proportion de "bas risque" est faible, le contrat d'équilibre est toujours (z_B^*, z_H^*) proposé dans un cadre non participatif.

- Dans le même ordre d'idée, Mimra et Wambach (2010) ont montré qu'un équilibre au sens de Rothschild et Stiglitz existait quelque soit la configuration lorsque les assurés prennent

Graphique 12



en compte la possibilité pour un assureur de faire faillite (sans changer la définition de l'équilibre). Encore une fois, cette modification rend l'écrémage non profitable. En effet, le départ des "bas risque" d'un contrat rassemblant détériore alors l'utilité espérée des "haut risque" en augmentant la probabilité de faillite de l'assureur proposant ce contrat. Par le même mécanisme que précédemment, le contrat supposé écrémer les "bas risque" devient alors attractif pour les "haut risque" et aucun écrémage n'est possible. On a donc comme équilibre (au sens de Rothschild et Stiglitz) : (z_B^*, z_H^*) si $\lambda_B \leq k$ et $(\overline{z_B}, \overline{z_B})$ si $\lambda_B > k$

Chapitre 4

Le risque moral

L'analyse précédente ne considère pas les actions que les assurés peuvent mettre en place afin de réduire individuellement le risque auquel ils font face. De telles actions sont pourtant primordiales dans l'analyse de la tarification de l'assurance puisqu'elles vont impacter l'espérance de profit de l'assureur, mais aussi parce que la souscription d'un contrat d'assurance peut modifier le comportement des assurés quant à ces actions. On considèrera ici plus particulièrement des actions qui peuvent réduire (i) la taille du sinistre D (par exemple le port de la ceinture de sécurité ou l'installation d'extincteurs), auquel cas on parlera d'auto-assurance, ou (ii) la probabilité du sinistre p (par exemple l'ABS, rouler prudemment) auquel cas on parlera d'auto-protection. Bien que comparables du point de vue de l'assuré, ces deux activités ont des conséquences différentes pour l'assureur. Si la fonction de coût de l'assureur est affectée par les activités d'auto-protection, elle ne l'est pas par celles d'auto-assurance, dans la mesure où on suppose que le montant du sinistre est observé ex-post par la compagnie.

4.1 L'auto-assurance et ses conséquences

Soit c le coût des activités d'auto-assurance, dont l'impact sur le montant du dommage est décrit par la fonction $D(.)$. Nous supposerons $D'(c) < -1$ (i.e. $(c + D(c))' < 0$) et $D'' > 0$, c'est-à-dire que les activités d'auto-assurance sont à rendements positifs mais décroissants.

Sans assurance, l'espérance d'utilité d'un ménage est alors :

$$V_0(c) = pu(R - D(c) - c) + (1 - p)u(R - c)$$

qui est concave sous les hypothèses précédentes. Le niveau d'auto-assurance c_0 choisit par le ménage non assuré est alors donné par la condition du premier ordre :

$$V_0{}'(c_0) = -pu'(R - D(c_0) - c_0)(1 + D'(c_0)) - (1 - p)u'(R - c_0) = 0$$

C'est-à-dire

$$-(D'(c_0) + 1) = \frac{1 - p}{p} \times \frac{u'(R - c_0)}{u'(R - D(c_0) - c_0)}$$

Considérons maintenant le cas où notre ménage peut acheter de l'assurance au prix unitaire π (par unité de couverture). En désignant par c son niveau d'auto-assurance et par q la quantité d'assurance qu'il achète, son espérance d'utilité vaut :

$$V(q,c) = pu(\underbrace{R - D(c) - c + (1-\pi)q}_{\equiv A}) + (1-p)u(\underbrace{R - c - \pi q}_{\equiv N})$$

Les décisions optimales de l'assuré sont alors déterminées par :

$$\begin{cases} \frac{\partial V}{\partial c}(q^*,c^*) & = pu'(A)(1+D'(c^*)) + (1-p)u'(N) = 0 \\ \frac{\partial V}{\partial q}(q^*,c^*) & = (1-\pi)pu'(A) - \pi(1-p)u'(N) = 0 \end{cases}$$

c'est-à-dire

$$-(D'(c^*)+1) = \frac{1-p}{p} \times \frac{u'(N)}{u'(A)} = \frac{1-\pi}{\pi}$$

$D(\cdot)$ étant convexe, on en déduit que c^* est une fonction croissante de π. Autrement dit, les dépenses d'auto-assurance augmentent (et la couverture diminue) quand l'assurance devient plus chère. **L'auto-assurance est donc un substitut de l'assurance**. Remarquons pour terminer que quand π vaut 1, q^* est égal à zéro (on voudrait même $q = -\infty$) et on retrouve le cas sans assurance étudié plus haut $c^* = c_0$. La croissance de c^* en π implique donc que pour tout π inférieur à 1, c^* est inférieur à c_0. **L'accès à un marché d'assurance diminue donc l'incitation à s'auto-assurer**.

4.2 Auto-protection et risque moral

En influant sur la probabilité de sinistre, les activités d'auto-protection influent sur l'espérance de coût d'un contrat pour l'assureur. L'idéal serait donc d'inclure dans les contrats des dispositions incitant les assurés à entreprendre effectivement ces activités qui réduisent la fréquence des sinistres. Cependant la plupart du temps, les activités d'auto-protection (par exemple : conduire prudemment) ne sont pas observables par l'assureur. Il est alors difficile de discerner (même avec du scoring) entre les problèmes d'anti-sélection (caractéristiques cachées) et de "risque moral" (action cachée).

On s'intéresse ici à un contexte de risque moral "pur", en supposant tous les assurés identiques. Le seul problème pour l'assureur est alors d'inciter ses assurés à diminuer leur probabilité d'accident en accomplissant des activités d'auto-protection qui ont un coût (psychologique ou monétaire). On supposera également ces activités inobservables, aussi bien ex-ante que ex-post.

Le seul instrument que peut alors utiliser la compagnie est le niveau de franchise, qui doit réaliser un compromis acceptable entre le désir de couverture de l'assuré et le souci de la compagnie d'inciter à cet effort.

On décrira ici le modèle le plus simple dans lequel le ménage a le choix de faire ou non, pour un coût en terme d'utilité c, un effort d'auto-protection $e = 1$ qui réduit la probabilité d'accident de $\overline{p}$ à $\underline{p}$. Alors l'utilité d'un ménage non assuré s'écrira :

$$V_0(e) = p(e)u(R-D) + (1-p(e))u(R) - c(e)$$

avec

$$e \in \{0,1\},\ p(0) = \overline{p} > p(1) = \underline{p} \text{ et } c(0) = 0 < c(1) = c$$

En considérant un contrat d'assurance z, composé d'une prime Π et d'un remboursement (couverture) en cas de sinistre q, l'utilité d'un assuré sera :

$$V(e,\Pi,q) = p(e)u(R - D - \Pi + q) + (1 - p(e))u(R - \Pi) - c(e)$$

Si l'effort était parfaitement observable, l'assureur pourrait le faire intervenir dans un contrat. Si on se restreint aux contrats non déficitaires ($\Pi \geq p(e)q$), le contrat optimal (de 1er rang) pour l'assuré est caractérisé par l'assurance complète ($q = D$) et la tarification actuarielle ($\Pi = p(e)q$).

Dans ce cas, un assuré fera l'effort d'auto-protection si :

$$\begin{aligned} & u(R - \underline{p}D) - c > u(R - \overline{p}D) \\ \Leftrightarrow\ & c < c^* = u(R - \underline{p}D) - u(R - \overline{p}D) \end{aligned}$$

Au contraire, un ménage non assuré fera l'effort si :

$$\begin{aligned} & \underline{p}u(R - D) + (1 - \underline{p})u(R) - c > \overline{p}u(R - D) + (1 - \overline{p})u(R) \\ \Leftrightarrow\ & c < \overline{c} = (\overline{p} - \underline{p})\left[u(R) - u(R - D)\right] \end{aligned}$$

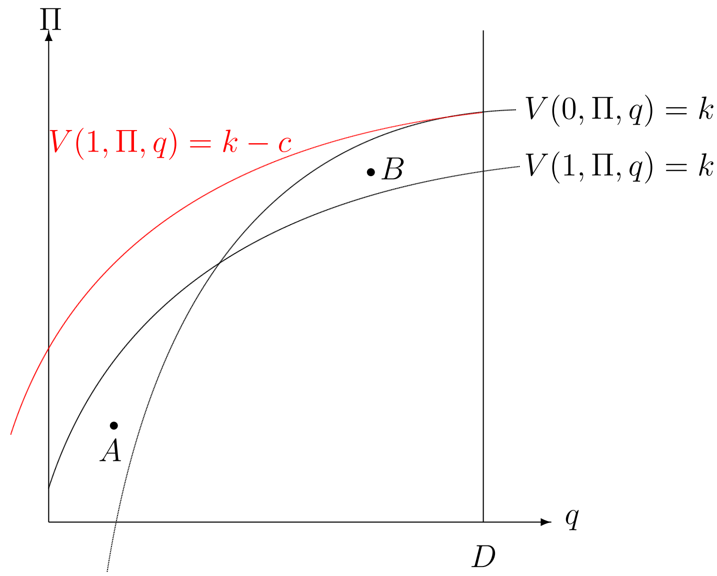
On peut remarquer qu'il n'y a aucune relation systématique entre $\overline{c}$ et c^*. Autrement dit, l'accès à un marché d'assureur peut ou peut ne pas augmenter les incitations à l'effort. Pour s'en convaincre, posons $\Delta = \overline{c} - c^*$. On a alors (comme $u(\cdot)$ est concave et d'après l'inégalité de Jensen) :

$$\begin{aligned} \lim_{\underline{p}\to 0} \Delta &= u(R - \overline{p}D) - \left[\overline{p}u(R - D) + (1 - \overline{p})u(R)\right] > 0 \text{ et} \\ \lim_{\overline{p}\to 1} \Delta &= \left[\underline{p}u(R - D) + (1 - \underline{p})u(R)\right] - u(R - \underline{p}D) < 0 \end{aligned}$$

Supposons maintenant que $c < \min\{\overline{c}, c^*\}$, et plaçons nous dans le cas où e n'est pas observable par l'assureur. Recherchons l'allure des préférences d'un assuré dans le plan (q,Π).
Pour cela on doit d'abord considérer l'ensemble des contrats z qui donnent à l'assuré le même niveau d'utilité k qu'il fasse l'effort ou non, c'est-à-dire $\{(\Pi,q)/V(1,\Pi,q) = k \text{ ou } V(0,\Pi,q) = k\}$. Ensuite, il faudra étudier pour chaque contrat dans cet ensemble, si l'assuré a intérêt à faire l'effort ou pas.
On remarque d'abord qu'en $q = D$ (couverture complète) l'utilité brute du coût de l'effort est indépendante du niveau d'effort : $V(e,\Pi,D) + c(e) = u(R - \Pi)\ \forall e$. Ainsi, en $q = D$, $V(1,\Pi,D) = V(0,\Pi,D) - c$. Par ailleurs, on sait d'après le chapitre 1, que la pente de la courbe $V(1,\Pi,q) = k$ est égale à $\underline{p}$ en $q = D$ alors que celle de $V(0,\Pi,q) = k$ est égale à $\overline{p}$. Enfin, on montre facilement toujours en utilisant le chapitre 1 (équation (*)), que la pente des courbes d'indifférence est croissante en p $\left(\left(\frac{\partial \Pi}{\partial q}\Big|_{V(\Pi,q)=k}\right)'_p > 0\right)$. La pente de la courbe $V(0,\Pi,q) = k$ est donc supérieure pour tout niveau de couverture q à celle de la courbe $V(1,\Pi,q) = k$.

Ainsi, la courbe d'indifférence $V(1,\Pi,q)=k$ sera d'abord dessus puis dessous $V(0,\Pi,q)=k$.

Graphique 13



Pour des faibles niveaux de couverture (point A), l'assuré préfèrera donc faire l'effort ($V(1,\Pi,q)>V(0,\Pi,q)$) alors que si la couverture est élevée, son espérance d'utilité sera supérieure si il ne fait pas l'effort (point B). On a donc la proposition suivante :

Proposition. *Pour chaque niveau de prime (Π), il existe un niveau de couverture seuil ($L(\Pi)$) tel que :*

- *l'assuré fera un effort si la couverture proposée est inférieure à ce seuil ($e=1$ si $q\leq L(\Pi)$) et*
- *ne le fera pas si la couverture proposée est supérieure ($e=0$ si $q>L(\Pi)$)*

De plus la fonction $L(\Pi)$ est croissante : plus la prime est élevée, plus le niveau de couverture au dessous duquel l'assuré fera l'effort est grand.

Démonstration. *Posons*

$$\begin{aligned} A(q,\Pi) &\equiv V(0,\Pi,q)-V(1,\Pi,q) \\ &= c-(\overline{p}-\underline{p})\left[u(R-\Pi)-u(R-\Pi+q-D)\right] \end{aligned}$$

L'assuré fait l'effort si et seulement si A est négatif. Or A est une fonction continue et strictement croissante de q. On peut donc définir $L(\Pi)$ par la relation :

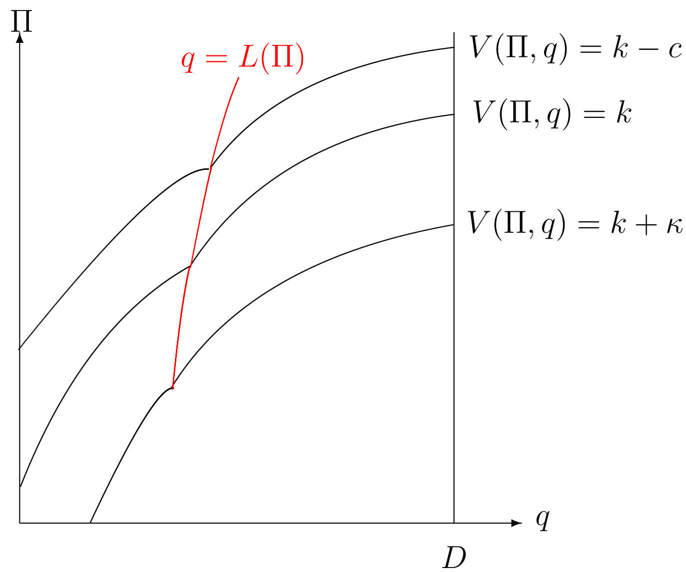
$$A(L(\Pi),\Pi)=0$$

On en déduit le résultat recherché. La croissance de $L(.)$ découle du théorème des fonctions implicites :

$$\begin{aligned}
& \frac{\partial A}{\partial q}(L(\Pi), \Pi) > 0 \textit{ déjà remarqué} \\
& \frac{\partial A}{\partial \Pi}(L(\Pi), \Pi) = -(\overline{p} - \underline{p})[u'(R - \Pi + L(\Pi) - D) - u'(R - \Pi)] < 0 \\
\Rightarrow \quad & \frac{dL}{d\Pi} = -\frac{\frac{\partial A}{\partial \Pi}(L(\Pi), \Pi)}{\frac{\partial A}{\partial q}(L(\Pi), \Pi)} > 0
\end{aligned}$$

Par ailleurs, le raisonnement précédent nous permet d'obtenir les courbes d'indifférences suivantes et de remarquer qu'elles possèdent des points anguleux.

Graphique 14



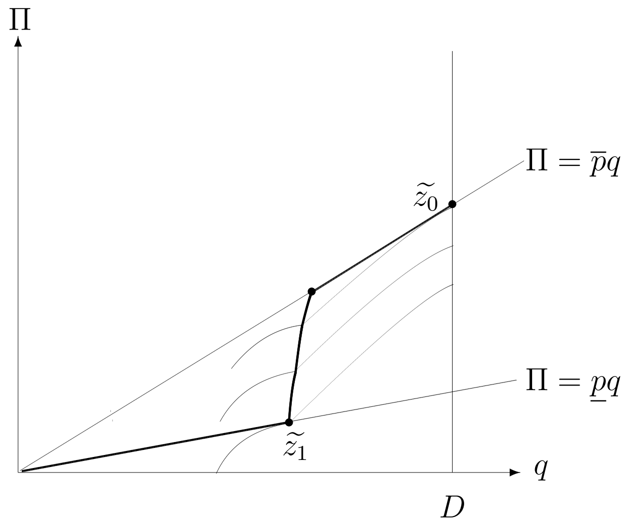
Maintenant que nous en savons plus sur les courbes d'indifférence des assurés, étudions les contrats optimaux dans un contexte où l'effort est inobservable.

Dans une économie où tous les assurés sont identiques, il est naturel de se restreindre aux allocations anonymes, qui correspondent à offrir le même contrat d'assurance à tous. D'après la proposition précédente, un tel contrat $z = (q, \Pi)$ ne sera réalisable que si :

$$\begin{cases} \Pi \geq \underline{p}q & \text{et} \quad q \leq L(\Pi) \quad \text{(contrats incitant à l'effort), ou} \\ \Pi \geq \overline{p}q & \text{et} \quad q \leq D \quad \text{(contrats n'incitant pas à l'effort)} \end{cases}$$

Comme les courbes d'indifférence, l'ensemble des contrats réalisables a donc aussi des points anguleux.

Graphique 15



Parmi ces contrats réalisables, on s'intéresse maintenant à celui qui est du point de vue des assurés optimal au second rang, c'est-à-dire au contrat qui maximise l'utilité d'un assuré ($V = \max_{e\in\{0,1\}} V(e,\Pi,q)$) sur l'ensemble des contrats réalisables. Ce serait le contrat proposé par des assureurs en concurrence parfaite si ils n'observaient pas l'effort fourni par les assurés. En utilisant les notions précédentes, on obtient facilement la proposition suivante.

Proposition. *Il existe une valeur seuil de côût de l'effort $\widetilde{c}$ (inférieure à c^*), telle que le contrat optimal de second rang $\widetilde{z}$ est caractérisé par :*

- *une assurance complète et une absence d'incitation à l'effort si le coût de l'effort dépasse ce seuil : $\widetilde{z} = \widetilde{z}_0 = (D, \overline{p}D)$ et $e = 0$ si $c > \widetilde{c}$*
- *une assurance partielle et de l'incitation à l'effort si le coût de l'effort est inférieur à ce seuil. Dans ce cas, le contrat optimal est le meilleur contrat (du point de vue des assurés) incitant à l'effort au taux actuariel. On a donc, si $c \leq \widetilde{c}$, $e = 1$ et $\widetilde{z} = \widetilde{z}_1 = (\underline{q}, \underline{\Pi})$ avec $(\underline{q}, \underline{\Pi})$ tels que $\underline{q} = L(\underline{\Pi})$ et $\underline{\Pi} = \underline{pq}$*

Démonstration. *$\widetilde{z}$ est forcément sur la frontière inférieure de l'ensemble des contrats réalisables. Cette frontière est constituée de 3 morceaux (nous avons supposé que $c < \overline{c}$ et donc que $L(0) < 0$)*

- *une partie de la droite $\Pi = \underline{p}q$ sur laquelle le maximum de V est situé en $\widetilde{z}_1$, intersection avec $q = L(\Pi)$*
- *une partie de la courbe $q = L(\Pi)$ sur laquelle le maximum de V est situé en $\widetilde{z}_1$*
- *une partie de la droite $\Pi = \overline{p}q$ sur laquelle le maximum de V est situé en $\widetilde{z}_0 = (D, \overline{p}D)$*

La proposition découle alors de la comparaison de $V(\widetilde{z}_0)$ et $V(\widetilde{z}_1)$, en remarquant que par construction en $\widetilde{z}_1$, l'assuré est indifférent entre faire l'effort et pas :

$$\begin{aligned} V(\widetilde{z}_1) &= \underline{p}u[R - D + q_1(1-\underline{p})] + (1-\underline{p})u(R - \underline{p}q_1) - c \\ &= \overline{p}u[R - D + q_1(1-\underline{p})] + (1-\overline{p})u(R - \underline{p}q_1) \end{aligned}$$

En particulier, q_1 est défini implicitement par :

$$\frac{c}{\overline{p}-\underline{p}} = u(R-\underline{p}q_1) - u(R-D+q_1(1-\underline{p}))$$

Le membre de droite de cette égalité étant décroissant par rapport à q_1, q_1 est lui aussi une fonction décroissante de c $\left(\frac{\partial q_1}{\partial c} < 0\right)$. Or nous avons :

$$V(\widetilde{z}_0) > V(\widetilde{z}_1) \Leftrightarrow u(R-\overline{p}D) > \overline{p}u[R-D+q_1(1-\underline{p})] + (1-\overline{p})u(R-\underline{p}q_1)$$

Le membre de droite de cette inégalité étant une fonction croissante de q_1, on a démontré la proposition. On a en effet, $\frac{\partial[V(\widetilde{z}_0)-V(\widetilde{z}_1)]}{\partial c} = \frac{\partial[V(\widetilde{z}_0)-V(\widetilde{z}_1)]}{\partial q_1}\frac{\partial q_1}{\partial c} > 0$, c'est-à-dire $V(\widetilde{z}_0) > V(\widetilde{z}_1)$ pour de grandes valeurs de c (on peut facilement voir que c'est le cas en $c = c^*$ pour lequel, par définition, les assurés préfèrent même $\widetilde{z}_0$ à $(\underline{p}D, D)$) et $V(\widetilde{z}_0) < V(\widetilde{z}_1)$ pour de petites valeurs de c (c'est notamment le cas en $c = 0$, où les agents font l'effort quelque soit le contrat).

On peut résumer les résultats précédents comme suit

coût		$\widetilde{c}$ c^*	
Premier rang	Couverture complète Effort maximal		Couverture complète Effort nul
Second rang	Couverture partielle Effort maximal	Couverture complète Effort nul	

Ainsi lorsque le coût de l'auto-protection est assez faible, la couverture partielle permet d'inciter l'assuré à l'effort optimal. En revanche si le coût de l'auto-protection est trop élevé, faire participer l'assuré à son risque est trop onéreux, il est alors préférable de l'assurer complètement quitte à ce qu'il abandonne toute activité d'auto-protection.

On peut par ailleurs montrer que lorsque les assureurs peuvent se garantir l'exclusivité (un assureur ne pourra alors pas assurer un client ayant déjà contracté auprès d'une autre compagnie), il existe un équilibre en contrat et un seul. Il est formé d'un contrat unique, le contrat de second rang précédent. L'hypothèse d'exclusivité nous permet d'éviter des situations dans lesquelles malgré la concurrence, les assureurs peuvent faire un profit positif (avec un effort maximal des assurés) en se plaçant sur la courbe $q = L(\Pi)$ (et $z \neq \widetilde{z}_1$)

4.3 Le risque moral ex-post : le cas des fraudes à l'assurance

On a considéré jusqu'ici le cas d'actions pouvant être entreprises par l'agent avant la réalisation du sinistre (ex-ante). Cependant, on peut également imaginer que le fait d'être assuré favorise des comportements ex-post (après la survenance d'un sinistre) visant à augmenter la taille de la perte (déclarée) voire à déclarer des sinistres non survenus. On parle alors de fraude à l'assurance ou de risque moral ex-post.

L'existence de telles fraudes et leur ampleur est par essence difficile à mesurer. Cependant, il a été estimé par diverses études que la fraude à l'assurance représentait entre 10% et 20% des sinistres déclarés. "The Insurance Information Institute" a même estimé que les fraudes à l'assurance ont couté aux assureurs américains près de 30 millions de dollars en 2004.

D'un point de vue théorique, la fraude à l'assurance devient possible dès que l'assuré a la possibilité de falsifier une déclaration que l'assureur ne peut pas facilement vérifier. Si par exemple, un assuré peut sans coût falsifier un sinistre et qu'il est impossible (ou infiniment coûteux) pour l'assureur de vérifier le sinistre ; un assuré (non contraint par des considérations morales) aura toujours intérêt à déclarer le sinistre donnant lieu au remboursement le plus élevé. Anticipant cela, l'assureur ne proposera qu'un montant unique de couverture, quelque soit la taille du sinistre. Si, au moins, la survenance d'un sinistre est vérifiable le montant de cette couverture sera positif. Dans le cas contraire, le marché de l'assurance disparaît totalement.

Afin de rendre le problème plus intéressant et plus réaliste, on se concentrera ici sur deux scenarii moins extrêmes. On étudiera d'une part le cas où la falsification par l'assuré se fait sans coût mais la vérification par l'assurance est possible (mais coûteuse) ; et d'autre part le cas où la vérification est impossible mais la falsification est coûteuse.

Vérification coûteuse et audit déterministe

On reprend le modèle précédent en rendant la taille de la perte aléatoire.

On considère donc un assuré au revenu R faisant face à un risque de sinistre D, distribué selon la fonction de répartition $F(D)$ (de support $[0, \overline{D}]$). On suppose bien entendu que l'événement $D = 0$ peut être atteint avec une probabilité non nulle, c'est-à-dire que D suit une distribution mixte avec une masse de probabilité en 0.

On suppose ici que la taille du sinistre D est connue uniquement de l'assuré mais que l'assureur peut la découvrir en payant un coût c. On parlera alors d'audit. L'assureur n'a donc pas la même information si il audite ou non. Si il décide de ne pas auditer le sinistre, il n'a d'information que sur la taille du sinistre déclarée par l'assuré, noté $\widehat{D}$. La couverture sera alors notée $q_N(\widehat{D})$. Au contraire, en cas d'audit, l'assureur connaît à la fois $\widehat{D}$ et D, la vraie valeur du sinistre. La couverture pourra dans ce cas dépendre à la fois de D et de $\widehat{D}$, et donc inclure une forme de punition si $D \neq \widehat{D}$. On notera alors la couverture : $q_A(D, \widehat{D})$. Comme précédemment, Π représentera la prime demandée par l'assureur.

Considérons d'abord le cas d'un audit déterministe, c'est-à-dire le cas où pour chaque niveau de sinistre (déclaré), l'assureur doit décider si il audit ou non. On note M l'ensemble des contrats audités. Dans ce cas, on obtient d'abord le résultat suivant

Lemme. *N'importe quel contrat est dominé au sens de Pareto (du point de vue de l'utilité espérée et du profit espéré) par un contrat incitatif (où $\widehat{D} = D$) pour lequel :*

- *les sinistres non audités sont tous couverts de la même façon : $q_N(D) = q_0 \ \forall D \in \overline{M}$*
- *les sinistres audités donnent lieu à une couverture plus élevée que les sinistres non audités $q_A(D, \widehat{D}) = q_A(D) > q_0 \ \forall D, \widehat{D} \in M$*

On peut démontrer le Lemme précédent en partant de n'importe quel contrat $\delta^0 = (\ M^0\ ,\ q_N^0(\widehat{D})\ ,\ q_A^0(D, \widehat{D})\ ,\ \Pi^0\ )$. Considérons tout d'abord l'espace $\overline{M^0}$. A l'intérieur de cet espace, l'assuré ne sera pas audité (et le sait). Notons $q_0 = \max_{D \in \overline{M^0}} q_N(D)$ et $D_0 = \arg \max_{D \in \overline{M^0}} q_N(D)$. Ainsi, q_0 est la couverture maximale qui peut être obtenue sans être audité et D_0 est le sinistre à déclarer pour obtenir cette couverture. Alors, il est clair que pour tout sinistre donnant lieu à une couverture inférieure à q_0, l'assuré a intérêt à déclarer D_0. Les mêmes paiements que pour δ^0 peuvent donc être obtenus en concervant la même prime Π^0 avec $q_N(D) = q_0 \forall D \in \overline{M}$ et $\overline{M} = \overline{M^0} \cup \{D / q_A(D, D) \leq q_0\}$, c'est-à-dire en diminuant les coûts d'audit. Par ailleurs, on peut facilement aboutir à un mécanisme incitatif (c'est-à-dire n'entraînant aucune fraude) dans la zone d'audit en fixant $q_A(D) = \max_{\widehat{D}} q_A^0(D, \widehat{D}) \ \forall D, \widehat{D} \in M$ (dans cette zone, le sinistre sera nécessairement audité et donc l'assureur en connaîtra toujours la vraie valeur). Il est cependant à noter que le mécanisme ne fonctionne que si, bien qu'il n'y ait jamais de fraude, l'assureur s'engage (de manière crédible) à auditer toutes les pertes dans l'espace M.

La force du Lemme précédent réside dans le fait qu'on puisse alors se limiter à l'analyse de tels contrats incitatifs. Ainsi, dans le cas de la concurrence, le contrat optimal sera le contrat (M , q_0 , $q_A(D)$, Π) qui maximise l'utilité espérée d'un assuré :

$$V = \int_M u(R - \Pi - D + q_A(D, D))dF(D) + \int_{\overline{M}} u(R - \Pi - D - q_0)dF(D)$$

sous la contrainte de participation de l'assureur (profit espéré non négatif)

$$\Pi - \int_M [q_A(D, D) - c]\, dF(D) - \int_{\overline{M}} q_0 dF(D) \geq 0$$

et la contrainte d'incitation (à dire la vérité) : $q_A(D, D) > q_0 \ \forall D \in M$.

On peut alors facilement montrer que :

(i) du fait de la concavité de $u(\cdot)$, il est optimal d'offrir une couverture minimale et constante quand les pertes sont petites et de couvrir plus largement les fortes pertes : $M =]m, \overline{D}]$ avec $m \in [0, \overline{D}]$, et

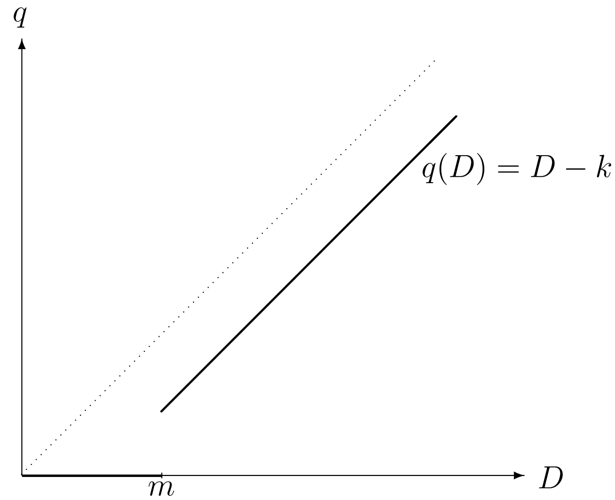
(ii) dans le cas où le sinistre est vérifié, la perte **marginale** est totalement assurée : $q_A(D) = D - k \ \forall D \in M$. En effet, la vérification systématique fait que l'assureur n'a pas besoin d'inciter l'assuré dans l'espace M.

Par ailleurs, on comprend facilement que si la survenance d'un sinistre ($D > 0$) n'est pas observable alors il est optimal de fixer $q_0 = 0$. On a donc le résultat suivant :

Proposition. *En présence d'audit déterministe, un contrat d'assurance optimal présente les propriétés suivantes :*

- *seules les pertes supérieures à un certain montant sont couvertes et systématiquement auditées ($M =]m, \overline{D}]$ et $q(D) = 0$ si $D \leq m$).*
- *quand le sinistre est couvert, la perte marginale est parfaitement couverte ($q(D) = D - k$ si $D > m$, avec $0 < k < m$)*

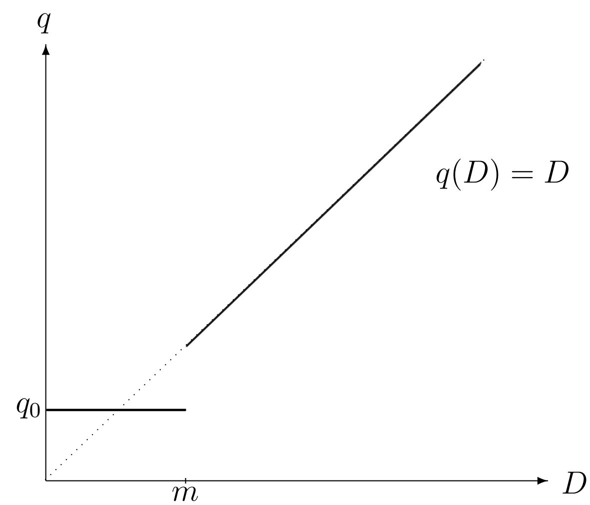
Graphique 16



Le fait que la perte totale ne soit pas totalement assurée ($k > 0$) provient du programme de maximisation. En effet, il en ressort que l'utilité marginale quand la perte est compensée doit être égale à l'espérance d'utilité marginale (ceci est dû au fait qu'une augmentation de la couverture dans un état doit être compensée par une augmentation de la prime dans tous les états). Comme les pertes ne sont pas compensées pour $D < m$, on sait que la richesse ex-post est alors inférieure à $R - \Pi$. Ainsi l'espérance d'utilité marginale est supérieure à $u'(R - \Pi)$ puisque $u(\cdot)$ est concave, et $R - \Pi - D + q_A(D)$ doit être inférieur à $R - \Pi$.

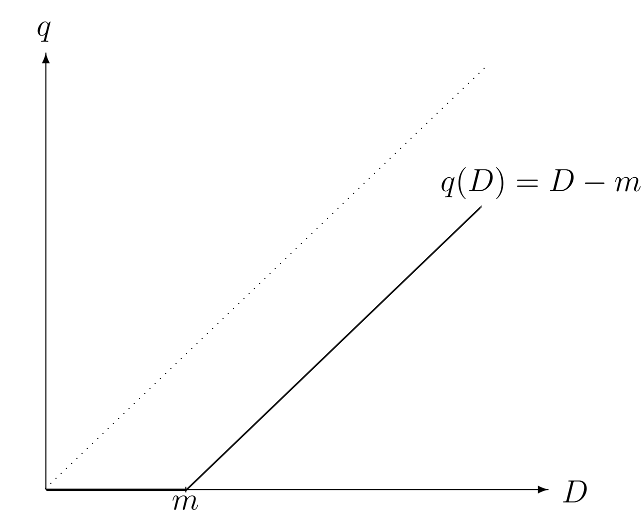
Remarque. • *Si il est possible de vérifier sans coût si un sinistre a eu lieu ou non ($D > 0$) alors $q(0) = 0$, $q(D) = q_0 > 0$ $\forall 0 < D \leq m$ et $q(D) = D$ $\forall D > m$, avec $0 < q_0 < m$. Comme, l'assureur n'est plus "contraint" à ne pas couvrir les faibles pertes ($D < m$), l'assurance peut être complète dans le cas où les sinistres sont vérifiés.*

Graphique 17



- *On retrouve un contrat classique avec franchise* ($q(D) = \sup\{0, D - m\}$ *avec* $m > 0$ *et* $M =]m, \overline{D}]$) *si on ajoute la possibilité pour l'assuré d'augmenter intentionnellement la taille du sinistre (et que l'assureur ne sait pas distinguer la partie du sinistre créée intentionnellement). En effet, on remarque facilement que les deux contrats décrits plus haut sont alors sujets à des augmentations intentionnelles de sinistre, notamment du fait de leur discontinuité.*

Graphique 18



Vérification coûteuse et audit aléatoire

On a considéré jusqu'ici que l'assureur devait choisir un ensemble de montants de sinistres déclarés qu'il s'engageait à toujours auditer. On peut également envisager qu'il choisit pour chaque niveau de sinistre déclaré une certaine probabilité d'audit $\gamma(\widehat{D})$ (ou ce qui revient au même une certaine part de sinistres à auditer). On ne peut alors plus avoir $q_A(D,\widehat{D}) = q_A(D)$ $\forall \widehat{D}$ puisque dans ce cas la fraude serait bénéfique en espérance. Il est donc nécessaire, pour que l'assuré ait intérêt à dire la vérité, que la fraude soit punie. Évidemment, le plus simple pour l'assureur serait d'infliger au fraudeur une punition infinie, ce qui garantirait l'absence de fraude. Cependant, cela est souvent impossible, notamment pour des raisons d'engagement limité (on ne peut pas demander à l'assuré plus que ce qu'il a). On fait donc l'hypothèse simplificatrice que $q_A(D,\widehat{D}) = -B$ $\forall \widehat{D} \neq D$. Sous cette hypothèse (on a en fait surtout besoin que la punition ne dépende pas de $\widehat{D}$), on peut montrer que, comme dans le cas d'un audit déterministe, l'attention peut être réduite aux contrats incitatifs. Ainsi, on se limite aux contrats tels que :

$$\begin{aligned}(1-\gamma(D))u(R-\Pi-D+q_N(D)) + \gamma(D)u(R-\Pi-D+q_A(D,D))\\ \geq (1-\gamma(\widehat{D}))u(R-\pi-D+q_N(\widehat{D}) + \gamma(\widehat{D})u(R-\Pi-D-B)\\ \forall D,\widehat{D} \in [0,\overline{D}]^2\end{aligned} \tag{IC}$$

Alors, en situation de concurrence, le contrat optimal est obtenu en maxisant l'espérance d'utilité de l'assuré :

$$V = \int_0^{\overline{D}} \left[(1-\gamma(D))u(R-\Pi-D+q_N(D)) + \gamma(D)u(R-\Pi-D+q_A(D,D))\right] dF(D)$$

sous les contraintes d'incitation (IC) et la contrainte de participation de l'assureur :

$$\Pi - \int_0^{\overline{D}} \left[(1-\gamma(D))q_N(D) + \gamma(D)\left(q_A(D,D)+c\right)\right] dF(D) \geq 0$$

La complexité de ce programme vient du fait qu'on ne peut pas identifier quelles contraintes d'incitation saturent à l'optimum. Cependant, supposer une fonction d'utilité CARA $\left(u(C) = -\frac{1}{\alpha}e^{-\alpha C}\right)$ simplifie grandement le problème. Dans ce cas le facteur $-\frac{1}{\alpha}e^{-\alpha(R-\Pi-D)}$ disparaît de (IC). En termes moins formels, le recours à une fonction d'utilité CARA fait disparaitre l'effet richesse du choix de frauder ou non, c'est à dire du choix entre les loteries : $[q_N(D),(1-\gamma(D)); q_A(D,D),\gamma(D)]$ et $[q_N(\widehat{D}),(1-\gamma(\widehat{D})); -B,\gamma(\widehat{D})]$. Une conséquence directe de ces simplifications est qu'il est alors suffisant, pour inciter l'assuré à dire la vérité, que la contrainte (IC) soit satisfaite pour le niveau de couverture le plus bas. C'est généralement le cas en $D=0$ où $q_N(0) = q_A(0,0) = 0$. La contrainte d'incitation devient (en remplaçant $\widehat{D}$ par D) donc : $\forall D \in [0,\overline{D}]$,

$$u(R-\Pi) \geq (1-\gamma(D))u(R-\pi+q_N(D)) + \gamma(D)u(R-\Pi-B) \tag{IC.CARA}$$

Cela revient à avoir :

$$\gamma(D) = \frac{u(R-\Pi+q_N(D)) - u(R-\Pi)}{u(R-\Pi+q_N(D)) - u(R-\Pi-B)}$$

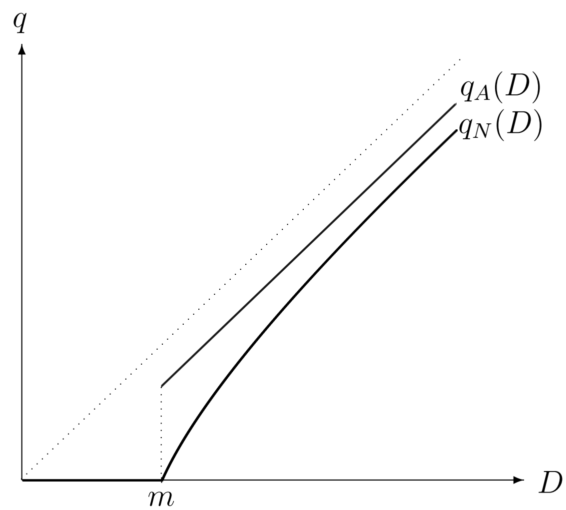
C'est à dire à avoir une probabilité d'audit $\gamma(D)$ croissante avec le niveau de couverture $q_N(D)$. Il est par ailleurs possible de montrer [7] qu'on obtient dans le cas CARA une généralisation du résultat obtenu avec audit déterministe :

Proposition. *Si les assurés ont une fonction d'utilité CARA et que la punition B est assez grande en cas de fraude* [8] *alors le contrat optimal est de la forme :*

- $q_A(D) = q_N(D) = 0$ *si* $D \leq m$
- $q_A(D) = D - k$ *et* $q_N(D) = D - k - \eta(D)$ *si* $D > m$ *avec* $\eta(D) > 0$, $\eta'(D) < 0$, $\eta(m) = m - k$ *et* $\eta(D) \to 0$ *quand* $D \to \infty$

avec $m > 0$ *et* $0 < k < m$

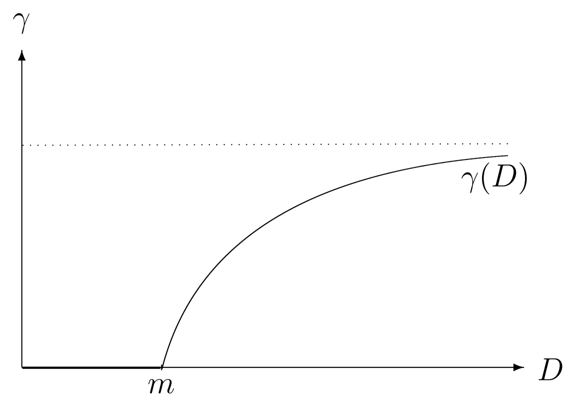
Graphique 19



On peut d'abord remarquer que **le contrat optimal possède une franchise**. Pour de petits sinistres, il n'est pas optimal d'offrir une couverture puisque cela impliquerait d'avoir à payer le coût d'audit avec une certaine probabilité. Pour des sinistres plus grands, il y aura **deux types de couverture** : une si le sinistre n'est pas audité et une autre, plus grande, si il l'est. Dans cette dernière, le sinistre marginal sera parfaitement assuré. Cela vient du fait que $q_A(\cdot)$ n'a pas de rôle incitatif d'après (IC.CARA). Ainsi, dans ce cas, il est optimal que le risque soit entièrement transmis à l'assureur. La couverture $q_N(D)$ est par ailleurs inférieure à $q_A(D)$. Cela vient du fait que – via la contrainte d'incitation – une augmentation de $q_N(D)$ augmente la probabilité d'audit $\gamma(D)$ ce qui a pour effet d'augmenter – via la contrainte de participation de l'assureur – la prime (au delà de l'effet d'une augmentation similaire de $q_A(D)$).

7. cf. M.-C. Fagart et P. Picard (1999), "Optimal Insurance Under Random Auditing", The Geneva Papers on Risk and Insurance Theory, 24, 29-54
8. sinon on revient dans le cas déterministe

Graphique 20



Remarque. *L'hypothèse d'utilité CARA simplifie grandement les choses du fait que l'aversion au risque ne dépende alors pas de la taille de la perte. Si, au contraire, l'aversion au risque de l'assuré est décroissante avec sa richesse, alors son incitation à frauder est plus grande pour les petites pertes (il est moins averse au risque de se faire prendre). Il est dans ce cas optimal d'offrir une couverture positive même pour les petites pertes afin de réduire cette incitation.*

Falsification coûteuse

Une autre partie de l'analyse économique de la fraude à l'assurance, suppose qu'il est impossible à l'assureur de vérifier une fraude, mais qu'il est coûteux à l'assuré de faire une fausse déclaration. Formellement, on supposera qu'un assuré ayant subi un sinistre D devra payer un coût $c(\widehat{D}-D)$ pour déclarer un sinistre $\widehat{D}$, avec $c(0)=0$, $c'>0$ et $c''>0$. Dans ce cas, et en reprenant les notations précédentes, l'espérance d'utilité d'un assuré ayant subi un sinistre D choisissant de déclarer un sinistre $\widehat{D}(D)$ s'écrit :

$$V=\int_0^{\overline{D}} u\left(R-\Pi-D+q\left(\widehat{D}(D)\right)-c\left(\widehat{D}(D)-D\right)\right)dF(D)$$

L'assuré choisira alors $\widehat{D}(D)$ de manière optimale, c'est-à-dire $\widehat{D}(D)$ tel que :

$$u'\left(R-\Pi-D+q\left(\widehat{D}(D)\right)-c\left(\widehat{D}-D\right)\right).\left[q'\left(\widehat{D}(D)\right)-c'\left(\widehat{D}(D)-D\right)\right]=0$$

L'assuré choisira donc de "gonfler" son sinistre tant que l'utilité marginale à agir ainsi dépasse le coût marginal que la fraude représente.

Plusieurs caractéristiques du contrat optimal peuvent être obtenues grâce à l'équation précédente.

- Si $c'(\cdot)<1$ (c'est-à-dire si gonfler le sinistre d'un euro coûte moins d'un euro) alors l'assurance offerte par le contrat optimal n'est pas complète ($q'(\widehat{D})<1\ \forall D$)

- Si $c'(0) = 0$ alors la fraude à l'assurance ne peut être évitée qu'en offrant une couverture fixe (indépendante du montant de la perte). Si on veut augmenter la couverture pour les sinistres élevés alors des fraudes apparaîtront nécessairement.
- Si $c'(0) = \delta > 0$ alors le contrat optimal est de la forme $q(D) = q_0 + \delta D$ avec $q_0 > 0$. Ainsi les petites pertes sont sur-assurées alors que comme $q'(D) = \delta < 1$, les gros sinistres ont tendance à être sous-assurés.

Extensions et évidences empiriques

Il est évident que les modèles présentés précédemment possèdent de nombreuses limitations et peuvent être complexifiés dans de nombreuses directions. Parmi celles qui ont été le plus étudiées, l'existence de raisons "morales" poussant l'assuré à ne pas frauder, peut être résolue relativement facilement. Les choses se compliquent cependant quand on suppose que les assurés ont des comportements différents face à la fraude et qu'il existe des assurés "honnêtes" et d'autres "opportunistes". On a alors un modèle avec à la fois de l'aléa moral et de l'anti-sélection.

Comme on l'a déjà évoqué, l'estimation de l'influence réelle de la fraude à l'assurance est par essence difficilement mesurable. Cependant certaines études ont essayé d'analyser dans quelle mesure de telles fraudes peuvent être repérées dans les données. Pour cela elles se sont basées sur une hypothèse simple : si la fraude a une probabilité assez élevée de réussir, alors on devrait observer des pertes plus élevées quand (toutes choses égales par ailleurs) la franchise est plus élevée. Les résultats sont alors plutôt significatifs. C'est notamment le cas d'une étude sur l'assurance automobile menée au Canada[9]. Il apparaît en effet que lorsqu'il n'y a pas de témoins sur le site de l'accident, les sinistres déclarés sont de montant entre 24.6% et 31.8% supérieur quand la franchise est de 500 \$ par rapport au cas où elle est de 250 \$ (alors que la différence n'est pas significative lorsqu'un témoin est présent). Ce résultat semble bien dû à la présence de fraudes (et non à de l'anti-sélection) puisqu'il est fortement lié à la présence ou non d'un témoin.

9. G. Dionne et R. Gagné (2001), "Deductible Contracts Against Fraudulent Claims : Evidence from Automobile Insurance", *Review of Economics and Statistics*, 83(2) 290-301.

Chapitre 5

Extensions et exercices

5.1 Extensions du modèle de Mossin

Défaut dans le modèle de Mossin

On considère le modèle standard de Mossin (revenu R, perte D avec probabilité p, fonction d'utilité Von Neumann Morgenstern $u(.)$) sans taux de chargement ($\lambda = 0$). Soit q la couverture payée par l'assureur en cas de perte. Cependant, on suppose maintenant que cette couverture ne sera payée qu'avec une probabilité σ, c'est-à-dire qu'avec probabilité $1 - \sigma$, la compagnie est en faillite et ne rembourse pas.

1. Calculer la prime actuariellement juste en prenant en compte la probabilité de défaut.
2. Expliquer intuitivement pourquoi ce n'est pas réaliste.
3. En supposant que la prime effective est la prime actuariellement juste calculée précédemment, calculer le choix optimal de couverture par les assurés (NB : la prime étant fixée, il n'est pas nécessaire d'étudier les courbes d'indifférence comme dans le modèle du cours).
4. Montrer qu'à l'optimum, l'utilité marginale dans un des états de la nature peut être obtenue comme combinaison linéaire des utilités marginales dans les autres états de la nature.
5. En déduire si l'assurance est complète à l'optimum.
6. Comment peut-on expliquer le résultat précédent ?
7. On peut montrer que dans le cas présent, une augmentation de l'aversion au risque n'augmente pas nécessairement la couverture optimale. Expliquer intuitivement pourquoi.

Exceptions dans le modèle de Mossin

On considère le modèle standard de Mossin (revenu R, perte D avec probabilité p, fonction d'utilité Von Neumann Morgenstern $u(.)$) sans taux de chargement ($\lambda = 0$). Soit q la couverture payée par l'assureur en cas de perte. Malheureusement, il y a beaucoup d'exceptions dans le contrat. Ainsi, en cas de perte, la couverture n'est payée qu'avec une probabilité σ. Avec probabilité $1 - \sigma$, on tombe dans une exception et la perte ne sera pas couverte. Cependant dans ce cas, l'assureur remboursera sa prime à l'assuré (ce qui rend cet exercice différent du précédent).

1. Calculer la prime actuariellement juste.
2. Écrire les conditions du premier et du second ordre du programme de maximisation d'un assuré lorsque la prime est actuariellement juste.
3. En déduire si l'assurance est complète à l'optimum.
4. Comparer avec le résultat de l'exercice précédent et commenter.

5.2 Demande d'assurance et risque exogène

Considérons un individu au revenu R ayant une fonction d'utilité $u(.)$ croissante et concave et faisant face à deux risques indépendants :
- *perdre un montant D avec probabilité p*
- *perdre un montant K avec probabilité θ.*

Une compagnie d'assurance se propose d'assurer au taux π le premier risque (q unités de couverture contre ce risque peuvent alors être achetées au prix $\Pi = \pi.q$). Par contre le deuxième risque est supposé inassurable (par exemple parce qu'il pas diversifiable). Le but de cet exercice est d'analyser l'effet de l'existence de ce risque exogène sur la demande d'assurance. On rappelle qu'en l'absence de ce deuxième risque, la couverture optimale q_0^* est définie par :

$$\frac{u'(R - D + (1-\pi)q_0^*)}{u'(R - \pi q_0^*)} = \frac{\pi}{p}\frac{1-p}{1-\pi}$$

1. Écrire l'espérance d'utilité atteinte pour un niveau q de couverture en présence du deuxième risque.
2. Écrire l'équation que doit vérifier q_1^*, couverture optimale en présence du deuxième risque.
3. Montrer que, lorsque la prime est actuarielle ($\pi = p$), $q_1^* = q_0^*$
4. Montrer que c'est également le cas ($q_1^* = q_0^*$) lorsqu'il existe un taux de chargement positif ($\pi > p$), si $u(C) = -\frac{1}{\alpha}e^{-\alpha C}$ (fonction d'utilité CARA)
5. En faisant un développement limité à l'ordre 1 au voisinage de 0 en K montrer que pour K suffisamment petit :
   - $q_1 > q_0$ si l'aversion au risque est décroissante (DARA)
   - $q_1 < q_0$ si l'aversion au risque est croissante (IARA)

5.3 Information génétique

On considère le modèle standard de Mossin (revenu R, perte D avec probabilité p, fonction d'utilité Von Neumann Morgenstern $u(.)$). Ici la perte D est associée au "coût" d'une maladie associée à un gène. Le gène possède deux formes, une forme favorable L : la probabilité de développer la maladie est p_L, et une forme défavorable H où la probabilité est p_H. A priori la proportion de type H est égale à b, celle de type L est donc $(1-b)$. Sans dépistage génétique, les individus ne savent pas de quel type ils sont.

1. Un assureur privé peut assurer contre cette maladie. Quel est le tarif actuariel proposé pour une couverture complète ? (l'assureur n'a pas d'information sur les types mais connaît p_L, p_H, b et D)

2. Quelle est l'utilité atteinte ?

On suppose maintenant que chaque individu peut, sans coût, réaliser un test génétique qui lui donne son type. On suppose d'autre part que l'assureur peut utiliser ce test pour tarifer mais ne peut exiger la réalisation du test pour un client qui souhaite s'assurer.

L'assureur présente a priori 3 tarifs. Un tarif si le test est L : prime = π_L*, un tarif sans test : prime=*π_M *et un tarif H : prime =* π_H*. Avec a priori* $\pi_L \leq \pi_M \leq \pi_H$

3. Un individu ayant réalisé le test a-t-il intérêt à en montrer le résultat à son assureur ?
4. En conséquence, quels tarifs sont effectivement choisis ?
5. Un individu a-t-il intérêt à réaliser le test ?
6. Que valent les primes des tarifs effectivement choisis si la tarification est actuarielle ?
7. Que vaut alors l'utilité atteinte ex-ante (avant test) ?
8. Que peut-on en déduire sur la valeur de l'information génétique ?

5.4 Le cas des risques de santé (utilité bidimensionnelle)

Dans le cas particulier des risques de santé, on ne peut pas raisonnablement limiter l'analyse aux seules pertes monétaires liées à la maladie. Aussi, on modélise dans ce cas des fonctions d'utilité à deux variables $u(w, h)$ *dépendant à la fois du revenu* w *et de l'état de santé* h*. On supposera que la fonction d'utilité est croissante et concave en chacun des arguments :* $u_1^{'}(\cdot) > 0$, $u_2^{'}(\cdot) > 0$, $u_{11}^{''}(\cdot) < 0$ *et* $u_{22}^{''}(\cdot) < 0$. *On s'intéressera ici particulièrement au rôle joué par la dérivé croisée* $u_{12}^{''}(\cdot)$ *(ou* $u_{21}^{''}(\cdot)$*).*

On considère le cas d'un individu fassant face à un risque de santé. Si il est en bonne santé, son revenu vaut R *et son état de santé est noté* S_H*. Il peut tomber malade avec probabilité* p*. Dans ce cas, il subit à la fois une perte de revenu (représentant le coût du traitement) et une détérioration de son état de santé. Sans assurance son revenu vaut alors* $R - D$ *et son état de santé est noté* $S_B < S_H$*.*

Risque de santé et assurance

On considère tout d'abord un modèle à une période dans lequel l'individu peut acheter une assurance qui couvre tout ou partie du coûte du traitement. Plus précisemment, il peut acheter au taux actuariel ($\Pi = p.q$)*, une assurance couvant* q *€ du coût de traitement si il tombe malade.*

1. Écrire l'espérance d'utilité atteinte pour un individus choisissant une couverture q.
2. Écrire l'équation que doit vérifier q^*, la couverture optimale.
3. Montrer que l'individu choisit une couverture complête $q = D$ si la fonction d'utilité est séparable : $u(w, h) = f(w) + g(h)$
4. Dans le cas général $u(w, h)$, sous quelle condition sur $u_{12}^{''}(\cdot)$, l'individu souhaite-t-il s'assurer plus que complètement. **Interpréter**.

Risque de santé et prévention

*On considère maintenant un cas **sans assurance** dans lequel l'individu peut faire un effort de prévention qui réduit le risque de tomber malade. On se place pour cela dans un cadre **à deux périodes** dans lequel :*

- *En première période, l'individu est en bonne santé : $h = S_H$. Il peut mettre en place un effort de prévention dont le coût **monétaire** est noté e. Si il choisit un niveau de prévention e (continu) son revenu est $R - e$ (le modèle est donc différent de celui vu en cours). Cet effort de prévention va réduire la probabilité d'occurrence de la maladie $p(e)$ (avec $p'(\cdot) < 0$ et $p''(\cdot) > 0$) à la période suivante.*
- *En seconde période, si l'individu ne tombe pas malade (avec probabilité $1 - p(e)$), son revenu est R et son niveau de santé S_H. Par contre, s'il tombe malade (avec probabilité $p(e)$), il subit un coût de traitement D (son revenu est $R - D$) et son état de santé devient $S_B < S_H$.*

1. Écrire l'espérance d'utilité intertemporelle de l'individu considéré (pour simplifier on supposera qu'il n'y a pas d'escompte entre les périodes)
2. Écrire l'équation qui définit son choix optimal de prévention e^* sous la forme $F(e^*, D, S_B, S_H) = 0$
3. En utilisant les dérivées partielles de $F(\cdot)$ déterminer comment ce niveau de prévention varie avec D et S_B. Interpréter.
4. On souhaite maintenant savoir comment e^* varie avec S_H. Montrer que si $u''_{12}(\cdot) < 0$ alors e^* est croissant en S_H. **Interpréter**.