

UNIVERSITÉ DE MONTRÉAL

ASYMÉTRIE D'INFORMATION ET
OFFRE DE SERVICES FRAUDULEUX

PAR
HERVÉ DÉRY
DEPARTEMENT DE SCIENCES ÉCONOMIQUES
FACULTÉ DES ARTS ET DES SCIENCES

MÉMOIRE PRÉSENTÉ À LA FACULTÉ DES ÉTUDES SUPÉRIEURES
EN VUE DE L'OBTENTION DU GRADE DE
MAÎTRE ES SCIENCES (M.Sc.)

(NOVEMBRE 1984)

Centre de documentation

MAR 06 1985

Sciences économiques, U. de M.



TABLE DES MATIÈRES

	<u>PAGE</u>
LISTE DES TABLEAUX	iv
SOMMAIRE	v
INTRODUCTION	1
CHAPITRE I - UNE REVUE DE LA LITTÉRATURE	6
1.1 Analyse théorique.....	8
1.1.1 Analyse graphique.....	8
1.1.2 Le médecin sans éthique.....	9
1.1.3 Le médecin avec éthique.....	14
1.1.4 Le producteur incertain.....	18
1.1.5 Comportement du consommateur.....	20
1.2 Analyse empirique.....	24
1.2.1 Difficultés théoriques de la vérification empirique.....	24
1.2.1.1 Influence réciproque entre l'offre et la demande.....	24
1.2.1.2 Elasticités de demande.....	25
1.2.1.3 Les facteurs de production.....	26
1.2.1.4 Fonction de production aléatoire....	27
1.2.1.5 Mobilité des consommateurs.....	28
1.2.2 Application empirique.....	29
1.2.2.1 Création de demande par l'offreur...	29
CHAPITRE II - ANALYSE DE L'HYPOTHÈSE D'OFFRE EXCÉDENTAIRE	34
2.1 Asymétrie d'information et optimalité.....	36
2.1.1 Atteinte à l'optimum de Pareto.....	36
2.1.2 Niveau de fraude optimal: achat d'un expert	39
2.2 Hypothèse d'offre excédentaire.....	41
2.3 Des modèles prototypes pour l'analyse.....	47
2.3.1 Prototype à un seul bien.....	47
2.3.2 Prototype séquentiel avec productions jointes.....	49
2.3.3 Prototype analytique: le cas des garagistes	53
2.3.3.1 Prototype à un seul bien.....	53
2.3.3.2 Prototype séquentiel à productions jointes.....	54

	<u>PAGE</u>
2.4 Généralisation des résultats.....	56
2.5 Productions jointes et hypothèse d'offre excédentaire.....	57
2.6 Productions jointes et économies d'envergure en concurrence pure et parfaite.....	59
2.7 Le producteur face à un prix endogène.....	66
2.8 Le prix du temps de consommation et les possibilités de fraude.....	67
2.8.1 Le producteur et le prix de temps de consommation.....	68
2.8.2 Le consommateur et le prix des temps de consommation.....	70
2.8.3 Temps de consommation et offre en production jointe.....	73
2.9 Effet de l'assurance sur la fraude en production jointe.....	76
2.9.1 Risque moral versus offre de services frauduleux.....	77
2.9.2 Le producteur avec assurance en production jointe.....	82
2.9.3 Le consommateur sous l'assurance.....	84
2.9.3.1 Prix du temps de consommation et productions jointes.....	88
CHAPITRE III - RÉGLEMENTATION DANS DEUX MARCHÉS IMPORTANTS..	91
3.1 La réglementation sur le marché de la réparation automobile.....	91
3.1.1 La Loi sur la protection du consommateur...	91
3.1.2 Décret sur les salariés de garage de la région de Montréal.....	95
3.1.3 Réglementation et offre de services frauduleux en production jointe.....	96
3.2 La réglementation sur le marché des services médicaux.....	100
3.2.1 Les incitatifs financiers.....	104
CONCLUSION	117

	<u>PAGE</u>
ANNEXES	
ANNEXE MATHÉMATIQUE	
Modèle I - Modèle à un bien.....	123
Modèle II - Modèle séquentiel avec productions jointes.....	124
Modèle III- Tarification optimale.....	125
Modèle IV - Le garagiste: prototype à un bien...	127
Modèle V - Le garagiste: prototype séquentiel avec productions jointes.....	129
ANNEXE JURIDIQUE: Protection du consommateur et réglements d'application.....	130
BIBLIOGRAPHIE	136
REMERCIEMENTS	140

LISTE DES TABLEAUX

		<u>PAGE</u>
TABLEAU 2.1	Prototype d'offres de services: simulation.....	51
TABLEAU 3.1	Coût moyen des contacts-patients et coût moyen des examens selon les groupes d'omnipraticiens étudiés et la classe de paiement.....	111
TABLEAU 3.2	Coût des examens en proportion du coût des contacts-patients, nombre d'autres actes par contact-patient et pourcentage de variation selon les groupes d'omnipraticiens et la classe de paiement.....	112
TABLEAU 3.3	Coût moyen des autres actes et pourcentage de variation selon les groupes d'omnipraticiens étudiés et la classe de paiement.....	113

SOMMAIRE

Dans ce mémoire, nous nous intéressons aux fondements micro-économiques des problèmes d'allocation des ressources, en particulier celui de la fraude. La notion de fraude est limitée à une offre de services qu'un consommateur-expert refuserait. L'asymétrie d'information entre producteurs et consommateurs peut motiver les offreurs à avoir des comportements frauduleux. Citons à titre d'exemple les relations patient-médecin et garagiste-consommateur.

Notre mémoire est divisé en trois chapitres. Dans le premier chapitre, nous présentons une revue de la littérature concernant l'hypothèse de création de demande par l'offreur et la fraude. Une bonne partie de cette littérature est interprétée en fonction de la relation patient-médecin. Il ressort de cette littérature que le producteur individuel qui a la possibilité de créer une demande pour ses services peut éviter (ou à tout le moins restreindre) une baisse de prix suite à une augmentation du nombre de producteurs per capita (déplacement de la courbe d'offre vers la droite) en entraînant une surconsommation de ses services.

Les nombreuses difficultés théoriques de la vérification empirique ne remettent pas en cause la plausibilité de l'hypothèse de création de demande par l'offreur mais bien la mesure de cette hypothèse. A ce niveau, l'étude de Fuchs (1978) (application aux médecins à l'aide de données américaines) est particulièrement convaincante et surmonte les principales embûches théoriques. Ces difficultés sont identifiées comme étant: l'influence réciproque entre l'offre et la demande, les élasticités de demandes différentes selon le

territoire et fonctions de la concentration de médecins (l'offreur), les coûts des facteurs de production différenciés, la nature aléatoire de la fonction de production et la mobilité des consommateurs.

Dans cette revue de la littérature, nous présentons également un modèle (Darby et Karni, 1973) afin d'expliquer les motivations d'un producteur à offrir des services inutiles à un consommateur mal informé. A l'optimum, le producteur égalise le revenu marginal d'un service supplémentaire frauduleux au coût marginal en terme de pertes de revenus présentes et futures.

Dans le second chapitre, nous reconsidérons le modèle de Darby et Karni (1973) et étudions l'importance de l'hypothèse d'offre excédentaire dans la justification de la fraude. L'offre excédentaire est définie comme étant une situation dans laquelle le producteur individuel désirerait vendre un plus grand nombre de services non-frauduleux. Cette analyse nous amène à la proposition 1.

Proposition 1: La simple observation que les carnets de rendez-vous ne sont pas vides ne signifie pas nécessairement qu'il n'y a pas d'excès de capacité. Il est possible qu'il y ait eu de la fraude.

Donc, le concept d'excès de capacité (offre excédentaire) pour justifier la fraude est pertinent seulement dans une version ex ante. Ce que l'on observe sur le marché, c'est une situation ex post et il est possible que l'offre de services frauduleux ait permis de remplir les carnets de rendez-vous.

Darby et Karni (1973) supposent que les offreurs produisent un seul bien. Nous étudions le lien entre cette hypothèse et celle

d'offre excédentaire et démontrons que cette dernière peut être tributaire de la première. Pour ce faire, nous développons un modèle théorique d'offre de services frauduleux en situation d'économie d'envergure avec des productions jointes. Le producteur offre séquentiellement deux services différents qui sont joints par la fonction de coût. Lorsque le premier service est offert et accepté, le coût marginal de production du second diminue. Pour fins d'analyse, nous développons des prototypes avec des fonctions de probabilités de refus et de coûts particulières. Tous les résultats sont généralisés par la suite. Nous démontrons la proposition suivante.

Proposition 2: Lorsque la production jointe de deux services amène une diminution du coût de production du second, le prix de ce service doit diminuer de façon proportionnelle à cette diminution de coût lorsqu'ils sont offerts conjointement. Ceci afin d'éviter une offre de services frauduleux imputable aux économies d'envergure et une surconsommation de ces services.

Si la tarification optimale¹ que nous suggère la proposition 2 n'est pas respectée, il devient alors possible de justifier la fraude sans excès de capacité. D'où la proposition 3:

¹ Notons que dans un monde parfaitement concurrentiel (pleine information), les économies d'envergure peuvent justifier économiquement les productions jointes et la tarification sera optimale telle que définie par la proposition 2. Nous serions à un optimum de Pareto.

Proposition 3: Dans un contexte de production jointe, l'hypothèse d'offre excédentaire ne représente pas une condition nécessaire à l'offre de services frauduleux. Il peut y avoir offre de services frauduleux même s'il y a pleine capacité.

Par la suite, nous intégrons le temps de consommation dans notre modèle. Dans le type de marché qui nous intéresse, les économies d'envergure sont généralement associées à une diminution du temps requis pour effectuer le traitement ou la réparation. En vertu de notre analyse du comportement du consommateur (modèle séquentiel de furetage) lorsque le temps de travail est requis pour consommer, nous pouvons considérer que, toutes choses étant égales par ailleurs, la probabilité de refus est moins élevée pour un service joint. Le modèle du producteur prévoit alors une augmentation des possibilités de fraude. Cependant, ceci ne justifie pas une relation négative entre le revenu du consommateur et l'offre de services frauduleux.

L'introduction de l'assurance dans notre modèle est également une étape importante de ce travail. Les services médicaux sont souvent consommés sous l'assurance. Après avoir fait le lien entre un problème particulier à l'assurance, le risque moral, et la fraude, nous démontrons qu'une augmentation de la couverture d'assurance aura un impact plus important si nous tenons compte du temps de consommation. Ainsi, par extension, l'introduction de l'assurance - loin de limiter l'offre de services inutiles des productions jointes - favorise particulièrement la fraude sous cette condition.

Enfin, au chapitre III, nous appliquons notre modèle théorique à deux marchés importants en considérant l'effet de la réglementation québécoise sur l'offre de services frauduleux en situation de production jointe. Ces deux marchés sont la réparation des automobiles et les services médicaux.

Le marché de la réparation des automobiles est un marché qui demeure relativement peu réglementé au Québec si ce n'est que par la Loi de la protection du consommateur. L'Office de la protection du consommateur (opération auto-témoin), l'Association pour la protection des automobilistes (en se substituant aux consommateurs lésés) et la Loi sur les petites créances (en diminuant le coût des poursuites judiciaires) sont autant de facteurs qui améliorent l'efficacité de la Loi. Dans ce domaine, la vigilance et le comportement responsable du consommateur demeurent ses meilleures garanties.

Le marché des services médicaux québécois, tant qu'à lui, est très réglementé et se caractérise par un régime d'assurance-maladie universel. Nous nous intéressons particulièrement aux incitatifs financiers implicitement inclus dans la structure tarifaire.

L'analyse de deux articles² traitant du sujet confirme nos conclusions théoriques. Le profil de pratique des médecins du Québec est influencé par les incitatifs financiers incorporés dans la structure tarifaire. Les productions jointes permettent de mieux "rentabiliser une visite". Le modèle séquentiel d'offre de services frauduleux en situation de production jointe avec économies d'envergure représente bien le comportement de ceux qui offrent des services

² Boutin (1980) et Contandriopoulos (1980).

inutiles dans l'unique but de maintenir ou augmenter leur revenu répondant ainsi positivement à l'incitation financière implicite à la tarification.

Nous ne croyons pas que tous les producteurs (médecins ou garagistes dans notre exemple) sont des fraudeurs. Au contraire, les études semblent plutôt nous laisser croire que la majorité ne l'est pas. Il s'agit plutôt d'étudier le comportement de ceux qui le sont.

INTRODUCTION

Lorsqu'il y a surplus de l'offre dans un certain marché, on peut s'attendre à une baisse de prix du bien ou du service en vertu de la loi de l'offre et de la demande. En effet, une offre excédentaire signifie que la courbe d'offre s'est déplacée vers la droite dans le temps de façon plus appréciable que la courbe de demande. Il devrait en résulter une baisse de prix.

Ce résultat fondamental que nous prédit le modèle néoclassique concurrentiel suppose la souveraineté du consommateur maximisant son utilité. Or, il semblerait que cette condition ne soit pas respectée dans certains marchés où le consommateur moyen est souvent incapable de juger et d'apprécier adéquatement l'utilité qu'il retire de sa consommation sans que cela occasionne des coûts supplémentaires. Ainsi, l'impuissance du client à apprécier adéquatement le produit qu'il consomme nous amène à croire que l'une des règles fondamentales de l'efficacité économique par laquelle le consommateur souverain obtient son objet de consommation au moindre coût s'en trouve brimée.¹

¹ "the theory rest, however, on strong assumption about information and rationality on the part of the consumer. In the case of medical care the model assumes that the consumer knows (1) what this health problem is; (2) what forms of care are appropriate; (3) the relative competence of all available suppliers to provide care, and (4) the total price of a treatment provided by each possible supplier the unreality of these assumptions is readily apparent, and explains why consumers in fact delegate to suppliers the responsibility for deciding what care, and how much, should be consumed"

Dans ce mémoire, nous nous préoccupons des marchés de confiance dans lesquels le consommateur doit généralement s'en remettre à l'avis d'experts afin d'éclairer sa décision. Nous pouvons citer, à titre d'exemple, les relations patient-médecin ou garagiste-consommateur. Dans ces marchés, l'asymétrie d'information entre producteurs et consommateurs peut motiver les offreurs à avoir des comportements frauduleux. Les coûts réels du furetage sont élevés.

La notion de fraude dans ce travail est limitée. Il y a offre de services frauduleux lorsque le producteur offre un service qui serait refusé si le client était un expert. Par contre, il y a offre de services inutiles si la consommation de ces services n'augmente pas l'utilité du consommateur (indépendamment du prix). Enfin, il y a activité de création de demande lorsque le producteur déplace la courbe de demande vers la droite en donnant une information inadéquate au consommateur (ce qui exclu le cas d'agents parfaits). Pour chaque prix, la quantité demandée augmente sous l'hypothèse de création de demande. Tous ces concepts sont étroitement interreliés. Ainsi, le concept de création de demande est beaucoup plus large que celui de la fraude. L'offre de service frauduleux, tant qu'à elle, se traduira toujours par une activité de création de demande. Par contre, la réciproque n'est pas vraie et certaines activités de création de demande ne sont pas frauduleuses. Un producteur peut offrir un service inutile sans qu'il le sache: soit que sa fonction de production est aléatoire ou soit qu'il est incompétent (Dionne 1980). De plus, tout

service frauduleux n'est pas nécessairement inutile. Certains services peuvent être utiles mais ne pas valoir le prix qu'il faut déboursier à l'achat. Dans ces conditions, si le consommateur avait été expert, il n'aurait pas acheté ce service puisqu'il lui est possible d'allouer son revenu de façon plus efficace. On n'achète pas tous les biens et services qui nous sont agréables.

A titre d'exemple, nous pouvons considérer le cas des services médicaux. Le service sera utile lorsque l'impact marginal du service (noté H) sur l'utilité du consommateur (noté u) est positif. Si l'état de santé du consommateur est noté HS, nous pouvons caractériser ainsi cet impact marginal: (Evans et Wolfson 1980)

$$\frac{\partial u}{\partial H} = \frac{\partial u}{\partial H} \Big|_{HS} + \frac{\partial u}{\partial HS} \cdot \frac{\partial HS}{\partial H}$$

Seulement le terme $\partial HS/\partial H$ est issu du savoir médical (relation spécifiquement technique) et un service peut être - d'un point de vue strictement médical - à prescrire ($\partial HS/\partial H > 0$) et entraîner une désutilité pour le consommateur. En effet, la désutilité de consommation ($\partial u/\partial H \Big|_{HS} < 0$) peut être trop forte et renverser l'utilité que procure une meilleure santé.

La littérature économique nous fournit un modèle théorique (Darby et Karni 1973) caractérisant le problème d'offre de services frauduleux à un consommateur mal informé. Une hypothèse importante utilisée par ces auteurs consiste à supposer que l'offre de services frauduleux survient lorsqu'il y a excès de capacité. L'offre excédentaire est définie comme une situation dans laquelle le prix du marché

est tel qu'à ce prix, le producteur individuel désirerait vendre une plus grande quantité de services non frauduleux. De plus, ils supposent que les offreurs produisent un seul bien.

L'objectif général de ce travail étant de mieux comprendre les fondements micro-économiques des problèmes d'allocation des ressources, en particulier celui de la fraude, nous allons donc, entre autres, étudier le lien entre ces deux hypothèses et démontrer que la première est tributaire de la seconde. Il est important de noter que l'objectif de ce travail n'est pas de développer un modèle représentant le comportement des producteurs tels les médecins. C'est pourquoi nous faisons l'hypothèse de la maximisation du profit. Il ne s'agit pas de dire que tous les médecins et les garagistes sont des fraudeurs mais plutôt d'étudier le comportement de ceux qui le sont.

Pour ce faire, nous avons structuré notre mémoire en trois chapitres. Dans le premier chapitre, nous présentons une revue de la littérature concernant l'hypothèse de création de demande et de fraude. A l'aide d'articles choisis, nous faisons ressortir les principaux modèles théoriques et les principales difficultés théoriques que pose la vérification empirique du phénomène. Cette littérature nous permet d'aborder les principales vérifications empiriques de l'hypothèse de création de demande par l'offreur avec prudence et circonspection.

Au chapitre II, nous étudions plus spécifiquement le comportement des producteurs frauduleux et analysons l'hypothèse d'offre excédentaire. Nous étudions l'importance de l'hypothèse d'offre excédentaire dans le processus de détermination du niveau de fraude et du lien étroit qui existe entre cette hypothèse et le fait que les modèles existants supposent que le producteur offre un seul type de bien. Pour

ce faire, nous élaborons un modèle séquentiel d'offre de services frauduleux - en situation de production jointe avec économies d'envergure - à un consommateur mal informé. A l'aide de ce modèle à caractère purement micro-économique, nous analysons l'importance du prix du temps de consommation sur les possibilités de fraude. De plus, nous adaptons le modèle à un contexte d'assurance généralisée tel que nous le connaissons au Québec relativement à l'assurance-maladie.

Enfin, dans le chapitre III, nous appliquons notre modèle théorique à deux marchés importants à savoir le marché des services médicaux et le marché des réparations automobiles. L'objectif visé n'est pas d'effectuer le calcul économique de la réglementation mais seulement d'analyser les effets théoriques de la réglementation québécoise sur l'offre de services frauduleux en production jointe dans ces deux marchés importants.

CHAPITRE I - UNE REVUE DE LA LITTÉRATURE

Dans cette partie du travail, nous allons nous pencher sur la littérature concernant l'hypothèse de création de demande par l'offreur. Nous allons particulièrement nous attarder à analyser l'effet de cette hypothèse sur les prix. Dans un premier temps, nous allons présenter une représentation graphique du phénomène. Nous retrouvons ce type d'exercice dans plusieurs articles dont celui de Reinhardt (1978) dans lequel l'analyse est présentée de façon particulièrement claire et complète. Par la suite, nous allons présenter le modèle de Pauly (1980) qui analyse le comportement du médecin sans éthique pour ensuite voir l'effet de l'introduction d'une variable supplémentaire dans la fonction d'utilité du médecin. Cette variable supplémentaire est le niveau d'information que le médecin transmet au patient. Puis, nous allons présenter dans ses grandes lignes le modèle de Evans (1974) et son extension (Sloan et Feldman 1978). Ces auteurs ont supposé que le médecin agit en certitude.

Le médecin, lorsqu'il influence la demande peut être en incertitude face à la réaction du consommateur. Nous allons donc par la suite supposer que le médecin maximise son espérance d'utilité de profit ce qui va nous permettre de déterminer le critère d'offre de services frauduleux. Pour ce faire, nous allons utiliser le modèle de Darby et Karni (1973). Pour expliquer le comportement de suretage du consommateur qui est différent de celui présenté couramment dans la littérature sur la recherche du prix et de la qualité des biens, nous allons présenter le modèle séquentiel de suretage tel que développé par G.

Dionne (1979). Ce modèle va nous permettre de mieux comprendre le fonctionnement de ce type de marché et de justifier certaines hypothèses du modèle de Darby et Karni (1973).

Fort de cette littérature théorique, nous allons par la suite aborder la vérification empirique de l'hypothèse de création de demande. Nous allons nous attarder aux difficultés théoriques de cette vérification empirique. Nous allons donc traiter des problèmes que posent l'influence réciproque entre l'offre et la demande (Contandriopoulos, Dionne et Tessier 1983), les élasticités de demande différentes selon les régions (Pauly 1980 et Satterthwaite 1979), les problèmes relatifs aux facteurs de production (Auster et Oaxaca 1981, Contandriopoulos et al. 1983), l'incertitude au niveau de la fonction de production des médecins (Wenner, Barnes et Zubkoff 1982) ainsi que la mobilité des consommateurs (Contandriopoulos et al. 1983).

Enfin, nous allons par la suite discuter des applications empiriques. Nous allons d'abord présenter un type d'étude particulier qui a pour but d'évaluer l'utilité des opérations. Nous allons donner un exemple des deux méthodes d'analyse possibles: la méthode rétrospective (groupe des Teamsters: 1960) et la méthode prospective (McCarthy 1974).

Par la suite, nous allons nous intéresser plus spécifiquement aux études portant sur la création de demande de façon plus explicite. Après avoir considéré brièvement l'étude de Lewis (1969) nous allons présenter deux études mieux structurées et plus pertinentes: celles de Fuchs (1978) et Pauly (1980). Ce dernier arrive à des résultats différents de Fuchs (1978) en utilisant la même source de données. Nous allons essayer de comprendre le pourquoi de cette divergence.

1.1. Analyse théorique

1.1.1. Analyse graphique

Malgré le fait qu'il semble y avoir excès de capacité (surplus de l'offre) sur le marché des services chirurgicaux aux Etats-Unis (comme semble vouloir le démontrer plusieurs études (Schneider en 1976, Blackstone en 1974, Sossys et surtout Nickerson en 1976)), il semblerait que le prix des interventions ne soit pas à la baisse. Ceci peut être interprété comme étant dû au fait que les producteurs aient pu profiter de la situation d'asymétrie d'information et créer leur propre demande.

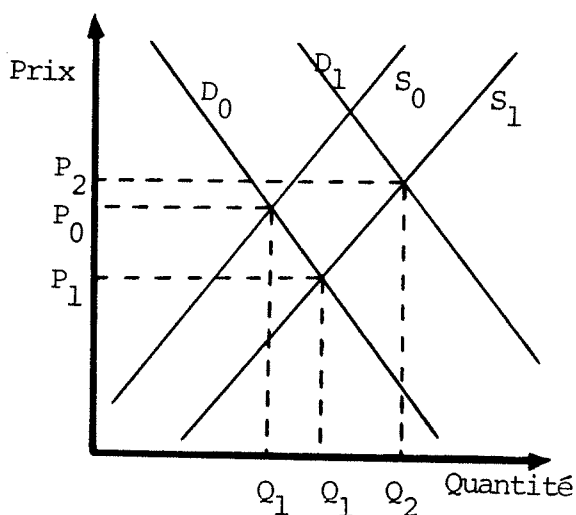


Figure 1.1

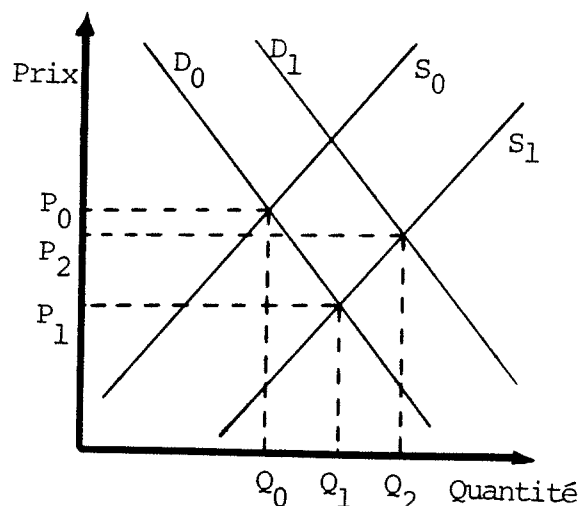


Figure 1.2

Ainsi, sur la figure 1.1 les médecins ont pu éviter la baisse de prix de P_0 à P_1 en déplaçant la courbe de demande de D_0 à D_1 et même faire augmenter le prix au-dessus de ce qu'il était (soit à P_2 P_0 P_1) avant l'augmentation du nombre de médecins per capita.

Par contre, sur la figure 1.2., nous considérons une situation où le producteur crée de la demande afin d'éviter que le prix ne chute de façon trop appréciable. Par rapport à la situation

initiale (P_0Q_0) le prix a diminué mais cette baisse est moins importante qu'elle ne l'aurait été si les producteurs n'avaient eu aucune influence sur la demande. Enfin, notons que le modèle néoclassique prévoit une augmentation des quantités échangées (de Q_0 à Q_1) suite à une augmentation du nombre de médecin per capita et ce dans les deux cas à l'étude. Le modèle de création de demande prévoit une augmentation encore plus forte.

La principale raison invoquée dans la littérature pour expliquer cette influence du médecin sur la demande de services une fois que le patient a déjà eu un contact avec lui est une relation d'agents imparfaite entre le patient et le médecin (Pauly 1980). Cette forme de relation a été justifiée de plusieurs façons: le manque d'expérience du consommateur puisqu'il ne fréquente que rarement ce genre de marché, la relation de confiance patient-médecin (Monsma 1970), le coût réel d'acquisition d'information dû au coût élevé du furetage (Dionne 1980) et la couverture d'assurance (Dionne 1980). L'éthique professionnelle empêche toute forme de publicité sur les prix alors que celle sur la compétence des offreurs demeure très limitée (Monsma 1970). De toute manière, il semble que les consommateurs soient portés à identifier la qualité au prix ce qui signifie qu'une baisse de prix pourrait être interprétée comme un manque de compétence et entraîner une diminution de la demande (Fuchs 1974).

1.1.2. Le médecin sans éthique (Pauly 1980, Anderson et al. 1981)

Dans le modèle de Pauly (1980), le médecin maximise une fonction d'utilité (u) qui est fonction du revenu (Y) et du loisir

(L). La quantité demandée est fonction des prix (P) et du niveau d'information que donne le médecin au patient (A). De plus, Pauly (1980) considère que le coût d'opportunité du loisir est constant pour le médecin. D'ailleurs, Pauly (1980) dit à ce sujet:

"In order to simplify the exposition, it will be assumed that physician time is available at a constant opportunity costs".¹

Cette hypothèse permet donc à Pauly (1980) de ramener le problème à une maximisation du revenu net (différence entre le revenu total et les coûts totaux):

$$\text{Max}_{P,A} Y = PQ(P,A) - CT(Q, (P,A))$$

Les conditions de premier ordre nous donnent respectivement les deux équations suivantes:

$$(1.1) \quad \frac{\partial Y}{\partial P} = \frac{\partial Q}{\partial P} \left[P - \frac{\partial CT}{\partial Q} \right] + Q = 0$$

Puisque $\partial Q/\partial P$ est négatif et différent de zéro, il faut que le prix soit supérieur au coût marginal afin de respecter cette condition de premier ordre. La deuxième équation:

$$(1.2) \quad \frac{\partial Y}{\partial A} = \frac{\partial Q}{\partial A} \left[P - \frac{\partial CT}{\partial Q} \right] = 0$$

Or, puisque nous savons que le prix est différent du coût marginal, cette condition de premier ordre implique que $\partial Q/\partial A = 0$

Donc, selon ce résultat, le médecin va choisir un A optimal (A* sur la figure 1.3) tel que $\partial Q/\partial A$ soit égal à zéro, ce qui signifie qu'il va donner le moins d'information possible au patient et

¹ Pauly (1980), p. 49.

qu'il va lui offrir beaucoup de services. Si le nombre d'offreurs per capita augmente, il n'y aura pas de création de demande supplémentaire. On peut donc prédire que le prix baissera.

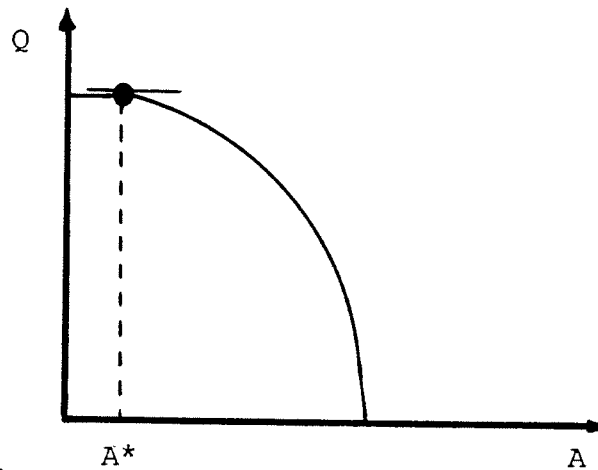


Figure 1.3

Afin de mieux comprendre ce phénomène, nous allons utiliser l'équation de détermination du prix que nous fournit Anderson, House et Ormiston (1981). Soit l'équation:

$$P = P(Q, A(W, R), R)$$

Les nouvelles variables W et R signifient respectivement la charge de travail et le nombre de médecins per capita. À partir de cette équation, nous pouvons écrire la dérivée partielle suivante:

$$(1.3) \quad \frac{\partial P}{\partial R} = \frac{\partial P}{\partial R} \Big|_{\bar{A}} + \frac{\partial P}{\partial A} \cdot \frac{\partial A}{\partial R}$$

En vertu du modèle néoclassique, nous savons que $\frac{\partial P}{\partial R} \Big|_{\bar{A}}$ est négatif. Pour ce qui est du signe de $\frac{\partial P}{\partial A}$, Anderson et al.

(1981) considèrent qu'il est négatif par hypothèse².

De plus, les auteurs considèrent que $\partial A/\partial R$ est également négatif. Leur argument pour en déduire que $\partial A/\partial R$ est négatif est le suivant:

"However, there appears to be some question in the literature regarding the signing of $\partial A/\partial R$. Following the Satterthwaite (1979) argument that as more physicians enter a market area it becomes easier to convince patients to over-consume medical services, we have $\partial A/\partial R < 0$; i.e. an increase in the density ratio leads to a higher level of demand creation activity³".

Dû à ces hypothèses, il devient clair que $\partial P/\partial R$ est de signe indéterminé pour ces auteurs et que la création de demande est une condition nécessaire mais non suffisante pour que le prix des services médicaux augmente suite à une augmentation du nombre de médecins per capita dans le temps. Par contre, s'il y a baisse de prix, elle sera moins importante que s'il n'y avait pas eu création de demande puisque $\partial P/\partial A \cdot \partial A/\partial R > 0$.

Nous notons que ce résultat est cohérent avec notre analyse graphique.

Pauly (1980), tant qu'à lui, obtient que $\partial Q/\partial A = 0$ (tiré de l'équation 1.2) ce qui signifie implicitement que $\partial A/\partial R = 0$ et donc par conséquent que $\partial P/\partial R = \partial P/\partial R|_{\bar{A}} < 0$.

² En fait, les auteurs interprètent la variation de A comme étant les activités de création de demande et ils considèrent que $\partial A/\partial R$ est positif, ce qui donne le même résultat. Cette uniformisation a été généralisée.

³ Anderson et al., p. 128

En fait, selon ce modèle de Pauly (1980), les médecins offrent le maximum de services et profitent de leur pouvoir de création de demande au maximum. Alors sous ces conditions, s'il y a une augmentation du nombre de médecins per capita - puisqu'il n'est plus possible pour ces derniers de se créer une nouvelle demande - ils devront subir une baisse de prix pour leurs services.

Ce résultat est tributaire de deux hypothèses posées par Pauly (1980). La première est la constance du coût d'opportunité du travail. Si tel n'est pas le cas, suite à une augmentation du nombre de médecins per capita la charge de travail du médecin peut diminuer - ce qui entraînerait une baisse du coût d'opportunité du travail - et le médecin pourrait se créer de la demande supplémentaire. Une autre manière (quoique moins élégante) de poser le problème serait d'imposer une contrainte supplémentaire au problème de maximisation qui fixerait un nombre d'heures de travail maximum. Lorsque cette contrainte est effective, la création de demande pourrait augmenter suite à une augmentation du nombre de médecins per capita.

La deuxième hypothèse importante de ce modèle est l'absence d'autres variables que le revenu et le loisir dans la fonction d'utilité du médecin. Si tel n'est pas le cas, (par exemple en incorporant l'éthique) il pourrait y avoir un coût en terme de bien-être à créer de la demande (ou à ne pas donner d'information). Les conclusions seraient différentes. C'est ce type d'analyse que nous abordons à la section suivante.

1.1.3. Le médecin avec éthique (Pauly 1980, Evans 1974, Sloan et Feldman 1978)

Pauly (1980) discute également des effets de l'éthique sur le niveau de services offerts et sur l'hypothèse de création de demande. L'équation suivante permet de caractériser un médecin maximisant le revenu, le loisir (puisque le coût d'opportunité du loisir est considéré comme constant, il n'est pas nécessaire de l'inclure en tant que tel dans la fonction d'utilité) et l'éthique (puisque $\partial u/\partial A > 0$):

$$\max u = u(Y(A), A)$$

où $\partial u/\partial Y > 0$, $\partial u/\partial A > 0$ et $\partial Y/\partial A < 0$

Les signes de ces dérivées partielles nous signifient que le médecin retire de la satisfaction à une augmentation de son revenu et il retire également de la satisfaction à donner de la bonne information au patient (et par conséquent à ne pas lui mentir) puisqu'il a de l'éthique. Mais, son revenu est relié négativement avec le niveau de bonne information qu'il donne au patient ce qui crée une relation d'arbitrage. Le médecin donnera de l'information au patient tant que cela va augmenter son utilité. Ce niveau d'information ne sera pas maximum à l'optimum puisqu'il affecte son revenu auquel il n'est pas indifférent. Ce comportement rationnel est caractérisé par cette condition du premier ordre de maximisation:

$$\frac{\partial u}{\partial A} = \frac{\partial u}{\partial Y} \cdot \frac{\partial Y}{\partial A} + \frac{\partial u}{\partial A} \Big|_{\bar{Y}} = 0$$

En substituant le résultat de l'équation (1.2), nous obtenons:

$$\frac{\partial u}{\partial A} = \frac{\partial u}{\partial Y} \left[P - \frac{\partial Q}{\partial A} - \frac{\partial CT}{\partial Q} \cdot \frac{\partial Q}{\partial A} \right] + \frac{\partial u}{\partial A} \Big|_{\bar{Y}} = 0$$

Après un réarrangement des termes et quelques manipulations élémentaires, nous obtenons l'équation (1.4)⁴:

$$(1.4) \quad \frac{\partial u / \partial A}{\partial u / \partial Y} = \frac{\partial Q}{\partial A} \left[P - \frac{\partial CT}{\partial Q} \right]$$

Puisque nous savons que le prix est supérieur au coût marginal et que le taux marginal de substitution entre le niveau d'information (ou de vérité) donné au patient et le revenu est positif, il faut que la dérivée $\partial Q / \partial A$ soit négative à l'optimum afin de respecter l'équation 1.4. Ce qui signifie donc que, comparativement à l'autre solution, le médecin va donner plus d'information au patient, va davantage lui dire la vérité dû au fait que mentir ou ne pas donner d'information engendre un coût en termes d'utilité.

Ce phénomène est transposé sur la figure 1.4. Lorsqu'un médecin donne le maximum d'information à ses patients, il fait face à la

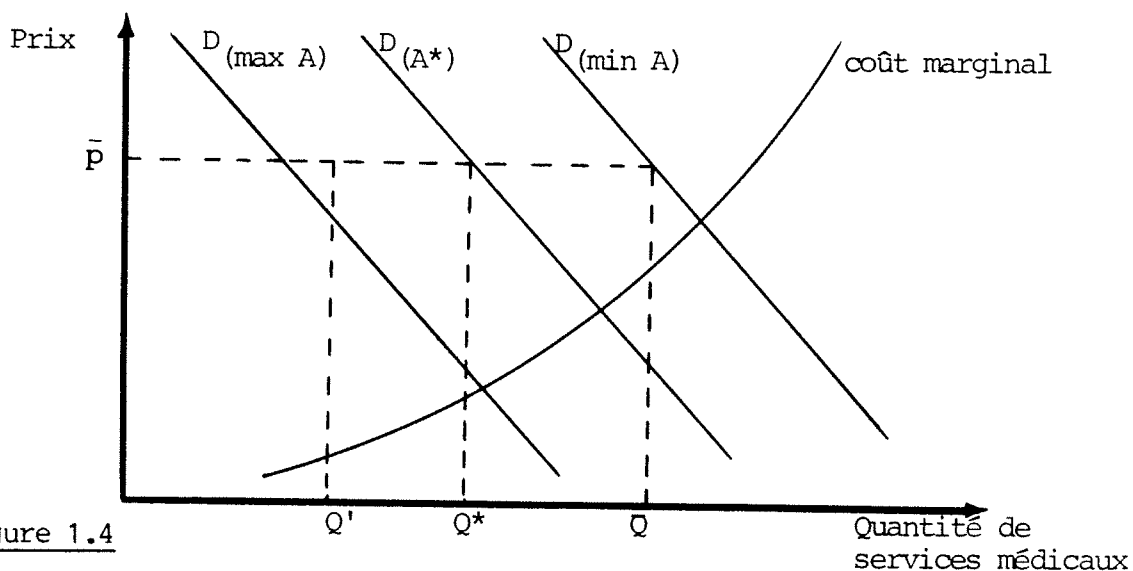


Figure 1.4

⁴ Pauly (1980) présente les conditions d'équilibre sans les fonder. Dionne et Contandriopoulos (1982) reformulent au complet le problème afin de fonder ces conditions. Cette partie du texte ainsi que la figure 1.4 et son interprétation sont tirées de l'article de Dionne et Contandriopoulos (1983).

demande $D(\max A)$. Il peut ne donner aucune information (et même mentir) ce qui lui permet de créer de la demande et faire face à la demande $D(\min A)$. Puisqu'il est coûteux pour le médecin de mentir mais qu'il aime voir son revenu augmenter, il va se créer une relation d'arbitrage et le médecin va choisir un A^* (A optimal) et faire face à la demande $D(A^*)$.

Supposons que le prix soit fixé à \bar{p} . Le médecin, s'il n'avait pas d'éthique produirait \bar{Q} mais puisqu'il a un coût à mentir, il produira Q^* . Si le nombre de médecins per capita augmente, la courbe de demande auquel fait face le médecin se déplace vers la gauche. Si le médecin garde le même A^* , il produira Q' . Le médecin va donc se créer de la nouvelle demande puisque:

"At this point (au point Q'), the physician's income is less so he may be willing to sacrifice some accuracy to recoup part of the loss income. (Ordinarily he would not wish reduce accuracy so much as to restore his original level of income.) Moreover, at Q' , there will also be a bigger gain from increasing Q by one unit, since $P - \partial CT/\partial Q$ is greater. If $\partial Q/\partial A$ is constant or increases as Q decreases, then the net return from changing A will also have increased"⁵.

Evans (1974) s'était également intéressé au problème de création de demande. Sloan et Fledman (1978) reprennent le modèle original de Evans (1974) et en font des extensions. Dans ce modèle, le médecin maximise une fonction d'utilité (u) qui est fonction du revenu

5 Pauly (1980), p. 53

($u_Y > 0$), de la charge de travail ($u_W < 0$) et de l'influence du médecin sur la demande du patient ($u_D < 0$) ce qui signifie implicitement que le médecin a de l'éthique. Il va utiliser son pouvoir de création de demande jusqu'à ce que l'utilité marginale de cette augmentation de revenu soit égale à la désutilité que procure l'augmentation de sa charge de travail et les activités de création de demande.

Nous avons donc le système d'équation suivant:

$$u = u(Y, W, D)$$

$$w = r \cdot F(P, D)$$

$$y = r \cdot F(P, D) \cdot P - C(w)$$

où r représente le ratio population/médecin⁶;

$C(.)$ la fonction de coût; et

P le prix du service.

Substituant les deuxième et troisième équations dans la première nous obtenons la fonction d'utilité à minimiser par rapport à P et à D . A partir des conditions de premier ordre, ($\partial u / \partial P = 0$ et $\partial u / \partial D = 0$) Sloan et Feldman (1978) utilisent la technique de dérivation totale pour obtenir les équations (1.5) et (1.6):

$$(1.5) \quad \frac{dR}{dr} = \frac{-u_{Pr} u_{PD} - u_{PD} u_{DR}}{\Delta}$$

$$(1.6) \quad \frac{dD}{dr} = \frac{-u_{PP} u_{Dr} - u_{Pr} u_{PD}}{\Delta}$$

⁶ Les auteurs utilisent R pour désigner le ratio pop./m.d. mais pour s'assurer une certaine uniformité, nous allons garder R pour désigner le nombre de médecins per capita ($r = 1/R$).

où Δ est le déterminant hessien qui est positif (condition de second ordre), u_{PP} et u_{DP} sont des dérivées secondes négatives. De plus, sous des restrictions plausibles au sujet des fonctions $F(\cdot)$ et $W(\cdot)$, nous pouvons considérer que u_{PP} et u_{Pr} sont positives. Par contre, le signe de u_{Dr} est indéterminé.

Considérons les deux cas possibles:

Premier cas:

$$\text{Si } u_{Pr} < 0 \text{ alors, } \frac{dP}{dr} \leq 0 \text{ et } \frac{dD}{dr} \leq 0$$

Deuxième cas:

$$\text{Si } u_{Dr} > 0 \text{ alors, } \frac{dP}{dr} > 0 \text{ et } \frac{dD}{dr} > 0$$

Reinhardt (1978) nous informe sur la plausibilité de rencontrer ces cas. Le second cas considéré serait improbable. Sous l'hypothèse de création de demande une augmentation du nombre de médecin per capita peut entraîner une hausse des honoraires mais ce n'est pas toujours le cas. Mais, s'il y a baisse de prix, nous pouvons considérer que celle-ci a été moins importante que ce qu'aurait prédit le modèle néoclassique qui postule l'indépendance des courbes d'offre et de demande.

1.1.4. Le producteur incertain (Darby et Karni 1973)

Jusqu'à maintenant, lorsque l'on considérait que le producteur n'avait pas d'éthique, la solution optimale était très simple: il profitait le plus possible de sa position privilégiée offrant par con-

séquent beaucoup de services (dont plusieurs inutiles). En fait, le client est plus ou moins bien informé et le producteur peut perdre des profits en voulant trop exagérer. Darby et Karni (1973) ont développé un modèle afin d'expliquer les motivations d'un producteur à offrir des services inutiles à un consommateur mal informé d'une part et afin de déterminer le montant optimal de fraude (si le consommateur avait été un spécialiste, il aurait refusé ces services) d'autre part. Ce modèle s'applique bien au marché des services médicaux dû au coût élevé d'acquisition d'information pour les consommateurs et dû à la nature du produit. De plus, il y a incertitude puisque le producteur ne sait pas si le patient va accepter ou non l'offre de services inutiles. Voici donc les principaux paramètres de ce modèle.

Soit un producteur maximisant la valeur présente de profit (noté π) provenant d'un consommateur:

$$(1.7) \quad \pi = [1 - F(Q)] [PQ - C(Q)] + [1 - G(Q)] \cdot V$$

où P est le prix du service;

Q la quantité de services offerte;

V est la valeur présente pour le producteur des profits futurs anticipés;

$C(Q)$ est la fonction de coût;

$F(Q)$ est l'estimation par le producteur de la fonction de probabilité cumulative que le client refuse le service. Par hypothèse $F'_Q \geq 0$;

$G(Q)$ est la probabilité que le consommateur ne revienne pas dans le futur, $G'_Q > 0$ par hypothèse.

Le niveau de service total prescrit est obtenu en maximisant sur Q :

$$(1.8) \quad \frac{\partial \pi}{\partial Q} = [1 - F(Q)] [P - C'_Q] - F'_Q [PQ - C(Q)] - G'_Q V = 0$$

d'où:

$$[1 - F(Q)] [P - C'_Q] = F'_Q [PQ - C(Q)] + G'_Q V$$

La condition d'équilibre de premier ordre égalise le revenu marginal d'un service supplémentaire frauduleux au coût marginal en terme de pertes de revenus présentes et futures.

Les auteurs s'intéressent au signe de $\partial Q / \partial V$ en utilisant la technique de différenciation totale:

$$\frac{\partial Q}{\partial V} = \frac{G'_Q}{\Delta}$$

où Δ est la condition de deuxième ordre (négatif).

Puisque G'_Q est positif, alors $\partial Q / \partial V$ est négatif. Ainsi, plus la valeur présente pour un producteur des profits futurs anticipés d'un consommateur est élevée, moins celui-ci a de chance (lire probabilité) d'être fraudé.

1.1.5. Comportement du consommateur (Darby et Karni 1973, Dionne 1979)

Le consommateur ne connaît pas beaucoup le service pour lequel il fait du furetage et il peut rencontrer des producteurs intéressés à exploiter cette situation d'asymétrie d'information en offrant des services inutiles. Son comportement de furetage est donc différent de ceux présentés couramment dans la littérature sur la recherche du prix et des qualités des biens. Pour expliquer le comportement du consommateur sur le type de marché qui nous intéresse, le modèle séquentiel de furetage tel que présenté par Darby et Karni (1973) et Dionne (1979) nous semble particulièrement adéquat.

L'événement s'étant produit (maladie ou bris automobile), le consommateur choisit son producteur au moyen de visites successives jusqu'à ce qu'il rencontre des conditions qu'il juge satisfaisantes. Le consommateur utilise cette fonction d'espérance d'utilité comme stratégie de furetage.

Avant de se présenter sur le marché, le consommateur estime le niveau de services requis et les coûts correspondants à partir de son information a priori. Son expérience passée sur le marché ainsi que les connaissances qu'il a acquises hors du marché (bien de réputation) vont également affecter son comportement. Soit:

$$(1.9) \quad Eu(K) = \int_0^R u [\bar{K} - \ell(Q)] H(\ell(Q), I) d\ell(Q)$$

où $Eu(K)$ est l'espérance d'utilité du consommateur;

\bar{K} est sa richesse initiale;

K est sa richesse finale;

$\ell(Q)$ est le coût du service s'il accepte;

$\ell(0)$ est le coût du diagnostic;

R est la limite supérieure de $\ell(Q)$ [$0 \leq R \leq \infty$];

H est une fonction de densité quelconque;

I est le niveau d'information a priori du consommateur.

Dans ce genre de modèle, il n'y a pas de courbe de demande individuelle et la courbe de demande de marché est l'agrégation des stratégies optimales des consommateurs.

Alors que la consommation de la plupart des biens n'utilise pas du temps de travail, il n'en est pas toujours ainsi pour la consommation des services médicaux où la réparation des automobiles. Il en est de même pour les activités de furetage.

Or, Dionne (1979) évalue les possibilités de fraude en fonction du coût réel du furetage. C'est donc le temps utilisé pour les activités de furetage qui est pertinent à son analyse. Concept qu'il introduit en redéfinissant K :

$$\bar{K} = wT$$

où, T est la quantité de temps de travail total d'un individu pour une période donnée.

w est le taux de salaire par unité de temps de travail.

Le consommateur se présente donc sur le marché et se fait offrir $\ell(\dot{Q})$. Sa décision suivra les critères suivants:

$$A) \quad u(\bar{K} - \ell(\dot{Q})) \geq Eu(K)$$

Cela coûte moins cher que ce que le consommateur avait anticipé. Il acceptera le service offert, il ne se sent pas fraudé et il reviendra dans le futur au besoin.

$$B) \quad Eu(K - \ell(0) - wT) \leq u(\bar{K} - \ell(\dot{Q})) < Eu(K)$$

Il se sent fraudé mais accepte quant même le présent service étant donné le coût fixe du diagnostic ($\ell(0)$) et le prix du temps du furetage wT (t représente le temps total utilisé pour un diagnostic). Il ne reviendra pas dans le futur.

$$C) \quad u(\bar{K} - \ell(\dot{Q})) < Eu(K - \ell(0) - wt)$$

L'individu se sent fraudé et refuse le service et ne reviendra pas. L'utilité certaine de cette offre est inférieure à l'espérance d'utilité d'aller voir un autre producteur (payant ainsi le prix du

temps du furetage) et de payer à celui-ci le coût fixe relié au diagnostic.

La représentation graphique de ce processus de décision pour un individu qui est neutre face au risque donne:

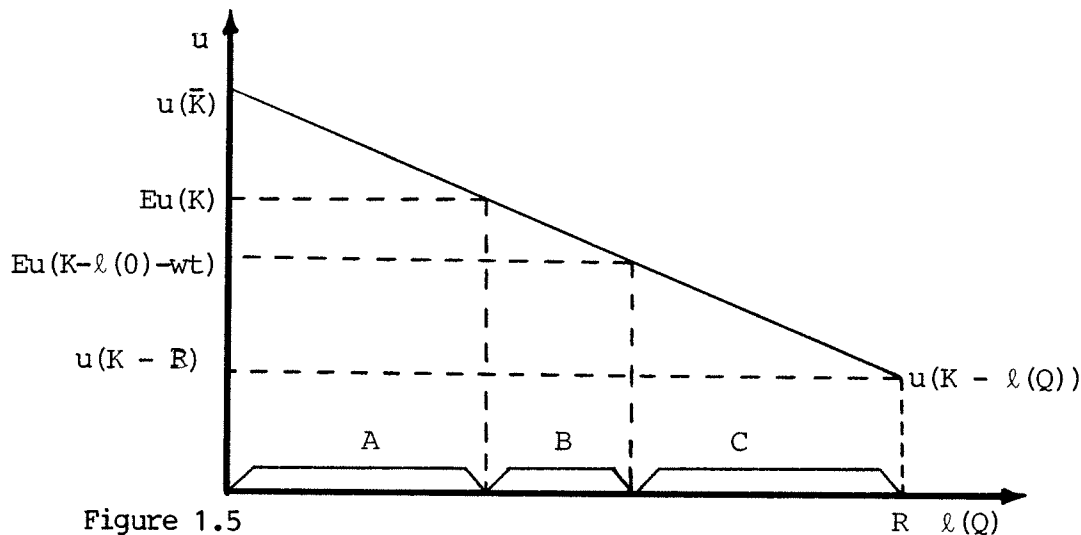


Figure 1.5

Dans le modèle de Darby et Karni (1973), le producteur estimait subjectivement les probabilités que le consommateur refuse le service ou ne revienne pas dans le futur. Ce faisant, il évalue les régions B et A respectivement.

De plus, ce modèle de comportement du consommateur nous permet de justifier les hypothèses sur les dérivées F'_Q et G'_Q . Nous pouvons voir que plus le producteur offre un niveau de service important, plus il a de chances de se retrouver dans les régions B et C. C'est pourquoi, nous pouvons considérer que F'_Q et G'_Q sont positives.

1.2. Analyse empirique

1.2.1. Difficultés théoriques de la vérification empirique

Avant d'aborder les principales études empiriques en tant que telles, nous allons nous intéresser aux principales difficultés d'ordre théorique rencontrées dans les études empiriques ayant pour objectif de vérifier si les producteurs créent leur propre demande. Les auteurs s'intéressent donc généralement à l'influence du nombre de médecins per capita (ressources) sur le prix des traitements. Ce type d'étude est souvent effectué en coupe instantanée dû en particulier à la disponibilité des données. Il y a néanmoins un certain nombre de difficultés qui se posent.

1.2.1.1. Influence réciproque entre l'offre et la demande

La dépendance entre l'offre et la demande peut également être interprétée comme un réponse de l'offre à la demande (Contandriopoulos, Dionne et Tessier 1982). Ainsi, lors d'études économétriques, il faut tenir compte des caractéristiques des patients puisqu'elles affectent le risque de maladie et par conséquent la demande (Wennberg, Barnes et Zubkoff 1982). Il est intéressant de noter que ce faisant, nous allons tenir compte indirectement de la qualité des services puisque les caractéristiques des patients influencent également le type de service demandé. Newhouse (1978) a rappelé que l'analyse néoclassique postule la même qualité de services d'une région géographique à une autre. Il faut tenir compte de ce postulat.

1.2.1.2. Elasticités de demande

Si les élasticités de demande varient d'une région géographique à une autre, il est possible de constater une relation positive entre le nombre d'offreurs per capita et le prix d'équilibre sans qu'il y ait création de demande. Ce résultat est représenté sur la figure 1.6 où l'on a superposé deux marchés spatialement différents ayant des élasticités de demande différentes. Malgré le fait que l'offre (S_1 et S_2) per capita soit plus forte dans le marché 2, le prix y est plus élevé et il n'y a pas de création de demande (voir Pauly 1980).

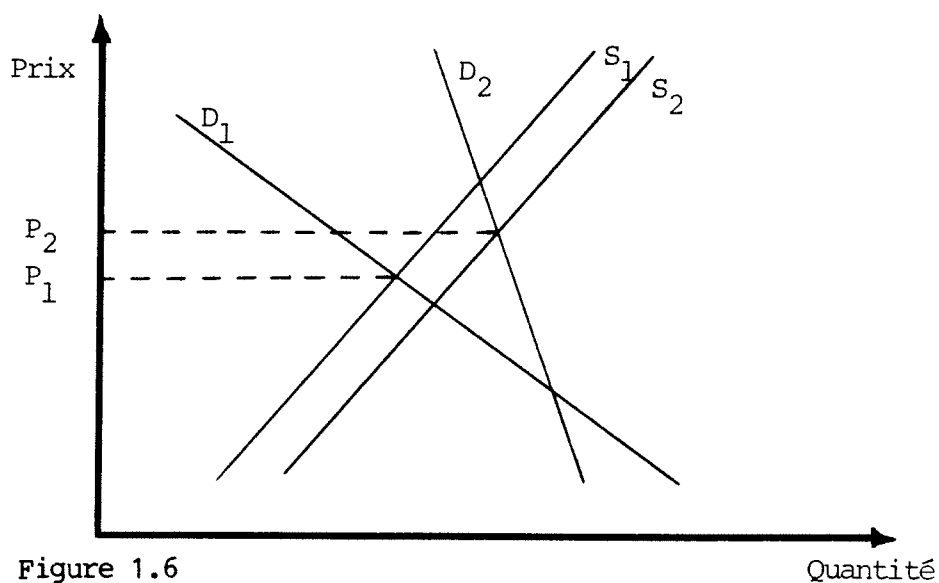


Figure 1.6

Par contre, selon Satterthwaite (1979), si l'on est en présence d'un bien de réputation, l'élasticité de la demande va être affectée par le nombre d'offreurs. Appliquant ce principe au marché qui nous intéresse, il en résulte que la demande serait plus inélastique dans les régions où le nombre de médecins per capita est élevé dû au fait que dans celles-ci, la chance de bien connaître la

réputation d'un médecin est plus faible.⁷

1.2.1.3. Les facteurs de production

Les facteurs de production que doit utiliser le producteur peuvent induire l'économètre en erreur de deux façons.

D'une part, pour qu'il y ait création de demande (du moins en courte période), il faut que les facteurs de production ne soient pas contraignants. Cette remarque de Auster et Oaxaca (1981) est appliquée par Contandriopoulos et al. (1983) dans le cas où il y a une contrainte de lits d'hôpitaux effective. Les auteurs concluent ainsi sur le sujet:

"Cette remarque s'applique également pour les modèles de Sloan et Feldman (1978) et de Reinhardt (1978). Si les hôpitaux sont utilisés à pleine capacité, un accroissement du nombre de chirurgiens per capita ne devrait pas augmenter le nombre d'interventions per capita ceteris paribus. Graphiquement, l'offre de services chirurgicaux ne se déplacerait pas vers la droite suite à une augmentation du nombre de chirurgiens per capita"⁸.

Enfin, notons que cette remarque est valable surtout en courte période. En longue période, le nombre de lits d'hôpitaux peut augmenter.

D'autre part, Auster et Oaxaca (1981) font remarquer que l'économètre doit s'assurer qu'il n'y a pas de différences importantes dans les coûts des facteurs de production entre les régions d'analyse.

⁷ C'est ce qui permet à Anderson (1981) (voir section 1.1.2) d'arguer que lorsque le nombre de médecins par capita augmente, il devient plus facile de convaincre les clients à surconsommer ce qui entraîne une augmentation de la création de demande.

⁸ Contandriopoulos, Dionne et Tessier (1983) p. 738

En effet, une différence de prix pourrait s'expliquer par une différence dans les coûts de production et non par la création de demande. Nous croyons que cette remarque est très valable mais il ne faut pas oublier que ce qui est pertinent, ce sont les coûts de production que doivent payer les producteurs. Ce genre de problème se posera principalement lorsqu'il y a des différences de régime de santé entre les régions. Le prix des facteurs de production peut être similaire mais ne peut pas être encouru par les mêmes agents.

1.2.1.4. Fonction de production aléatoire

Selon Wenner, Barnes et Zubkoff (1982), la théorie de la création de demande sous-estime l'incertitude des médecins. En fait, la fonction de production est aléatoire et le médecin pose le diagnostic et choisit le traitement qui en découle en situation incertaine.

Nous croyons que suite à cette critique, trois remarques s'imposent:

- pour que l'incertitude cause des problèmes, il faut qu'il y ait des différences entre le vécu de l'incertitude dans une région par rapport à une autre. Il est tout de même possible que le vécu de l'incertitude soit ressenti différemment selon la région et diminue lorsque le nombre de médecin per capita augmente;
- lorsqu'il y a un consensus dans la profession relatif à un certain traitement par exemple, nous pouvons considérer que le médecin n'a pas à prendre une décision en incertitude. Cela est vrai même si plus tard ce consensus ne tient plus;

- il est toujours plus intéressant de discuter des cas rares et problématiques alors qu'une forte proportion de cas sont traités de façon relativement certaine et routinière par le médecin. Malgré tout, il demeure toujours une certaine forme d'incertitude même dans les diagnostics les plus simples.

Malgré le fait que la notion de services frauduleux - telle que définie dans ce travail - soit conceptuellement relativement simple, il n'en demeure pas moins que les difficultés pratiques que pose la définition de la fraude sont importantes dans un monde où la fonction de production du producteur peut être aléatoire (Pauly (1979)).

1.2.1.5. Mobilité des consommateurs

Contandriopoulos et al. (1983) modifient l'analyse en tenant compte de la mobilité des patients. Dans un tel contexte, il devient nécessaire de distinguer les interventions subies par les résidents d'un comté des interventions effectuées dans le comté. De plus, l'introduction de la mobilité des patients ne nous permet plus de conclure que l'hypothèse de création de demande est vérifiée lorsque l'on observe empiriquement un lien positif entre les prix monétaires et l'offre de services.

Ainsi, dans un contexte de mobilité des patients et de création de demande, on peut s'attendre à une relation positive entre les ressources médicales et les interventions effectuées. Le lien avec les interventions subies est ambigu et il en est de même pour les prix.

1.2.2. Application empirique (Marché des soins chirurgicaux)

1.2.2.1. Création de demande par l'offreur

Lewis (1969) fut l'un des premiers à se préoccuper de la vérification empirique de la création de demande. En comparant les onze régions administratives du Kansas à l'aide des données de 1965 et ce pour six types d'opérations électives, il établit un lien statistique entre la quantité d'interventions chirurgicales et le nombre de médecins qui opèrent par capita. Selon Lewis (1969) le nombre d'opérations électives per capita effectuées dans une région donnée est fonction du nombre de lits d'hôpitaux (per capita) dans la région, du nombre de chirurgiens et de médecins qui opèrent (per capita) et du nombre de chirurgiens certifiés (per capita). L'auteur obtient des résultats significatifs pour quatre types d'opérations mais ces résultats peuvent aussi bien être interprétés comme un déplacement des individus vers les ressources disponibles que par l'influence de l'offre sur la demande. D'ailleurs, pour vérifier cette dernière hypothèse, le type d'études utilisées par Lewis (1969) ne peut jamais être concluant puisqu'une relation positive entre le nombre de producteurs per capita et le nombre d'opérations per capita est également compatible avec la théorie néoclassique (voir section 1.1.1, figures 1.1 et 1.2).

Néanmoins, sous l'hypothèse de création de demande, les quantités produites augmentent de façon plus importante et au détriment des opérations électives. Or, c'est dans le cas des appendicectomies que

l'auteur obtient les résultats les plus probants. C'est pour ce type d'opération qu'il y a le moins de désagrément et de risque pour le patient (surtout si l'opération est produite conjointement). Les coûts en terme d'éthique sont peut-être moins importants pour cette opération. Malgré ce genre de considérations, les études de ce genre ne sont guère convaincantes.

Par contre, l'étude que présente Fuchs (1978) est beaucoup mieux structurée. Afin d'éliminer le problème d'influence réciproque entre l'offre et la demande, Fuchs (1978) a utilisé la méthode des équations simultanées. Ainsi, l'auteur identifie les variables théoriques qui peuvent influencer l'offre mais non la demande et utilise l'offre prédite dans l'estimation de la demande. Fuchs (1978) tient compte de la mobilité des patients entre les zones urbaines et non-urbaines (dans une même division) et il suppose qu'il n'y a pas de mobilité des patients inter-divisions. Puisque la variable intra-division n'a pas donné de résultats significatifs, il est probable que cette hypothèse (non mobilité entre les divisions) n'est pas trop forte.

L'équation d'offre est la suivante:

$$S^* = \beta_0 + \beta_1 \text{NONMET} + \beta_2 \text{NRMET} + \beta_3 \text{HOTEL} + \beta_4 \% \text{WYTE} + \beta_5 Q^{**} + e$$

où S^* est le nombre de chirurgiens pour 100 000 habitants;

NONMET est une variable dichotomique indiquant qu'il s'agit d'une zone rurale;

NRMET est le % de personnes dans une zone rurale demeurant dans une ville de 50 000 habitants et plus. Pour les zones urbaines, NRMET est égal à zéro;

HOTEL est le revenu par capita en hôtel et motel. Cette variable se veut une "proxy" de l'attrait d'une région;

%WYTE est le pourcentage de blanc;

Q** est le nombre d'opérations par 10 000 habitants.

Le principal résultat est:

	NONMET	NRMET	HOTEL	%WYTE	Q*
Coefficient	-28	18	.17	.27	-.002
Test "t"	(4.8)	(1.9)	(3.6)	(2.2)	(0.5)

Seule la variable Q* n'est pas significativement différente de zéro (test à 95%). Ainsi, la demande potentielle d'une région n'a pas d'effet sur la localisation des chirurgiens.

L'équation pour la demande, tant qu'à elle, est de cette forme:

$$Q^{**} = \beta_0 + \beta_1 \hat{S}^* + \beta_2 \hat{METS} + \beta_3 INC + e$$

où Q** est le nombre d'opérations par 100 000 habitants;

\hat{S}^* est le nombre prédit de chirurgiens par 100 000 habitants;

\hat{METS} est le nombre prédit de chirurgiens dans une zone urbaine divisé par la population totale de la division correspondante. Cette variable est égale à zéro pour les zones urbaines;

INC est le revenu réel per capita.

Voici le principal résultat:

	S*	\hat{METS}	INC
Coefficient	60	30	230
Test "t"	(3.1)	(2.0)	(0.6)

Puisque S* est significativement différent de zéro (test à 95%), l'hypothèse selon laquelle un changement exogène dans l'offre des chirurgiens affecte la demande pour les opérations chirurgicales (ceteris paribus) n'est pas rejetée.

Fuchs (1978) a également vérifié que l'influence de l'offre sur la demande est élevée pour les classes d'éducation plus faibles et pour les opérations les plus électives (moins nécessaires et moins urgentes). De plus, grâce à d'autres régressions, l'auteur obtient un effet positif de l'offre sur le prix des opérations.

La contribution de Fuchs (1978) est importante. Il semble confirmer que le comportement des agents économiques dans ce marché ne cadre pas avec le modèle néoclassique. Il renforce la plausibilité de l'hypothèse de création de demande par l'offreur. Il est malheureux que l'auteur n'ait pas vérifié si les élasticités de demande ne sont pas trop différentes entre les comtés puisque Auster et Oaxaca (1981) ont déjà vérifié que l'étude de Fuchs (1978) a passé la difficulté théorique que peut présenter les facteurs de production.

Pauly (1980) a utilisé les mêmes données que Fuch (1978) pour ses estimations mais il les utilise différemment. Au lieu de les agréger, il utilise les données individuelles ce qui lui permet d'aller

chercher beaucoup de variables de caractéristiques d'une part et d'éliminer également le problème d'influence réciproque entre l'offre et la demande d'autre part. En effet, l'auteur se sert de données individuelles pour l'utilisation des services et de données par zone pour mesurer les ressources. Or, il est raisonnable de croire que le nombre de chirurgiens per capita ne répond pas à des demandes individuelles. Pauly (1980) ne tient pas compte du problème que peut poser la mobilité des patients.

L'auteur n'obtient pas de résultats significatifs pour expliquer les taux de chirurgiens à l'aide des variables caractérisant les disponibilités des ressources (quels que soient les niveaux d'éducation des patients).

Il explique ainsi ces résultats: puisqu'il se sert de données individuelles, il peut explicitement tenir compte de l'état des status de santé d'une part et d'autre part, il a éliminé les patients ayant des revenus inférieurs au seuil de pauvreté (parce que leurs dépenses sont couvertes par Medicaid). Ce faisant, il éliminait les individus peu instruits. Or, Fuchs (1978) obtenait les résultats les plus probants pour ces personnes.

De plus, selon Dionne et Contandriopoulos (1982), en utilisant toutes les interventions Pauly (1980) a pu noyer l'effet de création de demande qui se manifeste surtout pour certaines interventions électives à risque peu élevé pour le patient de préférence peu instruit.

Nous croyons que Pauly (1980) en ne tenant pas compte de ce résultat ait jeté le bébé avec l'eau du bain.

CHAPITRE II - ANALYSE DE L'HYPOTHESE D'OFFRE EXCEDENTAIRE

Dans ce chapitre, nous nous intéressons de façon spécifique à la relation micro-économique entre le producteur potentiellement frauduleux et le consommateur mal informé. Après avoir repris brièvement le modèle à un bien de Darby et Karni (1973) afin de discuter de l'optimum de Pareto dans un tel contexte et du niveau de fraude optimal en considérant l'achat d'un expert, nous étudions l'importance de l'hypothèse d'offre excédentaire pour justifier la fraude. L'offre excédentaire est définie comme étant une situation dans laquelle le prix du marché est tel qu'à ce prix, le producteur individuel désirerait offrir une plus grande quantité de services non-frauduleux. En intégrant une contrainte de capacité dans le modèle de base, nous démontrons graphiquement que l'hypothèse d'offre excédentaire est une condition nécessaire à l'offre de services frauduleux dans le modèle à un bien de Darby et Karni (1973) lorsqu'elle est définie de façon ex ante.

Par la suite, nous reprenons ce modèle simple auquel nous spécifions une fonction de probabilité particulière ainsi qu'une fonction de coût spécifique. Après avoir dérivé la fonction d'offre de ce modèle, nous introduisons les productions jointes grâce à un modèle séquentiel d'offre de services. La comparaison entre les deux modèles prototypes nous permet d'analyser l'effet des productions jointes en économies d'envergure sur l'offre de services frauduleux. Nous appliquons également les prototypes à un marché où la facture présentée au client est l'addition de deux composantes: le prix de la main d'oeuvre et le prix des pièces. Ce type de pratique est généralement associé au

marché de la réparation des automobiles. La généralisation des résultats nous amène à discuter de l'importance de l'hypothèse d'offre excédentaire dans un contexte de production jointe.

Un retour en arrière nous amène à discuter de la possibilité des productions jointes dans un monde parfaitement concurrentiel où il n'existe aucun problème d'information. Nous nous situons dans un contexte d'économies d'envergure. Par la suite nous supposons qu'il y a asymétrie d'information sur l'utilité et la qualité du service. Dans ce nouveau contexte de double asymétries d'information, le producteur individuel a une certaine influence sur les prix. Nous avons donc un producteur face à un prix endogène.

Toutes choses étant égales par ailleurs, une diminution du temps requis pour la consommation du service sera perçue comme une baisse de prix. Dans la section qui suit, nous intégrons cette caractéristique au modèle d'offre de services frauduleux. Cette extension prend une signification toute particulière dans un contexte de production jointe où le coefficient d'économies d'envergure est généralement associé à une baisse du temps de production et de consommation par extension.

Enfin, ce chapitre se termine par l'introduction de l'assurance dans notre modèle d'offre de services frauduleux. En effet, dû au caractère aléatoire de l'événement, la consommation du service est souvent effectuée sous l'assurance. Avant d'aborder l'étude de l'effet de l'assurance sur la fraude en production jointe, nous établissons le lien entre l'offre de services frauduleux et un problème d'allocation des ressources imputable à l'assurance soit le risque moral.

2.1. Asymétrie d'information et optimalité

2.1.1. Atteinte à l'optimum parétien

Darby et Karni (1973) nous présentent un modèle d'offre de services inutiles à un consommateur mal informé. Soit un producteur maximisant la valeur présente de profit (noté π) provenant d'un consommateur.

$$\pi = [1 - F(Q)] [PQ - C(Q)] + [1 - G(Q)] \cdot V$$

où P est le prix du service;

Q est la quantité de services offerts;

V est la valeur présente pour le producteur des profits futurs anticipés;

C(Q) est la fonction de coût;

F(Q) est l'estimation par le producteur de la fonction de probabilité cumulative que le client refuse le service. Par hypothèse $F'_Q \geq 0$;

G(Q) est la probabilité que le consommateur ne revienne pas dans le futur, $G'_Q \geq 0$ par hypothèse.

La condition de premier ordre du problème tel que posé par Darby et Karni (1973) nous garantit l'égalité entre le revenu marginal d'un service supplémentaire frauduleux et le coût marginal en termes de pertes de revenus présentes et futures. Ainsi, le producteur rationnel offrira des services jusqu'à ce qu'il gagne à offrir un service supplé-

mentaire (profit monétaire net) soit égal à ce qu'il perd en terme de ventes présentes et futures. Réécrivons cette condition d'optimalité (revue de littérature; équation (1.8)):

$$(2.1) \quad [1 - F(Q)] [P - C'_Q] = F'_Q [PQ - C(Q)] + G'_Q V$$

De cette équation, nous pouvons en déduire l'équation (2.2):

$$(2.2) \quad P = C'_Q + \frac{F'_Q [PQ - C(Q)] + G'_Q V}{[1 - F(Q)]}$$

En situation concurrentielle, il y a atomisticité des agents et le producteur individuel ne tient pas compte de la perte future de profits provenant d'un consommateur insatisfait ($V = 0$).

De plus, s'il y a parfaite information, la fonction $F(Q)$ et sa dérivée F'_Q seront nulles. Ainsi, sous ces conditions, le producteur offrira ses services jusqu'à ce que le prix soit égal au coût marginal de production ($P = C'_Q$). Il s'agit là d'une condition d'optimalité de Pareto.

Par contre, s'il y a asymétrie d'information, l'estimation par le producteur de la fonction de probabilité que le client refuse le service n'est pas égale à zéro [$0 < F(Q) < 1$]. De plus, en vertu de notre revue de la littérature étudiant le comportement du consommateur sur un tel marché, nous pouvons faire l'hypothèse qu'il y a augmentation de la probabilité de refus s'il y a plus de services offerts ($F'_Q > 0$). Il en sera de même pour la fonction $G(Q)$. La valeur des profits futurs (noté V) sera positive s'il n'y a pas parfaite concurrence.

Dû à ces considérations, l'équation (2.2) nous assure qu'à l'optimum, le prix sera supérieur au coût marginal de production (noté C'_Q). L'une des conditions fondamentales de l'optimum de Pareto n'est plus respectée. Notons que ce résultat est compatible avec l'analyse de Pauly (1980) telle que rapportée dans notre revue de la littérature (1.1.2.).

En situation concurrentielle $V = 0$ et s'il y a asymétrie d'information, l'équation (2.1) peut s'écrire de la forme suivante¹:

$$\frac{P - C'_Q}{PQ - C(Q)} = \frac{F'_Q}{1 - F(Q)}$$

Ainsi, il y aura maximisation des profits lorsque la part des profits totaux que le producteur retire de la dernière unité vendue est égale à la contribution de cette production marginale en termes d'augmentation de la probabilité de refus de tous les services sur la probabilité d'acceptation.

2.1.2. Niveau de fraude optimal: achat d'un expert

Nous considérons qu'il y a fraude et par extension création de demande lorsque le service aurait été refusé si le consommateur avait été un expert. Conséquemment à cette définition, Darby et Karni (1973) considèrent qu'il y aura offre de services frauduleux si l'équation (2.3) est respectée:

$$(2.3) \quad [1 - F(\tilde{Q})] [P - C'_{\tilde{Q}}] > F'_{\tilde{Q}} [P\tilde{Q} - C(\tilde{Q})] + G'_{\tilde{Q}} \cdot V$$

¹ Ce résultat est tiré de l'article de Darby et Karni (1973).

où \tilde{Q} est le montant acheté si le consommateur avait été un expert. Pour qu'il y ait offre de services frauduleux, il faut qu'à la marge, il soit rentable d'offrir de tels services. Jusqu'à \tilde{Q} , les services offerts ne sont pas frauduleux. Il faut donc qu'à \tilde{Q} , le revenu marginal d'offrir des services frauduleux soit supérieur au coût marginal en termes de pertes de profits présentes et futures.

S'il y a égalité entre le revenu marginal et le coût marginal de production à \tilde{Q} , il n'y aura pas d'offre de services frauduleux et le problème d'asymétrie d'information n'aura pas de répercussion sur l'allocation des ressources. Enfin, il est possible que le producteur ne se rende pas à \tilde{Q} dans l'offre de services puisqu'à \tilde{Q} , le coût marginal serait supérieur au revenu marginal. Il y aurait sous-consommation des services due au problème d'asymétrie d'information. Cette situation qui découlerait d'une méfiance extrême des consommateurs est peu probable quoique théoriquement possible.

Le phénomène est représenté graphiquement par la figure 2.1. Le coût marginal de la fraude est croissant en Q puisque plus l'offre de services frauduleux est élevée, plus les pertes de profits présentes (dûes aux refus) et futures sont élevées. Le revenu marginal d'offrir de la fraude, quant à lui, est décroissant et ce pour deux raisons. D'une part, puisque le producteur est passif face aux prix et que le coût marginal est croissant, la différence entre le prix et le coût marginal est décroissante (respect de la condition de second ordre). D'autre part, la probabilité de refus augmente lorsque le niveau de services augmente. L'inverse se produit pour ce qui est de la

probabilité d'acceptation².

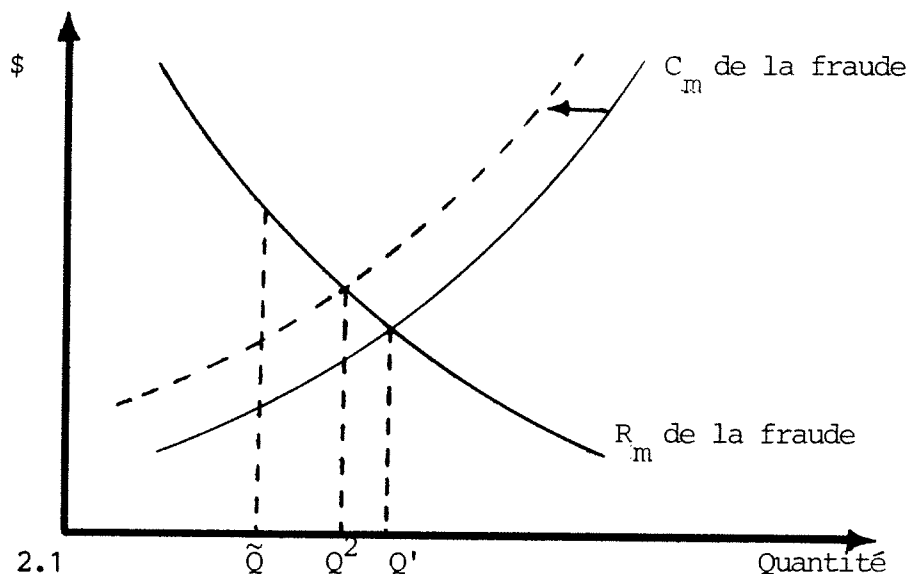


Figure 2.1

A \tilde{Q} , le revenu marginal (R_m) d'offrir de la fraude est supérieure au coût marginal (C_m). Le niveau de service offert sera Q' dont $Q' - \tilde{Q}$ est frauduleux.

Ainsi, si la valeur des profits futurs anticipés d'un consommateur augmente, le coût marginal de la fraude augmente également. Les consommateurs pour qui V est élevé ont donc moins de chance d'être fraudé ($Q^2 - \tilde{Q} < Q' - \tilde{Q}$). A titre d'exemple, l'automobiliste en tourisme a plus de chances (lire probabilités) de se voir offrir des services frauduleux lors d'une panne que le client régulier du même garagiste. Cette analyse est compatible avec les résultats obtenus par Darby et Karni (1973) utilisant la différenciation totale (1.1.4).

$$2 \quad R_m \text{ de la fraude} = (1 - F(Q))(P - C'_Q)$$

$$\frac{\partial R_m}{\partial Q} = -F'_Q(P - C'_Q) + (1 - F(Q))(-C''_Q) \leq 0$$

$$C_m \text{ de la fraude} = F'_Q(PQ - C(Q)) + G'_Q V$$

$$\frac{\partial C_m}{\partial Q} = F''_Q(PQ - C(Q)) + F'_Q(P - C'_Q) + G''_Q V \geq 0$$

De plus, si le producteur ressent un coût en terme d'éthique à offrir des services frauduleux, l'effet sera le même que précédemment. Le niveau optimal de fraude va diminuer. Cette interprétation est également compatible avec les modèles présentés dans la revue de la littérature.

2.2. Hypothèse d'offre excédentaire

Nous savons qu'à l'optimum, le prix sera supérieur au coût marginal de produire le service (noté C'_Q ce qui est différent du C_m de la fraude). Sous les hypothèses de la concurrence parfaite, il y a atomisticité des agents et le producteur individuel ne tient pas compte des pertes de profits futures provenant d'un consommateur insatisfait (G'_Q et V sont alors égaux à zéro). Dans ce cas, la différence entre le prix et le coût marginal de production sera de: $\frac{F'_Q [PQ - C(Q)]}{[1 - F(Q)]}$. Ce terme (tiré de l'équation 2.2) représente le coût marginal supplémentaire d'un service dû au problème d'asymétrie d'information. Il s'agit de la contribution de la dernière unité offerte aux pertes de profits attribuables au fait que le client peut détecter l'activité frauduleuse et se retirer. Il y a un risque associé à l'offre d'un service frauduleux supplémentaire.

Nous pouvons représenter graphiquement cette condition d'équilibre. Sur la figure 2.2, à l'optimum, la quantité offerte est inférieure à Q'' . Le point Q' respecte cette condition. Ce résultat

justifie l'obligation de poser l'hypothèse de surplus de capacité de l'offre afin de retrouver de la fraude. Rappelons que la notion d'offre excédentaire est définie ici comme étant une situation dans laquelle le prix du marché est tel qu'à ce prix, un producteur individuel désirerait offrir une plus grande quantité de services non frauduleux.

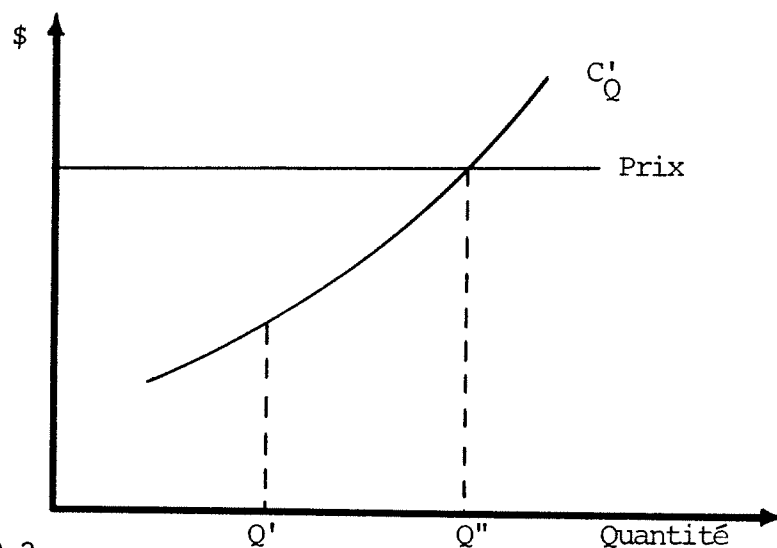


Figure 2.2

S'il n'y avait pas d'excès de capacité de l'offre, le producteur n'aurait alors aucun avantage à offrir un service supplémentaire frauduleux et il produirait la quantité Q'' . Le producteur n'a pas intérêt à tronquer un service non frauduleux par un service frauduleux puisqu'il y a un coût supplémentaire en termes de pertes de profits présentes et futures relié à la notion de fraude. Pour chaque service frauduleux vendu, il devrait refuser d'offrir un service non-frauduleux à un autre client. Ceci est d'autant vrai si le producteur subit un coup supplémentaire à offrir des services inutiles en termes de perte d'utilité due à l'éthique.

Un premier réflexe pour expliquer que les prix demeurent élevés serait de croire que l'on se trouve dans un monde de rationnement quantitatif. Dans un tel monde de déséquilibre avec rationnement, les prix sont rigides en courte période ne réagissant pas rapidement et parfaitement aux excès de l'offre ou de la demande. Les ajustements étant d'ordre quantitatif, une modification dans la relation demande-offre aura un impact sur les carnets de rendez-vous, les files d'attente, les heures de travail etc. Néanmoins, l'école du déséquilibre admet que, en longue période, il y aura des pressions sur les prix. En effet, selon Malinvaud (1980):

"A long terme, les biens, les services ou les types de travail qui sont de plus en plus demandés, mais dont l'offre ne peut suivre le rythme d'évolution de la demande à coûts réels constants, tendent sans aucun doute à voir leurs prix augmenter par rapport à ceux qui sont placés dans la situation inverse; mais cette augmentation se produit essentiellement à cause des différences dans l'importance relative des hausses de prix lorsque celles-ci ont lieu, ce qui habituellement ne se manifeste pas à intervalles très rapprochés."³

Mais, il n'en demeure pas moins que selon Malinvaud (1980), "les ajustements par les quantités sont beaucoup plus apparents et déterminants dans la courte période que les ajustements par les prix"⁴. Il s'agirait là non seulement d'un fait institutionnel mais également d'un comportement rationnel justifié par l'incertitude, les coûts de transaction et d'information etc. En effet, lorsqu'il y a un écart entre l'offre et la demande sur un marché, il existe des possibilités d'échanges mutuellement avantageux. Mais, "si ces possibilités

³ Malinvaud (1980), p. 45.

⁴ Malinvaud (1980), p. 46.

ne sont pas saisies c'est justement en raison de coûts qui s'y opposent: coût de recherche, coût de transaction, coût de rupture de certains contrats explicites ou implicites, coût des changements les plus divers."⁵

Ainsi, pour pallier ces coûts, les agents économiques voient des avantages mutuels à maintenir entre eux des liens durables, plus ou moins explicitement codifiés ce qui a pour effet de créer l'inflexibilité des prix et de justifier des prix élevés. Dans un tel contexte, l'hypothèse de concurrence parfaite n'est pas respectée puisque la demande à laquelle fait face chaque vendeur n'est pas parfaitement élastique.

Par contre, on peut réagir et dire que le marché fonctionne comme dans un modèle avec des prix flexibles mais parce qu'il y a asymétrie d'information et par extension possibilités de fraude, les prix demeurent élevés. Dans un tel contexte, l'inflexibilité des prix à la baisse et le rationnement qui en découle est la résultante du problème d'asymétrie d'information qui confère au producteur une capacité à offrir des services inutiles. Le producteur va préférer offrir des services frauduleux plutôt que de diminuer les prix de façon trop importante dans la mesure où le problème d'information lui permet d'influencer la demande. La capacité à frauder des producteurs rationnés empêche le marché d'être parfaitement efficace.

Il n'en demeure pas moins qu'il y a un coût supplémentaire associé à la notion de fraude (pertes de profits dues à l'augmentation

⁵ Malinvaud (1980), p. 17.

de la probabilité de refus, coût en terme d'éthique etc.). D'où l'importance de l'hypothèse d'offre excédentaire dans la justification micro-économique de la fraude.

Mais, la définition de l'excès de capacité est telle qu'il devient problématique de vérifier empiriquement s'il y a excès de capacité ou non. C'est ce qui nous amène à formuler la proposition suivante:

Proposition 1: La simple observation que les carnets de rendez-vous ne sont pas vides ne signifie pas nécessairement qu'il n'y a pas d'excès de capacité. Il est possible qu'il y ait eu de la fraude.

Afin de démontrer cette proposition, il est nécessaire de modifier quelque peu nos outils d'analyse. Nous allons donc considérer que le producteur est contraint et ne peut pas offrir plus que Q_{pc} (pleine capacité). A Q_{pc} le revenu marginal d'offrir des services frauduleux est ramené à zéro. Pour offrir un service supplémentaire, il doit sacrifier un autre service à un consommateur. Nous avons donc les trois possibilités suivantes:

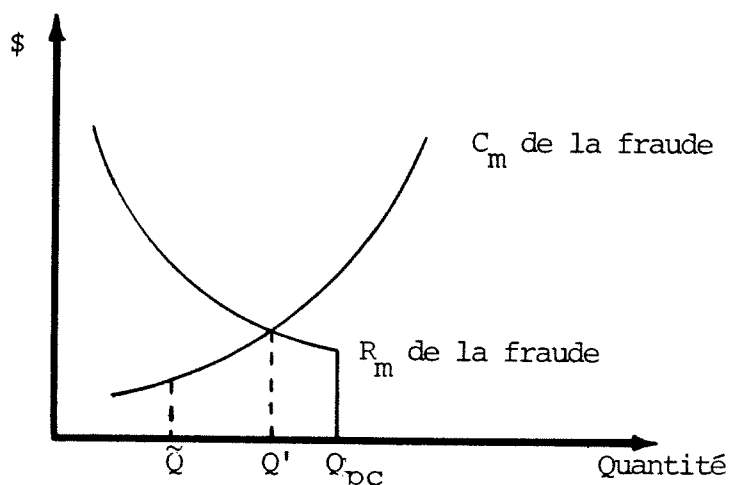


Figure 2.3

A \tilde{Q} il y a offre excédentaire et la fraude est rentable à la marge jusqu'à Q' . Il y a $Q' - \tilde{Q}$ d'offre de services frauduleux et surplus de capacité sur le marché.

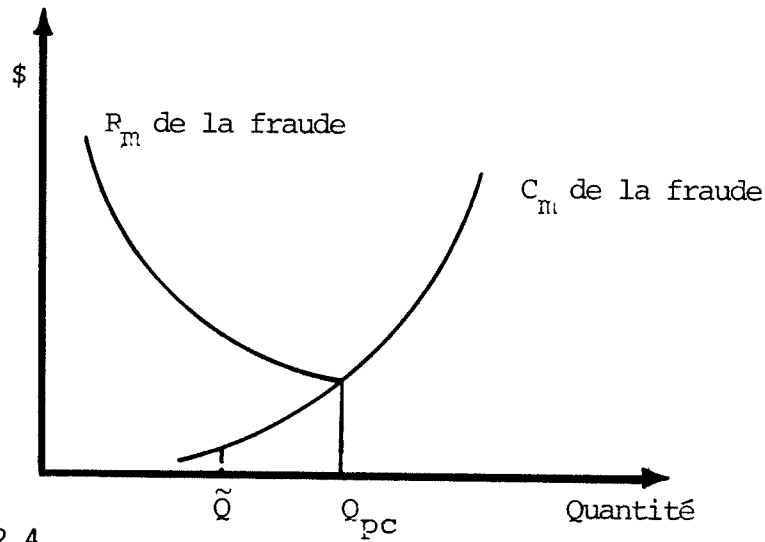


Figure 2.4

Il y a offre de services frauduleux de \tilde{Q} à Q_{pc} . Les carnets de rendez-vous sont complets mais le marché est saturé. Il y a offre excédentaire (référence "ex ante" donc à \tilde{Q}).

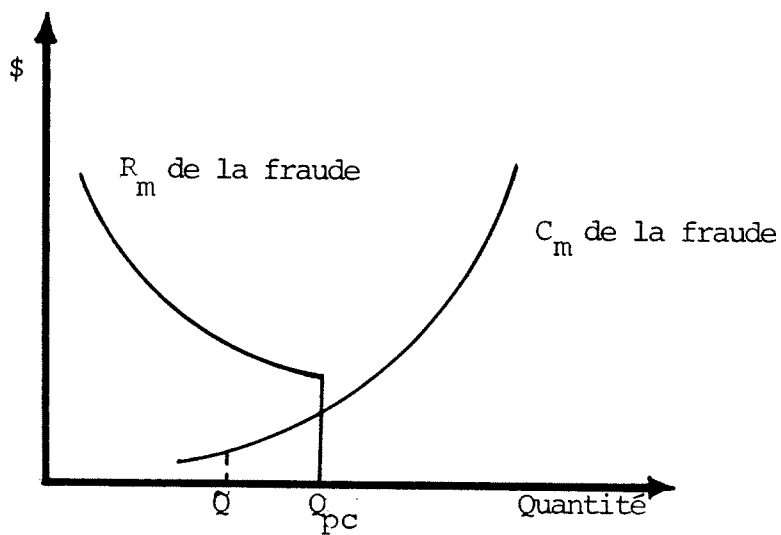


Figure 2.5

Il y a offre de services frauduleux de \tilde{Q} à Q_{pc} . Les carnets de rendez-vous sont complets et de nouveaux producteurs pourraient produire à pleine capacité. Le niveau de fraude augmenterait simplement.

Les deuxième et troisième cas (figures 2.4 et 2.5) démontrent la proposition 1. Dans le modèle de Darby et Karni (1973), il faut qu'il y ait des carnets de rendez-vous sans fraude non-complets pour justifier la fraude. L'inverse n'est pas vrai. Si l'on observe que les carnets de rendez-vous sont complets, cela n'implique pas qu'il n'y a pas de fraude dans le marché. Les cas 2 ou 3 ont pu se produire.

2.3. Des modèles prototypes pour l'analyse

2.3.1. Prototype à un seul bien

Soit un producteur dans un marché parfaitement concurrentiel (donc $V = 0$) maximisant une fonction d'espérance de profit. Il doit estimer la probabilité que le patient accepte son offre. Pour ce faire, il considère la fonction de répartition uniforme

$$F(Q) = \frac{Q}{\bar{Q}}$$

où Q est le niveau de service offert et \bar{Q} un niveau de service tel que le client va certainement refuser le service (probabilité égale à un).

La fonction de coût est de la forme suivante:

$$C(Q) = Q^2$$

Le producteur maximise la fonction suivante:

$$\max_Q \pi = [1 - F(Q)] [PQ - C(Q)]$$

donc,

$$\max_Q \pi = [1 - \frac{Q}{\bar{Q}}] [PQ - Q^2]$$

Le niveau de service total optimal est obtenu en maximisant cette fonction de profit sur Q . Soit la condition de premier ordre suivante:

$$\frac{\partial \pi}{\partial Q} = [1 - \frac{Q}{\bar{Q}}] [P - 2Q] - \frac{1}{\bar{Q}} [PQ - Q^2] = 0$$

Après quelques manipulations, nous retrouvons l'équation sous la forme suivante⁶:

$$(2.4) \quad \frac{\partial \pi}{\partial Q} = \frac{3}{\bar{Q}} Q^2 - [2 + \frac{2P}{\bar{Q}}] Q + P = 0$$

La fonction d'offre sera explicitée en réécrivant (2.4) de la façon suivante:

$$Q^* = \frac{\bar{Q}}{3} + \frac{P}{3} \pm \frac{1}{3} \sqrt{\bar{Q}^2 + P^2 - P\bar{Q}}$$

L'offre de service sera donc fonction du prix et de \bar{Q} . Le producteur, en estimant les probabilités que le client accepte son offre estime quel est le \bar{Q} pour chaque patient. La décision d'offrir tel niveau de service sera fonction de cette estimation a priori.

⁶ Toutes les manipulations relatives à ce problème sont à l'annexe mathématique: modèle I.

2.3.2. Prototype séquentiel avec productions jointes

Jusqu'à maintenant, nous avons toujours considéré que le producteur produisait un seul service. Or, en fait, celui-ci offre toute une gamme de services plus ou moins reliés entre eux. Nous allons maintenant lever cette hypothèse (un seul service) et considérer que le producteur offre deux services différents mais reliés par la fonction de coût. Le producteur peut offrir les deux services séparément mais s'il les offre conjointement, le coût de production du deuxième service diminue. Au lieu de coûter Q_2^2 , il ne coûte à produire que vQ_2^2 où v est le coefficient d'économies d'envergure ($0 < v < 1$).

A titre d'exemple, nous pouvons considérer dans le domaine médical l'ablation des amydales et des végétations ou encore, l'appendicectomie qui est souvent pratiquée conjointement avec une autre opération interne. Il devient alors moins coûteux de produire l'appendicectomie par rapport à une situation où cette intervention était effectuée seule. Il y a donc économies d'envergure.

Le modèle est dit séquentiel puisque nous supposons que le premier service (Q_1) a été offert et accepté. C'est alors que le producteur doit évaluer quel niveau il doit offrir de Q_2 . Il maximise l'espérance de profit. Étant neutre au risque, le problème se pose:

$$\max_{Q_1, Q_2} \pi = P_1 Q_1 - Q_1^2 + \left[1 - \frac{Q_2}{\bar{Q}_2}\right] [P_2 Q_2 - v Q_2^2]$$

Nous nous intéressons à la condition de premier ordre régissant Q_2 ; la maximisation sur Q_2 nous donne:

$$\frac{\partial \pi}{\partial Q_2} = \left[1 - \frac{Q_2}{\bar{Q}_2}\right] [P_2 - 2vQ_2] - \left[\frac{1}{\bar{Q}_2} P_2 Q_2 - vQ_2^2\right] = 0$$

Après quelques manipulations, nous retrouvons l'équation sous la forme suivante⁷:

$$\frac{\partial \pi}{\partial Q_2} = \frac{3v}{\bar{Q}_2} Q_2^2 - \left[2v + \frac{2P_2}{\bar{Q}_2} \right] Q_2 + P_2 = 0$$

La fonction d'offre sera explicitée en réécrivant cette équation de la façon suivante:

$$(2.6) \quad Q_2^* = \frac{\bar{Q}^2}{3} + \frac{P_2}{3v} \pm \frac{1}{3v} \sqrt{v^2 \bar{Q}_2^2 + P_2^2 - vP_2 \bar{Q}_2}$$

Il est intéressant de comparer les équations (2.5) et (2.6) à l'aide d'une simulation. Nous allons donc donner des valeurs numériques aux paramètres \bar{Q} et v (respectivement 30 et 0.5). Ceci signifie donc que le producteur estime que s'il offre 30 unités, le client refusera le service avec la probabilité de refus égale à 1. Si le service est produit conjointement, les coûts de production sont de moitié.

Il est important de noter que dans ce type de modélisation, les prix sont exogènes le producteur individuel n'ayant pas d'influence sur leurs déterminations. Le consommateur ayant un comportement donné (fonction de répartition uniforme où \bar{Q} est égal à 30), les simulations permettent de déterminer les quantités offertes par un producteur selon différents prix. Lorsque nous passons d'un prix à un autre, il s'agit de deux régimes différents et le consommateur ne fait pas face aux différents prix. De plus, ce passage d'un prix à un autre dépasse le cadre de ces modèles. Enfin, il ne faut certes pas attribuer aux différents prototypes (et a posteriori aux simulations) les vertus de révéler sans ambiguïté les propriétés du monde réel. Soit les résultats selon différents prix:

⁷ Toutes les manipulations relatives à ce problème sont à l'annexe mathématique: modèle II.

TABLEAU 2.1
PROTOTYPES D'OFFRES DE SERVICES: SIMULATION

P	$\bar{Q} (= \bar{Q}_2)$	Q^*	Q_2^* (si $v = .5$)	Q_2^*/Q^*
1	30	0.46	0.986	2.13
3	30	1.46	2.83	1.94
12	30	5.28	8.83	1.67
24	30	8.83	12	1.36
30	30	10	12.67	1.27
33	30	10.46	12.92	1.24
300	30	14.60	14.81	1.02
600	30	14.81	14.91	1.01

Tel qu'attendu, le producteur n'offrira jamais 30 unités de services à un consommateur puisqu'il les refuserait. Il est également intéressant de constater que le ratio Q_2^*/Q^* diminue lorsque le prix et les quantités offertes augmentent.

A la limite, la diminution de moitié des coûts de production n'a presque plus d'effet sur la quantité offerte. Il n'y a donc presque plus de fraude imputable aux économies d'envergure. Il y a deux explications à ce phénomène. D'une part, lorsque les prix sont bas, les coûts de la fraude en termes de pertes de profits ainsi que la probabilité de refus le sont aussi. D'autre part, les économies d'envergure permettent alors de faire augmenter le profit unitaire de façon plus importante (en terme relatif) que lorsque les prix sont élevés. Enfin, l'étude du tableau 2.1 nous suggère une tarification optimale. C'est ce qui nous amène à formuler la proposition 2.

Proposition 2 : Lorsque la production jointe de deux services amène une diminution du coût de production du second, le prix de ce service doit diminuer de façon

proportionnelle à cette diminution de coût lorsqu'ils sont offerts conjointement. Ceci afin d'éviter une offre de services frauduleux imputable aux économies d'envergure et une surconsommation de ces services.

Cette proposition implique que si on multiplie chaque prix (P) dans l'équation (2.5) par $1/v$, la solution Q^* sera identique à celle de l'équation (2.6). En corollaire, nous pouvons dire que si nous multiplions chaque prix dans l'équation (2.6) par v , nous allons obtenir la solution de l'équation (2.5).

Démontrons la première partie de cette proposition. Pour ce faire, multiplions chaque prix de l'équation (2.5) par $1/v$ ce qui donne:

$$\frac{\bar{Q}}{3} + \frac{P}{3v} \pm \frac{1}{3} \sqrt{\frac{\bar{Q}^2}{v^2} + \frac{P^2}{v^2} - \frac{P\bar{Q}}{v}} = \bar{Q} + \frac{P}{3v} \pm \frac{1}{3v} \sqrt{v^2 \bar{Q}^2 + P^2 - vP\bar{Q}}$$

Nous retrouvons l'équation (2.6).

C.Q.F.D.

Pour démontrer la deuxième partie de la proposition, multiplions l'équation (2.6) par v ce qui donne:

$$\begin{aligned} & \frac{\bar{Q}_2}{3} + \frac{P v}{3v} \pm \frac{1}{3v} \sqrt{v^2 \bar{Q}_2^2 + P^2 v^2 - v^2 P \bar{Q}_2} \\ &= \frac{\bar{Q}}{3} + \frac{P}{3} \pm \frac{1}{3} \sqrt{\frac{1}{v^2} [v^2 \bar{Q}_2^2 + P^2 v^2 - v^2 P \bar{Q}_2]} \\ &= \frac{\bar{Q}_2}{3} + \frac{P}{3} \pm \frac{1}{3} \sqrt{\bar{Q}_2^2 + P^2 - P \bar{Q}_2} \end{aligned}$$

Nous retrouvons donc l'équation (2.5).

C.Q.F.D.

La proposition 2 nous amène donc à discuter de la tarification optimale. Ainsi, le prix du service (Q_2) devrait diminuer dans la même proportion que le coefficient d'économie d'envergure. Ainsi, si les coûts de production sont diminués de moitié lorsqu'il y a productions jointes, le prix du service devrait également diminuer de moitié.⁸ Une telle tarification est compatible avec l'optimum parétien⁹ et sera obtenue dans un marché parfaitement concurrentiel sans asymétrie d'information. Dans ce cas, le prix reflèterait bien les coûts en ressources pour produire le service.

2.3.3. Prototype analytique: le cas des garagistes

2.3.3.1. Prototype à un seul bien

Un exemple de marché où l'on retrouve le type de problème d'information qui nous intéresse est celui de la réparation des automobiles. L'étude de la tarification dans ce marché est intéressante. Généralement, la facture présentée au client est l'addition de deux composantes: le prix de la main-d'oeuvre et le prix des pièces. Reformulons notre prototype simple à un seul bien de manière à tenir compte de cette particularité.

Soit la fonction à maximiser:

$$\max_Q \pi = \left[1 - \frac{Q}{Q}\right] [(P - P_x)Q - (Q^2 + C \cdot Q)]$$

⁸ Cette interprétation découle directement de la proposition 2. Afin de ne pas alourdir le texte, la démonstration formelle est disponible en annexe: modèle III.

⁹ Cette assertion est démontrée à la section 2.6: productions jointes et économies d'envergure en concurrence pure et parfaite.

où P est le prix de la main-d'oeuvre pour produire (taux horaire x nombre d'heures);

P_x est le prix de la pièce nécessaire à la production d'une unité de services;

C est le coût de la pièce pour le producteur.

Le niveau de service total optimal est obtenue en maximisant sur Q :

$$\frac{\partial \pi}{\partial Q} = \left[1 - \frac{Q}{\bar{Q}}\right] [P + P_x - 2Q - C] - \frac{1}{\bar{Q}} [(P + P_x)Q - (Q^2 + CQ)] = 0$$

Après quelques manipulations élémentaires (reproduites en annexe mathématique: modèle IV), cette condition devient:

$$(2.7) \quad \frac{\partial \pi}{\partial Q} = \frac{3}{\bar{Q}} Q^2 - \left[\frac{2P + 2P_x - 2C}{\bar{Q}} + 2 \right] Q + P + P_x - C = 0$$

Gardons en réserve cette condition de premier ordre du maximum du prototype simple pour fins de comparaison avec sa contrepartie tirée du modèle séquentiel avec productions jointes.

2.3.3.2. Prototype séquentiel à productions jointes

Appliquons les transformations précédentes au prototype séquentiel à productions jointes que nous avons développé plus haut.

La fonction à maximiser devient:

$$\text{Max}_Q \pi = P_1 Q_1 + Q_1^2 + \left[1 - \frac{Q_2}{\bar{Q}_2}\right] [vP_2 + P_x] Q_2 - (vQ_2^2 + CQ_2)$$

Le prix de la main-d'oeuvre est également assigné du coefficient d'économies d'envergure (noté v). Ceci découle du fait que ce prix est le taux horaire (qui est fixe) multiplié par le nombre d'heures travaillées (qui est flexible). Le coefficient v ($0 < v < 1$) représente la diminution du temps requis pour effectuer la réparation Q_2 dû au fait que cette réparation est effectuée conjointement avec Q_1 . La référence étant le temps requis pour effectuer la réparation seulement. La condition de premier ordre est la suivante:

$$\frac{\partial \pi}{\partial Q} = \left[1 - \frac{Q_2}{\bar{Q}_2}\right] [vP_2 - 2vQ_2 + (P_x - C)] - \frac{1}{\bar{Q}_2} [(vP_2 + P_x)Q_2 - (vQ_2^2 + CQ_2)] = 0$$

Après quelques manipulations élémentaires (reproduites en annexe V), cette condition devient:

$$(2.8) \quad \frac{\partial \pi}{\partial Q} = \frac{3v}{\bar{Q}} Q_2^2 - \left[\frac{2vP_2 + (2P_x - 2C)}{\bar{Q}_2} + 2v \right] Q_2 + vP_2 + (P_x - C) = 0$$

Nous constatons que l'équation (2.8) est différente de l'équation (2.7). Afin de rendre les équations (2.7) et (2.8) semblables, il faudrait multiplier le terme $(P_x - C)$ par v puisque ce faisant, on pourrait mettre v en facteur. La fonction d'offre issue de ce problème serait donc la même. L'expression $P_x - C$ représente le profit sur les pièces vendues. Si ce profit est nul, les solutions (2.7) et (2.8) sont identiques. Généralement, cette expression est positive et nous pouvons justifier la fraude par les productions jointes dues au fait que le mécanisme des prix le permet. Pour que tel ne soit pas le cas, il faudrait que les profits réalisés sur la vente des pièces diminuent de façon proportionnelle à la diminution du temps nécessaire à la réparation due aux productions jointes. Si tel n'est pas le cas, nous pouvons justifier la fraude sans offre excédentaire.

2.4. Généralisation des résultats

Jusqu'à maintenant, nous avons utilisé une fonction de coût particulière ainsi qu'une fonction de probabilité simple. Nous allons maintenant démontrer que la proposition 2 est indépendante de ces hypothèses de travail et qu'elle est vérifiable dans un cadre beaucoup plus général.

Considérons un modèle séquentiel avec productions jointes où la fonction de coût ($C(Q)$) et la fonction de probabilité de refus ($F(Q)$) sont quelconques. Soit:

$$(2.9) \quad \max_Q \pi = P_1 Q_1 - C(Q_1) + [1 - F(Q)] [PQ - vC(Q)]$$

La condition de premier ordre nous donne:

$$\frac{\partial \pi}{\partial Q} = [1 - F(Q)] [P - vC'_Q] - F'_Q [PQ - vC(Q)] = 0$$

Multiplions chaque prix de cette condition d'équilibre par un coefficient (α), nous retrouvons:

$$(2.10) \quad \frac{\partial \pi}{\partial Q} = [1 - F(Q)] [\alpha P - vC'_Q] - F'_Q [\alpha PQ - vC(Q)] = 0$$

Si, comme nous le suggère la proposition 2, nous posons que $\alpha = v$, alors nous obtenons l'équation suivante:

$$\frac{\partial \pi}{\partial Q} = v[1 - F(Q)] [P - C'_Q] = vF'_Q [PQ - C(Q)]$$

En divisant les deux côtés de l'égalité par v , nous retrouvons la condition de premier ordre de Darby et Karni (1973) (dans le cas où il y a parfaite concurrence).

2.5. Productions jointes et hypothèse d'offre excédentaire

L'introduction des productions jointes nous amène donc à reconsidérer l'importance de l'hypothèse d'offre excédentaire dans la justification de la fraude. Cette hypothèse qui constituait une condition nécessaire à l'offre de services frauduleux lorsque l'on considérait que le producteur offrait un seul type de bien perd cette propriété dans un contexte de production jointe. L'offre excédentaire devient un facteur susceptible d'influencer les activités de création de demande au même titre que les deux autres principaux facteurs couramment cités dans la littérature: à savoir le mode de rémunération et la couverture d'assurance.

C'est ce qui nous amène à formuler ainsi la proposition 3.

Proposition 3 : Dans un contexte de production jointe, le respect de l'hypothèse d'offre excédentaire ne représente pas une condition nécessaire à l'offre de services frauduleux. Il peut y avoir offre de services frauduleux même s'il y a pleine capacité.

Afin de démontrer cette proposition, nous allons utiliser l'équation (2.10) représentant la condition de premier ordre dans le contexte de production jointe et interpréter \tilde{Q} comme étant le niveau de service acheté par un expert. Raisonnant comme Darby et Karni (1973), nous savons qu'il y aura offre de services frauduleux lorsque l'inégalité (2.11) est respectée:

$$(2.11) \quad [1 - F(\tilde{Q})] [\alpha P - vC'_{\tilde{Q}}] > F'_{\tilde{Q}} [\alpha P\tilde{Q} - vC(\tilde{Q})]$$

Pour qu'il y ait offre de services frauduleux, il faut qu'au niveau de service accepté par un expert (\tilde{Q}) le revenu marginal d'offrir des services frauduleux soit supérieur au coût marginal de cette offre. Lorsque l'on considérait un seul bien, il fallait que le prix soit supérieur au coût marginal de production afin de respecter l'équation (2.3) d'où l'importance de l'hypothèse d'offre excédentaire.

Dans le contexte de production jointe représenté par l'équation (2.11) le prix peut être égal au coût marginal de production ($P = C'_Q$) tout en respectant la condition d'optimalité; il faut par contre que le coefficient α soit supérieur à v . Si la tarification est optimale au sens de la proposition 2 (alors $\alpha = v$), l'hypothèse d'offre excédentaire redevient nécessaire afin de justifier la surconsommation du service qui nous intéresse.

Il est important de noter que tous les services offerts en production jointe ne sont pas nécessairement frauduleux. En effet, les économies d'envergure peuvent justifier économiquement les productions jointes dans un marché concurrentiel parfaitement informé. Nous allons maintenant ouvrir une parenthèse afin de déterminer les conditions nécessaires afin de justifier les productions jointes et caractériser l'emploi optimal des facteurs de production dans ces conditions et la tarification en vigueur.

Considérons, pour quelques instants du moins (section 2.6), un monde parfaitement concurrentiel où il n'existe aucun problème informationnel.

2.6. Productions jointes et économies d'envergure en concurrence pure et parfaite

Nous retrouvons une situation de production jointe lorsqu'une entreprise produit au moins deux biens interreliés sur le plan technique. De plus, les deux produits peuvent être obtenus dans des proportions différentes au cours du procédé de production.

Si les deux produits étaient toujours obtenus dans un rapport fixe, l'analyse ne requerrait plus un traitement particulier. Il s'agirait dans ce cas de redéfinir le bien produit et l'analyse pour un seul bien s'appliquerait.

L'interdépendance d'ordre technique se reflétera sur la fonction de coût. Pour que les productions jointes soient économiquement rentables, il faut que l'équation (2.12) soit respectée.

$$(2.12) \quad C(Q_1, Q_2) < C(Q_1, 0) + C(0, Q_2)$$

Ainsi, l'équation 2.12 sera respectée lorsque le coût de produire les biens Q_1 et Q_2 simultanément est inférieur à les produire séparément. Si le signe de l'inégalité est inversé, nous retrouvons une situation de déséconomies d'envergure et il devient plus rentable de produire les biens séparément. Nous ne retrouverions alors aucune production jointe dans le processus de fabrication.

Graphiquement, le concept d'économies d'envergure est représenté par la figure 2.6¹⁰ Si le vecteur des coûts nécessaires afin de produire les deux biens ensembles (représenté par l'axe

¹⁰ Cette représentation graphique est tirée de Baumol, Panzar et Willig (1982).

$C(Q_1, Q_2)$ est inférieur au plan OAB qui rejoint l'origine et les points $C(Q_1, 0)$ et $C(0, Q_2)$ (représentant les coûts pour produire les deux biens séparément), nous sommes dans une situation où il y a des économies d'envergure. Sur la figure 2.6, ce point critique est représenté par le point D.

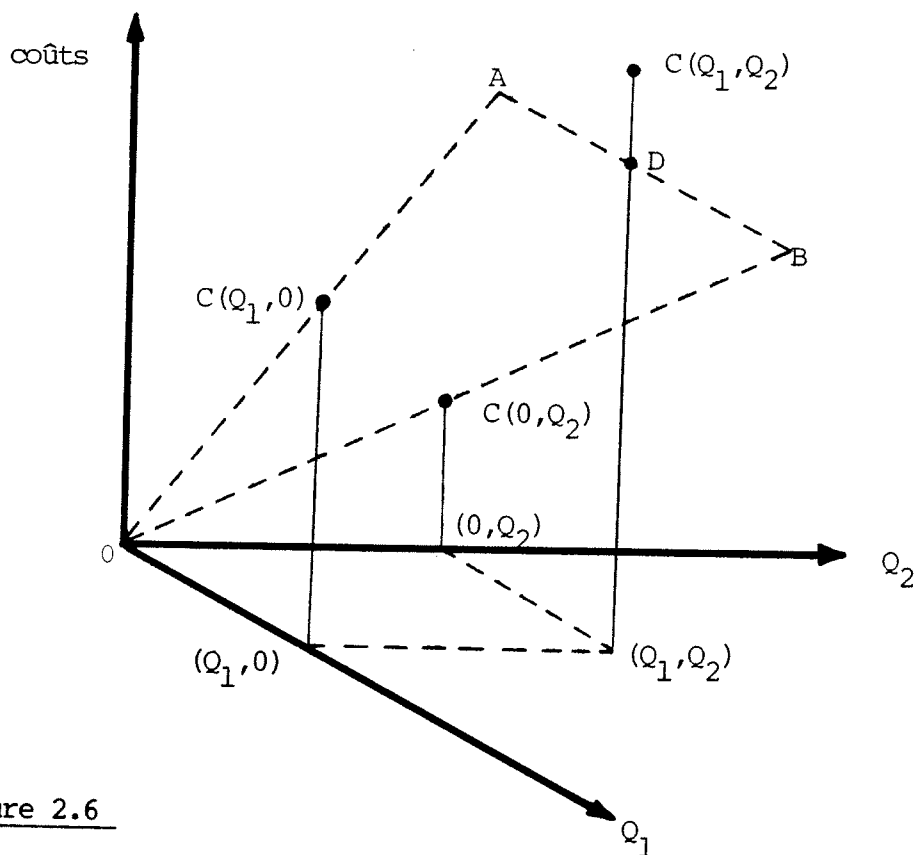


Figure 2.6

Baumol, Panzar et Willig (1982) démontrent que dans un contexte de concurrence parfaite (avec pleine information), la présence d'économies d'envergure est une condition nécessaire et suffisante aux productions jointes à l'équilibre.

En effet, une firme ne pourrait pas produire en longue période dans une situation de déséconomie d'envergure dans un marché parfaitement concurrentiel où il existe la libre entrée des entreprises. Elle se ferait "sortir du marché" par les autres entreprises plus efficaces qui produisent les biens indépendamment. De plus, si des économies d'envergure importantes sont réalisables, l'entreprise produisant un seul des biens liés à la production d'un autre par la fonction de coût se ferait également "sortir du marché" par les firmes à productions jointes. Pour Baumol, Panzar et Willig (1982):

"In competitive equilibrium, the firms comprising the industry produce the total industry output at least industry cost. In particular, no firm's output can be produced by two or more firms at a smaller total cost"¹¹

Henderson et Quandt (1978) caractérisent l'utilisation optimale d'un facteur de production (noté x) afin de produire deux outputs (respectivement Q_1 et Q_2) en production jointe par un producteur rationnel en concurrence parfaite. Soit P_1 , P_2 et P_x les prix des outputs et de l'input. La fonction de profit (noté π) à maximiser sera donc:

$$\max_{Q_1, Q_2} \pi = P_1 Q_1 + P_2 Q_2 - P_x \cdot x$$

La quantité d'inputs utilisée peut être exprimée en fonction de la quantité produite. Nous retrouvons alors:

$$\max_{Q_1, Q_2} \pi = P_1 Q_1 + P_2 Q_2 - P_x \cdot f(Q_1, Q_2)$$

¹¹ Baumol, Panzar et Willig (1982) p. 247.

La condition de premier ordre d'un maximum nous donne:

$$\frac{\partial \pi}{\partial Q_1} = P_1 - P_x f'_{xQ_1} = 0 \quad \text{alors,} \quad P_1 = P_x f'_{xQ_1}$$

$$\frac{\partial \pi}{\partial Q_2} = P_2 - P_x f'_{xQ_2} = 0 \quad \text{alors,} \quad P_2 = P_x f'_{xQ_2}$$

De ces conditions de premier ordre, nous en déduisons:

$$\frac{P_1}{P_2} = \frac{f'_{Q_1}}{f'_{Q_2}} = \text{TmT} \quad (\text{taux marginal et transformation})$$

Or, puisque $x = f(Q_1, Q_2)$, en vertu de la règle sur la dérivation des fonctions inverses, nous savons que:

$$\frac{\partial Q_1}{\partial x} = \frac{1}{f'_{Q_1}} \quad \text{et} \quad \frac{\partial Q_2}{\partial x} = \frac{1}{f'_{Q_2}}$$

$$\text{Donc,} \quad \frac{P_1}{P_2} = \frac{\partial Q_2 / \partial x}{\partial Q_1 / \partial x} = \frac{f'_{Q_1}}{f'_{Q_2}}$$

$$\text{Alors,} \quad P_1 \partial Q_1 / \partial x = P_2 \partial Q_2 / \partial x = P_x$$

À l'optimum, la valeur de la productivité marginale de l'input pour la production de chaque output doit être égale au prix de l'input.

Une hypothèse importante est implicitement utilisée par ces auteurs. Cette hypothèse, nous pouvons la formuler ainsi: les biens, même s'ils sont produits conjointement, peuvent être vendus séparément. Cette hypothèse est très importante et n'est pas toujours respectée lorsqu'il y a production de services. Dans le type de marché qui nous intéresse, il y a précisément des économies d'envergure

dues au fait que les deux produits sont produits et vendus au même consommateur simultanément.

Ainsi, dans cette situation, des services qui pourraient être produits conjointement avec économies ne le seront pas. Soit qu'il s'agisse de deux consommateurs différents, soit qu'il s'agisse du même consommateur mais à deux périodes de temps différentes.

Afin de détourner ce problème, nous allons utiliser un modèle d'offre séquentiel de services. Un consommateur parfaitement informé se présente sur un marché parfaitement concurrentiel et annonce son intention d'acheter le service Q_1 , au prix P_1 qui coûterait C'_{Q_1} au producteur. Il se fait alors offrir le service Q_2 qui pourrait être produit conjointement. Pour le producteur - le client acceptant Q_1 - le coût de produire Q_2 diminue et représente seulement vC'_{Q_2} où v représente le coefficient d'économie d'envergure ($0 < v < 1$). Nous pouvons caractériser ce problème. Soit le producteur maximisant le profit (noté π);

$$\max \pi = P_1 Q_1 - C(Q_1) + P_2 Q_2 - vC(Q_2)$$

Nous nous intéressons à la condition de premier ordre régissant Q_2 ; la maximisation sur Q_2 nous donne:

$$\frac{\partial \pi}{\partial Q_2} = P_2 - vC'_{Q_2} \quad \text{donc,} \quad P_2 = vC'_{Q_2}$$

Ainsi, dans un marché parfaitement concurrentiel sans asymétrie d'information, si les coûts de production sont diminués de moitié lorsqu'il y a production jointe (si $v = \frac{1}{2}$), le prix du service devra également diminuer de moitié. Cette condition à l'optimum de Pareto assure que le prix reflète bien les coûts en ressources pour produire

un service supplémentaire.

Il est important de faire le lien entre ce modèle séquentiel d'offre de services et le modèle séquentiel d'offre de services frauduleux développé plus haut. En fait, il s'agit de deux contextes totalement différents.

Dans un cas, nous nous trouvons dans un marché parfaitement concurrentiel où il n'y a pas d'asymétrie d'information. Le producteur offre un service à un consommateur parfaitement informé qui peut évaluer l'utilité qu'il retire de la consommation du service offert. L'offreur ne peut pas influencer le consommateur qui refuserait le service s'il était offert à un prix supérieur à vC'_{Q_2} en faveur d'un concurrent. La consommation du bien Q_2 est plus élevée que s'il n'y avait pas eu possibilité d'économies d'envergure puisque le prix diminue sous cette condition.

Dans l'autre cas, la situation est totalement différente. Il y a asymétrie d'information et le producteur peut être tenté d'essayer de tirer profit de cette situation en offrant des services que le consommateur aurait refusé s'il avait été un expert. Le producteur est en situation incertaine puisque le consommateur peut se sentir fraudé et refuser l'offre. Si la tarification est optimale au sens que nous l'avons défini plus haut, le producteur offrira un service supplémentaire frauduleux seulement si son carnet de rendez-vous est incomplet. Par contre, si la tarification n'est pas optimale, alors le producteur peut réagir à l'incitation financière et substituer des services frauduleux à des services non frauduleux profitant ainsi de la carence du marché (puisque $\alpha > v$).

Nous pouvons ainsi justifier la fraude sans excès de capacité de l'offre. Néanmoins, nous notons qu'au prix offert par le marché pour un service offert conjointement, le producteur voudrait offrir en plus grand nombre ces services dotés d'un coefficient d'économies d'envergure avantageux. En ce sens, le producteur est encore rationné et l'analyse présentée au début de ce travail demeure valide. Il devient alors pertinent de s'interroger au sujet du type de tarification que nous fournira un marché libre.

Jusqu'à maintenant, dans ce mémoire, nous avons supposé qu'il y avait asymétrie d'information uniquement sur l'utilité du service. Dans un tel contexte, un marché libre devrait amener une tarification optimale puisque le consommateur refuserait de consommer un service supplémentaire à un prix plus élevé que αP (où $\alpha > v$) puisque dû à la concurrence, il va pouvoir consommer le service supplémentaire (qu'il soit frauduleux ou non) à ce prix efficace. Lorsque nous supposons qu'il y a seulement asymétrie d'information sur l'utilité du service, nous supposons qu'il y a pleine information sur les prix et la qualité du service d'une part et que le consommateur peut aller acheter le service chez un autre producteur sans coût supplémentaire d'autre part (pas de frais de diagnostic, pas de coût de transport, etc...).

Ces hypothèses ne sont pas toujours respectées et il est possible que le producteur individuel ait une certaine influence sur les prix.¹² Dans la section suivante, nous analysons le comportement d'un producteur potentiellement frauduleux dans une telle situation.

¹² Il est possible que l'asymétrie d'information sur les prix soit liée à l'asymétrie d'information sur l'utilité du service. Nous ne sommes pas dans un monde "parfait" où les prix optimaux véhiculeraient toute l'information. Le principe de la "vérité des prix" du modèle néo-classique n'est pas respecté.

2.7. Le producteur face à un prix endogène

Supposons que le prix n'est plus exogène et que le producteur peut charger un prix plus élevé à un client ou lui offrir un service frauduleux. Nous pouvons représenter ce double comportement en maximisant l'équation 2.13 (fonction de profit) sur Q et P .¹³ Le modèle demeure à offre séquentiel et productions jointes.

$$(2.13) \quad \max_{P,Q} \pi = P_1 Q_1 - C(Q_1) + [1 - F(P,Q)] [\alpha P Q - vC(Q)]$$

La maximisation sur Q et P nous donne:

$$(2.14) \quad \frac{\partial \pi}{\partial Q} = [1 - F(P,Q)] [\alpha P - vC'_Q] - F'_Q [PQ - vC(Q)] = 0$$

$$(2.14) \quad \frac{\partial \pi}{\partial P} = [1 - F(P,Q)] [\alpha Q] - F'_P [\alpha P Q - vC(Q)] = 0$$

Alors que l'équation 2.14 est similaire à la condition de premier ordre issue du modèle d'offre séquentiel de services frauduleux à prix fixe (exogène), l'équation 2.15 caractérise le comportement du producteur face à la possibilité d'augmenter ses prix.

La condition d'équilibre de premier ordre (équation 2.15) égalise le revenu marginal en termes de quantité que le producteur espère vendre ($[1 - F(P,Q)] \alpha Q$) au coût marginal en termes de pertes de revenus nettes suite à l'augmentation du prix ($F'_P [\alpha P Q - C(Q)]$).

La valeur de la dérivée F'_P dépendra de plusieurs facteurs dont le niveau d'asymétrie d'information sur les prix, la capacité du consommateur à juger de la qualité du service offert, le coût réel du furetage et la relation de confiance consommateur-producteur.¹⁴

¹³ Dionne (1979) effectue une généralisation semblable à son modèle d'offre de services frauduleux à un bien et assurance.

¹⁴ En intégrant le processus d'acquisition d'information du consommateur dans son modèle séquentiel de furetage, à l'aide de la statistique bayésienne, Dionne (1979) vérifie qu'il y a un lien positif entre la confiance qu'a le consommateur envers le marché et les possibilités de fraude.

Alors que la consommation de la plupart des biens n'utilise pas du temps de travail, il n'en est pas de même pour la consommation des soins médicaux et pour la réparation des automobiles. Cette caractéristique propre à ce type de service prend une signification particulière dans un contexte d'économies d'envergure. Il n'est alors plus possible de considérer ce facteur fixe. Toutes choses étant égales par ailleurs, une diminution du temps requis pour la consommation sera perçue comme une baisse de prix pour le consommateur. Dans la section suivante, nous intégrons cette caractéristique au modèle d'offre de services frauduleux.

2.8. Le prix du temps de consommation et les possibilités de fraude

Dans les modèles de ce genre que nous retrouvons dans la littérature, tel que celui exposé dans Dionne (1979), le temps de furetage est important tandis que l'étude du temps de consommation n'est pas intéressante. Dans ces modèles, l'étude de ce facteur n'est pas pertinente parce qu'il n'est pas fonction des autres variables. Le temps de consommation est exogène.

Par contre, dans le modèle que nous développons tel n'est pas le cas. En effet, les économies d'envergure qui sont en cause dans le type de marché qui nous intéresse sont généralement associées à une diminution du temps requis pour effectuer le traitement ou la réparation. Or, sur ces marchés, nous retrouvons des services offerts en situation de production jointe ou non. Il en découle donc que le facteur temps de consommation ne peut plus être considéré comme étant exogène.

En effet, lorsqu'un consommateur accepte un service joint, le temps de travail requis pour consommer le service (temps d'attente dans le cas de la réparation automobile) est généralement moins long que si le service était consommé seul. Il est donc clair que dans ce contexte, il devient pertinent d'étudier comment le temps de consommation peut affecter les possibilités de fraude.

Nous anticipons qu'une diminution du temps de consommation ait un effet semblable à une diminution de coût dans le processus de décision du consommateur et qu'ainsi, le producteur estime que si le temps de consommation augmente, la probabilité que le client refuse le service augmente également. Il devrait en résulter une augmentation de l'offre de services frauduleux.

Avant de passer à la modélisation de l'introduction de cette nouvelle variable dans le modèle, il est important de noter que malgré cette extension, nous ne voulons aucunement nier l'importance du temps de furetage comme élément prépondérant pour justifier l'asymétrie d'information sur ce type de marché et la fraude qui peut en découler ainsi que la contribution importante qu'amène l'étude de cet élément central.

2.8.1. Le producteur et le prix du temps de consommation

Le producteur peut estimer que la fonction de probabilité cumulative que le client refuse le service offert soit fonction du prix du temps requis pour consommer le service. Nous pouvons donc modifier en conséquence le modèle d'offre de services frauduleux. Les profits (noté π) provenant d'un consommateur sont:

$$(2.16) \quad \max_Q \pi = [1 - F(Q, \tilde{t} w)] [PQ - C(Q)]$$

où \tilde{t} représente le temps de travail requis pour consommer un service (Q) et w le taux de salaire. Le niveau de service total optimal est obtenu en maximisant (2.16) sur Q :

$$(2.17) \quad \frac{\partial \pi}{\partial Q} = [1 - F(Q, \tilde{t} w)] [P - C'_Q] - F'_Q [PQ - C(Q)] = 0$$

De façon générale, l'équation (2.17) peut être représentée par:

$$L(Q, \tilde{t} w) = 0$$

Or, en vertu du théorème des fonctions implicites, nous savons que:

$$\frac{\partial L}{\partial Q} dQ + \frac{\partial L}{\partial \tilde{t} w} d\tilde{t} w = 0$$

Dans le premier terme de cette équation, nous retrouvons la condition de second ordre du problème de maximisation (noté $D < 0$).

Donc, nous avons:

$$D dQ + \frac{\partial L}{\partial \tilde{t} w} d\tilde{t} w = 0$$

que nous pouvons résoudre pour obtenir:

$$D dQ = \tilde{t} \left[-F'_{\tilde{t} w} (P - C'_Q) - F'_{Q \tilde{t} w} (PQ - C(Q)) \right] d\tilde{t} w = 0$$

Il est maintenant possible de dégager la relation suivante:

$$\frac{dQ}{d\tilde{t} w} = \frac{F'_{\tilde{t} w} (P - C'_Q) + F'_{\tilde{t} w Q} (PQ - C(Q))}{D < 0}$$

Le signe de $\frac{dQ}{d\tilde{t} w}$ dépend du signe du numérateur. Ainsi, si $F'_{\tilde{t} w} > 0$ et $F'_{\tilde{t} w Q} \geq 0$, alors, une augmentation du temps de consommation

diminue les possibilités de fraude. Inversement, une diminution du temps de consommation augmente les possibilités de fraude. Si $F'_{\tilde{t}w} > 0$, cela signifie que le producteur estime que si le temps de consommation augmente, la probabilité que le client refuse le service augmente également.

Nous devons étudier en détail la fonction $F(Q, \tilde{t}w)$ afin de statuer sur le signe de $dQ/d\tilde{t}w$. Ce qui nous amène à considérer le comportement du consommateur lorsque le prix du temps de consommation n'est pas nul.

2.8.2. Le consommateur et le prix du temps de consommation

Pour étudier le comportement du consommateur, nous utilisons le modèle séquentiel de furetage. L'évènement s'étant produit, le consommateur se présente sur le marché et doit évaluer l'offre du premier producteur qu'il rencontre. Soit une fonction d'espérance d'utilité utilisée comme référence en guise de stratégie de furetage.

$$(2.18) \quad EU(K) = \int_0^R u[\bar{K} - \ell(Q)] H(\ell(Q), I) d\ell(Q)$$

où $EU(K)$ est l'espérance d'utilité du consommateur;

\bar{K} est sa richesse initiale;

K est sa richesse finale;

R est la limite supérieure de $\ell(Q)$;

$\ell(Q)$ est le coût du traitement;

H est une fonction de densité quelconque (nous allons dorénavant écrire $H(.,.)$ pour $H(\ell(Q), I)$);

I est le niveau d'information à priori du consommateur;

Q est la quantité de traitements ou de services offerts. Il

s'agit donc de la quantité de Q_1 et/ou de Q_2 offerte.

Nous pouvons introduire le prix du temps de consommation en redéfinissant $\ell(Q)$:

$$\ell(Q) = PQ + \tilde{t}wQ$$

où $\tilde{t}w$ est le prix unitaire du temps (consommation d'une unité de Q)

Nous pouvons réécrire (2.14):

$$Eu(K) = \int_0^R U(K - PQ - \tilde{t} - \tilde{t}wQ)H(.,.),I) d\ell(Q)$$

Le consommateur se fait offrir $\ell(Q)$. S'il accepte, son coût en terme d'utilité sera $u(\bar{K} - \ell(Q))$. S'il refuse, il utilise son temps pour aller voir ailleurs ($K = wt$ où t représente le temps du diagnostic). La décision sera donc prise selon le critère suivant. Il acceptera si:

$$u(\bar{K} - \ell(Q)) \geq Eu(\bar{K} - wt)$$

Il refusera dans le cas contraire (si $<$).

La représentation graphique de ce processus de décision pour un individu qui est neutre au risque donne:

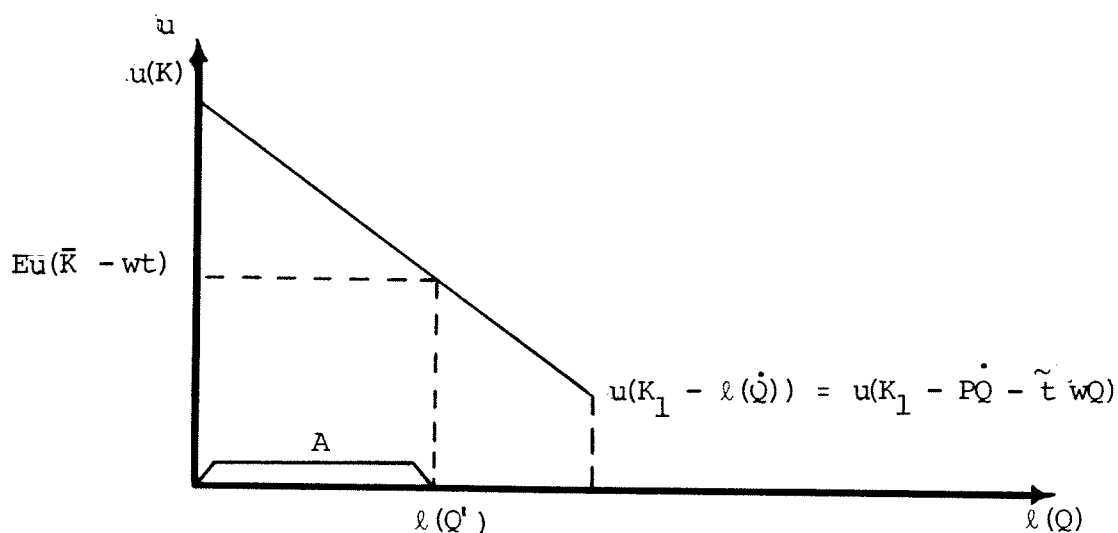


Figure 2.7

Le consommateur acceptera l'offre de service si elle se situe dans la région A . La frontière de la région est déterminée par:

$$Eu(\bar{K} - wt) = u(\bar{K} - PQ - \tilde{t}wQ)$$

Au point $\lambda(Q')$, nous avons:

$$u(\bar{K} - PQ' - \tilde{t}wQ) = u(\bar{K} - E(PQ) - E(\tilde{t}wQ) - wt)$$

Donc,

$$PQ' + \tilde{t}wQ' = E(PQ) + E(\tilde{t}wQ) - E(\tilde{t}wQ) - wt$$

Ce qui implique que:

$$PQ' = E(PQ) + E(\tilde{t}wQ) - \tilde{t}wQ + wt$$

$$(2.19) \quad \frac{\partial (PQ')}{\partial \tilde{t}} = -wQ < 0$$

Les possibilités de fraude augmentent avec une diminution du temps de consommation. Du point de vue du consommateur, le coût d'une même offre diminue. Nous pouvons donc considérer que $F'_{\tilde{t}w} > 0$. Si le temps de consommation diminue, la probabilité que le client refuse le service augmente. Nous avons vu dans la partie traitant du producteur que $F'_{\tilde{t}w} > 0$ signifie que $\partial Q / \partial \tilde{t}w$ est négatif. Le producteur préfère offrir des services dont le temps requis pour les consommer est relativement peu élevé, la probabilité qu'ils soient refusés étant plus faible.

2.8.3. Temps de consommations et offre en production jointe

Dans le type de marché qui nous intéresse, les économies d'envergure sont généralement associées à une diminution du temps requis pour effectuer le traitement ou la réparation. Ainsi, lorsqu'un consommateur accepte un service joint, le temps de travail requis pour consommer ce service est généralement moins long que si le service était consommé seul.

Ainsi, en vertu de notre analyse du comportement du consommateur lorsque du temps de travail est requis pour consommer (section 2.7.2.), nous pouvons considérer que - toutes choses étant égales par ailleurs - la probabilité de refus est moins élevée pour un service joint.¹⁵ Le modèle du producteur (2.7.1.) prévoit que dans une telle situation, les possibilités de fraude augmentent.

Dans notre revue de la littérature portant sur la vérification empirique de l'hypothèse de création de demande (1.2.2.2.), nous avons constaté un lien entre le revenu du consommateur (et par conséquent le taux de salaire) et les possibilités de

¹⁵ En fait, le consommateur peut ne pas être indifférent entre le temps de consommation et le temps de travail. Nous pouvons en tenir compte en réécrivant le modèle en multipliant w par une variable exogène: β . Pour justifier une hausse de l'offre de services joints imputable aux économies de temps de consommation que les économies d'envergures entraînent, il est seulement nécessaire que $\beta w > 0$ (l'interprétation de βw étant la suivante: il s'agit du coût encourue pour la consommation du service pendant une unité de temps). Dans la littérature économique portant sur le prix du temps, les études économétriques révèlent généralement que les consommateurs accordent une fraction du taux de salaire horaire au prix du temps. Il est raisonnable de croire que $0 < \beta < 1$.

fraude.¹⁶ Compte tenu du résultat exprimé par l'équation 2.19, un premier réflexe nous amènerait à croire que l'introduction du prix du temps de consommation dans le modèle nous permet de justifier théoriquement ce résultat sans être dans l'obligation de postuler un niveau d'asymétrie d'information et/ou un comportement qui varierait avec le revenu.¹⁷

Mais, tel n'est pas le cas. En effet:

$$\frac{\partial PQ'}{\partial w} = E(\tilde{t}Q) - \tilde{t}Q + t$$

Pour justifier théoriquement une relation négative entre le revenu du consommateur et l'offre de services frauduleux, il faudrait réunir les deux conditions suivantes:

$$\tilde{t}Q > E(\tilde{t}Q)$$

$$\tilde{t}Q - E(\tilde{t}Q) > t$$

Or, nous avons postulé un manque d'information de la part du consommateur seulement par rapport à la quantité qu'il désire consommer. Sous l'hypothèse de pleine information sur les prix et le temps de consommation, la première condition ne sera pas respectée. A un niveau d'information équivalent, on devrait donc obtenir une relation positive entre l'offre de services frauduleux et le revenu du consommateur, le coût réel du furetage étant plus élevé pour les consommateurs à haut taux de salaire. Ainsi, pour justifier une

¹⁶ Il ne s'agirait pas de dire qu'il y aura toujours une relation positive entre le revenu du consommateur et les possibilités de fraude mais de justifier théoriquement un résultat empirique. Le coût du furetage augmente également avec le revenu.

¹⁷ Quoique cette interprétation demeure envisageable. Rappelons-nous que Fuchs (1978) a vérifié que l'influence de l'offre sur la demande est plus élevée pour les classes d'éducation plus faibles.

hypothétique relation positive entre l'influence de l'offre sur la demande et le revenu des consommateurs, il faut envisager soit un niveau d'asymétrie d'information sur les besoins fonction du revenu, soit une double asymétrie d'information (ou encore les deux). Ceci afin de compenser le coût réel du furetage plus élevé pour les consommateurs à revenus plus élevés.

De plus, il est intéressant de noter que généralement, la diminution relative du temps de consommation attribuable au fait que le service soit offert et produit conjointement est plus importante que le coefficient d'économie d'envergure (noté v). Deux raisons peuvent justifier cette assertion.

La première justification est reliée au phénomène d'indivisibilité que l'on retrouve généralement sur le marché du travail. A titre d'exemple, considérons le cas d'un travailleur qui doit effectuer une visite à son médecin ou encore faire effectuer une réparation à son automobile et qu'il soit dans l'obligation d'y consacrer une demi-journée. Si le temps requis pour effectuer le premier service ne requiert que deux heures, il est alors possible que l'acceptation d'un deuxième service n'entraîne pas de pertes de temps de travail additionnelles.¹⁸

La seconde justification, (qui s'applique principalement au marché des services médicaux), est reliée à la notion de convalescence. En effet, lorsqu'il est question d'opérations chirurgicales par exemple, les coûts relativement élevés en termes de pertes de revenus d'emploi sont imputables au temps nécessaire à la convalescence. Or, souvent, une intervention nécessite peu ou pas de

¹⁸ Toutefois, il est possible d'imaginer des contre-exemples où il se produit le phénomène inverse dû aux indivisibilités.

pertes de temps supplémentaires si elle est pratiquée en même temps qu'une autre intervention.

Nous pouvons citer, à titre d'exemple, le cas de l'appendicectomie qui peut nécessiter plusieurs jours de convalescence si elle est pratiquée seule. Si l'opération est produite conjointement à une autre opération interne, elle n'augmente pas le temps de rétablissement.

Nous pouvons également justifier une tarification non optimale par une réglementation. C'est ce qui peut se produire dans un régime d'assurance-santé tel que l'on retrouve au Québec. La tarification est le résultat d'une négociation entre l'assureur (l'Etat) et les producteurs. Mais, la littérature économique nous apprend que l'assurance peut également amener le consommateur à modifier son comportement. C'est pourquoi, avant d'aborder l'étude des effets de l'assurance sur la fraude, il est nécessaire d'étudier le lien entre le risque moral et l'offre de services frauduleux à l'aide de quelques textes choisis. C'est ce que nous ferons dans la première partie de la section suivante.

2.9. Effet de l'assurance sur la fraude en production jointe

Le nombre de producteurs per capita n'est pas le seul facteur susceptible d'influencer les activités de création de demande. Dans la littérature économique, la couverture d'assurance est également fréquemment citée à ce titre.

Nous allons donc maintenant situer la notion d'offre de services frauduleux par rapport à celle de risque moral à l'aide de

cette littérature qui est généralement appliquée au marché des services médicaux. Le risque moral est un phénomène de mauvaise allocation des ressources sous l'assurance que l'on peut rencontrer lorsque l'assureur ne peut pas observer les actions de l'assuré à des coûts raisonnables. Il peut donc y avoir consommation abusive.

2.9.1. Risque moral versus offre de services frauduleux

Dans un contexte de pleine assurance avec carte blanche,¹⁹ le consommateur désiera consommer tous les services utiles (noté $\partial u / \partial H > 0$ où u est le niveau d'utilité et H le niveau de service) puisque le prix de référence pour le consommateur rationnel maximisant son utilité est celui qu'il est amené à déboursier. Puisque le prix de référence pour le consommateur (dans le cas d'une pleine assurance, celui-ci est égal à zéro) est inférieur au coût marginal social (Cm_s), la quantité demandée de services médicaux (noté H) augmentera (de q_0 à q_1 sur la figure 2.8)

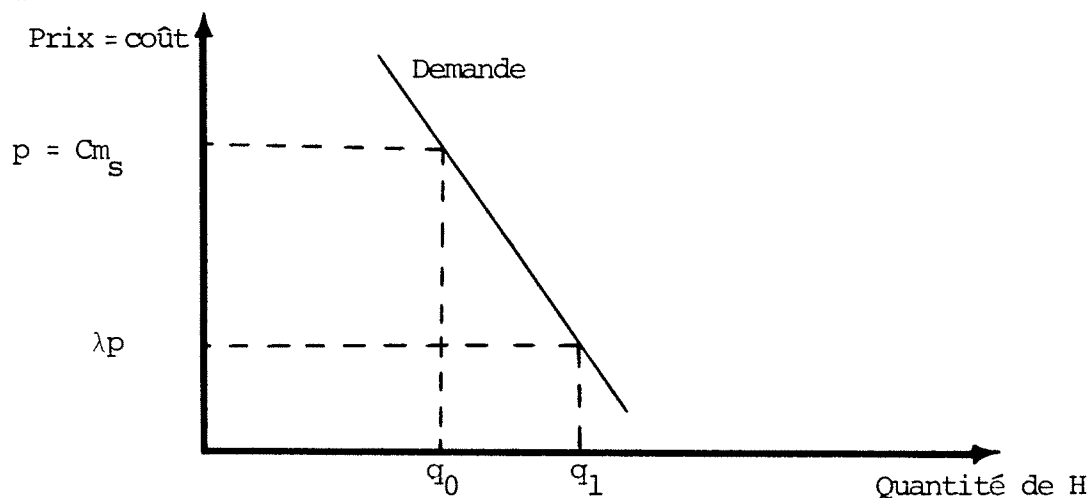


Figure 2.8

¹⁹ L'assuré peut consommer la quantité qu'il désire à un prix monétaire nul. Il n'y a pas de déductible ni de co-paiement. Par contre, le prix du temps de consommation est à la charge du consommateur et il est égal à λp où $0 < \lambda < 1$.

Cette consommation additionnelle qui découle de l'assurance est due à un problème d'information entre assureur et assuré (Pauly 1968). Ce problème de risque moral est dû au fait que l'assureur ne peut observer séparément l'état de la nature et le comportement de l'assuré qui peut influencer l'événement aléatoire par ses activités. Nous pouvons distinguer deux formes de risque moral:

- diminution des activités d'auto-protection pour l'assuré. La probabilité que l'événement se produise augmente sous l'assurance.
- augmentation des dépenses lorsque l'événement s'est produit (qui peut être due à la réduction des activités de marché ou encore à l'influence directe de l'assuré sur le montant de la perte).

Enfin, il est très important de noter que toute augmentation de consommation de services médicaux associée à une augmentation de la couverture d'assurance n'est pas nécessairement due à un problème de risque moral. En effet, l'assurance permet une réallocation efficace des revenus puisqu'elle amène un effet-revenu (Marshall 1976).

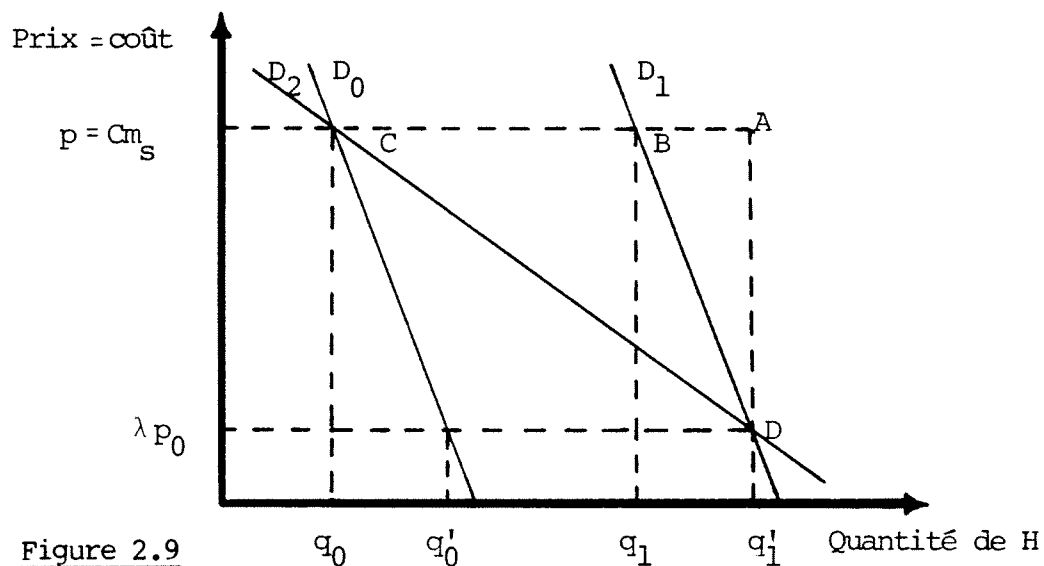


Figure 2.9

La figure 2.9 va nous permettre de représenter le phénomène. Initialement, les consommateurs paient le prix p_0 qui est égal au coût marginal social et consomment la quantité q_0 de services médicaux. Suite à l'implantation d'un régime de pleine assurance avec carte blanche, la quantité consommée augmente jusqu'à q'_1 puisque le prix auquel fait face le consommateur est égal λp_0 ($\lambda > 0$, le coût du temps de consommation étant positif).

Un premier réflexe serait de considérer la courbe de demande D_2 comme étant effective. Ainsi, sous cette condition, la consommation abusive imputable au risque moral serait de $q'_1 - q_0$ et la perte sociale qui y serait associée serait de DAC.

Mais, l'assurance procure également un effet-revenu aux consommateurs puisqu'elle permet aux individus de consommer plus. Il en résulte donc un important problème d'identification.²⁰ Sur la figure 2.10, nous constatons qu'au prix p_0 , les consommateurs auraient acheté q_1 unités de services médicaux si leurs revenus le leur avaient permis. Les courbes de demandes pertinentes pour mesurer l'ampleur du problème de risque moral sont donc les courbes D_0 et D_1 puisqu'elles n'incluent qu'un effet-prix. La consommation abusive imputable au risque moral serait alors seulement de $q'_1 - q_1$ et la perte sociale qui y est associée de ABD. Le passage de q_0 à q_1 constitue seulement une réallocation efficace des ressources due à l'assurance.

²⁰ Dans le cadre de ce travail, nous ne faisons qu'aborder les problèmes reliés au risque moral et à l'assurance afin de situer la consommation abusive par rapport à d'autres types de problèmes d'information que nous analysons. Cette partie du texte est inspirée de deux travaux beaucoup plus complets traitant du sujet: Dionne (1979) et St-Michel (1983).

Jusqu'à maintenant, dans cette section traitant du risque moral, nous avons implicitement considéré que le médecin est un agent neutre du processus de consommation des services médicaux. Nous avons ainsi pu mettre en évidence l'asymétrie d'information propre au modèle principal-agent (assureur-assuré) et caractériser le risque moral propre aux assurés.

Nous allons maintenant lever cette hypothèse implicite et supposer que l'intermédiaire (en l'occurrence le médecin) n'est pas indifférent à une variation de la couverture d'assurance et que son comportement en est affecté. Nous pouvons représenter graphiquement le phénomène (St-Michel 1983).

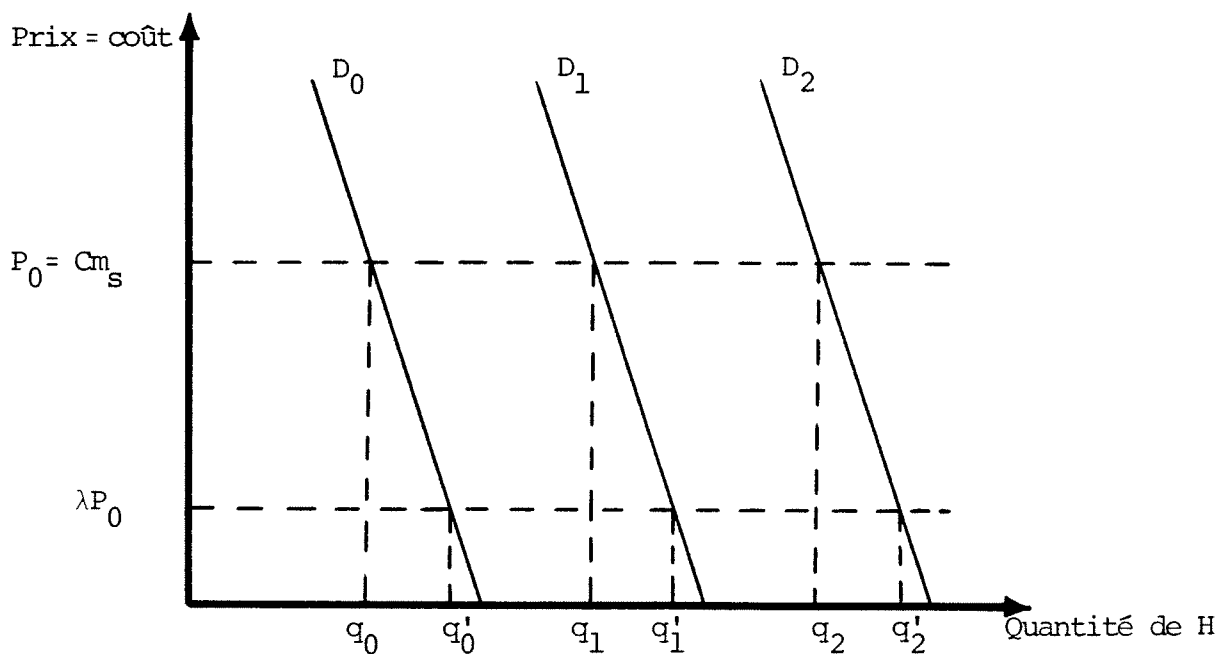


Figure 2.10

Pour une maladie donnée, l'individu non assuré fait face à la demande D_0 et consomme q_0 . Suite à l'introduction d'un régime de pleine assurance avec carte blanche, l'individu assuré fait maintenant face à la demande D_1 et consomme q_1 . Nous avons déjà statué au sujet de l'interprétation de cette hausse de la consommation: de q_0 à q_1 ; il s'agit d'une réallocation efficace des ressources due à l'assurance tandis que de q_1 à q'_1 ; il s'agit de consommation abusive due au risque moral propre aux assurés. Par contre, si le médecin n'est plus neutre face à l'assurance de son patient, il peut y avoir création de demande et la demande pour ce dernier correspond à D_2 (figure 2.10).²¹ Le passage de D_1 à D_2 est entièrement assimilé au risque moral propre à l'attitude médicale et l'augmentation de la consommation de q'_1 à q_2 est rendue possible par l'asymétrie d'information dans la relation médecin-assuré. L'assurance peut donc avoir une incidence sur le niveau de fraude tel que défini dans ce travail et le lien entre la notion de fraude et celle de risque moral propre à l'attitude médicale est direct. En effet,

"le risque moral dû à l'offre est la consommation accrue attribuable à l'asymétrie d'information dans les relations assureur-médecin et médecin-assuré; or puisque la fraude correspond à une offre de

²¹ Dionne (1979 et 1980) démontre qu'un producteur rationnel maximisant son profit ne sera pas neutre face à l'assurance. Pour ce faire il introduit l'assurance dans le modèle d'offre de services frauduleux de Darby et Karni (1973) et il adapte son modèle séquentiel de furetage à un contexte d'assurance. Nous ne présentons pas cette contribution importante dans cette revue de la littérature parce que, en 2.9.2 et 2.9.3 nous allons utiliser la même méthodologie afin d'analyser l'effet de l'assurance sur la fraude en économie d'envergure. Nous allons ainsi justifier le passage de D_1 à D_2 .

services qui serait refusée par un consommateur informé de manière parfaite, étant donné une couverture d'assurance, le risque moral s'identifie complètement à de l'offre frauduleuse."²²

Par contre, la réciproque n'est pas vraie;

"Le risque moral afférent aux actions uniques du médecin s'associerait donc intégralement à de la fraude alors que celle-ci est observable en l'absence de couverture d'assurance (Darby et Karni 1973, Dionne 1979)"²³

Fort de cette littérature qui nous a permis de situer l'offre de services frauduleux dans un contexte d'assurance, nous pouvons maintenant adapter le modèle d'offre de services frauduleux avec productions jointes à un contexte d'assurance. Il ne s'agit pas de déterminer le choix optimal d'une couverture d'assurance puisque notre analyse est limitée au choix du niveau de services lorsque l'événement s'est produit. La couverture d'assurance est exogène (elle peut être imposée par l'État).

2.9.2. Le producteur avec assurance en production jointe

Le producteur peut estimer que la fonction de probabilité cumulative que le client refuse le service offert soit fonction du niveau d'assurance du consommateur. Nous pouvons modifier en conséquence notre modèle d'offre de services en production jointe.

²² Saint-Michel (1983) p. 170.

²³ Saint-Michel (1983) p. 176.

L'équation (2.9) - représentant le modèle général - sera donc modifiée et la valeur présente des profits (noté π) provenant d'un consommateur assuré est représentée par l'équation (2.20):

$$(2.20) \quad \max \pi = P_1 Q_1 - C(Q_1) + [1 - F(Q, s)] [PQ - vC(Q)]$$

où s représente le niveau d'assurance du consommateur. Les autres variables conservent les mêmes définitions que précédemment.

Le niveau de service total optimal est obtenu en maximisant (2.20) sur Q :

$$(2.21) \quad \frac{\partial \pi}{\partial Q} = [1 - F(Q, s)] [P - vC'_Q] - F'_Q [PQ - vC(Q)] = 0$$

De façon générale, l'équation (2.21) peut être représentée par:

$$G(Q, s) = 0$$

Or, en vertu du théorème des fonctions implicites, nous savons que:

$$\frac{\partial G}{\partial Q} dQ + \frac{\partial G}{\partial s} ds = 0$$

Dans le premier terme de cette équation, nous retrouvons la condition de second ordre du problème de maximisation (noté $D < 0$).

Donc, nous avons:

$$DdQ + \frac{\partial G}{\partial s} ds = 0$$

que nous pouvons résoudre pour obtenir:

$$D dQ + [-F'_S (P - vC'_Q) - F'_{QS} (PQ - vC(Q))] ds = 0$$

La relation entre Q et s sera donc:

$$\frac{dQ}{ds} = \frac{F'_S (\alpha P - vC'_Q) + F'_{QS} (\alpha PQ - vC(Q))}{D}$$

Le signe de dQ/ds dépend du signe du numérateur. Ainsi, si $F'_S < 0$ et $F'_{QS} \leq 0$, alors, l'assurance augmente les possibilités de fraude pour le producteur. Si $F'_S < 0$, cela signifie que le producteur estime que si son client augmente son assurance, sa probabilité de refuser un nouveau service sera plus faible. Puisque nous pouvons obtenir toute une combinaison de résultats possibles, nous devons étudier en détail la fonction $F(Q,s)$ afin d'être en mesure de statuer sur le signe de dQ/ds . Pour ce faire, nous allons étudier le comportement du consommateur sous l'assurance.

2.9.3. Le consommateur sous l'assurance

Pour analyser le comportement du consommateur sous l'assurance, nous adaptons le modèle séquentiel de furetage au contexte de l'assurance. Soit une fonction d'espérance d'utilité servant de stratégie de furetage au consommateur:²⁴

²⁴ La méthodologie utilisée dans cette section du mémoire est inspirée de Dionne (1979 et 1980). Celui-ci analyse différents régimes d'assurance. Ces analyses nous permettent de constater l'importance du coût du temps du furetage dans le niveau de fraude en concurrence parfaite. L'auteur effectue également plusieurs autres extensions du modèle de base. D'une part, il laisse tomber l'hypothèse de neutralité au risque au profit de celle de l'aversion au risque. D'autre part, il s'intéresse au lien (qui s'avère positif) entre la confiance du consommateur envers le marché et les possibilités de fraude.

$$Eu(K, \theta) = \int_0^R u(\bar{K} - \ell(Q) + s)H(.,.) d\ell(Q)$$

$$\text{e.q.} \quad s = \theta(\ell(Q))$$

$$\text{et} \quad 0 < \theta < 1$$

θ étant la fraction de la perte couverte par l'assurance.

Dans le cas d'un modèle linéaire,²⁵ cette relation peut se réécrire:

$$Eu(K, \theta) = \int_0^R u(\bar{K} - (1 - \theta)\ell(Q)) H(.,.) d\ell(Q)$$

$$\text{e.q.} \quad 0 < \theta < 1$$

Afin d'analyser l'effet de l'assurance sur les possibilités de fraude, on doit vérifier ce qu'il advient de la région d'acceptation (la région A sur les figures 2.8 et 2.11) lorsque l'on modifie le niveau d'assurance (θ).

Au point $\ell(Q')$, nous avons l'équation (2.22)

$$(2.22) \quad u[\bar{K} - (1 - \theta)\ell(Q')] = u[\bar{K} - (1 - \theta)E\ell(Q) - wt]$$

A la limite, l'utilité certaine que le consommateur retire de la consommation du niveau de service offert en termes de perte de richesse est égale à l'espérance d'utilité d'aller consulter un autre offreur et payer le coût associé au temps de furetage (w_t). Ce qui implique que:

$$(1 - \theta)\ell(Q') = (1 - \theta)[E\ell(Q')] + w_t$$

²⁵ L'assurance est imposée (exogène) puisqu'un individu neutre au risque n'achèterait pas d'assurance.

Donc,

$$(2.23) \quad \lambda(Q') = E \lambda(Q') + \frac{wt}{(1-\theta)}$$

$$(2.24) \quad \frac{\partial \lambda(Q')}{\partial \theta} = \frac{wt}{(1-\theta)^2} > 0$$

Puisque la frontière se déplace vers la droite lorsque θ augmente, la probabilité de refus du consommateur diminue ($F'_s < 0$). Les possibilités de fraude augmentent, donc, $\partial Q/\partial s > 0$ (voir section 2.8.2). Nous pouvons représenter graphiquement ce résultat:

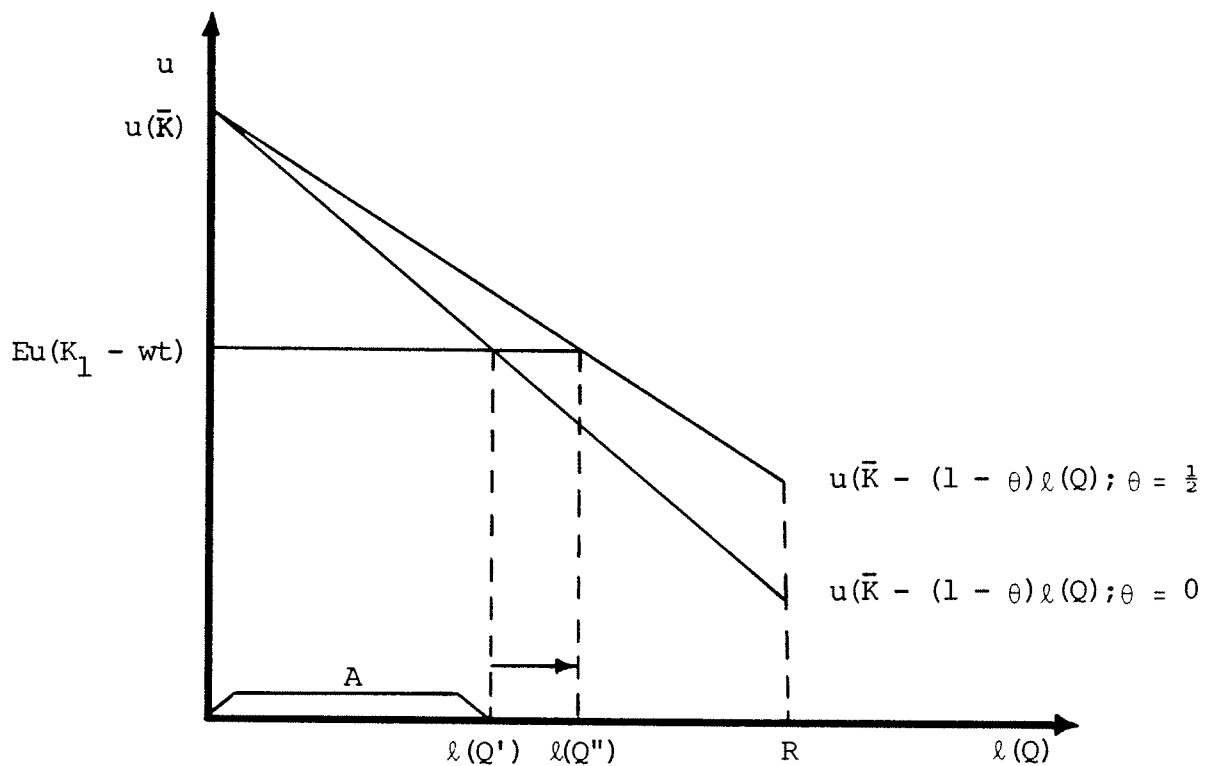


Figure 2.11

Introduisons le prix du temps de consommation en redéfinissant $\lambda(Q)$:

$$\lambda(Q) = PQ + \tilde{t}wQ$$

Considérons la couverture d'assurance:

$$\theta = 1 \quad \text{pour } PQ$$

$$\theta = 0 \quad \text{pour } \tilde{t}_Q wQ$$

Nous pouvons réécrire l'équation (2.22) afin de tenir compte du temps de consommation:

$$(2.25) \quad u(\bar{K} - \tilde{t}wQ' - (1 - \theta)PQ') = u(\bar{K} - (1 - \theta)E(PQ) - wt - E(\tilde{t}wQ))$$

Donc,

$$\tilde{t}_Q wQ' + (1 - \theta)PQ' = (1 - \theta)E(PQ) + wt + E(\tilde{t}wQ)$$

$$(1 - \theta)PQ' = (1 - \theta)E(PQ) + wt + E(\tilde{t}wQ) - \tilde{t}wQ$$

Ce qui nous permet d'obtenir:

$$(2.26) \quad PQ' = E(PQ) + \frac{wt + E(\tilde{t}wQ) - \tilde{t}wQ}{(1 - \theta)}$$

$$(2.27) \quad \frac{\partial PQ'}{\partial \theta} = \frac{wt + E(\tilde{t}wQ) - \tilde{t}wQ}{(1 - \theta)^2}$$

Comparons les équations (2.24) et (2.27) afin d'évaluer l'effet de la prise en compte du temps de consommation sur les possibilités de fraude.²⁶ Le temps de consommation augmentera les possibilités de fraude seulement si $E(\tilde{t}wQ) > \tilde{t}wQ$; c'est-à-dire si le consommateur croit que le temps de consommation requis pour consommer l'offre de service de ce producteur est moins élevé que le temps requis pour consommer une autre offre. Ce résultat est intéressant et il est obtenu parce que nous comparons deux contextes différents. Le nouveau contexte (prix du temps de consommation non nul) favorisera essentiellement les producteurs qui font des offres peu coûteuses en temps. Pour ces producteurs $E(\tilde{t}wQ) > \tilde{t}wQ$; une augmentation de la couverture d'assurance aura un impact plus important sur la probabilité d'acceptation de leurs clients dans un monde où le prix du temps de consommation est positif.

2.9.3.1. Prix du temps de consommation et productions jointes

Comme nous l'avons mentionnée plus haut (section 2.8.3., temps de consommation et offre en production jointe), lorsqu'il y a des productions jointes, le temps requis pour consommer le service supplémentaire est moindre par rapport à une situation où ce service serait produit indépendamment.

Nous pouvons réécrire (2.25) lorsque Q_2 est produit conjointement:

$$u(\bar{K} - (1 - \psi)\tilde{t}wQ' - (1 - \theta)PQ_2') = u(\bar{K} - (1 - \theta)E(PQ_2')) - wt - E(\tilde{t}wQ_2)$$

²⁶ pour obtenir l'équation 2.24, nous ne tenons pas compte du temps de consommation. Alors, $Q(Q')$ est égal à PQ' .

où ψ représente le coefficient de réduction en temps due aux productions jointes ($0 < \psi < 1$). Donc,

$$(1 - \theta)(PQ'_2) = (1 - \theta)E(PQ'_2) + wt + E(\tilde{t}wQ_2) - \psi(\tilde{t}wQ_2)$$

Ce qui nous permet d'obtenir:

$$(2.28) \quad PQ'_2 = E(PQ'_2) + \frac{wt + E(\tilde{t}wQ_2) - \psi\tilde{t}wQ_2}{(1 - \theta)}$$

$$(2.29) \quad \frac{\partial PQ'_2}{\partial \theta} = \frac{wt + E(\tilde{t}wQ_2) - \psi\tilde{t}wQ_2}{(1 - \theta)^2}$$

En comparant les équations 2.26 et 2.28 nous constatons que les économies d'envergure ont vraisemblablement pour effet d'augmenter la région d'acceptation (puisque $2.28 > 2.26$)²⁷

De la même manière, une comparaison des équations 2.27 et 2.29 nous permet de constater que l'assurance aura plus d'impact sur l'offre de services produits en économies d'envergure que sur les services produits indépendamment dans un monde où le prix du temps de consommation n'est pas nul. Il est même possible d'imaginer une situation où le coefficient de réduction en temps de consommation dû aux productions jointes (noté ψ) est tel qu'il renverse le signe de la dérivée partielle.²⁸

27 Quoiqu'il est possible d'imaginer un consommateur n'ayant absolument aucune information sur le temps normalement requis pour consommer le service. Ce consommateur pourrait estimer que $E(\tilde{t}wQ)$ est toujours égal à $\tilde{t}wQ$ (qu'il soit offert conjointement ou non). Dans ce cas limite, les économies d'envergure n'auraient pas d'influence sur la région d'acceptation et sur l'effet de l'assurance.

28 Nous rencontrerions ce cas si $\tilde{t}wQ > wt + E(\tilde{t}wQ)$ dans (2.27) et $\psi\tilde{t}wQ_2 < wtE(\tilde{t}wQ_2)$ dans (2.29). A notre avis, cette situation est peu probable puisque la première condition ne sera généralement pas respectée. Même si $\tilde{t}wQ$ était supérieur à $E(\tilde{t}wQ)$, les coûts d'aller voir un autre producteur (wt) sont élevés.

Ainsi, l'introduction de l'assurance, loi de limiter l'offre de services frauduleux en production jointe favorise particulièrement la fraude sous cette condition. Les consommateurs sont prêts à accepter une grande quantité de services joints sous l'assurance puisque les coûts pour les consommer en terme de temps sont relativement faibles tandis que le coût réel du furetage est élevé.

CHAPITRE III - RÉGLEMENTATION DANS DEUX MARCHÉS IMPORTANTS

Dans ce chapitre, nous appliquons notre modèle théorique à deux marchés importants en considérant l'effet de la réglementation québécoise dans ces marchés sur l'offre de services frauduleux en situation de production jointe. Ces deux marchés sont la réparation des automobiles et les services médicaux.

3.1. La réglementation sur le marché de la réparation automobile

Le marché de la réparation automobile demeure un marché relativement peu réglementé. Nous retrouvons la réglementation sous deux formes. La Loi sur la protection du consommateur réglemente ce marché ainsi que le décret sur les salariés de garages de la région de Montréal. De plus, certaines institutions telles l'Association pour la protection des automobilistes (A.P.A.) et l'Office de la protection du consommateur (O.P.C.) peuvent, par leurs actions, influencer le marché.

3.1.1. La Loi sur la protection du consommateur

La Loi sur la protection du consommateur (Québec, 1^e novembre 1980) consacre une section complète (la section IV) aux contrats relatifs aux automobiles et aux motocyclettes, dont une sous-section traite plus spécifiquement de la réparation d'automobile et de motocyclette. Cette sous-section, représentée par les articles 167 à 181,¹ constitue une partie importante de la réglementation du marché

¹ Ces articles de la Loi sur la protection du consommateur sont reproduits en entier en annexe (annexe juridique) ainsi que leurs règlements d'application.

de la réparation automobile. Nous pouvons résumer ainsi les articles 167 à 181 de la Loi sur la protection du consommateur:

- Le commerçant doit, pour une réparation de plus de 50\$:
 - fournir une évaluation écrite;
 - fournir une facture détaillée de la réparation effectuée;
 - remettre les pièces remplacées au consommateur s'il l'a exigé au moment où il a demandé de faire la réparation.
- Le commerçant doit faire connaître au consommateur ses droits (affichage) et le tarif horaire de la main d'oeuvre.
- Une réparation d'automobile est garantie pour 3 mois ou 5,000 kilomètres selon le premier terme atteint.

De plus, le marché de la réparation automobile est également assujéti aux articles de la section IV qui traite du louage de services à exécution successive. L'article 222 de cette section sanctionne la pratique interdite sur la qualité d'un bien. Les sous-sections c) et e) sont particulièrement intéressantes et se lisent comme suit:

"Aucun commerçant, manufacturier ou publicataire ne peut faussement, par quelque moyen que ce soit:

- c) prétendre qu'un bien ou un service a été fourni
- e) prétendre qu'un bien ou un service est nécessaire pour changer ou effectuer une réparation."

La principale contribution de la Loi sur la protection des consommateurs est d'obliger le producteur (appelé commerçant dans la Loi) à tarifier seulement pour les services rendus dont le consommateur avait préalablement accepté l'offre.

Le respect de cette condition est toujours implicite à tous les modèles présentés dans ce mémoire (à commencer par celui de Darby et Karni 1973). Le non-respect de cette hypothèse pourrait entraîner un type de fraude qui déborde la définition que nous lui avons donnée dans notre travail, la fraude se limitant à une offre de services qu'aurait refusée un consommateur-expert. Lorsque de tels services frauduleux sont produits, nous faisons toujours l'hypothèse que le consommateur en avait accepté les termes préalablement.

L'obligation pour le commerçant de fournir une évaluation écrite et d'afficher le taux horaire de la main d'oeuvre facilite le furetage du consommateur en lui permettant de comparer plus facilement différentes offres de services. En améliorant l'efficacité du furetage et en en diminuant le coût réel puisque l'information devient plus facilement accessible, le niveau d'offre de services frauduleux optimal diminue puisque la probabilité que le consommateur refuse le service et sa dérivée ($F(Q)$ et F'_Q) augmentent. De plus, le coût d'une telle mesure est négligeable.

L'article 222e interdit l'offre de services dont la relation technique n'est pas positive. Le service doit améliorer l'état du véhicule en étant techniquement utile. Le respect de cet article limiterait considérablement la fraude puisque tous les services dont l'utilité technique n'est pas positive peuvent être considérés comme frauduleux; mais ne l'éliminerait pas complètement. Nous avons vu, en

effet, que certains services pouvaient être techniquement utiles mais économiquement inutiles et par conséquent frauduleux.

De plus, dû en particulier au problème d'asymétrie d'information, un tel article de loi demeure très difficile d'application. A ce niveau, l'Office de la protection du consommateur joue un rôle important en particulier grâce à l'opération auto-témoin. L'opération consiste à identifier les garagistes qui font l'objet de plaintes de la part des consommateurs et dont l'honnêteté est déjà sérieusement mise en doute. Par la suite, un employé de l'OPC conduit un véhicule défectueux préalablement marqué (pour fins de preuve) chez ces garagistes. Enfin, les garagistes dont les services sont jugés malhonnêtes (environ la moitié des visites mais n'oublions pas que l'échantillon n'est pas représentatif) sont accusés d'avoir enfreint au moins l'article 222 e) ou 222 c) de la Loi sur la protection du consommateur, en ayant prétendu, à tort, qu'un bien ou un service avait été nécessaire pour effectuer une réparation ou encore qu'un bien ou service avait été fourni. L'OPC en rendant l'opération permanente et en rendant les résultats publics contribue à opérationnaliser la Loi. La menace devient plus crédible et les coûts d'offrir des services inutiles augmentent. La fraude devrait diminuer.

L'Association pour la protection des automobilistes utilise également la Loi sur la protection du consommateur pour poursuivre devant les tribunaux certains garagistes. En se substituant à certains consommateurs mal informés et en publiant de l'information (par exemple, publication d'une liste de garages recommandés), l'APA contribue également à améliorer l'efficacité de la loi en rendant la

menace plus crédible. De plus, l'introduction des cours de petites créances, en diminuant le coût des poursuites judiciaires, tend également à faire diminuer la fraude.

3.1.2. Décret sur les salariés de garages de la région de Montréal

Les salariés des garages de la région de Montréal sont soumis au décret 184 (février 1950) qui a été modifié par le Décret modifiant le décret relatif aux employés de garages dans l'île de Montréal en vertu de la Loi sur les décrets de convention collective (L.R.Q., c. D-2, d.8) et du décret 1693-82. Ces articles de loi ont été adoptés le 7 juillet 1982 et sont entrés en vigueur le 28 juillet 1982, date de publication finale dans la Gazette officielle du Québec.

Cette réglementation est régie par le comité paritaire de l'industrie de l'automobile de Montréal et du district. Les faits saillants des décrets sont de trois ordres:

- Imposition de prix planchers pour les employés (salaires minimums);
- Classification des travailleurs en deux catégories: compagnon (1^e, 2^e ou 3^e classe) et apprenti (1^e semestre, 2^e semestre, 2^e année ou 3^e année);
- Réglementation du ratio apprenti/compagnon. Pour chaque métier dans lequel un employeur utilise les services de compagnons de première ou de deuxième classe, l'employeur a le droit d'accepter un maximum de deux apprentis par trois compagnons de première ou deuxième classe. S'il n'y a qu'un ou deux compagnons de première ou de deuxième classe dans un métier, il a le droit d'embaucher un apprenti dans ce métier. Les apprentis travaillent aux mêmes heures et dans le même édifice que les compagnons.

Alors qu'il semblerait que la première contrainte soit non effective (les salaires offerts par le marché (à Montréal) sont plus élevés que les prix planchers) la deuxième serait plus contraignante

pour plusieurs garagistes. La réglementation du ratio apprenti/compagnon a comme effet de garantir au consommateur une compétence minimale. Les problèmes reliés à la nature aléatoire de la fonction de production en sont moins importants. Par ailleurs, une telle contrainte, lorsqu'elle est effective, n'est généralement pas sans coût. Il s'agit, en quelque sorte, d'une barrière à l'entrée déguisée.

Il est possible qu'il soit plus efficace de ne pas réglementer le ratio apprenti/compagnon mais plutôt d'obliger les travailleurs à identifier dans quelle classe ils se situent. Ainsi, le consommateur aurait une information précieuse à un coût faible.

Enfin, nous notons qu'il s'agit de droits acquis. Il n'y a pas d'examen ou de test périodique à subir ou encore d'obligations en termes de temps de travail pour s'assurer que le travailleur se "recycle". Evidemment, de telles mesures ne seraient pas sans coûts.

3.1.3. Réglementation et offre de services frauduleux en production jointe

Une étude exhaustive de la réglementation sur le marché de la réparation automobile nous apprend qu'il y a peu de mécanismes en place pour contrer l'offre de services frauduleux en production jointe.

D'une part, le décret des salariés de garages de la région de Montréal n'aborde pas ce type de considération que ce soit de façon directe ou indirecte. D'autre part, la Loi sur la protection du consommateur, si elle n'aborde pas la question directement, peut influencer le niveau de fraude dû aux productions jointes.

L'article 222 e) de la Loi qui interdit à un producteur de prétendre, faussement, qu'un service ou une pièce est nécessaire pour effectuer une réparation s'applique également, s'il y a lieu, aux productions jointes.

La Loi sur la protection du consommateur (article 173) oblige le garagiste à respecter un certain type de tarification. Celui-ci doit, afin de respecter la Loi, remettre une facture en deux composantes: la main d'oeuvre (nombre d'heures multiplié par le taux horaire) et le prix des pièces. La Loi confirme notre perception du fonctionnement du marché et la pertinence d'utiliser un modèle séquentiel d'offre de service tel que développé au chapitre II. Notre modèle avec productions jointes appliqué à un tel marché prédit qu'il peut y avoir offre de services frauduleux même si l'hypothèse d'excès de capacité de l'offre n'est pas respectée. Le client ayant déjà accepté un premier service, le producteur sera incité à offrir un deuxième service qui est frauduleux afin de bénéficier des économies d'envergure. En effet, le premier service ayant été accepté, il devient moins coûteux de produire le second et le profit que le producteur réalise sur la vente des pièces est le même quel que soit le temps requis pour installer la pièce. Le système de tarification tel qu'instauré par la loi sur la protection du consommateur est donc non optimal puisqu'il incorpore un incitatif financier.²

L'article 173 a également pour effet de rendre illégal le travail à forfait. Dans une telle politique de "prix tout compris", les agents acceptent une convention par laquelle il est stipulé un prix

² A moins que le profit sur la vente des pièces soit nul.

fixé par avance d'une manière invariable. Un garagiste pourrait, dans un tel système, offrir et publiciser grandement un forfait très peu élevé afin d'attirer la clientèle dans son garage pour ensuite "au besoin", lui offrir d'autres services - utiles ou non selon les cas - afin de rentabiliser la visite. Nous pouvons prévoir que dans un tel système la fraude due aux productions jointes augmenterait. En effet, les garagistes les moins honnêtes seraient en mesure d'offrir les forfaits les plus bas puisque, utilisant pleinement leur capacité de frauder, ils pourraient ensuite rentabiliser un plus grand nombre de visites en offrant des services supplémentaires frauduleux conjointement. Un plus grand nombre de consommateurs seraient alors orientés vers les garagistes "à éviter". Dans un tel contexte, ce marché serait inefficace et certains garagistes pourraient même être contraints à offrir des services frauduleux sous peine de se faire "sortir du marché".

Enfin, il ne faut pas confondre un tel scénario avec le phénomène de la spécialisation de certains garages dans le but de diminuer les coûts de production. Ce phénomène est particulièrement important pour les réparations en guise d'entretien (usure normale de certaines pièces) de l'automobile où la fonction de production n'a rien d'aléatoire. A titre d'exemple, nous pouvons citer la pose de systèmes d'échappement et de silencieux.

De tel spécialistes effectuent relativement peu de productions jointes et le désir de tarifier à forfait peut être motivé par la plus grande facilité qu'offre un tel contexte afin de publiciser leurs offres. Dans ce cas, le travail à forfait peut être doublement intéressant puisqu'en facilitant le furetage des consommateurs, il entraîne des économies d'échelle.

Notre modèle d'offre de services frauduleux en situation de production jointe (modifié afin de s'appliquer à la structure de tarification québécoise (section 2.3.3.2)) justifie l'offre de services frauduleux lorsque le deuxième service est accompagné de la vente d'une pièce. Or, la Loi sur la protection du consommateur permet au client de se faire remettre la pièce changée sur demande (article 175). Le client qui demande de se faire remettre toutes les pièces défectueuses a alors moins de chance (lire plus faible probabilité) de se faire offrir des services inutiles. D'une part le producteur peut interpréter ce geste comme étant une méfiance de la part du client et considérer que la probabilité cumulative que le client refuse le service ou qu'il ne revienne pas dans le futur est plus élevée ainsi que leurs variations en fonction de la quantité offerte.

D'autre part, le risque de poursuite devant les tribunaux en cas d'activité frauduleuse découverte (en vertu de l'article 222 e) de la Loi) est plus crédible puisque le consommateur a en main un élément essentiel à la preuve.³

Notons que le garagiste peut refuser de remettre la pièce si celle-ci est échangée contre une pièce réusinée ou remise à neuf (article 175 b). En général, dans de tel cas, le consommateur peut racheter sa pièce à un coût relativement bas. Le producteur, face à un individu ayant adopté un tel comportement, peut anticiper que le coût de la fraude dans ce cas est relativement élevé.

³ La probabilité de perdre un éventuel procès est également plus forte.

Le consommateur averti, face à un producteur qui refuserait de vendre la pièce usagée à un prix raisonnable, pourrait alors utiliser cette information et refuser le service offert ou ne pas revenir dans le futur si le coût du furetage est trop élevé.

Nous croyons donc que l'article 175 de la Loi sur la protection du consommateur régissant la remise de la pièce usagée a pour effet de limiter l'offre de services frauduleux tel que définie dans ce travail. Cet effet doit être particulièrement important pour les services qui font l'objet de notre extension du modèle de base puisque ce sont ceux-ci qui se prêtent le mieux à la caractéristique (importance de la pièce) des services frauduleux visés par cet article.

Le moyen qui serait le plus efficace pour enrayer la fraude due aux économies d'envergure serait de séparer le marché de la réparation automobile du marché de la vente des pièces. Ainsi, l'incitation financière fournie par cette dernière activité serait éliminée. Mais, une telle mesure entraînerait des coûts de transaction qui pourraient s'avérer prohibitifs.

3.2. La réglementation sur le marché des services médicaux

Au Québec, le marché des services médicaux est très réglementé. Les principales caractéristiques de l'organisation des services de santé sont:⁴

⁴ Les informations sont tirés de: Contandriopoulos (1980), notes de cours ECN6640 (1983) et la Loi sur les services de santé et les services sociaux (Loi 27: 1980)

- La médecine est une profession exclusive;
- Les libertés professionnelles des médecins (libertés thérapeutiques, du choix du lieu et du mode de pratique et liberté du choix de la spécialité) sont garanties (avec certains incitatifs par contre);
- Les médecins sont en général payés à l'acte et ils assument les coûts de pratique privée;
- Liberté du médecin de choisir son patient;
- Système général d'assurance-maladie avec tiers payant (pleine assurance avec carte blanche);
- Système hospitalier public et gratuit (financement par fonds publics);
- Droit de toute personne de recevoir des services de santé et des services sociaux adéquats sur les plans à la fois scientifique, humain et social, avec continuité et de façon personnalisée, compte tenu de l'organisation des ressources des établissements qui dispensent ces services (Loi 27; article 4).

Dans le cadre de ce travail, nous n'analysons pas en profondeur toute la réglementation existant sur le marché des services médicaux et ses effets. Nous nous limitons à étudier les mécanismes en place pour contrer ou du moins limiter l'offre de services frauduleux.

De plus, nous considérons deux textes. Dans le premier, l'auteur (Contandriopoulos, 1980) étudie l'influence des considérations économiques dans les décisions de diagnostics et thérapeutiques des médecins. Le deuxième auteur (Boutin, 1980) illustre d'une façon globale l'influence des stimulants économiques sur le comportement des médecins.

Au début de ce travail, nous avons discuté brièvement d'un problème relié à une couverture d'assurance comme celle que l'on retrouve au Québec. En effet, puisque le consommateur fait face à un

prix monétaire nul, il est rationnel pour celui-ci d'accepter tous les services utiles (l'utilité qu'il retire de la consommation doit seulement être supérieure à la désutilité de consommation).⁵ Or, pour la société, un tel niveau de consommation est trop élevé puisqu'une partie de celle-ci est abusive due au problème de risque moral.

Dans l'optique de ce travail, puisque le consommateur se réfère à un prix monétaire égal à zéro, toute réglementation qui enraierait l'offre de services inutiles éliminerait l'offre de services frauduleux. D'un point de vue social, il pourrait quand même y avoir abus. Enfin, si un service nécessaire est défini sans égard à l'utilité de l'assuré (techniquement ou médicalement utile: $\partial HS/\partial H > 0$), il peut y avoir offre de services inutiles et frauduleux même si l'assuré ne paie pas le service due à la désutilité de consommation ($\partial u/\partial H|_{\overline{HS}} < 0$).

Ces nuances au sujet de la notion d'utilité et de fraude dans un marché pleinement assuré nous permettent de justifier quelques articles du code de déontologie des médecins (1982) qui renferme, en quelque sorte, les règlements d'applications de la Loi médicale (1982). Ainsi, il y est écrit (article 2.03.21) que "le médecin ne doit fournir un soin ou donner une ordonnance de médicaments ou de traitements que si ceux-ci sont médicalement nécessaires." Nous pouvons considérer que cet article signifie au médecin qu'il doit offrir seulement des services qui vont entraîner une amélioration de l'état de santé du patient ($\partial HS/\partial H > 0$).

⁵ Nous ne tenons pas compte du prix du temps de consommation.

consommateur ($\partial u/\partial H$), il faut s'en remettre à l'article 2.03.23 du même code qui dit que "le médecin doit refuser sa collaboration ou sa participation à tout acte médical qui irait à l'encontre de l'intérêt du patient".

Le respect de cette condition nécessite beaucoup plus d'information que la précédente qui ne nécessitait que de l'information d'ordre technique. Le respect de l'article 2.03.23 demande de l'information très personnelle afin que le médecin s'assure de ne pas offrir de services économiquement inutiles. Une relation de confiance entre offreur et demandeur devient alors nécessaire. L'article 2.03.09 le confirme: "le médecin doit chercher à établir une relation de confiance mutuelle entre lui-même et son patient et s'abstenir d'exercer sa profession d'une façon impersonnelle."

La régie de l'assurance maladie, tant qu'à elle, va influencer l'offre de services inutiles de façon indirecte. En effet, la régie considère le profil moyen de chaque spécialité ce qui lui permet d'édicter des normes. Lorsque les revendications d'un médecin ne sont pas dans les normes, la régie forme un comité appelé à porter un jugement sur le profil de pratique de ce médecin. Cette pratique, sans éliminer toute fraude, peut néanmoins permettre de limiter les cas extrêmes.

La Loi sur les services de santé et les services sociaux rend le chef de département clinique responsable des services offerts par les médecins dans leurs départements en édictant les normes de la bonne pratique. En effet, l'article 71.1 de la Loi dit que: "Sous l'autorité du directeur des services professionnels du centre hospitalier, le chef de département clinique:

- 1) coordonne, sous réserve de l'article 112, les activités professionnelles des médecins et des dentistes dans son département et, dans la mesure prévue par le plan d'organisation du centre hospitalier, gère les ressources de son département;
- 2) élabore, pour ces départements, des règles d'utilisation des ressources du centre hospitalier; ces règles peuvent prévoir des sanctions administratives pour, notamment, limiter ou suspendre le droit d'un médecin ou d'un dentiste d'utiliser les ressources du centre hospitalier".

Un tel système, afin d'être opérationnel pour enrayer les services inutiles qui pourraient se produire devrait prévoir un important mécanisme d'acquisition d'information par le chef de département. De plus, il est possible que l'autorité du chef de département (qui est un confrère de travail) soit déficiente. Le système, tel qu'établit présentement, permettrait seulement de limiter les cas d'exagération trop flagrants.

3.2.1. Les incitatifs financiers: (Contandriopoulos 1980, Boutin 1980)

Notre cadre théorique le prédit; le profil de pratique des médecins du Québec peut être influencé par les incitatifs financiers incorporés dans la structure tarifaire. C'est la vérification de cette prédiction qui sera l'objectif de l'article de Contandriopoulos (1980). Pour ce faire, l'auteur compare les profils de pratique des médecins québécois en 1971 et 1973. L'année 1971 coïncidant avec le

début du régime d'assurance-maladie, il est permis de croire que les médecins n'avaient pas encore adapté leur pratique aux incitatifs financiers incorporés dans les nouveaux tarifs de remboursement.

L'auteur se devait de mesurer l'évolution du profil de pratique pour en faire ressortir le rôle plus ou moins important des stimulants économiques qu'il fallait également mesurer. Pour ce qui est de la mesure de l'évolution du profil de pratique, l'auteur a utilisé l'importance relative en pourcentage d'un acte par rapport à la somme totale qui était réclamée à la régie. La détermination du caractère plus ou moins lucratif des différents actes - nécessaires à la mesure des stimulants économiques - s'est effectuée de la façon suivante: division du tarif payé par la régie par la valeur relative de chaque acte (établie par la California Medical Association).

Les résultats sont présentés en trois étapes distinctes caractérisant le cheminement de l'étude. Dans un premier temps, l'auteur analyse l'évolution des actes réclamés par les médecins entre 1971 et 1973. Cette analyse indique "qu'il y a eu des modifications excessivement importantes dans la structure des actes réclamés par les médecins dans les trois premières années du régime d'assurance-maladie".⁶

Dans un second temps, afin de mieux cerner cette évolution et de l'expliquer, l'utilisation d'un modèle de régression où la variable dépendante est un indice d'évolution était requise. Les variables explicatives sont les caractéristiques du médecin et de sa pratique. L'interprétation des résultats nous indique "que ce sont les médecins

⁶ Contandriopoulos, (1980), p. 279

dont le revenu en 1973 est le plus élevé qui ont le plus modifié leur pratique... et plus le revenu était élevé en 1971 moins les médecins ont été incité à modifier leur profil de pratique".⁷

Outre la différence dans le sens de l'influence des variables revenus (revenu 71 négatif et revenu 73 positif), le fait saillant de cette seconde étape est la faible valeur explicative du modèle économétrique utilisé (R^2 légèrement inférieur à 10%). Or, étant donné que les caractéristiques des demandeurs de soins furent très stables de 1971 à 1973, il faut regarder du côté des caractéristiques du système de soins et tout particulièrement dans les incitations financières incorporées dans la structure tarifaire pour mieux expliquer l'évolution du profil de pratique des médecins. C'est ce que fera Contandriopoulos (1980) à la troisième étape.

D'une première analyse sommaire des différentes spécialités, "il se dégage que les médecins peuvent se subdiviser en deux grands groupes en ce qui concerne l'évolution du profil de pratique entre 1971 et 1973".⁸ A savoir:

- un premier groupe qui a favorisé les actes rémunérateurs au détriment d'autres qui l'étaient moins (médecine, pédiatrie, psychiatrie, médecine interne, chirurgie, gynéco-obstétrique et la médecine générale);
- un deuxième groupe qui a simplement accru de façon importante le volume des actes posés même si le tarif moyen de ces actes devait diminuer (radiologues, physiatres et rhumatologues et les anesthésistes).

⁷ Contandriopoulos, (1980), p. 283.

⁸ Contandriopoulos, (1980), p. 286.

Par la suite, le calcul pour chaque spécialité du coefficient de corrélation simple entre la variation dans l'importance relative d'un acte et le caractère plus ou moins lucratif de celui-ci révèle des relations significativement positives (seuil à 95%) pour les omnipraticiens, les pédiatres et les médecins des spécialités médicales. Pour ces trois catégories, l'hypothèse est confirmée. Enfin, l'étude d'actes particuliers est intéressante.

L'un des facteurs étudiés est l'effet du personnel para-médical sur l'offre de services. Certains actes peuvent être posés entièrement ou partiellement par du personnel auxiliaire. L'auteur, Contandriopoulos, conclut ainsi sur le sujet.

"quand la structure tarifaire prévoit la rémunération de la composante technique d'un acte (radiologie, physiothérapie), les médecins tendent à exploiter très énergiquement cette possibilité en déléguant à des techniciens le soin de produire des actes qu'ils supervisent".⁹

Notre analyse du lien qui existe entre l'offre de services frauduleux et l'hypothèse d'excès de capacité de l'offre nous permet d'expliquer ce résultat. En effet, l'emploi du personnel para-médical par le médecin n'est pas sujet à la contrainte de pleine capacité de façon aussi importante que le médecin. La courbe représentant le revenu marginal de la fraude n'est donc pas cassée et si la situation exprimée par la figure 2.5 prévaut pour l'emploi de ce personnel,

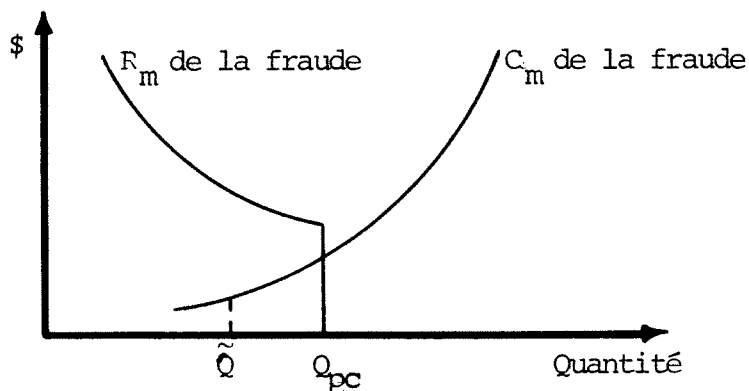
⁹ Contandriopoulos, (1980), p. 290.

l'offre de services frauduleux pourra augmenter.¹⁰ Puisque, en pratique, le médecin doit consacrer une partie de son temps à gérer ces employés, il y a une réallocation des ressources du médecin. La contrainte de temps est beaucoup moins importante dans ce cas puisqu'il peut superviser plusieurs personnes.

Enfin, Contandriopoulos (1980) s'intéresse à la tarification effective lorsque le patient consomme plusieurs (au moins deux) services simultanément. Selon l'auteur, l'un des facteurs pour expliquer l'évolution observée dans la fréquence de certains actes serait le rendement de la visite. En effet,

"dans ce cas, le médecin posera des actes diagnostiques et thérapeutiques additionnels au cours de la visite, actes dont la justification médicale n'est pas toujours évidente...même s'ils sont relativement peu rémunérateurs, le coût marginal de les poser pour le médecin est minime à l'occasion de la visite. Le cas le plus classique serait l'injection de substances sclérosantes dans les veines des membres inférieurs. Chez les omnipraticiens, cet acte s'est accru de 202% en importance relative, soit de 243,000 à 811,000; chez les

10 Reprenons la figure 2.5 et son analyse:



Il y a offre de services frauduleux de \tilde{Q} à Q_{pc} . Les carnets de rendez-vous sont complets et de nouveaux producteurs pourraient produire à pleine capacité. Le niveau de fraude augmenterait simplement.

chirurgiens il s'est accru de 77%, en gynico-obstétrique de 1,385% (2,000 à 23,000) et chez les anesthésistes de 1 000% (23 000 à 333 000)".¹¹

Dans le même ordre d'idée, l'auteur conclut que "la capacité des médecins de rentabiliser un contact avec un patient dépend de l'arsenal des actes mineurs qu'il peut ajouter à la consultation".¹²

Le lien entre cette conclusion de Contandriopoulos (1980) et le modèle séquentiel en situation de production jointe que nous avons développé est direct. Lorsque l'auteur parle d'actes mineurs que le médecin peut ajouter à la consultation, nous sommes alors dans un contexte de production jointe. De plus, lorsqu'il constate que le coût marginal de poser ces actes complémentaires pour le médecin est minime à l'occasion de la visite, l'hypothèse de la présence d'économies d'envergure est vérifiée.

Enfin, sachant que le prix que doit payer le consommateur pour ces actes est invariable¹³, le doute que transmet Contandriopoulos (1980) sur la justification médicale de ces actes confirme le fait que nous nous trouvons bien dans un contexte d'asymétrie d'information où le producteur tente de tirer profit d'une situation privilégiée en

11 Contandriopoulos, (1980), p. 287.

12 Contandriopoulos, (1980), p. 290.

13 Nous sommes dans un contexte de pleine assurance avec carte blanche. Le consommateur n'est pas affecté directement par une variation dans la structure tarifaire. Cette situation due à l'assurance généralisée peut faire en sorte que le patient accepte l'offre de services supplémentaires. Le coût supplémentaire en temps de consommation est généralement faible pour ces services consommés à l'occasion d'une visite.

offrant des services inutiles. L'étude de Contandriopoulos (1980) contribue à justifier un modèle avec productions jointes dans un contexte d'asymétrie d'information pour expliquer et prédire l'allocation des ressources à l'intérieur de la profession médicale. En effet, les économies d'envergure engendrent des incitatifs financiers face auxquels les producteurs réagissent en modifiant leur profil de pratique, le système de tarification n'étant pas neutre.

L'étude de Boutin (1980), tant qu'à elle, traite spécifiquement de la dispensation des actes complémentaires à l'examen. L'auteur utilise les données de la régie de l'assurance maladie du Québec. Ces données, reproduites aux tableaux 3.1, 3.2 et 3.3 permettent à l'auteur d'effectuer trois types d'analyses.

Dans un premier temps, il analyse l'évolution des indicateurs en fonction des classes de paiements chez les médecins ayant exercé en 1975 en utilisant les données figurant dans la 1^e colonne (pour chacun des indicateurs). Il se dégage de cette analyse que "le coût moyen par contact-patient augmente au fur et à mesure qu'on monte dans les classes de paiement et ceci est essentiellement dû à la dispensation plus fréquente d'actes complémentaires à l'examen".¹⁴ Cette relation positive entre le niveau de productions jointes et le revenu explique un coût moyen plus élevé par contact-patient et une proportion moindre du coût du contact-patient provenant de l'activité principale (les examens) pour les omnipraticiens à plus haut revenu.

¹⁴ Boutin, (1980), p. 257.

TABLEAU 3.1

COÛT MOYEN DES CONTACTS-PATIENTS ET COÛT MOYEN DES EXAMENS
SELON LES GROUPES D'OMNIPRATICIENS ÉTUDIÉS ET LA CLASSE DE PAIEMENT

Classe de paiement	Coût moyen des contacts-patients			Coût moyen des examens		
	Omnipraticiens 1975		Omnipraticiens 1976	Omnipraticiens 1975		Omnipraticiens 1976
	\$ (1975)	Δ %	Δ %	\$ (1975)	Δ %	Δ %
\$1-19 999	8.91	15,0	23,4	7.07	12,9	25,7
20 000-39 999	8.93	7,3	19,8	7.00	6,6	24,8
40 000-59 999	9.02	4,9	16,4	6.95	4,7	22,5
60 000-79 999	9.33	4,8	13,0	6.93	5,3	21,6
80 000-99 999	9.42	3,1	9,5	6.86	3,8	19,9
100 000 et plus	10.95	1,9	-7,1	7.03	2,8	12,6
Ensemble	9.31	5,3	13,1	6.95	5,8	22,0

SOURCE: Données produites par la Régie de l'assurance maladie du Québec, Service de la Recherche et des Statistiques et reproduites par Boutin (1980).

TABLEAU 3.2

COÛT DES EXAMENS EN PROPORTION DU COÛT DES CONTACTS-PATIENTS,
 NOMBRE D'AUTRES ACTES PAR CONTACT-PATIENT ET POURCENTAGE DE VARIATION SELON LES
 GROUPES D'OMNIPRATICIENS ÉTUDIÉS ET LA CLASSE DE PAIEMENT

Classe de paiement	Coût des examens en proportion du coût des contacts-patients			Nombre d'autres actes par contact-patient*		
	Omnipraticiens 1975		Omnipraticiens 1976	Omnipraticiens 1975		Omnipraticiens 1976
	N (1975)	Δ %	Δ %	N (1975)**	Δ %	Δ %
\$1-19 999	.793	-1,8	1,9	.192	25,0	26,1
20 000-39 999	.784	-0,6	4,1	.177	17,5	-0,9
40 000-59 999	.771	-0,1	5,3	.218	8,7	-3,2
60 000-79 999	.743	-0,4	7,6	.266	6,8	-3,1
80 000-99 999	.728	0,7	9,5	.398	3,8	-8,6
100 000 et plus	.642	0,9	21,1	.735	-2,6	-32,1
Ensemble	.747	0,4	7,9	.294	3,7	-11,7

* Excluant les actes éliminés dans la nouvelle entente.

** Pour fins de comparabilité, les données de 1975 ont été ajustées pour tenir compte de la modification intervenue en 1976 dans le mode de dénombrement de certains actes de chirurgie mineure.

SOURCE: Données produites par la Régie de l'assurance-maladie du Québec, Service de la Recherche et des Statistiques et reproduites par Boutin (1980).

TABLEAU 3.3

COÛT MOYEN DES AUTRES ACTES ET POURCENTAGE DE VARIATION SELON LES GROUPES
D'OMNIPRATICIENS ÉTUDIÉS ET LA CLASSE DE PAIEMENT

Classe de paiement	Coût moyen des autres actes*		
	Omnipraticiens 1975		Omnipraticiens 1976
	\$ (1975)**	Δ %	Δ %
\$1-19 999	8.38	-0,2	4,0
20 000-39 999	9.75	-5,3	14,2
40 000-59 999	8.56	-3,5	8,5
60 000-79 999	8.22	-2,6	-0,4
80 000-99 999	5.78	-2,9	-2,1
100 000 et plus	4.89	3,9	-10,3
Ensemble	7.27	0,0	8,0

* Excluant les actes éliminés dans la nouvelle entente.

** Pour fins de comparabilité, les données de 1975 ont été ajustées pour tenir compte de la modification intervenue en 1976 dans le mode de dénombrement de certains actes de chirurgie mineure.

SOURCE: Données produites par la Régie de l'assurance-maladie du Québec, Service de la Recherche et des Statistiques et reproduites par Boutin (1980).

De plus, le recours plus important aux actes complémentaires s'effectue par la dispensation d'actes moins coûteux. En effet, le coût moyen des autres actes décroît avec la classe de revenu passant de 8.38\$ à 4.89\$ selon la classe.

Cette première analyse confirme donc l'une des conclusions de Contandriopoulos (1980). Le médecin peut améliorer la rentabilité d'une visite en rajoutant des actes mineurs à la consultation. La tarification était telle, en 1975, que les économies d'envergure que le médecin récoltait en effectuant des productions jointes lui permettaient d'améliorer sa situation financière.

Par la suite, Boutin (1980) étudie la variation dans le profil de pratique entre 1975 et 1976; période pendant laquelle il n'y a eu aucune variation de tarif. Ces variations sont présentées dans la deuxième colonne des tableaux 3.1 et 3.2. Ainsi, les diverses composantes du coût du contact patient augmentent de 1975 à 1976 mais ces augmentations sont inversement reliées au revenu du médecin.

"En effet, de la même manière qu'il s'avère de plus en plus difficile d'accroître le nombre des contacts-patients au fur et à mesure que celui-ci s'élève, de la même manière est-il difficile d'accroître le coût moyen par contact-patient au fur et à mesure qu'on y atteint un niveau élevé."¹⁵

Dans cette optique, rappelons que le modèle d'offre de services frauduleux prédisait que le coût marginal de la fraude était une fonction croissante par rapport à la quantité offerte tandis que le revenu marginal de la fraude diminuait lorsque la quantité de services offerts augmentait (voir la figure 2.1).

¹⁵ Boutin, (1980), p. 261.

Enfin, Boutin (1980) s'intéresse à l'effet de la deuxième entente sur le recours aux actes complémentaires. Rappelons qu'en 1976, une nouvelle entente est intervenue entre le ministère des Affaires Sociales et les fédérations des médecins. Cette entente amène - entre autres - plusieurs changements importants d'ordre financier. Les hausses de tarif contenues dans la deuxième entente n'étant pas uniformes pour tous les types d'examens et d'actes, les prix relatifs de ceux-ci ont variés. De plus, 26 actes diagnostiques et thérapeutiques (ce qui représente environ le tiers des actes complémentaires) effectués lors d'un examen et qui étaient rémunérés distinctement dans les ententes précédentes sont maintenant inclus dans le tarif de l'examen. L'un des objectifs de ces mesures était de restreindre le paiement d'actes complémentaires. Afin d'analyser l'effet de la nouvelle entente sur le recours aux autres actes, Boutin (1980) utilise les 3^e colonne des tableaux 3.1 et 3.2 qui fournissent l'évolution des indicateurs entre 1976 et 1977. Cette analyse

"a permis de constater une réduction importante des écarts entre les classes de paiements, et ceci, en grande partie parce que la nouvelle entente produisait un abaissement notable de la dispensation d'actes complémentaires à l'examen chez les médecins des classes de paiement les plus élevées, soit ceux qui justement en dispensaient le plus."¹⁶

En restreignant les possibilités de productions jointes, la nouvelle entente a eu pour résultat d'amoindrir la différence entre les classes de revenus quant au coût moyen par contact-patient, le coût des examens et la proportion du coût des contacts-patients provenant des examens. Ainsi, par exemple, alors que le coût moyen des visites au

¹⁶ Boutin, (1980), p. 263.

médecin continuait d'augmenter (de 13.1% en moyenne) en général, cette augmentation a été de 23.4% pour la classe de paiement à plus faible revenu tandis que la classe de paiement la plus élevée a vu ce coût moyen diminuer de 7.1% pendant la période. Il est à noter que la tendance à la standardisation des modes de pratique avait été amorcée avant la mise en application de la deuxième entente quoiqu'il semblerait que celle-ci ait contribué à accélérer le processus.

L'étude de Boutin (1980) confirme donc celle de Contandriopoulos (1980) en ce qui a trait à l'importance des productions jointes afin de rentabiliser une visite. Dans le contexte d'asymétrie d'information qui existe sur le marché des services médicaux, puisque l'existence des économies d'envergure est confirmée, le modèle séquentiel d'offre de services frauduleux avec productions jointes représente bien le comportement de ceux qui offrent des services inutiles dans l'unique but de maintenir ou augmenter leurs revenus répondant ainsi positivement à l'incitation financière implicite à la tarification.

CONCLUSION

Dans la littérature, on distingue trois facteurs principaux susceptibles d'influencer les activités de création de demande, soit le nombre de producteurs per capita, le mode de rémunération et la couverture d'assurance. Dans ce mémoire, nous avons concentré notre analyse au premier des facteurs précités et ce pour plusieurs raisons. Premièrement, il s'agit du principal facteur explicatif retenu par les auteurs intéressés à ce problème et surtout, il s'agit du facteur le plus universel. De plus, une augmentation du nombre d'offreurs (per capita) peut créer une offre excédentaire. Or, l'objectif général de ce travail est l'étude de l'importance de l'hypothèse d'offre excédentaire pour justifier la fraude.

L'un des résultats les plus importants de ce travail concerne la notion d'offre excédentaire pour justifier la fraude dans un modèle à un bien. En effet, la simple observation que les carnets de rendez-vous ne sont pas vides ne signifie pas nécessairement qu'il n'y a pas de fraude. Il peut y avoir excès de capacité. Ce qui est pertinent en termes de notion d'excès de capacité pour justifier la fraude, ce sont des carnets de rendez-vous non-complets sans fraude. Les carnets de rendez-vous peuvent être remplis par des services frauduleux. Il faut parler d'offre excédentaire "ex post" à la fraude.

Un autre résultat important est la justification de la fraude sans excès de capacité si l'on considère que le producteur peut offrir plusieurs services (au moins deux) à un même client et que ces services sont reliés avantageusement par la fonction de coût. Par

contre, si la tarification est optimale au sens de Pareto, toutes les économies d'envergure ainsi produites seront refilées aux consommateurs sous forme de baisse de prix et l'incitation à la fraude n'existe plus.

Ce qui est pertinent pour le consommateur, c'est le prix ressenti qui peut incorporer le prix du temps de consommation et le prix en termes de désutilité de consommation. Lorsque le producteur offre des services en situation de production jointe, ces prix vont généralement diminuer de façon plus importante que le prix monétaire même si la tarification est optimale (indivisibilité et surtout, notion de convalescence). L'estimation par le producteur de la probabilité de refus devrait diminuer. Notre modèle prévoit que dans ce cas, le niveau de services frauduleux augmenterait.

L'introduction de l'assurance est également une étape importante de ce mémoire puisque souvent les services qui nous intéressent sont consommés sous l'assurance. L'assurance couvre généralement seulement les pertes monétaires ce qui fait que les prix non-monétaires (prix du temps de consommation et désutilité de consommation) prennent un poids relatif dans la décision du consommateur plus importante par rapport à une situation sans assurance. Or, nous savons que les productions jointes affectent particulièrement ces coûts ressentis par le consommateur (notion de convalescence). En conséquence, l'assurance aura encore plus d'impact sur l'offre de services en production jointe avec économies d'envergure que sur les services produits indépendamment.

Le consommateur moyen qui se présente sur le marché de la réparation des automobiles au Québec peut compter sur quelques institutions qui, heureusement, peuvent le guider dans sa démarche de furetage et de consommation quoique dans ce domaine, la vigilance et le comportement responsable du consommateur demeurent ses meilleures garanties. Dans le marché des services médicaux, malgré une réglementation très lourde, la même règle générale s'applique. Malgré les nombreuses difficultés théoriques que pose la vérification empirique, l'étude de Fuchs (1978) semble confirmer que le comportement des agents économiques dans ce marché ne cadre pas avec le modèle néoclassique. Les études de Boutin (1980) et Contandriopoulos (1980) confirment nos conclusions théoriques. Le profil de pratique des médecins du Québec est influencé par les incitatifs financiers incorporés dans la structure tarifaire et les production jointes permettent de mieux "rentabiliser une visite". Le modèle séquentiel d'offre de services frauduleux en situation de production jointe avec économies d'envergure représente bien le comportement de ceux qui offrent des services inutiles dans l'unique but de maintenir ou augmenter leur revenu répondant ainsi positivement à l'incitation financière implicite à la tarification.

Ce qui ne veut pas dire que le modèle ne pourrait pas être raffiné. Aux contraire, on peut considérer plusieurs avenues de recherche. On pourrait, à titre d'exemple, réinterpréter le modèle d'offre de services frauduleux en production jointe en reconsidérant la valeur présente des profits futurs ou encore en incorporant la notion de bien de réputation dans le modèle. Une augmentation de la fraude entraînerait une perte de profits futurs provenant d'autres consommateurs potentiels.

De plus, la notion d'incertitude pourrait être beaucoup plus développée en tenant compte de la riscophobie des agents économiques.

Dans notre modèle d'offre de services frauduleux avec productions jointes, nous faisons l'hypothèse que les offres sont séquentielles. Ainsi, lorsque le producteur offre le deuxième service, le premier a été accepté mais la production n'en est pas commencée. Cette hypothèse est importante. Si c'était l'offre et la production des services qui étaient séquentielles, le modèle serait fortement différent. Dans ce cas, même en parfaite information, le producteur ferait face à la demande de marché - qui est généralement beaucoup moins élastique - entre le prix P et αP . Il y aurait une situation de duopole entre le producteur et le client et le prix pour le deuxième service serait de $\alpha P < P_2 < P$. La tarification optimale ($P_2 = \alpha P$) serait un cas particulier et rien nous laisse croire qu'un marché libre nous donnerait ce cas. Au contraire, le producteur ayant une certaine influence sur les prix, le niveau de fraude devrait augmenter. La notion de profits futurs serait importante parce qu'elle déterminerait en grande partie le "pouvoir de négociation" du consommateur.

À la lumière de ce qui précède, nous pouvons mieux comprendre l'importance qu'accorde la Loi sur la protection du consommateur à l'estimé écrit obligatoire. Néanmoins, le producteur peut toujours changer son offre après avoir commencé la réparation. La fonction de production est aléatoire. Surtout lorsque la réparation (ou l'opération) n'est pas commencée.

Les modèles que nous avons développés caractérisent un producteur potentiellement frauduleux face à un consommateur non-expert. Il s'agit donc d'une relation purement micro-économique. Par contre, si on peut justifier des disparités entre les carnets de rendez-vous de différents producteurs, certains peuvent être en situation d'offre excédentaire tandis que d'autres producteurs imposeraient une file d'attente à leurs clients. Sous ces conditions, il pourrait y avoir fraude sur le marché même si la tarification était optimale; la condition d'offre excédentaire n'étant pas respectée.

Enfin, nous ne croyons pas que tous les producteurs (médecins ou garagistes dans notre exemple) sont des fraudeurs. Au contraire, les études semblent plutôt nous laisser croire que la majorité ne l'est pas. Il s'agit plutôt d'étudier le comportement de ceux qui le sont.

ANNEXES

ANNEXE MATHÉMATIQUE

Modèle I: Modèle à un bien

$$\text{Max } \pi = \left[1 - \frac{Q}{\bar{Q}}\right] [PQ - Q^2]$$

Condition de premier ordre:

$$\frac{\partial \pi}{\partial Q} = \left[1 - \frac{Q}{\bar{Q}}\right] [P - 2Q] - \frac{1}{\bar{Q}} [PQ - Q^2] = 0$$

$$\frac{\partial \pi}{\partial Q} = P - \frac{PQ}{\bar{Q}} - 2Q + \frac{2Q^2}{\bar{Q}} - \frac{PQ}{\bar{Q}} + \frac{Q^2}{\bar{Q}} = 0$$

$$\frac{\partial \pi}{\partial Q} = P - 2Q - \frac{2PQ}{\bar{Q}} + \frac{3Q^2}{\bar{Q}} = 0$$

$$\frac{\partial \pi}{\partial Q} = P - Q \left[2 + \frac{2P}{\bar{Q}}\right] + \frac{3Q^2}{\bar{Q}} = 0$$

$$\frac{\partial \pi}{\partial Q} = \frac{3}{\bar{Q}} Q^2 - \left[2 + \frac{2P}{\bar{Q}}\right] Q + P = 0$$

Utilisant la formule: $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, nous retrouvons les racines

de la forme: $ax^2 + bx + c = 0$. Donc,

$$Q^* = \frac{2 + \frac{2P}{\bar{Q}} \pm \sqrt{\left(2 + \frac{2P}{\bar{Q}}\right)^2 - 4 \cdot \frac{3}{\bar{Q}} P}}{6 / \bar{Q}}$$

$$Q^* = \frac{2 + \frac{2P}{\bar{Q}} \pm \sqrt{4 + \frac{8P}{\bar{Q}} + \frac{4P^2}{\bar{Q}^2} - \frac{12P}{\bar{Q}}}}{6 / \bar{Q}}$$

$$Q^* = \frac{2 + \frac{2P}{\bar{Q}} \pm \sqrt{4 + \frac{4P^2}{\bar{Q}^2} - \frac{4P}{\bar{Q}}}}{6/\bar{Q}} = \frac{2 + \frac{2P}{\bar{Q}} \pm \sqrt{\frac{4}{\bar{Q}^2}(\bar{Q}^2 + P^2 - P\bar{Q})}}{6/\bar{Q}}$$

$$Q^* = \frac{2 + \frac{2P}{\bar{Q}} \pm \frac{2}{\bar{Q}} \sqrt{\bar{Q}^2 + P^2 - P\bar{Q}}}{6/\bar{Q}} = \frac{\bar{Q}}{6} \left[2 + \frac{2P}{\bar{Q}} \pm \frac{2}{\bar{Q}} \sqrt{\bar{Q}^2 + P^2 - P\bar{Q}} \right]$$

$$Q^* = \frac{\bar{Q}}{3} + \frac{P}{3} \pm \sqrt{\bar{Q}^2 + P^2 - P\bar{Q}}$$

C.Q.F.D.

Modèle II: Modèle séquentiel avec productions jointes

Q_1 est accepté et le producteur doit évaluer quel niveau il va offrir de Q_2 . Le coefficient de production jointe est v ($0 < v < 1$).

$$\max_{\pi} = P_1 Q_1 - Q_1^2 + \left[1 - \frac{Q_2}{\bar{Q}_2} \right] [P_2 Q_2 - v Q_2^2]$$

Condition de premier ordre: (pour Q_2)

$$\frac{\partial \pi}{\partial Q_2} = \left[1 - \frac{Q_2}{\bar{Q}_2} \right] [P_2 - 2vQ_2] - \frac{1}{\bar{Q}_2} [P_2 Q_2 - vQ_2^2] = 0$$

$$\frac{\partial \pi}{\partial Q_2} = P_2 - \frac{P_2 Q_2}{\bar{Q}_2} - 2vQ_2 + \frac{2vQ_2^2}{\bar{Q}_2} - \frac{P_2 Q_2}{\bar{Q}_2} + \frac{vQ_2^2}{\bar{Q}_2} = 0$$

$$\frac{\partial \pi}{\partial Q_2} = P_2 - 2vQ_2 - \frac{2P_2 Q_2}{\bar{Q}_2} + \frac{3vQ_2^2}{\bar{Q}_2} = 0$$

$$\frac{\partial \pi}{\partial Q_2} = P_2 - \left[2v + \frac{2P_2}{\bar{Q}_2} \right] Q_2 + \frac{3v}{\bar{Q}_2} Q_2^2 = 0$$

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad : \rightarrow \quad Q_2^* = \frac{2v + \frac{2P_2}{\bar{Q}_2} \pm \sqrt{\left(2v + \frac{2P_2}{\bar{Q}_2} \right)^2 - 4 \frac{3v}{\bar{Q}_2} P_2}}{6v/\bar{Q}_2}$$

$$Q^* = \frac{2v + \frac{2P_2}{\bar{Q}_2} \pm \sqrt{4v^2 + \frac{8vP_2}{\bar{Q}_2} + \frac{4P_2^2}{\bar{Q}_2} - \frac{12vP_2}{\bar{Q}_2}}}{6v/\bar{Q}_2}$$

$$= \frac{2v + \frac{2P_2}{\bar{Q}_2} \pm \sqrt{\frac{4}{\bar{Q}_2^2} (v^2 Q_2^2 + P_2^2 - 4P_2 \bar{Q}_2)}}{6v/\bar{Q}_2}$$

$$= \frac{2v + \frac{2P_2}{\bar{Q}_2} \pm \frac{2}{\bar{Q}_2} \sqrt{v^2 \bar{Q}_2^2 + P_2^2 - vP_2 \bar{Q}_2}}{6v/\bar{Q}_2}$$

$$= \frac{\bar{Q}_2}{3} + \frac{P_2}{3v} \pm \frac{1}{3v} \sqrt{v^2 \bar{Q}_2^2 + P_2^2 - vP_2 \bar{Q}_2}$$

C.Q.F.D.

Modèle III: Tarification optimale

Soit α un coefficient de tarification. Le modèle séquentiel avec productions jointes devient :

$$\text{Max } \pi = P_1 Q_1 - Q_1^2 + \left[1 - \frac{Q_2}{\bar{Q}_2} \right] [\alpha P_2 Q_2 - v Q_2^2]$$

Condition de premier ordre: (pour Q_2)

$$\frac{\partial \pi}{\partial Q_2} = \left[1 - \frac{Q_2}{\bar{Q}_2}\right] [\alpha P_2 - 2vQ_2] - \frac{1}{\bar{Q}_2} [\alpha P_2 Q_2 - vQ_2^2] = 0$$

$$\frac{\partial \pi}{\partial Q_2} = \alpha P_2 - 2vQ_2 - \frac{\alpha P_2 Q_2}{\bar{Q}_2} + \frac{2vQ_2^2}{\bar{Q}_2} - \frac{\alpha P_2 Q_2}{\bar{Q}_2} + \frac{vQ_2^2}{\bar{Q}_2} = 0$$

$$\frac{\partial \pi}{\partial Q_2} = \alpha P_2 - 2vQ_2 - \frac{2\alpha P_2 Q_2}{\bar{Q}_2} + \frac{3vQ_2^2}{\bar{Q}_2} = 0$$

$$\frac{\partial \pi}{\partial Q_2} = \alpha P_2 - Q_2 \left[2v + \frac{2\alpha P_2}{\bar{Q}_2}\right] + 3 \frac{vQ_2^2}{\bar{Q}_2} = 0$$

$$\frac{\partial \pi}{\partial Q_2} = \frac{3v}{\bar{Q}_2} Q_2^2 - \left[2v + \frac{2\alpha P_2}{\bar{Q}_2}\right] Q + \alpha P_2 = 0$$

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} : \rightarrow Q^* = \frac{2v + \frac{2\alpha P_2}{\bar{Q}_2} \pm \sqrt{\left(2v + \frac{2\alpha P_2}{\bar{Q}_2}\right)^2 - 4 \frac{3v}{\bar{Q}_2} \cdot \alpha P_2}}{6v/\bar{Q}_2}$$

$$= \frac{2v + \frac{2\alpha P_2}{\bar{Q}_2} \pm \sqrt{4v^2 + \frac{8vP_2\alpha}{\bar{Q}_2} + \frac{4\alpha^2 P_2^2}{\bar{Q}_2^2} - \frac{12v\alpha P_2}{\bar{Q}_2}}}{6v/\bar{Q}_2}$$

$$= \frac{2v + \frac{2\alpha P_2}{\bar{Q}_2} \pm \sqrt{\frac{4}{\bar{Q}_2^2} [v^2 \bar{Q}_2^2 + \alpha^2 P_2^2 - vP_2 \bar{Q}_2]}}{6v/\bar{Q}_2}$$

$$= \frac{2v + \frac{2\alpha P_2}{\bar{Q}_2} \pm \frac{2}{\bar{Q}_2} \sqrt{v^2 \bar{Q}_2^2 + \alpha^2 P_2^2 - v P_2 \alpha \bar{Q}_2}}{6v/\bar{Q}_2}$$

$$Q^* = \frac{\bar{Q}_2}{3} + \frac{\alpha P_2}{3v} \pm \sqrt{v^2 \bar{Q}_2^2 + \alpha^2 P_2^2 - v P_2 \alpha \bar{Q}_2}$$

Si $\alpha = v$:

$$Q^* = \frac{\bar{Q}_2}{3} + \frac{P_2}{3} \pm \frac{1}{3v} \sqrt{v^2 \bar{Q}_2^2 + v^2 P_2^2 - v^2 P_2 \bar{Q}_2}$$

$$= \frac{\bar{Q}_2}{3} + \frac{P_2}{3} \pm \frac{1}{3v} \sqrt{v^2 (\bar{Q}_2^2 + P_2^2 - P_2 \bar{Q}_2)}$$

Donc,

$$Q^* = \frac{\bar{Q}_2}{3} + \frac{P_2}{3} \pm \frac{1}{3} \sqrt{\bar{Q}_2^2 + P_2^2 - P_2 \bar{Q}_2}$$

Ce qui est similaire à la fonction d'offre issue du modèle à un bien (modèle I).

C.Q.F.D.

Modèle IV: Le garagiste: prototype à un bien

Soit la fonction à maximiser:

$$\text{Max } \pi = \left[1 - \frac{Q}{\bar{Q}}\right] [(P + P_x)Q - (Q^2 + C \cdot Q)]$$

où: P est le prix de la main d'oeuvre pour produire Q (taux de salaire x nombre d'heures travaillées);

P_x est le prix de la pièce nécessaire à la production d'une unité de service;

C est le coût de la pièce (pour le producteur).

Condition de premier ordre:

$$\frac{\partial \pi}{\partial Q} = \left[1 - \frac{Q}{\bar{Q}}\right] [P + P_x - 2Q - C] - \frac{1}{\bar{Q}} [(P + P_x)Q - (Q^2 - C \cdot Q)] = 0$$

$$\begin{aligned} \frac{\partial \pi}{\partial Q} &= P + P_x - 2Q - C - \frac{PQ}{\bar{Q}} - \frac{P_x Q}{\bar{Q}} + \frac{2Q^2}{\bar{Q}} + \frac{QC}{\bar{Q}} - \frac{PQ}{\bar{Q}} - \frac{P_x Q}{\bar{Q}} \\ &\quad + \frac{Q^2}{\bar{Q}} + \frac{CQ}{\bar{Q}} = 0 \end{aligned}$$

$$\frac{\partial \pi}{\partial Q} = \frac{2PQ}{\bar{Q}} - \frac{2P_x Q}{\bar{Q}} + \frac{2QC}{\bar{Q}} - 2Q + P + P_x - C + \frac{3Q^2}{\bar{Q}} = 0$$

$$\frac{\partial \pi}{\partial Q} = \frac{3}{\bar{Q}} Q^2 - \left[\frac{2P + 2P_x - 2c}{\bar{Q}} + 2 \right] Q + P + P_x - C = 0$$

C.Q.F.D.

Modèle V: Le garagiste: Prototype séquentiel avec productions jointes

$$\text{Max } \pi = P_1 Q_1 + Q_1^2 + \left[1 - \frac{Q_2}{\bar{Q}_2}\right] [(vP_2 + P_x)Q_2 - (vQ_2^2 + CQ_2)]$$

Condition de premier ordre: (Pour Q_2)

$$\begin{aligned} \frac{\partial \pi}{\partial Q_2} &= \left[1 - \frac{Q_2}{\bar{Q}_2}\right] [vP_2 - 2vQ_2 + (P_x - C)] \\ &\quad - \frac{1}{\bar{Q}_2} [(vP_2 + P_x)Q_2 - (vQ_2^2 + CQ_2)] = 0 \end{aligned}$$

$$\frac{\partial \pi}{\partial Q_2} = vP_2 + P_x - 2vQ_2 - C - \frac{vP_2 Q_2}{\bar{Q}_2} - \frac{P_x Q_2}{\bar{Q}_2} + \frac{2vQ_2^2}{\bar{Q}_2} + \frac{CQ_2}{\bar{Q}_2}$$

$$- \frac{vP_2 Q_2}{\bar{Q}_2} - \frac{P_x Q_2}{\bar{Q}_2} + \frac{vQ_2^2}{\bar{Q}_2} + \frac{CQ_2}{\bar{Q}_2} = 0$$

$$\frac{\partial \pi}{\partial Q_2} = \frac{-2vP_2 Q_2}{\bar{Q}_2} - \frac{2P_x Q_2}{\bar{Q}_2} + \frac{2CQ_2}{\bar{Q}_2} - 2vQ_2 + vP_2 + P_x - C + \frac{3vQ_2^2}{\bar{Q}_2} = 0$$

$$\frac{\partial \pi}{\partial Q_2} = \frac{3v}{\bar{Q}_2} Q_2^2 - \left[\frac{2vP_2 + (2P_x - 2c)}{\bar{Q}_2} + 2v \right] Q_2 + vP_2 + (P_x - C) = 0$$

C.Q.F.D.

ANNEXE JURIDIQUE: Protection du consommateur et règlements d'application

Nous reproduisons ici les articles 167 à 181 de la Loi sur la protection du consommateur traitant de la réparation d'automobile.

(1^e novembre 1980)

3.- Réparation d'automobile et de motocyclette

167. Aux fins de la présente sous-section, on entend par:

- a) "commerçant" une personne qui effectue une réparation moyennant rémunération;
- b) "réparation": un travail effectué sur une automobile à l'exception d'un travail prévu par règlement.

168. Avant d'effectuer une réparation, le commerçant doit fournir une évaluation écrite au consommateur. Le commerçant ne peut se libérer de cette obligation sans une renonciation écrite en entier par le consommateur et signée par ce dernier.

L'évaluation n'est pas requise lorsque la réparation doit être effectuée sans frais pour le consommateur.

Un commerçant ne peut exiger de frais pour faire une évaluation à moins d'en avoir fait connaître le montant au consommateur avant de faire l'évaluation.

169. S'il faut, pour fournir une évaluation, démonter en tout ou en partie une automobile ou une partie d'une automobile, la somme mentionnée en vertu de l'article 168 doit comprendre le coût de remontage au cas où le consommateur décide de ne pas faire effectuer la réparation et ceux de la main-d'oeuvre et d'un élément requis pour remplacer un objet non récupérable ou non réutilisable détruit lors du démontage.

170. L'évaluation doit indiquer:

- a) le nom et l'adresse du consommateur et ceux du commerçant;

- b) la marque, le modèle, et le numéro d'immatriculation de l'automobile;
- c) la nature et le prix total de la réparation à effectuer;
- d) la pièce à poser, en précisant s'il s'agit d'une pièce neuve, usagée, réusinée ou remise à neuf; et
- e) la date et la durée de validité de cette évaluation.

171. L'évaluation acceptée par le consommateur lie également le commerçant. Aucuns frais supplémentaires ne peuvent être exigés du consommateur pour la réparation prévue dans l'évaluation.

172. Le commerçant ne peut effectuer une réparation non prévue dans l'évaluation acceptée avant d'avoir obtenu l'autorisation expresse du consommateur.

Dans le cas où le commerçant obtient une autorisation orale, il doit la consigner dans l'évaluation en indiquant la date, l'heure, le nom de la personne qui l'a donnée et, le cas échéant, le numéro de téléphone composé.

173. Lorsqu'il a effectué une réparation, le commerçant doit remettre au consommateur une facture indiquant:

- a) le nom et l'adresse du consommateur et ceux du commerçant;
- b) la marque, le modèle et le numéro d'immatriculation de l'automobile;
- c) la date de la livraisons de l'automobile au consommateur et le nombre de milles ou de kilomètres indiqués au totalisateur de l'automobile à cette date;
- d) la réparation effectuée;
- e) la pièce posée en précisant s'il s'agit d'une pièce neuve, usagée, réusinée ou remise à neuf et son prix;

- f) le nombre d'heures de main-d'oeuvre facturé, le tarif horaire et le coût total de la main-d'oeuvre;
 - g) les droits exigibles en vertu d'une loi fédérale ou provinciale;
 - h) le total des sommes que le consommateur doit déboursées pour cette réparation; et
 - i) les caractéristiques de la garantie.
174. Lorsqu'une réparation est faite par un sous-traitant, le commerçant a les mêmes obligations que s'il l'avait lui-même effectuée.
175. Le commerçant doit, si le consommateur l'exige au moment où il demande de faire la réparation, remettre à ce dernier la pièce qui a été remplacée et ce, au moment où le consommateur prend livraison de son automobile sauf:
- a) si la réparation est faite sans frais pour le consommateur;
 - b) si la pièce est échangée contre une pièce réusinée ou remise à neuf; ou
 - c) si la pièce remplacée fait l'objet d'un contrat de garantie en vertu duquel le commerçant doit remettre cette pièce au manufacturier ou au distributeur.
176. Une réparation est garantie pour trois mois ou 5 000 kilomètres, selon le premier terme atteint. La garantie prend effet au moment de la livraison de l'automobile.
177. La garantie prévue à l'article 176 ne couvre pas un dommage qui résulte d'un usage abusif par le consommateur après la réparation.
178. L'acceptation de l'évaluation ou le paiement du consommateur n'est pas préjudiciable à son recours contre le commerçant en raison d'une absence d'autorisation préalable de la réparation, d'une malfaçon ou d'un prix qui excède, selon le cas, le prix indiqué dans l'évaluation ou la somme du prix indiqué dans l'évaluation et du prix lors de la modification autorisée.

179. Malgré l'article 441 du Code civil, le commerçant ne peut retenir l'automobile du consommateur:

- a) si le commerçant a omis de fournir une évaluation au consommateur avant d'effectuer la réparation; ou
- b) si le prix total de la réparation est supérieur au prix indiqué dans l'évaluation, à la condition que le consommateur paie le prix indiqué dans l'évaluation; ou
- c) si le prix total de la réparation est supérieur à la somme du prix indiqué dans l'évaluation et du prix convenu lors de la modification autorisée à la condition que le consommateur paie un prix égal à cette somme.

180. Un commerçant qui effectue la réparation d'automobiles doit, conformément aux exigences prescrites par règlement, afficher dans un endroit bien en vue de son établissement une pancarte informant les consommateurs des principales dispositions prévues dans la présente sous-section.

181. Les articles 167 à 175 et 177 à 180 s'appliquent, compte tenu des adaptations nécessaires, à la réparation d'une motocyclette adaptée au transport sur les chemins publics.

Une réparation d'une motocyclette adaptée au transport sur les chemins publics est garantie pour un mois. La garantie prend effet au moment de la livraison de la motocyclette.

Le règlement d'application de la Loi sur la protection du consommateur (21 sept. 1982) défini (article 73) le terme "réparation" utilisé dans le texte de la loi sur la protection du consommateur. Tant qu'à eux, les articles 74 à 77 du règlement d'application explicite l'article 180 de la loi relatif à l'affichage obligatoire par un commerçant qui effectue la réparation d'automobiles. Voici donc ces articles d'application:

73. Ne constitue pas une réparation au sens du paragraphe b de l'article 167 de la Loi:
- a) un travail dont le coût total, incluant le prix des pièces et le coût de la main-d'oeuvre, n'excède pas 50 \$;
 - b) l'installation, sur une automobile ou sur une motocyclette, de pneumatiques ou d'une batterie lorsque l'achat et l'installation font l'objet d'une même facture.

74. Un commerçant qui effectue la réparation d'automobile ou de motocyclette doit afficher, dans un endroit bien en vue de son établissement, une pancarte sur laquelle figurent les mentions suivantes:

"Avis aux consommateurs.

- I- En vertu de la Loi sur la protection du consommateur (L.R.Q., c. P-40.1), lorsqu'un consommateur fait effectuer une réparation d'automobile ou de motocyclette dans cet établissement, le commerçant doit, pour une réparation de plus de 50 \$:
 - a) fournir une évaluation écrite;
 - b) fournir une facture détaillée de la réparation effectuée;
 - c) remettre les pièces remplacées si le consommateur l'a exigé au moment où il a demandé de faire la réparation.
 - II- Tarif horaire de la main d'oeuvre:.....\$
 - III- Une réparation d'automobile est garantie pour 3 mois ou 5 000 kilomètres selon le premier terme atteint.
- Une réparation de motocyclette est garantie pour 1 mois, sans limite de kilométrage.
- La garantie prend effet au moment de la livraison de l'automobile ou de la motocyclette.
- Ces dispositions ne constituent qu'un résumé des droits du consommateur. Ces droits peuvent comporter des restrictions. Pour plus de détails, le consommateur aura avantage à consulter les articles 167 à 181 de la Loi sur la protection du consommateur et, au besoin, à communiquer avec l'Office de la protection du consommateur."

75. La pancarte doit être faite d'un matériau rigide. La surface sur laquelle figurent les mentions prescrites par l'article 74 doit être mate et de couleur blanche.
76. La pancarte doit respecter les dimensions minimales suivantes:
- a) hauteur: 100 centimètres;
 - b) largeur: 50 centimètres.
77. Le texte des mentions prescrites par l'article 74 doit être imprimé d'une couleur foncée et mate, en caractère typographique équivalent à l' **HELVÉTIKA DEMI-GRAS** d'au moins 36 points.

Enfin, le marché de la réparation d'automobile est également assujetti par les articles de la section IV qui est plus générales. L'article 222 de la loi sur la protection du consommateur est particulièrement intéressant. Celui-ci nous dit que:

222. Aucun commerçant, manufacturier ou publicitaire ne peut faussement, par quelque moyen que ce soit:
- a) invoquer une circonstance déterminée pour offrir un bien ou un service;
 - b) déprécier un bien ou un service offert par un autre;
 - c) prétendre qu'un bien ou un service a été fourni;
 - d) prétendre qu'un bien a un mode de fabrication déterminé;
 - e) prétendre qu'un bien ou un service est nécessaire pour changer une pièce ou effectuer une réparation;
 - f) prétendre qu'un bien ou un service est d'une origine géographique déterminés.
 - g) indiquer la quantité d'un bien ou d'un service dont il dispose.

BIBLIOGRAPHIE

- AMBRAHAM-FROIS, G. et BERREBI E. (1981), "La demande, face cachée de la production jointe", Revue économique, no. 6, (novembre), pp. 1154-1165.
- AHQ (1981), Loi sur les services de santé et les services sociaux, (décembre), 101 pp.
- ANDERSON, R.K., HOUSE, D. and ORMISTON, M.B. (1981), "A Theory of Physician Behaviour with Supplier-Induced Demand", Southern Economic Journal, Vol. 48, no. 1, pp. 124-134.
- ARROW, K.J. (1963), "Uncertainty and the Welfare Economics of Medical Care", American Economic Review, (December), pp. 941-973.
- AUSTER, R.D. and OAXACA, R.L. (1981), "Identification of Supplies Induced Demand with Health Care Sector", Journal of Human Resources, Vol. XVI, no. 3, pp. 327-342.
- BAUMOL W.J., PANZAR J.C. and WILLIG M.D. (1982), Contestable Markets and the Theory of Industry Structure, Harcourt Bruce Jouanovich, New-York, 510 p.
- BLACKSTONE, E.A. (1974), "Misallocation of Medical Resources: the Problem of Excessive Surgery", Public Policy, vol. 22 (Summer), pp. 329-352.
- BOUTIN, J.G. (1980), "Effets de certaines mesures de la deuxième entente des médecins omnipraticiens sur leur activité dans le cadre du régime d'assurance-maladie", Actualité économique, (avril-juin) no. 2, pp. 239-263.
- CLEMENS, E. (1980), "Price Discrimination and the Multiproduct Firm", Review of Economic Studies, 19, no. 48, 1950-1951, pp. 1-11; reproduit dans R.B. Hettlebower and G.W. Stocking, eds, Reading in Industrial Organization and Public Policy, Homewood II; American Economic Association, Irwin, 1958, pp. 262-276.
- COMITE PARITAIRE DE L'INDUSTRIE DE L'AUTOMOBILE DE MONTRÉAL ET DU DISTRICT (1982), PARLONS NOUS: le décret 184 des salariés de garages de la région de Montréal est modifié de le 28 juillet 1982, vol. 7, no. 2 (septembre) 4 p.
- COMITE PARITAIRE DE L'INDUSTRIE DE L'AUTOMOBILE DE MONTRÉAL ET DU DISTRICT (1983), PARLONS NOUS: l'apprentissage et la qualification, vol. 8, no. 1 (mars), 4 p.
- CONTANDRIOPOULOS, A.P. (1976), "Un modèle de comportement des médecins en tant que producteurs de services", Thèse de doctorat non publiée, Département de sciences économiques, Université de Montréal, (août), 294 p.

- CONTANDRIOPOULOS, A.P. (1980), "Stimulants économiques et utilisation des services médicaux", Actualité économique, (avril-juin), pp. 264-296.
- CONTANDRIOPOULOS, A.P. ET DIONNE G., (1983), ECN 6640, Notes de cours: Économies des services de santé, Département de sciences économiques, Université de Montréal (hiver).
- CONTANDRIOPOULOS, A.P., DIONNE, G. et TESSIER, G. (1983), "Mobilité des patients et modèles de création de demande: le cas du Québec" Actualité économique, Vol. 59 (décembre), pp. 730-752
- CORPORATION PROFESSIONNELLE DES MEDECINS DU QUÉBEC (1982), LOI 27: organisation professionnelle: Questions-réponses, document non publié, (octobre), 24 p.
- DARBY, M. and KARNI, E. (1973), "Free Competition and the Optimal Amount of Fraud", Journal of Law and Economics, Vol. 16, pp. 66-88.
- DIONNE, G. (1979), "Le risque moral et le furetage des consommateurs", Thèse de doctorat non publiée, Département de sciences économiques, Université de Montréal, (août), 280 p.
- DIONNE, G. (1980), "Analyse des effets de l'assurance et de la relation de confiance consommateur-producteur sur les possibilités d'abus des chirurgiens", Actualité économique, (avril-juin), pp. 211-238.
- DIONNE, G. and CONTANDRIOPOULOS, A.P. (1982), "Doctors and their Workshop: A Review Article", texte non publié, Département de sciences économiques, Université de Montréal, (décembre), 41 p.
- EDITEUR OFFICIEL, QUÉBEC, (1982), Code de déontologie des médecins, M-9, r. 4,, (21 septembre), 6 p.
- EDITEUR OFFICIEL, QUÉBEC (1981), Décret sur les salariés de garages de la région de Montréal: loi sur les décret et convention collective, D-2, r. 46, pp. 4-939 à 4-945.
- EDITEUR OFFICIEL, QUÉBEC (1982), Loi, médicale, M-9, (14 septembre), 12 p.
- EDITEUR OFFICIEL, QUÉBEC (1980), Loi sur la protection du consommateur, chapitre P-40.1, (1 novembre), 90 p.
- EDITEUR OFFICIEL, QUÉBEC (1982), Règlement d'application de la loi sur la protection du consommateur, P-40.1, r. 1, (21 septembre) pp. 8-397 à 8-440
- EDMONSTON, P. et ROSEMAN E., (1978), Le guide de survie du consommateur, les éditions Quize, Montréal, 365 p.

- EVANS, R.G. (1972), Price Formation in the Market for Physician Services, Prices and Incomes Commission, Ottawa, 131 p.
- EVANS, R.G. (1974), "Supplier-Induced Demand: Some Empirical Evidence and Implication", in Perlman, M. [Ed.], The Economics of Health and Medical Care (London: Macmillan), pp. 162-173.
- EVANS, R.G. and WOLFSON, A.D. (1980), "Faith, Hope and Charity: Health Care in the Utility Function", 41 p.
- FUCHS, V.R. (1974), Who Shall Live? Health, Economics and Social Choice, Basic Books Inco Publishers, New-York, 1974, 168 p.
- FUCHS, V.R. (1978), "The Supply of Surgeons and the Demand for Operations", Journal of Human Resources, Vol. XIII, Supplément, pp. 35-56.
- GREEN J. (1978), "Physician-Induced Demand for Medical Care", The Journal of Human Resources, XIII, (Supplement), pp. 21-34.
- HENDERSON, J.M., et QUANDT, R.E. (1978), Microéconomie, Dunod, Paris, 408 p.
- LEIJONHUFVUD A. (1973), On Keynesian Economics and the Economics of Keynes, Oxford University Press, New-York, p. 43.
- LEWIS, C.E. (1969), "Variation in the Incidence of Surgery", New England Journal of Medicine, (octobre), pp. 880-884.
- MALINVAUD E. (1980), Réexamen de la théorie du chômage, Calmann-Lévy, Paris, 188 p.
- MARSHALL, J.M. (1976), "Moral Hazard", American Economic Review, Vol. LXVI, no. 5, (December), pp. 880-890.
- MATTE, L. (1983), "L'opération auto-témoin, O.P.C.", Revue Protégez-vous, février, pp. 5-7.
- MCCARTHY, E.G. and WIDNER, G.W. (1974), "Effets of Screening by Consultants on Recommended Elective Surgical Procedures", New England Journal of Medicine, (December), pp. 1,331-1,335.
- MIGUE, J.L. et BELANGER (1972), Le prix de la santé, Editions Hurtubise, HMH Ltée., Montréal, 238 p.
- MINTZ J.M. (1981), A Note on Mutliproduct Economics of Scale and Economics of Scope, Economics Letters, North-Holland Publishing Company, pp. 29-33.
- MONSMA, M.N. Jr. (1970), "Marginal Revenue and the Demand for Physician Services" in Klarman, H.E. ed., Empirical Studies in Health Economics, The Johns Hopkins Press, Baltimore and London, 433 p.

- NEWHOUSE J.P. (1978), *The Economics of Medical Care: A Policy Perspective*, Addison-Wesley, Massachusetts, 116 p.
- NICKERSON, R.J. et AUTRES (1976), "Doctors who Perform Operation", New England Journal of Medicine, (October), Part A: october 24 pp. 421-926; Part B: october 28, pp. 982-989.
- PANZAR, J.C. and WILLIG, R.D. (1981), "Economics of Scope", American Economic Review, 71, no. 2, (may), pp. 268-272.
- PAULY, M.V. (1978), "Is Medical Care Different?", in Greenberg, W. (Ed.), Competition in the Health Care Sector, Aspen, Germantown, (Maryland), pp. 11-35.
- PAULY, M.V. (1979), "What is Necessary Surgery?", Milbank Memorial Fund Quarterly/Health and Society, Vol. 57, pp. 95-117.
- PAULY, M.V. (1980), *Doctors and their Workshops*, N.B.E.R., Chicago, 132 p.
- REINHARDT, U. (1978), "Parkinson's Law and the Demand for Physicians' Services", in Competition in the Health Care Sector, W. Greenberg (Ed.), Aspen Systems Corporation, Germantown, pp. 45-102.
- SATTERTHWAITE, M.A. (1979), "Consumer Information, Equilibrium Industry Price and the Number of Sellers", Bell Journal of Economics, Vol. 10, pp. 483-502.
- SCHNEIDE J.F. (1972), "Useful Work for Specialist", New England Journal of Medicine, Vol. 287 (27 juillet), p. 207.
- SLOAN, F. and FELDMAN, R. (1978), "Competition Among Physician", in Competition in the Health Care Sector, W. Greenberg (Ed.), A.S.C., Germantown, pp. 45-102.
- ST-MICHEL, P. (1983), *Analyse économique de l'impact d'un ticket modérateur sur la consommation des services de santé*, Mémoire de maîtrise non publié, Département de sciences économiques, Université de Montréal, (juin), 276 p.
- WENBERG, Z.E., BARNES, B.A., ZUBKOFF, M. (1982), "Professional Uncertainty and the Problem of Supplier-Induced Demand", Soc. Sc. Med., Vol. 16, pp. 811-824.
- ZWEIFEL, P. (1981), "Supplier-Induced Demand in a Model of Physician Behavior", in Van Der Gaag, J. and Perlman, M. (Eds.), Health, Economics and Health Economics, Amsterdam (North-Holland), pp. 245-271.

REMERCIEMENTS

Je sais gré au professeur Georges Dionne d'avoir accepté la direction pédagogique et scientifique de ce mémoire. Son sens aigu des responsabilités, son enthousiasme et sa rigueur intellectuelle en firent le directeur de recherche idéal et furent une source d'inspiration et de motivation continuelle.

Je remercie également les professeurs Marcel Boyer, André-Pierre Contandriopoulos et Abraham Hollander pour leur considération à différentes étapes de la recherche. Leurs commentaires et suggestions ont permis d'améliorer le contenu de ce travail.

Ce projet a bénéficié des contributions financières de l'Office de la protection du consommateur et du fond FCAC pour l'aide et le soutien à la recherche du Gouvernement du Québec.

Je remercie M. Charles Bergeron qui m'a rendu le service de réviser le manuscrit.

Je suis également reconnaissant envers mes parents pour leur soutien tout au long de mes études.

Finalement, je tiens à remercier mon épouse, Cara, pour son travail graphique et surtout pour ses encouragements et sa compréhension.

H. D.

