

Essai sur le risque moral An Essay on Moral Hazard

Richard Arnott

Volume 63, numéro 2-3, juin–septembre 1987

Incertain et information

URI : <https://id.erudit.org/iderudit/601411ar>

DOI : <https://doi.org/10.7202/601411ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)

1710-3991 (numérique)

[Découvrir la revue](#)

Citer cet article

Arnott, R. (1987). Essai sur le risque moral. *L'Actualité économique*, 63(2-3), 74–97. <https://doi.org/10.7202/601411ar>

Résumé de l'article

Cet essai présente quelques-uns des résultats obtenus dans un projet de recherche à long terme sur le risque moral entrepris par l'auteur et Joseph Stiglitz. Le message central est que la présence du risque moral modifie radicalement les propriétés positives et normatives de l'équilibre concurrentiel. En particulier, le risque moral peut générer des courbes d'indifférence non convexes et un ensemble de possibilités non convexe. Principalement à cause de ces non-convexités, la définition de l'équilibre concurrentiel devient subtile. Selon la définition choisie, à un extrême, l'équilibre peut ne pas exister et, à l'autre extrême, il se peut qu'il y ait une infinité d'équilibres. De plus, l'équilibre concurrentiel peut être caractérisé par le rationnement des achats d'assurances, des profits positifs, de l'inactivité du marché même s'il y a des gains à l'échange, et/ou de l'assurance aléatoire. Finalement, sous le risque moral, ni le premier ni le second théorème de l'économie du bien-être ne sont valides et les prix du marché ne reflètent pas les coûts d'opportunité sociaux.

ESSAI SUR LE RISQUE MORAL*

Richard ARNOTT
Queen's University

Cet essai présente quelques-uns des résultats obtenus dans un projet de recherche à long terme sur le risque moral entrepris par l'auteur et Joseph Stiglitz. Le message central est que la présence du risque moral modifie radicalement les propriétés positives et normatives de l'équilibre concurrentiel. En particulier, le risque moral peut générer des courbes d'indifférence non convexes et un ensemble de possibilités non convexe. Principalement à cause de ces non-convexités, la définition de l'équilibre concurrentiel devient subtile. Selon la définition choisie, à un extrême, l'équilibre peut ne pas exister et, à l'autre extrême, il se peut qu'il y ait une infinité d'équilibres. De plus, l'équilibre concurrentiel peut être caractérisé par le rationnement des achats d'assurances, des profits positifs, de l'inactivité du marché même s'il y a des gains à l'échange, et/ou de l'assurance aléatoire. Finalement, sous le risque moral, ni le premier ni le second théorème de l'économie du bien-être ne sont valides et les prix du marché ne reflètent pas les coûts d'opportunité sociaux.

An Essay on Moral Hazard. — This essay presents some of the results obtained in a long-term research project on moral hazard undertaken by the author and Joseph Stiglitz. The central message is that the presence of moral hazard radically alters the positive and normative properties of competitive equilibrium. In particular, moral hazard may give rise to nonconvex indifference curves (in the relevant space) and causes the feasibility set to be nonconvex. Principally because of these nonconvexities, there are subtleties in the definition of competitive equilibrium. Depending on the definition, at one extreme competitive equilibrium may non exist, while at the other there may be an infinity of equilibria. Furthermore, competitive equilibrium may be characterized by the rationing of insurance purchases, positive profits, inactivity of the market even when there are gains from trade, and/or random insurance. Finally, with moral hazard, neither the first nor the second theorems of welfare economics holds, and market prices do not correspond to social opportunity costs.

*Cet essai présente quelques-uns des résultats obtenus dans un projet de recherche à long terme sur le risque moral avec Joseph Stiglitz. Le soutien financier pour cette recherche par la National Science Foundation, l'Olin Foundation et le C.R.S.H.C. est grandement apprécié. Nous avons reçu beaucoup de commentaires utiles sur notre travail; ceux de Martin Hellwig ont été particulièrement pénétrants. Je remercie Georges Dionne et Suzanne Larouche Sidoti pour avoir amélioré la version originale de l'essai.

INTRODUCTION

Dans le traitement standard (Arrow-Debreu) du risque et de l'assurance, à l'équilibre concurrentiel, il y a un ensemble d'états de la nature caractérisés par des probabilités exogènes. Les états réalisés sont observables. L'assurance permet des transferts forfaitaires entre les états réalisés et ainsi n'entraîne pas de problèmes d'incitation ou (de façon équivalente) d'effets de substitution.

Le risque moral survient lorsque les états de la nature et les actions des individus ne sont pas observables par l'assureur¹. Ce qui est observable, c'est le fait qu'un accident particulier s'est produit. Dans ces circonstances, il n'y a pas de mécanisme par lequel l'assureur peut persuader un individu assuré à révéler soit l'état de la nature réalisé, soit ses activités d'autoprotection sans mentir. Ainsi les événements qui sont assurés sont des accidents ayant des degrés variables de sévérité, indépendamment de l'état de la nature et des actions de l'individu. La provision d'assurance contre ces événements influencera généralement les incitations de l'individu à produire de la prévention², c'est-à-dire entraînera des problèmes d'incitation ou des effets de substitution. Par conséquent, il y a une relation d'arbitrage entre le partage des risques et les incitations. C'est le problème du risque moral.

Le risque moral est important dans l'économie. Il survient lorsque le risque est présent, les individus sont risco-phobes et « l'effort » est coûteux à observer. Il n'est pas observé seulement dans les marchés privés d'assurance, mais aussi dans tous les autres contextes où l'assurance est fournie par les gouvernements, les institutions sociales comme les fondations charitables, les amis et la famille, ou dans les contrats principal-agent.

Durant la dernière décennie, Joseph Stiglitz et moi avons travaillé à un projet de recherche commun de long terme sur le risque moral. Cet essai présente quelques-uns de nos résultats. Le traitement sera illustratif, sélectif et non technique. Le message central est que la présence du risque moral modifie radicalement la nature de l'équilibre concurrentiel. En présence de risque moral :

- a) à une extrémité, l'équilibre peut ne pas exister ; à l'autre il peut y avoir une infinité d'équilibres ;
- b) quand l'équilibre existe, certains marchés d'assurance peuvent être inactifs même s'il y a une demande pour l'assurance ;

1. C'est la forme extrême du risque moral. Quand les états de la nature et/ou les actions des individus sont partiellement observables (avec un bruit), les problèmes associés au risque moral demeurent mais sont dilués.

Des contributions importantes sur la théorie du risque moral ont été faites par Arrow (1970), Spence et Zeckhauser (1971), Pauly (1974), Helpman et Laffont (1975), Mirrlees (1975), Marshall (1976), Holmstrom (1979) et Shavell (1979).

2. Ainsi, la provision d'assurance influera sur les probabilités des événements. Pour cette raison, on peut définir le risque moral lorsque la provision d'assurance influence les probabilités des événements.

- c) dans les marchés d'assurance actifs, l'équilibre peut être caractérisé par des profits positifs, du rationnement d'assurance, et/ou des primes et des remboursements aléatoires ;
- d) ni le premier ni le second théorème de l'économie du bien-être ne sont valides ;
- e) les prix du marché ne reflètent pas les coûts d'opportunité sociaux. En conséquence, les gains potentiels d'une intervention gouvernementale afin d'améliorer l'efficacité sont considérables.

Le risque moral survient à cause d'une forme particulière d'asymétrie d'information – les actions de l'assuré ne sont pas observables par l'assureur. *La sélection adverse* est un phénomène différent d'assurance causé par une autre forme d'asymétrie d'information – l'assureur ne peut pas identifier le risque d'un individu particulier bien qu'il connaisse la distribution des risques dans la population. Dans les contextes d'assurance du monde réel, les deux phénomènes sont toujours présents et se présentent simultanément. Dans cet essai, pour simplifier l'analyse, la sélection adverse sera absente par hypothèse. Les théoriciens ont fait des progrès assez limités en analysant le risque moral et la sélection adverse ensemble, ce qui a retardé l'investigation empirique de l'économie de l'assurance. Au cours de l'essai, des caractéristiques de la provision d'assurance dans le monde réel, qui peuvent être expliquées par le risque moral, seront mentionnées. Par contre, le lecteur doit garder à l'esprit qu'en absence de sélection adverse, de telles explications sont partielles et à titre d'essai.

Même si quelques exemples seront proposés, l'essai sera présenté à un niveau abstrait. Afin d'étoffer l'argumentation, le lecteur peut trouver utile de réfléchir en terme d'un exemple un peu usé – l'assurance personnelle contre les incendies. Il y a beaucoup d'activités que les locataires ou les propriétaires peuvent entreprendre pour réduire la probabilité d'un incendie ou pour réduire les dommages prévus au moment d'un incendie – ne pas jeter les cendres des cigarettes dans la corbeille à papier, ne pas fumer au lit, ne pas quitter lorsque le four de la cuisinière est allumé, éteindre les cendres dans la cheminée avant de se coucher, remplacer immédiatement les fils électriques usés, avoir un extincteur fonctionnant dans chaque chambre, acheter des matériaux à l'épreuve du feu pour la maison, faire en sorte qu'il y ait une sortie facile pour chaque chambre en cas d'incendie, etc. Toutes ces activités occasionnent des décisions économiques qui sont influencées par la provision d'assurance.

1. LE MODÈLE DE RÉFÉRENCE

Pendant la plus grande partie de cet essai, le modèle le plus simple avec risque moral sera utilisé. Chaque individu dans l'économie a une seule activité qui a deux conséquences possibles : soit qu'un accident avec dommage unique survient, soit qu'il ne survient pas (une autre interprétation est qu'un accident

a toujours lieu mais peut être grave ou non)³. Le contenu des sections 1-4 est basé sur Arnott et Stiglitz [1987a].

Un seul bien de consommation est produit dans l'économie. Le niveau de production d'un individu est fonction du fait qu'il ait un accident ou non. La probabilité d'accident est une fonction de sa prévention ou de son « effort » – un terme générique pour les nombreuses activités et les nombreuses dépenses d'autoprotection qu'un individu peut entreprendre. Les probabilités d'accident individuelles sont statistiquement indépendantes.

Les individus sont riscophobes et ainsi veulent s'assurer contre l'accident. Un assureur ne peut pas observer l'effort de ses clients. Par conséquent, la provision d'assurance est caractérisée par le risque moral, et l'assurance est offerte contre l'accident sans référence aux actions de l'assuré.

Soit y_1 et y_0 , la consommation sous les événements accident et non-accident, respectivement. Sans assurance :

$$y_0 = w ; \quad y_1 = w - d ;$$

où w est la production d'un individu lorsqu'il n'y a pas d'accident, d est le dommage dû à l'accident, et $w - d$ est la production quand un accident survient. Une police d'assurance fournit un remboursement net de la prime α si un accident se passe et, autrement, fait payer une prime β . Ainsi, sous l'assurance

$$y_0 = w - \beta ; \quad y_1 = w - d + \alpha. \quad (1.1)$$

La probabilité individuelle qu'un accident survienne p est une fonction de son effort e . Dans certaines parties de l'essai, on fait l'hypothèse que l'individu a le choix entre un nombre fini de niveaux d'effort, chacun correspondant à une technique de prévention d'accident différente. Il est de plus supposé qu'il n'y a pas de niveau d'effort pour lequel l'accident est certain et il n'est pas possible pour l'individu de causer un accident délibérément. Dans d'autres parties de l'essai, l'individu peut choisir son niveau de prévention d'un continuum de niveaux d'effort. Dans ce cas, $p(e)$ est la probabilité d'accident associée au niveau d'effort e , où $e = 0$ correspond à une absence complète d'effort. Il est supposé que $p(0) = \bar{p} < 1$, $p'(e) < 0$ et $p''(e) > 0$; plus d'effort réduit la probabilité d'accident et les unités successives d'effort réduisent la probabilité d'accident de moins en moins.

Dans le cas où les niveaux d'effort sont discrets, l'espérance mathématique de l'utilité individuelle est

$$EU^j = (1-p^j)U_0^j(y_0) + p^jU_1^j(y_1), \quad (1.2a)$$

3. Il y a une section de la littérature sur le risque moral qui traite d'un continuum de conséquences – Spence et Zeckhauser (1971), Mirrlees (1975), et Grossman et Hart (1983). Les deux derniers articles sont concentrés sur les non-convexités dans le choix individuel d'effort que le risque moral peut causer. Voir note 6 pour plus de détails.

où j identifie le niveau d'effort, $U_0^j(\bullet)$ est la fonction d'utilité lorsqu'il n'y a pas d'accident, et $U_1^j(\bullet)$ est la fonction d'utilité lorsqu'il y a un accident. La fonction d'espérance d'utilité est définie *séparable* si l'espérance d'utilité peut être écrite comme suit :

$$EU^j = (1-p^j)u_0(y_0) + p^j u_1(y_1) - e^j. \quad (1.2b)$$

Dans beaucoup de contextes, l'effort est ordinal ; la définition précédente suppose une cardinalisation particulière de l'effort – que l'effort est mesuré par la désutilité qu'il cause. Si l'« effort » occasionne des dépenses sur des biens de prévention d'accident, la fonction d'espérance d'utilité avec « effort monétaire » s'écrit :

$$EU^j = (1-p^j)U_0(y_0 - e^j) + p^j U_1(y_1 - e^j). \quad (1.2c)$$

La fonction d'utilité au moment d'un accident peut être différente de celle sans accident lorsque l'accident cause de la douleur ou change les goûts. Un accident occasionne un coût en terme d'utilité si $U_0^j(y) > U_1^j(y)$ pour tout j et y , et réduit l'utilité marginale si $(U_0^j(y))' > (U_1^j(y))'$ pour tout j et y , où ' indique une dérivée. Quand $U_0^j(y) = U_1^j(y)$ pour tout j et y , la fonction d'espérance d'utilité est indépendante des événements. Le choix approprié de la forme fonctionnelle de l'espérance d'utilité dépend du contexte d'analyse.

La terminologie décrivant la forme de la fonction d'espérance d'utilité s'applique aussi bien au cas avec un continuum de niveaux d'effort. La fonction générale d'espérance d'utilité dans ce cas est :

$$EU = (1-p(e))U_0(y_0, e) + p(e)U_1(y_1, e) \quad (1.2d)$$

où $\frac{\partial U_i}{\partial e} < 0$ et $\frac{\partial^2 U_i}{\partial e^2} \leq 0$. Si elle est séparable, $U_i(y, e) = U_i(y) - e$; si elle est indépendante des événements, $U_0(y, e) = U_1(y, e)$ pour tout y et e^4 , et si elle est séparable et indépendante des événements, $U_i(y, e) = u(y) - e$.

Dans certains contextes, le modèle est pertinemment interprété comme statique, dans d'autres contextes comme décrivant un état stationnaire⁵.

Un peu de discussion peut rendre la terminologie plus claire. La plupart des accidents occasionnent des coûts en terme d'utilité et réduisent l'utilité marginale. La mort est un exemple extrême. Mais un accident peut être bénéfique en terme d'utilité ; par exemple, un ouvrier devenu infirme suite à un problème de dos qui ne causera pas de douleur s'il ne travaille pas, peut être plus heureux avec une généreuse pension d'invalidité que dans une situation sans mal de dos (n'oubliez pas que « bénéfique en terme d'utilité » est défini en gardant constant la consommation dans les événements accident et non-accident). Un accident peut aussi augmenter l'utilité marginale de la richesse ; dans l'exemple précédent, cela

4. La littérature précédant 1980 a supposé que la fonction d'espérance d'utilité était indépendante des événements. Le premier article qui aborde le cas contraire est celui de Dionne (1982).

5. Cependant, sous cette interprétation, la possibilité que la tarification de l'assurance soit basée sur l'expérience passée de l'individu n'est pas traitée.

se passerait si la consommation et le loisir étaient complémentaires. La fonction d'espérance d'utilité associée aux « accidents » financiers est typiquement indépendante des événements.

Si l'accident peut avoir lieu seulement une fois et si l'effort précède l'événement et donc la consommation, la fonction d'espérance d'utilité est probablement presque séparable. Mais si la possibilité d'accident est continue, la consommation et l'effort sont probablement des substituts – l'effort réduit le loisir qui est complément à la consommation.

2. LA DÉCISION D'EFFORT DE L'INDIVIDU

L'intuition suggère que lorsque la couverture d'assurance augmente, l'assuré est moins prudent. Cette intuition, quoique normalement correcte, n'est pas toujours vérifiée.

Le problème de l'individu qui choisit son niveau d'effort est :

$$\text{Max } EU = (1-p(e))U_0(w-\beta, e) + p(e)U_1(w-d+\alpha, e) \quad (2.1)$$

La condition de premier ordre, avec effort positif, est :

$$\{-p'(U_0 - U_1)\} + \left[(1-p) \frac{\partial U_0}{\partial e} + p \frac{\partial U_1}{\partial e} \right] = 0 . \quad (2.2)$$

Le terme entre accolades est le bénéfice marginal de l'effort – la réduction de la probabilité d'accident occasionnée par une unité supplémentaire d'effort ($-p'$) fois l'augmentation en utilité d'une réduction d'une unité dans la probabilité d'accident ($U_0 - U_1$). Le terme entre crochets est l'espérance de coût marginal de l'effort – les désutilités marginales de l'effort pondérées par les probabilités des deux événements.

La condition de deuxième ordre est :

$$A \equiv -p''(U_0 - U_1) - 2p' \left[\frac{\partial U_0}{\partial e} - \frac{\partial U_1}{\partial e} \right] + \left[(1-p) \frac{\partial^2 U_0}{\partial e^2} + p \frac{\partial^2 U_1}{\partial e^2} \right] < 0. \quad (2.3)$$

A n'est pas nécessairement négatif. En conséquence, l'effort peut être discontinue en fonction des paramètres du contrat d'assurance et peut *augmenter*, de façon discontinue, en réponse à une augmentation du montant d'assurance fourni.

Même si la condition de deuxième ordre est toujours satisfaite, l'effort peut augmenter quand le montant d'assurance augmente. Pour démontrer ceci, considérons une augmentation du remboursement d'assurance pour une prime constante :

$$\frac{\partial e}{\partial \alpha} \Big|_{\beta} = - \frac{1}{A} \left[p' \frac{\partial U_1}{\partial y_1} + p \frac{\partial^2 U_1}{\partial e \partial y_1} \right] \text{ avec } e > 0 . \quad (2.4)$$

Cette expression est positive si $\frac{\partial^2 U_1}{\partial e \partial y_1}$ est suffisamment positif – si une augmentation de y_1 diminue considérablement la désutilité marginale de l'effort. Un exemple est un contrat de tenancier passé entre un propriétaire et des ouvriers agricoles dans un pays pauvre : si le propriétaire aide ses ouvriers qui ont souffert d'une mauvaise récolte, il peut prévenir leur sous-alimentation, ce qui améliorera leur productivité pour la prochaine récolte.

Cependant, avec l'utilité séparable, A est toujours négatif, ce qui implique qu'il y a un maximum unique pour la décision d'effort d'un individu⁶. De plus, une augmentation de la couverture d'assurance ne peut jamais occasionner une augmentation de l'effort. En outre, au-dessus d'un niveau d'assurance caractérisé par $\lim_{e \downarrow 0} p'(e)(-u_0 + u_1) - 1 = 0$, l'effort est nul – cette équation caractérise le lieu géométrique d'effort nul (LEN) qui, dans le cas de la séparabilité, divise l'ensemble pour lequel l'effort est positif dans (α, β) de l'ensemble pour lequel l'effort est nul.

Les résultats de cette section sont résumés dans la

PROPOSITION 1 : Normalement et toujours avec l'utilité séparable, les augmentations de la couverture d'assurance ne causent pas d'augmentations d'effort. Cependant, avec l'utilité non séparable, l'effort peut augmenter soit de façon continue ou discontinue suite à une augmentation de l'assurance.

3. LA FORME ÉTRANGE DES COURBES D'INDIFFÉRENCE SOUS LE RISQUE MORAL

De l'analyse de la décision d'effort de l'individu, on peut dériver une fonction (correspondance) reliant l'effort aux paramètres du contrat d'assurance, $e = e(\alpha, \beta)$. La substitution de cette correspondance dans (2.1) rend l'espérance d'utilité fonction uniquement de paramètres exogènes et des paramètres du contrat d'assurance, c'est-à-dire $V(\alpha, \beta) \equiv EU(e(\alpha, \beta), \alpha, \beta)$. Il sera très utile de conduire l'analyse dans l'espace $\alpha - \beta$.

La procédure, dans cette section, sera d'analyser la forme des courbes d'indifférence dans l'espace $\alpha - \beta$, d'abord pour le cas d'effort fixe, puis pour deux niveaux d'effort, et finalement, pour un continuum de niveaux d'effort.

6. Avant 1980, il était commun d'obtenir le contrat optimal entre un assureur (principal) et un assuré (agent) en maximisant la fonction objective du principal avec la condition de premier ordre du problème de l'agent qui choisit son niveau d'effort comme contrainte. Mirrlees (1975) et Grossman et Hart (1983) ont démontré que cette procédure – qui est appelée « l'approche du premier ordre » – n'est pas, en général, correcte à cause de la non-convexité possible du problème de l'agent et de la possibilité d'une solution en coin à son problème. Dans notre analyse, nous évitons cette erreur par un traitement explicite des possibilités de non-convexité dans le problème de l'agent et des solutions en coin.

3.1 Un seul niveau d'effort

De (1.2a) :

$$\left. \frac{d\beta}{d\alpha} \right|_{V_0^j} = - \left[\frac{\partial EU^j / \partial \alpha}{\partial EU^j / \partial \beta} \right] = \frac{(U_1^j)' p^j}{(U_0^j)' (1-p^j)} \equiv S^j > 0 \quad (3.1)$$

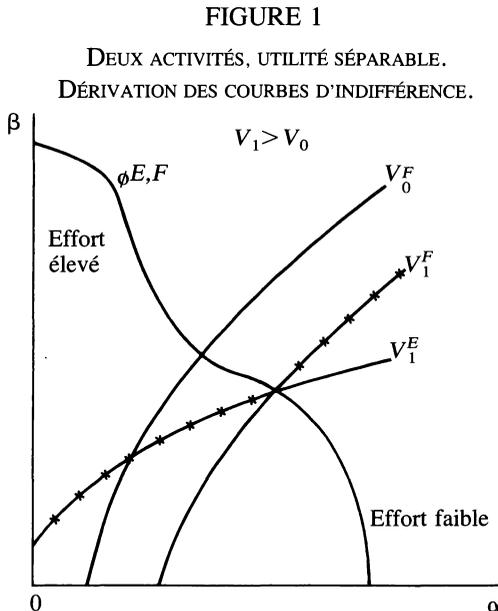
et

$$\left. \frac{d^2\beta}{d\alpha^2} \right|_{V_0^j} = - S^j (A_1^j + S^j A_0^j) < 0, \quad (3.2)$$

où $A_i^j \equiv \frac{(U_1^j)''}{(U_1^j)'}$ est le coefficient absolu de riscophobie sous l'événement i . Les courbes d'indifférence ont une pente positive parce que du point de vue de l'individu, le remboursement est un bien et la prime est un mal ; de plus, elles ont une courbure négative à cause de la riscophobie – lorsque la provision d'assurance augmente, l'individu est prêt à payer une augmentation plus petite de prime pour une augmentation fixe de remboursement.

3.2 Deux niveaux d'effort, utilité séparable

Supposons qu'il y a deux niveaux d'effort – faible (F) et élevé (E) ; avec le niveau d'effort faible, la probabilité d'accident est plus grande. Si la fonction d'utilité est séparable, la courbe d'indifférence correspondant au niveau d'effort faible est plus escarpée que la courbe d'indifférence correspondant au niveau d'effort élevé ; avec un effort plus faible, la probabilité d'accident est plus grande et ainsi l'individu est prêt à payer une plus grande augmentation de prime pour une augmentation fixe de remboursement.



Dans la section précédente, il a été démontré avec une utilité séparable et un continuum de niveaux d'effort, que l'effort diminue lorsque la quantité d'assurance augmente. Le résultat analogue avec deux niveaux d'effort est que l'individu choisit l'effort élevé en présence d'un montant d'assurance petit et l'effort faible avec un montant élevé. La frontière séparant la région d'effort élevé de la région d'effort faible est appelée le *lieu géométrique de changement*.

Les pièces sont maintenant assemblées pour construire une courbe d'indifférence avec effort variable. Passons à la figure 1.

Le lieu géométrique de changement est dénoté par $\phi^{E,F}$. V_1^E est la courbe d'indifférence correspondant à l'utilité U_1 quand l'individu choisit l'effort élevé, et V_1^F la courbe d'indifférence lorsque l'individu se limite à un effort faible. Les courbes d'indifférence plus basses correspondent à un niveau d'utilité plus haut. Au-dessous de $\phi^{E,F}$, l'individu choisit l'effort élevé et les courbes d'indifférence correspondant à l'effort élevé sont pertinentes; au-dessus de $\phi^{E,F}$, les courbes d'indifférence correspondant à l'effort faible sont pertinentes. Ainsi, V_1 , la courbe d'indifférence pour le niveau d'utilité U_1 avec effort endogène, coïncide avec V_1^E au-dessous du lieu géométrique de changement, et avec V_1^F au-dessus. Il est évident, à partir de la figure, que les courbes d'indifférence pour le cas de deux niveaux d'effort, avec effort endogène, ont une forme dentelée et ne sont jamais convexes.

3.3. Continuum de niveaux d'effort, utilité séparable

La forme d'une courbe d'indifférence, quand il y a un continuum de niveaux d'effort, est déterminée par deux forces. L'aversion au risque implique des courbes d'indifférence convexes (courbure négative). Cependant, lorsque la provision d'assurance augmente, l'effort baisse ; ce qui augmente la pente des courbes d'indifférence ; cet effet rend concaves les courbes d'indifférence. Quel effet domine? Cela dépend de la puissance relative des deux forces. L'algèbre élémentaire donne :

$$\frac{d^2\beta}{d\alpha^2} \Big|_{\bar{V}} = \{-S(A_1 + SA_0)\} + \frac{u'_1}{u'_0(1-p)^2} \frac{dp}{d\alpha} \Big|_{\bar{V}} \quad (3.3)$$

L'expression entre accolades donne l'effet de la riscophobie qui rend convexes les courbes d'indifférence ; le deuxième terme à droite reflète l'effet de désincitation à l'effort qui rend concaves les courbes d'indifférence. Le premier effet augmente avec le degré de riscophobie ; le second effet augmente avec la sensibilité de la probabilité d'accident au montant d'assurance fourni.

Le résultat important est que *lorsque l'effet de désincitation à l'effort domine l'effet de riscophobie, les courbes d'indifférence sont non convexes*. Il sera démontré que la non-convexité des courbes d'indifférence (dans l'espace $\alpha - \beta$) causée par le risque moral a des implications très significatives sur l'existence et la nature de l'équilibre concurrentiel.

PROPOSITION 2 : Avec effort constant, les courbes d'indifférence sont convexes dans l'espace des primes et des remboursements ; avec un nombre fini de niveaux d'effort, les courbes d'indifférence ont une forme dentelée et ne sont jamais convexes ; et avec un continuum de niveaux d'effort, les courbes d'indifférence peuvent être ou ne pas être convexes, selon les puissances relatives des effets de la riscophobie et de désincitation à l'effort.

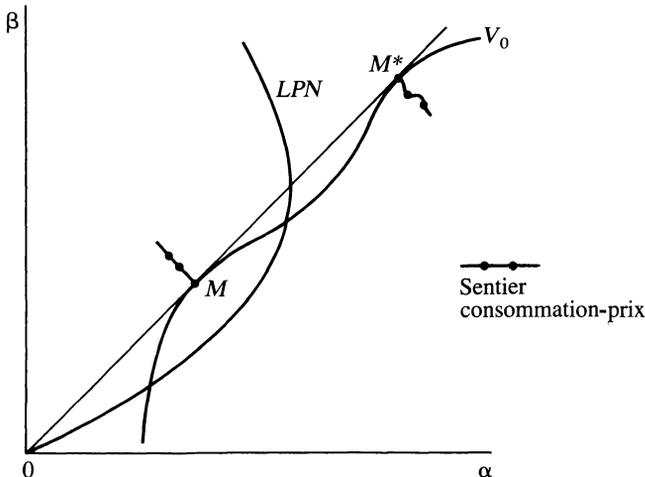
3.4. *Les sentiers de consommation-prix et de consommation-revenu peuvent être discontinus*

Les sentiers de consommation-prix et de consommation-revenu sont définis de la manière conventionnelle. Le prix d'assurance, q , est défini comme étant la prime payée par unité de remboursement, c'est-à-dire $q \equiv \frac{\beta}{\alpha}$. Ainsi un sentier consommation-prix est le lieu géométrique des points d'utilité maximale le long des lignes de prix $\beta = q\alpha$. Un sentier consommation-revenu est le lieu géométrique des points d'utilité maximale le long des lignes de prix ayant une pente q .

La non-convexité des courbes d'indifférence peut causer des sentiers consommation-prix et consommation-revenu discontinus. Ce résultat est illustré à la figure 2, qui montre une situation où le sentier consommation-prix est discontinu sur le lieu géométrique de profit nul (LPN), $(1 - p)\beta - p\alpha = 0$ (les caractéristiques de cette ligne seront étudiées dans la section suivante) une situation qui sera importante plus tard.

FIGURE 2

LA NON-CONVEXITÉ DES COURBES D'INDIFFÉRENCE PEUT CAUSER DES DISCONTINUITÉS DANS LE SENTIER CONSOMMATION-PRIX.



Arnott et Stiglitz [1982] décrivent les conditions qui déterminent où les sentiers consommation-prix et consommation-revenu sautent vers l'intérieur ou vers l'extérieur à un point de discontinuité. Dans le cas du sentier consommation-prix, ce qui importe est où (en terme de la figure 2) l'utilité augmente plus vite, au-dessous de M ou au-dessous de M^* , lorsque le prix de l'assurance est réduit. Ce résultat indique que l'existence et les caractéristiques des discontinuités dans les sentiers consommation-prix et consommation-revenu dépendent des caractéristiques globales plutôt que locales des fonctions d'espérance d'utilité et de la probabilité d'accident. Ainsi :

PROPOSITION 3: La non-convexité des courbes d'indifférence peut générer des discontinuités dans les sentiers consommation-prix et consommation-revenu. L'existence et les caractéristiques de ces discontinuités dépendent des caractéristiques globales des fonctions d'utilité et de la probabilité d'accident.

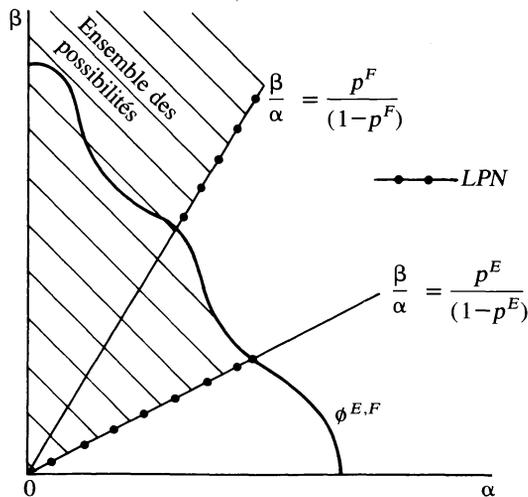
4. L'ENSEMBLE DES POSSIBILITÉS N'EST PAS CONVEXE

L'ensemble des possibilités est l'ensemble des contrats pour lesquels les profits sont non négatifs, c'est-à-dire (α, β) ont des caractéristiques selon lesquelles $(1 - p(e(\alpha, \beta)))\beta - p(e(\alpha, \beta))\alpha \geq 0$, et le *lieu géométrique de profit nul (LPN)* est le lieu géométrique des contrats dont les profits sont nuls.

4.1 *Utilité séparable*

Pour un niveau d'effort fixé à \tilde{e} , le lieu géométrique de profit nul est le rayon qui part de l'origine avec une pente de $\tilde{q} = p(\tilde{e})/(1-p(\tilde{e}))$. La méthode de construction du lieu géométrique de profit nul avec deux niveaux d'effort est analogue à celle employée pour construire les courbes d'indifférence dans ce

FIGURE 3
ENSEMBLE DE POSSIBILITÉS ET LIEU GÉOMÉTRIQUE DE PROFIT NUL,
DEUX ACTIVITÉS, UTILITÉ SÉPARABLE.



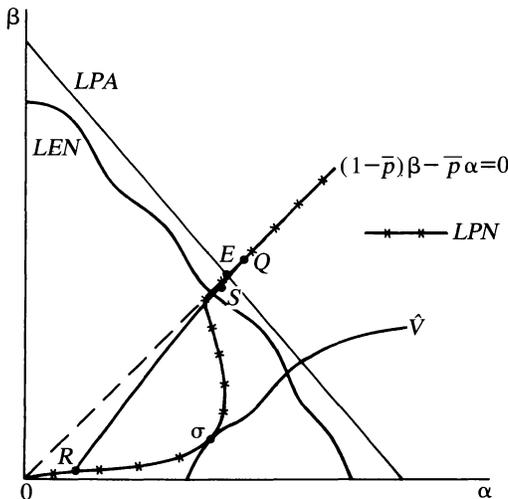
cas. Si nous nous référons à la figure 3, pour (α, β) au-dessous du lieu géométrique de changement, le lieu géométrique de profit nul correspondant à l'effort élevé est pertinent, et pour (α, β) au-dessus de $\phi^{E,F}$, le lieu géométrique de profit nul correspondant à l'effort faible est pertinent. L'ensemble des possibilités avec effort endogène est évidemment non convexe.

Avec un continuum de niveaux d'effort, la pente du lieu géométrique de profit nul est

$$\frac{d\beta}{d\alpha} \Big|_{LPN} = \begin{cases} \frac{p + (\beta + \alpha)(\partial p / \partial \alpha)}{(1-p) - (\beta + \alpha)(\partial p / \partial \beta)} & \text{avec } e > 0 \\ \frac{\bar{p}}{1-\bar{p}} & \text{avec } e = 0. \end{cases} \quad (4.1)$$

Deux caractéristiques de l'ensemble des possibilités sont d'intérêt particulier. D'abord, à chaque point sur le lieu géométrique de profit nul avec une pente positive et effort positif, le lieu géométrique est plus abrupt que le rayon reliant le point à l'origine. Deuxièmement, l'ensemble des possibilités est non convexe. Ce résultat peut être démontré plus facilement en utilisant la géométrie. Passons à la figure 4. Traçons la ligne d'effort nul (*LEN*) qui a été caractérisée à la section 2. Au-delà de la ligne d'effort nul, la *LPN* est donnée par $(1-\bar{p})\beta - \bar{p}\alpha = 0$. Ensuite, notons de (4.1), que la pente de la *LPN* à l'origine est $p(e(0, 0))/1-p(e(0, 0)) < \bar{p}/(1-\bar{p})$. Maintenant, prenons un point *Q* sur la *LPN* au-delà de la *LEN* et un autre point *R* sur la *LPN* près de l'origine. Étant donné qu'il est impossible pour toute la ligne joignant ces deux points d'être à l'intérieur de l'ensemble des possibilités, l'ensemble est non convexe.

FIGURE 4
AVEC UNE UTILITÉ SÉPARABLE,
L'ENSEMBLE DES POSSIBILITÉS EST NON CONVEXE.



Quelques autres caractéristiques du lieu géométrique de profit nul sont : i) il est continu ; ii) sa courbure est a priori indéterminée ; iii) il peut avoir un segment avec une pente négative, mais il ne peut pas avoir une pente nulle ; iv) sa pente peut être discontinue à la ligne d'effort nul.

4.2. Utilité non séparable

Parce que l'effort peut être discontinu dans les paramètres du contrat d'assurance lorsque l'utilité est non séparable, l'ensemble des possibilités peut être non connexe et il se peut qu'il y ait des segments sur la frontière de l'ensemble des possibilités avec des profits strictement positifs.

Les résultats établis jusqu'ici dans cette section sont résumés dans la

PROPOSITION 4 : Avec une utilité séparable, l'ensemble de possibilités est connexe, mais non convexe. Avec l'utilité non séparable, l'ensemble de possibilités peut ne pas être connexe et peut contenir des segments sur sa frontière où les profits sont strictement positifs.

4.3. Des résultats géométriques supplémentaires

Nous en sommes à l'étape d'un examen de l'existence et de la nature de l'équilibre concurrentiel. Mais nous avons besoin de quelques résultats additionnels.

Définissons la *ligne de pleine assurance (LPA)* comme étant le lieu géométrique de (α, β) pour lequel l'utilité marginale de la consommation est la même avec les événements accident et non-accident. Dans le cas de l'utilité séparable, la LPA est donnée par

$$u'_0(w-\beta) = u'_1(w-d+\alpha) . \quad (4.2)$$

Notez que lorsque la fonction d'utilité est aussi indépendante des événements, la ligne de pleine assurance est donnée par $w-\beta = w-d+\alpha$ où $d = \alpha-\beta$ — la consommation est de même niveau pour les deux événements.

Souvenez-vous qu'avec l'utilité séparable, la ligne d'effort nul est donné par

$$\lim_{e \downarrow 0} p'(e)(-u_0+u_1) - 1 = 0 \quad (4.3)$$

Les positions relatives de la *LEN* et de la *LPA* seront d'intérêt dans l'analyse qui suit. Si $\lim_{e \downarrow 0} p'(e) = -\infty$, donc $(-u_0+u_1) = 0$ le long de la *LEN* qui, avec l'utilité indépendante des événements implique que $u'_0 = u'_1$. Ainsi, quand $\lim_{e \downarrow 0} p'(e) = -\infty$, la *LPA* coïncide avec la *LEN*, et les deux lignes satisfont $d = \alpha-\beta$. Si, alternativement, $\lim_{e \downarrow 0} p'(e)$ est fini, alors $-u_0+u_1 < 0$, ce qui, avec l'utilité indépendante des événements, implique que $w-\beta > w-d+\alpha$ ou $d > \alpha + \beta$. Dans ce cas, la *LEN* reste strictement à l'intérieur de la *LPA* (ce cas est démontré à la figure 4).

5. EXISTENCE ET CARACTÉRISTIQUES DE L'ÉQUILIBRE CONCURRENTIEL

Une analyse complète de l'existence et des caractéristiques de l'équilibre concurrentiel en présence de risque moral est compliquée pour des raisons que nous allons maintenant expliquer. Même Arnott et Stiglitz [1987b] ne fournissent qu'une analyse partielle pour le cas spécial de l'utilité séparable et indépendante des événements. En conséquence, cette section touchera seulement des points d'intérêt spécial.

La procédure suivie par Arnott et Stiglitz [1987b] est de définir l'équilibre concurrentiel comme étant un équilibre de Nash dans les contrats admissibles, avec liberté d'entrée et de sortie, et d'examiner la nature de l'équilibre sous des définitions concurrentes de contrats admissibles.

5.1. *Quand les achats d'assurance sont observables, l'équilibre occasionne un rationnement de l'assurance et, possiblement, de l'assurance aléatoire*

La réaction intuitive d'un économiste formé dans la tradition Arrow-Debreu est de dire « Pourquoi est-il important d'aborder la définition des contrats acceptables? Dans les économies concurrentielles normales, chaque optimum de Pareto peut être décentralisé par un système de prix, avec des transferts forfaitaires. Pourquoi ce résultat ne s'applique-t-il pas aux économies avec risque moral? » En présence de risque moral, pour tout prix d'assurance, les assurés produiront moins de prévention que le niveau socialement optimal. Étant donné que les assurés produiront typiquement plus d'effort si, à un prix donné, moins d'assurance est fournie, une amélioration dans le sens de Pareto peut être normalement obtenue par le rationnement de la quantité d'assurance qu'un assuré peut acheter. En outre, si cela est possible, les assureurs concurrentiels fourniront de tels contrats.

Cette intuition est maintenant formalisée, en utilisant la géométrie développée dans les sections précédentes. La séparabilité de la fonction d'utilité est présumée. Revenons à la figure 4. Le contrat qui donne le maximum d'utilité aux individus, sous le risque moral et avec des profits non négatifs, est à $\sigma \equiv (\alpha^*, \beta^*)$, le point d'utilité maximum sur le lieu géométrique de profit nul, qui est le point de tangence entre LPN et \hat{V} . Il a été démontré auparavant qu'à chaque point sur la LPN où le lieu géométrique a une pente positive et où l'effort est positif, la pente de la LPN est plus grande que celle d'un rayon liant l'origine à ce point. Parce que σ satisfait ces deux conditions, le résultat implique que le contrat (α^*, β^*) , occasionne un rationnement au prix q^* .

Un assureur peut rationner les achats totaux d'assurance d'un client seulement s'il est capable d'observer ses achats auprès d'autres assureurs. La manière la plus simple est d'insister pour être l'assureur *exclusif* de chaque client. Une telle provision peut être mise en force seulement quand les achats totaux d'assurance d'un individu sont *observables*. Supposons qu'ils le sont. Alors (α^*, β^*) est le contrat d'équilibre. Il écarte tout autre contrat qui donne des profits non négatifs, mais ne peut pas être écarté lui-même.

PROPOSITION 5 : Quand les achats d'assurance sont observables, l'équilibre concurrentiel amène les assureurs à offrir *des contrats exclusifs* qui empêchent les assurés d'acheter de l'assurance d'un autre assureur. Le contrat exclusif d'équilibre est caractérisé par le point d'utilité maximum sur le lieu géométrique de profit nul et occasionne un rationnement au prix d'équilibre. À cet équilibre, l'effort est typiquement positif.

La proposition 5, qui a été établie par Pauly [1974], doit être qualifiée. Sa preuve considère, implicitement, seulement les contrats non aléatoires. En fait, il est possible que (α^*, β^*) soit écarté par un contrat exclusif qui rend aléatoires les primes et les remboursements. Rendre aléatoire peut être une amélioration dans le sens de Pareto à cause des non-convexités générées par le risque moral. Les conditions sous lesquelles les contrats exclusifs aléatoires dominent les contrats exclusifs non aléatoires sont présentées dans Arnott et Stiglitz [1987c].

Il faut noter que lorsque l'utilité est non séparable, l'équilibre avec des contrats exclusifs est au point d'utilité maximale dans l'ensemble des possibilités, qui peut être caractérisé par des profits positifs. La raison de ce résultat est que l'effort peut être discontinu dans les paramètres du contrat d'assurance. Si un assureur offre un contrat préférable au contrat d'équilibre, l'individu réduira son effort de façon discontinue, ce qui rendra le contrat non rentable.

5.2. *Quand seulement des contrats de prix sont acceptables, l'équilibre concurrentiel peut occasionner des profits positifs ou l'inactivité*

L'analyse de l'équilibre concurrentiel en présence de risque moral devient subtile quand les achats d'assurance totaux d'un individu ne sont pas observables. Quelles restrictions a priori sur l'ensemble des contrats acceptables sont justifiées? Est-ce qu'on permettrait des contrats donnant une couverture d'assurance négative, où un individu paie quand un accident a lieu? Est-ce que la définition d'équilibre exclurait les polices d'assurance qui sont offertes mais ne sont pas achetées? Avant de considérer ces questions, il sera instructif d'examiner les caractéristiques d'équilibre quand les assureurs sont restreints à offrir des contrats de prix dans lesquels les individus peuvent acheter autant d'assurance qu'ils le veulent aux prix cités.

Il est simple d'établir que *lorsque les assureurs sont restreints à offrir seulement des contrats de prix, l'équilibre existe toujours et, à l'équilibre, les achats d'assurance sont caractérisés par le point le plus bas sur le sentier consommation-prix compatible avec des profits non négatifs.*

Pauly [1974] a étudié *l'équilibre de prix actif avec profit nul*. Quand les achats d'assurance sont positifs et quand l'équilibre est sur le sentier consommation-prix :

$$q = \frac{\beta}{\alpha} = \frac{p(\partial U_1 / \partial y_1)}{(1-p)(\partial U_0 / \partial y_0)}, \quad (5.1)$$

et comme les assureurs ont des profits nuls,

$$\frac{\beta}{\alpha} = \frac{p}{1-p}. \quad (5.2)$$

Une comparaison entre (5.1) et (5.2) établit qu'à un équilibre de prix actif avec profit nul, la pleine assurance est fournie. À la section 4, il a été démontré qu'avec la pleine assurance, un individu peut produire soit un effort positif soit un effort nul. Dans les deux cas, un équilibre a lieu au point d'intersection entre le lieu géométrique de profit nul et la ligne de pleine assurance.

Cependant, un équilibre de prix peut être d'un autre type. À cause des possibles discontinuités du sentier consommation-prix causées par les non-convexités des courbes d'indifférence, un équilibre de prix peut être inactif ou peut être caractérisé par des profits positifs. La figure 2 illustre une situation où le sentier consommation-prix saute d'un côté à l'autre du lieu géométrique de profit nul sans le couper. Dans ce cas, le point M est l'équilibre parce qu'il est le point le plus bas sur le sentier consommation-prix compatible avec des profits non négatifs. En fait, à M , les profits sont strictement positifs! La raison qui explique des profits positifs à l'équilibre est qu'aux points de discontinuité du sentier consommation-prix, l'effort est discontinu dans les prix d'assurance. Si le prix d'assurance est baissé un peu au-dessous de q^M , l'individu réduit son effort de façon discontinue, ce qui rend l'assurance peu rentable.

L'équilibre de prix peut être aussi à l'origine. À un prix élevé d'assurance, l'individu préfère ne pas acheter d'assurance. Si les prix d'assurance sont baissés, l'individu, à un point donné, achète une grande quantité d'assurance et réduit son effort de façon discontinue, ce qui rend l'assurance peu rentable. Ainsi, l'assurance n'est pas achetée à l'équilibre et on dit que le marché est inactif.

PROPOSITION 6 : Quand seulement des contrats d'assurance de prix sont acceptables, l'équilibre a lieu au point le plus bas sur le sentier de consommation-prix compatible avec un profit nul. Il y a trois types d'équilibre de prix : a) l'équilibre de prix avec profit nul qui correspond à la pleine assurance; b) l'équilibre de prix avec des profits positifs; et c) l'équilibre de prix inactif.

5.3. La non-existence possible d'un équilibre de quantité

Dans la section précédente, les assureurs étaient restreints à offrir des contrats de prix. Mais avec les achats d'assurance non observables, il n'y a pas de bonnes raisons d'exclure les contrats de quantité (non exclusifs) qui spécifient un (α, β) , où de façon équivalente, une quantité α et un prix $q \equiv \beta/\alpha$ d'assurance. En fait, les assureurs peuvent préférer offrir des contrats de quantité à des contrats de prix parce qu'ils leur donnent un contrôle, du moins partiel et indirect, des achats totaux de leurs clients.

Quand tous les assureurs offrent des contrats de quantité, les caractéristiques de l'équilibre dépendent du nombre d'assureurs dans le marché et du contrat offert par chacun, qui ensemble constitue *la structure du marché*. Par exemple,

l'équilibre peut être différent si chaque individu achète la moitié de ses achats plutôt que le quart d'un assureur particulier. À cause de la complexité d'une analyse complète des équilibres de quantité, seulement un résultat sera présenté ici – la possibilité de la non-existence d'un équilibre de quantité.

Les contrats acceptables sont les contrats qui offrent une seule police de quantité. En outre, chaque assureur est autorisé à vendre seulement un contrat et seulement une unité de ce contrat à chaque individu.

Supposons que l'équilibre existe avec un seul assureur dans le marché, mais avec une possibilité d'entrée pour les autres assureurs. Pour simplifier, il est présumé que la fonction d'utilité est séparable. Si le contrat offert par l'assureur titulaire du marché donne un profit positif, un autre assureur peut faire un profit en offrant une petite police supplémentaire ; ainsi, il est nécessaire que l'équilibre soit sur le lieu géométrique de profit nul. Si le contrat de l'assureur titulaire du marché offre moins d'assurance que l'individu voudrait acheter au prix cité, un autre assureur peut faire un profit en offrant une petite police supplémentaire positive à un prix un peu plus haut. Au contraire, si le contrat de l'assureur titulaire fournit *plus* d'assurance que l'assuré voudrait acheter, un autre assureur peut gagner s'il offre une petite police supplémentaire négative à un prix un peu plus bas que celui du contrat du titulaire du marché. Ainsi, il est nécessaire qu'un contrat d'équilibre ait lieu sur le sentier de consommation-prix. Mais, comme nous l'avons déjà démontré, il est possible qu'il n'y ait pas de point d'intersection entre le lieu géométrique de profit nul et le sentier consommation-prix, dans lequel cas, il n'y a pas d'équilibre avec un seul assureur sur le marché.

5.4. *Une expansion de l'ensemble des contrats acceptables – la possibilité d'une infinité d'équilibres*

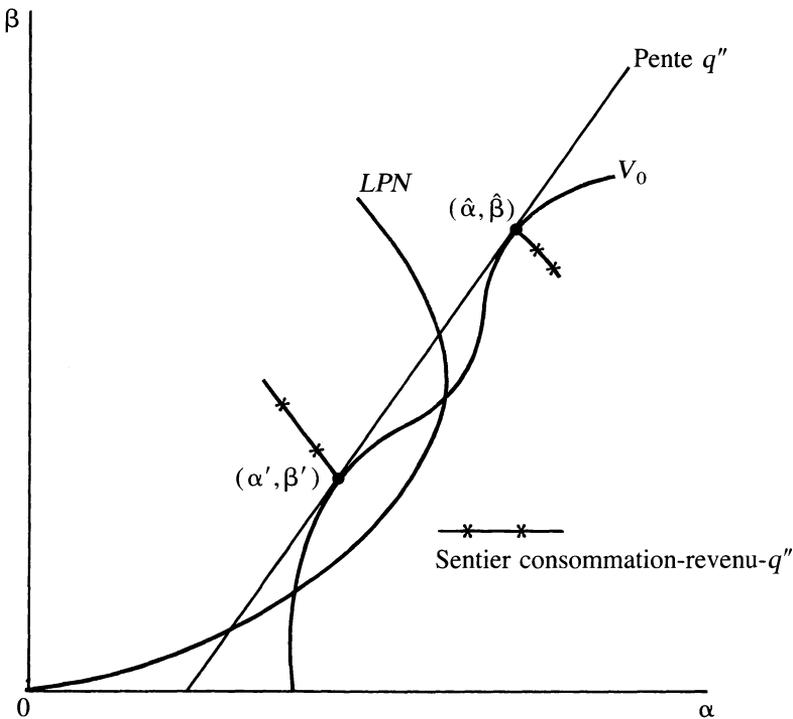
On a déjà considéré le cas où les achats d'assurance totaux d'un individu ne sont pas observables – celui où tous les assureurs offrent des contrats de prix et où tous les assureurs offrent un seul contrat de quantité avec une seule police d'assurance. Dans les deux cas, l'ensemble des contrats acceptables était restreint sans raison valable. Quand l'ensemble des contrats acceptables est élargi, permettant à chaque assureur d'inclure plusieurs polices d'assurance dans un contrat, de combiner des contrats de prix et de quantité, et d'offrir des polices d'assurance qui ne sont pas achetées à l'équilibre, on peut obtenir une infinité d'équilibres.

Ceci peut être démontré en utilisant la figure 5. Supposons qu'il y a un seul assureur dans le marché qui offre un contrat contenant une police de quantité (α', β') et une police de prix avec un prix q'' . (α', β') est le point le plus bas sur le sentier consommation-revenu correspondant au prix q'' à l'intérieur de l'ensemble des possibilités. Si le contrat de l'entreprise titulaire peut être dérangé, l'assuré achètera une quantité d'assurance quelconque, soit $(\tilde{\alpha}, \tilde{\beta})$, d'un autre assureur. Ainsi, sans manque de généralité, on peut supposer qu'un assureur, entrant dans le marché, offre un contrat de quantité. Aucun contrat d'entrée avec un prix plus élevé que q'' ne sera acheté, parce que le contrat de la firme titulaire

permet à l'individu d'acheter autant d'assurance qu'il veut au prix q'' . Par unité d'assurance, le contrat d'entrée qui permet les profits les plus grands a un prix seulement un peu plus bas que q'' . Si un contrat est offert à un tel prix, l'individu achètera (α', β') , de même que le contrat d'entrée et un montant quelconque d'assurance au prix q'' lui donnant des achats totaux correspondant à un point un peu au-dessous de $(\hat{\alpha}, \hat{\beta})$ sur le sentier consommation-revenu pour le prix q'' . Si $p(e(\hat{\alpha}, \hat{\beta}))/1-p(e(\hat{\alpha}, \hat{\beta})) > q''$, le contrat d'entrée perdra de l'argent. Parce que l'analyse ci-dessus n'a pas été fonction des conditions locales, s'il y a un tel équilibre, alors il y a une infinité d'équilibres analogues, chacun correspondant à un prix différent. Notez que même si la police de prix de l'assureur titulaire n'est pas achetée à l'équilibre, elle sert à décourager l'entrée. Une police utilisée de cette façon est appelée une police latente.

FIGURE 5

UNE ILLUSTRATION D'UNE INFINITÉ D'ÉQUILIBRES EN PRÉSENCE D'UNE EXPANSION DE L'ENSEMBLE DES CONTRATS ADMISSIBLES.



5.5. Commentaires

Les trois dernières sous-sections ont démontré que lorsque les assureurs ne peuvent pas observer les achats totaux de leurs clients, la nature de l'équilibre est sensible à la éfinition de l'ensemble des contrats acceptables. Les formes

de contrats que l'on observe dans la réalité dépendent, *inter alia*, de la connaissance des assureurs du marché, des coûts d'administration et du degré de sophistication intellectuelle des assurés. Parce que toutes ces considérations dépendent des conditions du temps et du lieu, les formes de contrats qui seraient admises dans l'ensemble de contrats acceptables dépendent du contexte.

Dans cette section, on a supposé que les assureurs n'avaient pas d'information sur les achats d'assurance de leurs clients à d'autres assureurs ou avaient une information complète. Hellwig [1983] a analysé des cas intermédiaires où la communication entre les assureurs est endogène. Il a démontré que si un équilibre existe, il ne sera pas caractérisé par une communication complète. Une autre approche de traitement de cas intermédiaires est de supposer que les assureurs peuvent obtenir des renseignements concernant les achats totaux de leurs clients, mais à un coût.

Qu'avons-nous appris de l'analyse précédente des contrats d'assurance observés dans la réalité? Tous les contrats individuels d'assurance normaux, sauf l'assurance contre la mort et ceux contre les accidents d'aviation commerciale, contiennent des clauses d'exclusivité. L'analyse est consistante avec cette observation, parce que le risque moral est probablement important dans presque toutes les formes d'assurance normale à l'exception des deux cas précités. D'autres caractéristiques comme l'absence de contrats avec assurance négative ou aléatoire, avec des polices latentes, et pour des accidents fréquents mais sans gravité, sont probablement causées par des coûts de transaction. Dans la littérature appliquée sur l'information imparfaite, on rencontre quelques fois l'assertion que l'absence de certains marchés peut être expliquée par la présence du risque moral et de la sélection adverse, sans référence aux coûts de transaction. De telles assertions semblent être sans fondement.

6. L'ASSURANCE HORS-MARCHÉ

L'analyse précédente était appliquée à *un marché* d'assurance, mais elle peut être facilement adaptée pour traiter d'autres formes d'assurance.

6.1. *Les rapports principal-agent*

Presque tous les rapports principal-agent sont caractérisés dans la réalité par le risque moral. Un exemple classique est le choix de contrat entre un propriétaire de terrain et un ouvrier (e.g. Stiglitz [1974]). Est-il efficace pour le propriétaire de i) louer le terrain à l'ouvrier ; ii) payer un salaire à l'ouvrier ; ou iii) partager la production avec l'ouvrier dans un contrat de partage des récoltes? La production du terrain dépend du temps incertain et de l'effort de l'ouvrier et ni l'un ni l'autre ne peuvent être observés par le propriétaire. Si le propriétaire loue le terrain, l'ouvrier est incité à travailler d'arrache-pied mais il est exposé à beaucoup de risques parce qu'il est le réclamatant résiduel. Au contraire, si le propriétaire paie un salaire à l'ouvrier, celui-ci est incité à être paresseux et négligent, parce qu'il peut prétendre qu'une production faible est causée par le mauvais temps. Ainsi, le métayage peut offrir la balance optimale entre le partage des risques et les diverses incitations.

Pour formaliser cette intuition, supposons qu'il y a deux niveaux de production et que la probabilité du niveau de production élevé croît lorsque l'effort de l'ouvrier augmente. Soit θ^E et θ^F , les niveaux élevé et faible de production ; soit y^E et y^F , les niveaux de consommation de l'ouvrier correspondants ; soit P le paiement moyen requis par le propriétaire et soit e l'effort. Avec les transformations des variables $\theta^E - P \leftrightarrow w$, $\theta^E - \theta^F \leftrightarrow d$, $y^E \leftrightarrow w - \beta$, et $y^F \leftrightarrow w - d + \alpha$, le modèle est le même que celui analysé dans cet essai sauf que le propriétaire recevra, en moyenne, une fraction de la production pour l'usage de son terrain. On obtient un résultat analogue à celui des contrats d'exclusivité, à savoir le métayage, si le propriétaire peut s'assurer que ses ouvriers n'obtiennent pas d'assurance contre la variabilité de la production d'une tierce personne.

6.2. *L'assurance sociale*

Presque toutes les formes d'assurance sociale sont influencées par le risque moral. Seulement un exemple sera présenté – le taux de remboursement (le ratio des paiements d'assurance sur le salaire) optimal d'assurance-chômage.

Le taux de remboursement influence une variété de marges – combien d'heures travaillées par semaine, la relation, dans le choix d'un emploi, entre le salaire et la satisfaction professionnelle, les décisions de démissionner et d'accepter une autre offre d'emploi, etc. L'exposé ici concernera une autre marge – l'effort de recherche d'emploi d'un chômeur. Les deux événements sont « trouver une position acceptable » et « ne pas trouver une position acceptable ». Un taux de remboursement plus haut donne plus d'assurance au chômeur contre l'incertitude de succès dans sa recherche d'emploi, mais parce que son effort n'est pas observable, baisse ses incitations au furetage. Le modèle est complètement analogue à celui présenté dans l'essai et permet d'obtenir un résultat analogue, c'est-à-dire que, typiquement, le remboursement sera partiel.

6.3. *À cause du risque moral, des institutions informelles qui diminuent l'efficacité peuvent se développer spontanément*

Cet exemple peut être utilisé pour illustrer un autre principe intéressant. Étant donné que l'assurance contre le chômage est rationnée, les membres d'une famille ou d'un groupe peuvent conclure une entente d'assistance mutuelle – ceux qui ont un emploi assistent financièrement les chômeurs. À cause de ces ententes les chômeurs ont moins d'incitation au furetage. Les membres d'une entente le savent, mais calculent que la plus grande partie de l'augmentation des bénéfices reçus du chômage par les membres de l'entente est payée par le gouvernement et principalement par la population entière. Ils ne tiennent pas compte du fait que si tous les citoyens calculent de la même manière, il y aura une augmentation des impôts ou une réduction des paiements du gouvernement. Il peut être démontré que, si les membres ne peuvent pas observer l'effort des autres membres, le bien-être de la société sera réduit par de telles ententes. La provision d'assurance dans de telles ententes amoindrit les bénéfices d'exclusivité. C'est un exemple

d'un phénomène général qui peut être d'importance. *Quand le risque moral existe, des institutions informelles hors-marché de provision d'assurance peuvent se développer spontanément, ce qui réduit l'efficacité des assurances fournies par le gouvernement et par le marché.* Le remède pour cette défaillance hors-marché est problématique.

7. ÉCONOMIE DU BIEN-ÊTRE AVEC RISQUE MORAL – UN ACCIDENT, UN BIEN

Est-ce que le gouvernement peut améliorer l'efficacité par son intervention lorsqu'il y a du risque moral? Pour répondre à cette question, il est important de spécifier quelles informations le gouvernement possède. Il sera présumé ici que le gouvernement possède les mêmes informations que les assureurs dans le marché.

Dans cette section, nous considérerons le cas traité jusqu'ici avec un seul accident et un seul bien.

Dans la section 5, il a été démontré que lorsque les achats d'assurance totaux d'un assuré sont observables, l'équilibre est efficace sous la réserve du risque moral.

Quand les achats d'assurance totaux d'un assuré ne sont pas observables, le traitement propre de l'économie du bien-être est problématique parce que les contraintes qui déterminent l'ensemble des contrats acceptables peuvent influencer les coûts et les possibilités d'intervention du gouvernement.

Le problème soulevé précédemment ne se présente pas lorsque les courbes d'indifférence sont convexes. Dans ce cas, sans intervention du gouvernement, l'équilibre existe toujours et a lieu au point d'intersection entre le lieu géométrique de profit nul et la ligne de pleine assurance. Considérons maintenant les effets d'un impôt appliqué aux assureurs sur les remboursements d'assurance et des revenus redistribués d'une façon forfaitaire aux assurés. L'impôt force les assureurs à offrir l'assurance à un prix plus élevé que le prix actuariel ce qui stimule l'effort. Il peut être démontré qu'avec un taux d'impôt adéquat, le gouvernement peut forcer l'économie à un optimum social contraint. Ainsi, *quand les achats d'assurance totaux des assurés ne sont pas observables et quand les courbes d'indifférence sont convexes, un optimum social contraint peut être atteint par la taxation des remboursements d'assurance.* Sous la réserve mentionnée dans le paragraphe précédent, un résultat similaire est valide quand les courbes d'indifférence ne sont pas convexes.

8. ÉCONOMIE DU BIEN-ÊTRE EN PRÉSENCE DE PLUSIEURS MARCHÉS

Cette section révisé quelques résultats présentés dans Arnott et Stiglitz [1986a].

Il a été démontré que lorsque les achats d'assurance sont observables en présence d'un seul bien et d'un seul type d'accident, l'équilibre concurrentiel est efficace sous la réserve du risque moral. Dans cette section, il sera démontré que même avec l'observabilité des achats d'assurance, l'équilibre concurrentiel sera inefficace, en général, s'il y a plus qu'un bien ou si les assurés n'achètent pas toutes leurs assurances contre tous les accidents d'un seul assureur.

L'explication générale de ce résultat générique concernant l'inefficacité est que la provision d'assurance en présence de risque moral implique que l'assuré recevra moins que tous les bénéfices sociaux pour ses soins. En conséquence, non seulement l'individu produira moins d'effort que le niveau socialement optimal, mais il y aura aussi une externalité induite par l'assurance qui causera des déviations entre les prix du marché et les coûts d'opportunité sociaux correspondants.

8.1. *Les prix du marché ne coïncident pas avec les prix d'ordre et les deux théorèmes de la théorie du bien-être sont invalides*

Considérons d'abord une économie avec un seul type d'accident, un seul bien qui prévient l'accident – les extincteurs, par exemple – un seul bien de consommation, et un seul type d'effort. Typiquement, la présence du risque moral fait que les assurés achètent trop peu du bien qui prévient l'accident et consacrent trop peu d'efforts, comparativement à la situation de premier rang où l'effort et les achats sont observables. Normalement, la subvention des extincteurs stimulera leur achat et réduira la probabilité d'accident. Ceci réduira la perte d'efficacité relative à la situation de premier rang et ainsi améliorera le bien-être social. Arnott et Stiglitz [1986a] développent ce point en dérivant quelques règles pour la taxation optimale dans ce contexte. Cette ligne de raisonnement implique, par exemple, que la sécurité industrielle serait subventionnée, et que l'alcool serait taxé pour décourager la conduite en état d'ébriété.

Cette ligne de raisonnement établit en plus, qu'à *l'équilibre concurrentiel dans une économie avec plusieurs marchés et en présence de risque moral, les prix du marché ne coïncident pas avec les prix d'ordre*. En conséquence, *ni le premier* (chaque équilibre concurrentiel est Pareto optimal) *ni le deuxième* (chaque optimum de Pareto peut être décentralisé comme un équilibre concurrentiel avec transferts forfaitaires) *théorème de l'économie du bien-être est valide*. Cela implique que les possibilités d'interventions gouvernementales pour améliorer le bien-être sont considérables. Cependant, pour établir les avantages véritables de l'intervention gouvernementale dans un contexte spécifique, il faut établir que les bénéfices excèdent les coûts.

8.2. *Il y a de l'inefficacité si les différentes polices d'assurance d'un individu ne sont pas fournies par un seul assureur*

Considérons maintenant une économie avec un seul bien mais deux types d'accidents qui sont indépendants statistiquement – l'incendie de la maison et l'invalidité occupationnelle, par exemple. Supposons que la technologie de la

provision d'assurance est telle que les assureurs se spécialisent dans les types d'assurances qu'ils vendent – dans le contexte qui nous intéresse, qu'un groupe d'assureurs fournit l'assurance contre les incendies et l'autre l'assurance contre les invalidités. Il y a deux sources de défaillance de marché distinctes. Premièrement, l'utilité sociale marginale des revenus peut être différente pour les deux types d'assureurs ; d'une façon différente, une amélioration dans le sens de Pareto peut être obtenue par le transfert d'un dollar d'un type d'assureur à un autre. Si on enlève un dollar aux assureurs contre les incendies, cela leur fait offrir de l'assurance à un prix supérieur au prix actuariel, ce qui stimule l'effort de leurs clients et réduit la perte d'efficacité associée au risque moral. En donnant le même dollar aux assureurs contre l'invalidité, on augmente la perte sèche due au risque moral, mais il n'y a pas de raison pour que les effets du transfert sur l'efficacité s'annulent. Ainsi, en général, le gouvernement peut améliorer l'efficacité par des transferts entre les types d'assurances.

Deuxièmement, même si les probabilités des deux types d'accidents sont statistiquement indépendantes, le niveau d'effort consacré à prévenir un type d'accident influencera le niveau d'effort pour prévenir l'autre type d'accident via la fonction d'espérance d'utilité.

Si nous éliminons l'hypothèse que la technologie de la provision d'assurance est telle que les assureurs se spécialisent dans des types d'assurance, ces résultats peuvent être réinterprétés en indiquant que tous les besoins d'assurance d'un individu, à tout moment dans le temps, devraient être fournis par un seul assureur, et, à l'équilibre concurrentiel, seront fournis par un seul assureur – une provision d'exclusivité étendue.

9. CONCLUSION

Nous croyons que cet essai a présenté un argument persuasif à l'effet que les non-convexités et les externalités qui sont créées par le risque moral changent qualitativement les caractéristiques de l'équilibre concurrentiel.

Le risque moral est présent jusqu'à un certain point dans toutes les activités économiques. Le fait que sa présence nous oblige à reformuler la plupart des méthodes d'analyse économique – qui ignorent le risque moral – dépend de son importance *quantitative*. Il n'existe pas d'évaluation de son importance parce que la recherche empirique nécessaire exige, tout comme son fondement théorique, un modèle général et dynamique qui traite du risque moral et de la sélection adverse en même temps. À notre avis, le développement d'un tel modèle est le projet de recherche qui a la plus grande priorité à l'ordre du jour de l'économie de l'assurance.

BIBLIOGRAPHIE

- ARNOTT, R. et J.E. STIGLITZ (1982), «Equilibrium in Competitive Insurance Markets. The Welfare Economics of Moral Hazard – Part I : Basic Analytics», Cahier de recherche 465, Institute for Economic Research, Queen's University.
- ARNOTT, R. et J.E. STIGLITZ (1986a), «The Welfare Economics of Moral Hazard», miméo.
- ARNOTT, R. et J.E. STIGLITZ (1986b), «Moral Hazard and Optimal Commodity Taxation», *Journal of Public Economics*, 29, pp. 1-24.
- ARNOTT, R. et J.E. STIGLITZ (1987a), «The Basic Analysis of Moral Hazard», miméo.
- ARNOTT, R. et J.E. STIGLITZ (1987b), «Equilibrium in Competitive Insurance Markets with Moral Hazard», miméo.
- ARNOTT, R. et J.E. STIGLITZ (1987c), «Randomization with Asymmetric Information», miméo.
- ARROW, K.J. (1970), *Essays in the Theory of Risk-Bearing*, London : North-Holland.
- DEBREU, G. (1959), *Theory of Value*, New York : John Wiley.
- DIONNE, G. (1982), «Moral Hazard and State Dependent Utility Function», *Journal of Risk and Insurance*, 49, pp. 405-423.
- GROSSMAN, S. et O. HART (1983), «An Analysis of the Principal-Agent Problem», *Econometrica*, 51, pp. 7-45.
- HELLWIG, M., «Moral Hazard and Monopolistically Competitive Insurance Markets», Cahier de recherche 105, University of Bonn.
- HELPMAN, E. et J.J. LAFFONT (1975), «On Moral Hazard in General Equilibrium», *Journal of Economic Theory*, 10, pp. 8-23.
- HOLMSTROM, B. (1979), «Moral Hazard and Observability», *Bell Journal of Economics*, 10, pp. 74-91.
- MARSHALL, J.M. (1976), «Moral Hazard», *American Economic Review*, 66, pp. 880-890.
- MIRRLIENS, J.A. (1975), «The Theory of Moral Hazard and Unobservable Behaviour – Part I», Nuffield College, Oxford, miméo.
- PAULY, M. (1974), «Overprovision and Public Provision of Insurance», *Quarterly Journal of Economics*, 88, pp. 44-62.
- SHAVELL, S. (1979), «On Moral Hazard and Insurance», *Quarterly Journal of Economics*, 93, pp. 541-562.
- SPENCE, M. et R. ZECKHAUSER (1971), «Insurance, Information, and Individual Action», *American Economic Review*, 61, pp. 380-387.
- STIGLITZ, J.E. (1974), «Incentives and Risk Sharing in Sharecropping», *Review of Economic Studies*, 41, pp. 219-255.