

**CONCURRENCE, INNOVATION ET
INTERNATIONALISATION:
LA DYNAMIQUE DE L'ECONOMIE FRANCAISE**

**CONTRAT CGP
Subvention 5.98**

Vol.1 : LA DYNAMIQUE CONCURRENTIELLE EN FRANCE

Vol. 2 : LES DYNAMIQUES DE L'INNOVATION

Vol 3. Annexes du Vol.1

**CREST-LEI,
URA 2200 (GRECSTA)
Responsable : A. PERROT**

**EUREQua
Université de Paris I, UMR 8594
Responsable : D. ENCAOUA**

**VOLUME 1
*LA DYNAMIQUE CONCURRENTIELLE EN FRANCE***

COMMISSARIAT GENERAL DU PLAN

Janvier 2000

Table des matières du volume 1

LA DYNAMIQUE CONCURRENTIELLE EN FRANCE

Synthèse

Première Partie: Une évaluation des structures de marché et de l'intensité de la concurrence.

I Les opérations de Fusions et Acquisitions transfrontalières: les cas américain et européen

II Une évaluation de l'intensité concurrentielle : unicité du prix et stratégies des entreprises dans l'Union Européenne

Deuxième Partie: La grande distribution et ses relations avec le secteur productif en France

I Les acteurs et le marché

II les relations entre producteurs et distributeurs

III Le mouvement de concentration dans la grande distribution

IV le développement des marques de distributeurs

Troisième Partie: Concurrence et concentration dans le secteur des assurances en France

I Une approche descriptive du secteur des assurances en France

II Analyse théorique

Quatrième Partie : Comment allouer et financer les missions de service universel dans un réseau ouvert à la concurrence ? Une évaluation du système français

I Les obligations de service universel et les options de régulation

II Le cadre d'analyse adopté

III Les résultats

Annexes

LA DYNAMIQUE CONCURRENTIELLE EN FRANCE

Rapport de synthèse

CREST-LEI

Janvier 2000

INTRODUCTION GENERALE

Il est de plus en plus souvent admis que le dynamisme des économies contemporaines, et *in fine*, le taux de croissance de ces économies, doivent beaucoup aux possibilités qu'ont les entreprises de se livrer, sur des marchés ouverts, une concurrence suffisamment intense. Selon cette thèse, les mécanismes concurrentiels, multiformes, qui seraient susceptibles de se développer sur de tels marchés permettraient d'accroître le surplus collectif par une meilleure allocation des ressources, conduiraient les consommateurs à rencontrer une offre de biens et services plus diversifiée en qualité et en variété et à des prix plus attractifs, et attireraient enfin les entrepreneurs par les opportunités de profits liées à l'expansion continue des marchés. Au contraire, un fonctionnement des marchés laissant s'ériger des barrières à l'entrée importantes, soit du fait d'une réglementation trop protectrice ou trop rigide, soit en raison de comportements de domination ou d'exploitation abusive de positions concurrentielles favorables, aurait des conséquences très négatives sur les prix, l'innovation, et plus généralement le bien-être des agents. Cet ensemble d'arguments met donc l'accent sur le rôle essentiel que jouent, pour le développement des économies, deux séries de facteurs : d'une part ceux qui agissent sur les structures de marché, d'autre part, ceux qui ont une incidence sur les comportements.

Une grande partie de ces facteurs est liée à l'interaction entre le fonctionnement des institutions, (entendues au sens large comme l'ensemble des réglementations et des législations pertinentes, les modalités de la mise en oeuvre de ces dispositions, etc...) et les comportements des agents économiques qu'elles visent. Ainsi, pour ne citer que quelques uns des exemples abordés par nos travaux, les mouvements de concentration ayant marqué particulièrement le paysage de la grande distribution française sont parfois vus comme une réponse des distributeurs à l'ensemble des dispositifs juridiques encadrant le développement de la grande distribution (contexte général de l'Ordonnance du 1^{er} décembre 1986, mais aussi lois Galland et Raffarin de 1996)¹. Autre exemple, dans le domaine tout différent des télécommunications la façon dont les obligations de service universel sont en France allouées entre différents opérateurs et financées ont une influence décisive sur l'intensité et la nature de la concurrence entre ces opérateurs.

¹ Voir sur ce point l'analyse de J. Philippe (DGCCRF et CREST-IEI), *Relations verticales et concurrence*, thèse de doctorat Université de Toulouse, décembre 1999.

Dans cette perspective, il est donc essentiel d'évaluer ce qu'on pourrait appeler « les performances concurrentielles » de l'économie française et plus précisément, d'apporter des éléments de réponse aux questions suivantes :

Les structures de marché, les modalités d'organisation et de surveillance de ces marchés et les comportements stratégiques qui s'y développent sont-ils de nature à freiner ou au contraire à favoriser le développement économique ? Les institutions et les comportements permettent-ils, en France, la constitution d'espaces économiques intégrés et ouverts, ou au contraire érigent-ils des barrières à l'entrée interdisant à la concurrence de produire tous ses effets positifs ? Comment les préoccupations d'équité (des citoyens, des entreprises) interfèrent-elles avec ces diverses questions ?

Les travaux présentés ici s'inscrivent directement dans cette problématique. Ils cherchent à cerner, dans quelques domaines clés de l'économie française, *les atouts et les entraves apportés au fonctionnement concurrentiel des marchés*, et à évaluer *la pertinence de certains choix législatifs ou réglementaires* en matière de concurrence en France.

Dans cette optique, nous avons choisi de privilégier quelques thèmes qui nous paraissent particulièrement significatifs. Après une partie transversale (I) consacrée aux mouvements de fusions et acquisitions et à une approche de l'intensité des mécanismes concurrentiels par la question de la convergence des prix, nous examinons plus spécifiquement trois domaines : la grande distribution (II), le secteur des assurances (III), et les télécommunications (IV). Dans ces trois secteurs d'activité, la France présente des spécificités institutionnelles ou historiques qui la démarquent souvent des autres marchés européens. Ces caractéristiques ont trait, dans tous les cas, aux relations entre les législations et les réglementations d'une part, et les agents économiques d'autre part.

Le « système d'explication » de nos résultats et la vision de l'économie française qui en découle, s'appuient ainsi sur l'analyse des interactions stratégiques entre les autorités publiques et les agents économiques, analyse qui constitue le point commun de tous les travaux présentés ici.

1) Une évaluation des structures de marché et de l'intensité de la concurrence :

La première partie de ce rapport a trait aux structures de marché et à une évaluation de l'intensité des mécanismes concurrentiels qui s'y manifestent. Nous apportons ici des éléments sur deux types de questions :

1) D'une part, quelles sont les caractéristiques des mouvements de concentration dans l'industrie française ? En particulier, on examine si les fusions & acquisitions dans lesquelles sont impliquées les entreprises françaises touchent particulièrement certains secteurs de l'économie, dans quelle mesure ces rapprochements concernent des entreprises nationales (opérations franco-françaises) ou des entreprises de nationalités différentes (opérations transfrontalières), quelle est la forme prise par ces opérations (fusion, acquisition d'actifs, prise de participation minoritaire ou majoritaire).

2) D'autre part, le degré de « perfection » de la concurrence peut être apprécié par l'intensité de la concurrence par les prix qui s'exerce sur les espaces économiques français et européen. En particulier, la « loi du prix unique », qui traduit le fait que les mécanismes concurrentiels devraient assurer la convergence des prix au sein d'un espace économique intégré, est-elle vérifiée entre régions françaises, d'une part, et entre régions françaises et régions européennes d'autre part ?

En théorie, un mécanisme concurrentiel dépourvu d'imperfections devrait conduire à une égalisation des prix pour des produits similaires. Une évaluation des différences de prix, entre pays d'une part et entre régions d'autre part, est donc un élément permettant de cerner la puissance de ces mécanismes de convergence. Nous explorons tout d'abord empiriquement cette question. Dans un deuxième temps, nous tentons de donner des fondements à la persistance des différentiels de prix : ceux-ci peuvent-ils avoir d'autres origines que des freins à la concurrence (différenciation des produits ? de la demande ?).

2) La grande distribution en France

Le rôle que joue la grande distribution dans les mécanismes concurrentiels n'est pas à démontrer : tout d'abord, les prix auxquels font face les consommateurs sont fixés (ou devraient l'être) par des acteurs de la distribution, et constituent donc aux yeux des acheteurs de biens de consommation un reflet tangible de l'intensité des processus concurrentiels. Ensuite, un nombre important de restructurations intervenues dans l'économie française (ou du moins certaines des plus visibles) concerne des entreprises de la grande distribution. Partant, nombre d'interrogations sur le rôle des concentrations dans le maintien d'une concurrence suffisante s'appuient sur le cas de la grande distribution.

En matière d'analyse économique, les « relations verticales » ont longtemps focalisé l'attention sur les liens contractuels entre producteurs et distributeurs : les questions liées à la discrimination, au refus de vente, aux primes de référencement, au déréférencement abusif et à

leurs conséquences sur le surplus collectif ont fait l'objet, jusqu'au début des années quatre-vingt-dix, de l'essentiel des travaux sur les relations entre producteurs et les distributeurs.

Du point de vue factuel, tous ces débats ont trouvé en France leur aboutissement dans la promulgation, en 1996, de deux lois (les lois dites Galland et Raffarin) dont l'objectif affiché était de « rétablir l'équilibre » perdu des relations commerciales entre les agents. L'ensemble des textes de loi encadrant le développement de la concurrence en France fait actuellement l'objet d'une réforme dont il est encore difficile d'évaluer l'ampleur².

L'analyse économique de la distribution s'intéresse maintenant surtout aux mouvements de concentration qui affectent le secteur du commerce, particulièrement en France. Cette partie de notre recherche donne tout d'abord une mesure de ces mouvements de concentration de la distribution, en essayant de marquer la place du cas français au sein de l'ensemble européen, et tente ensuite d'en donner les fondements théoriques. De ce point de vue, deux types de problèmes sont envisagés. Tout d'abord, on examine les raisons pour lesquelles les distributeurs reçoivent plus d'incitation à la fusion que les producteurs et on évalue les conséquences de ces restructurations vis-à-vis de différents critères. Ensuite, on s'intéresse à une autre caractéristique récente de la grande distribution : la montée en puissance des marques de distributeurs.

3) Concurrence et concentration dans le secteur des assurances en France

Le secteur des assurances, en Europe et plus particulièrement en France, se caractérise par une intensité très grande des mouvements de fusions. Nous nous attachons, dans cette partie, à donner quelques éléments factuels sur l'ampleur de ces mouvements, et essayons de donner des fondements, dans le cas français, à cette recherche de la grande taille, et à évaluer les effets de la réglementation prudentielle dans ce contexte.

Aux arguments traditionnellement invoqués pour expliquer le regroupement d'entreprises (économies d'échelle, recherche du pouvoir de marché, etc...) s'adjoint, dans ce secteur particulier, un mécanisme nouveau. Le principe de mutualisation des risques est en effet à la naissance d'un « effet de réseau de l'assurance », (ou effet taille) qui rend, à l'équilibre, le contrat proposé par une compagnie d'autant plus attractif que de nombreux individus lui sont rattachés. Cet effet taille renforce donc les incitations à la fusion.

² Les récentes « Assises de la distribution », qui se sont tenues le 13 janvier 2000 à Paris, avaient pour objectif de susciter le débat sur les rapports des fournisseurs avec la grande distribution.

Les travaux proposés ici, après avoir présenté les principaux éléments factuels relatifs à la concurrence sur le marché des assurances, examinent cet argument dans un contexte stratégique plus complexe qui prend en compte un élément essentiel du marché des assurances français : l'existence, aux côtés des compagnies d'assurance, de *mutuelles* qui relèvent d'un principe d'organisation différent, puisque les mutuelles reposent seulement sur un partage du risque sans engagement de capitaux propres. Au delà de cette spécificité, il convient donc de s'interroger sur les effets de la réglementation prudentielle en vigueur.

4) Allocation et financement des obligations de service universel dans un réseau ouvert à la concurrence.

Une part importante des réflexions récentes sur le bon fonctionnement de la concurrence est liée à l'ouverture de nouveaux marchés aux processus concurrentiels. Parmi ceux-ci, les anciens monopoles de réseaux (télécommunications, énergie) fournissent un vaste champ de réflexion. En effet, se posent ici, plus que sur d'autres marchés plus anciennement ouverts, non seulement des questions de politique de la concurrence liées aux risques d'abus de position dominante de la part des anciens monopoles, mais aussi des problèmes de réglementation proprement dite et d'articulation entre ces deux modes d'intervention de la puissance publique.

Relève en particulier de la réglementation le traitement des obligations de service universel. Dans la plupart des économies européennes en effet, la fin des monopoles a profondément remis en question le fonctionnement des services dits « publics » en réseaux, et en particulier la manière dont sont allouées et financées les obligations de service universel. Dans un monde ouvert à la concurrence, les phénomènes d'écémage rendent impossible le système de péréquation tarifaire qui constitue, en monopole, une manière simple de financer les déficits laissés par ces contraintes. Se pose dès lors la question de savoir quels sont les systèmes les plus efficaces (les moins coûteux collectivement) pour s'acquitter de ces obligations.

La France s'est dotée d'une réglementation qui repose sur deux principes : dans le secteur des télécommunications par exemple, les obligations de service universel sont confiées à l'ancien monopole, et elles sont financées par des taxes prélevées sur les trafics. On peut s'interroger sur le bien-fondé d'une telle régulation : d'autres systèmes, comme celui du « pay or play » (mis en oeuvre en Australie) ou des mécanismes d'enchères (comme dans le service postal allemand) peuvent être envisagés, et sont d'ailleurs mis en oeuvre dans d'autres pays. Nous visons ici à caractériser l'efficacité du système français parmi un ensemble de régulations

possibles et à évaluer du double point de vue collectif et redistributif, les conséquences de ces différentes régulations en matière de service universel.

PREMIERE PARTIE :

Une évaluation des structures de marché et de l'intensité de la concurrence

Introduction

L'analyse de la concurrence s'intéresse traditionnellement aux structures de marchés dans lesquelles elle voit un élément d'explication essentiel de la dynamique concurrentielle. Cependant, ces structures, loin d'être exogènes, sont le fruit de mouvements permanents de restructurations et de regroupements d'entreprises. Dans cette première partie, nous abordons deux questions qui permettent de cerner le paysage concurrentiel en France et de l'insérer dans le contexte européen. Tout d'abord, on examine, sur le plan empirique, quelles sont les caractéristiques des restructurations intéressant les entreprises françaises : ces mouvements de fusions-absorptions ont-ils des traits qui les distinguent de ceux observés au sein de l'Europe ? Quelle est la place des fusions trans-frontalières ?

Ensuite, on s'intéresse à l'évaluation empirique de la convergence des prix vers un prix unique au sein des espaces économiques français et européens.

I. Les concentrations dans l'industrie française.

Les fusions et acquisitions (F & A) trans-frontalières sont dans une large mesure motivées par les mêmes raisons et soumises aux mêmes contraintes que les opérations nationales: recherche de pouvoir de marché, de synergies, discipline des équipes dirigeantes, diversification restent évidemment parmi les principaux motifs des fusions, qu'elles se cantonnent à un cadre national ou qu'elles intéressent des entreprises appartenant à des pays différents. Toutefois, certaines explications sont spécifiques aux opérations impliquant deux entreprises de nationalités différentes : la recherche de la taille optimale supérieure à celle du marché national (i), l'exploitation d'une supériorité technologique ou de prix de facteurs plus avantageux (ii), le moyen de contourner des mesures politiques (iii), le moyen de suivre la clientèle à l'étranger (iv), la diversification internationale (v), l'assurance d'un accès à certains facteurs de production (vi) et évidemment les taux de change (vii).

(i) La recherche de la taille optimale vient sans doute au premier rang des explications précédentes ; une entreprise dont les possibilités de développement sont limitées à l'intérieur de son pays d'origine se tournera naturellement vers l'étranger.

Il est en effet possible que le marché national soit saturé, ou que le pays soit « trop petit » pour l'entreprise ; c'est par exemple le cas pour des groupes comme Fina naguère en Belgique ou

Shell aux Pays-Bas, la croissance externe à l'étranger leur permet de réaliser des économies d'échelle afin de rester compétitifs.

(ii) L'exploitation d'une supériorité technologique ou de prix de facteurs plus avantageux:

Certaines acquisitions sont réalisées dans le but de transférer de la technologie de pays avancés dans des domaines particuliers. Que ce soit l'entreprise dominante d'un point de vue technologique qui acquière l'entreprise la moins performante, ou au contraire l'entreprise la moins performante mais disposant d'une trésorerie abondante qui acquière la plus performante, la valeur combinée des deux entreprises sera supérieure à la somme de leurs deux valeurs séparées.

Comme la technologie, le prix des facteurs est un élément essentiel du coût de production. Le travail étant un facteur relativement peu mobile, les différences de productivité et /ou de prix du travail peuvent expliquer certaines acquisitions à l'étranger.

(iii) Le moyen de contourner des mesures politiques:

Le commerce international est fortement dépendant de mesures protectionnistes auxquelles peuvent avoir recours les gouvernements de façon conjoncturelle, comme l'illustre l'exemple récent des relations entre la France et la Grande-Bretagne à l'occasion de la « crise de la vache folle ». Dans ces conditions, l'acquisition d'entreprises à l'étranger peut être la réponse à la vulnérabilité des exportations comme des importations aux taxes, quotas et autres mesures protectionnistes.

(iv) Le moyen de suivre la clientèle à l'étranger :

Dans certains secteurs d'activité comme la banque ou l'assurance, l'implantation à l'étranger est un moyen de conserver les clients français à l'extérieur du pays, qui se tourneraient sinon vers les réseaux locaux. C'est une explication, partielle évidemment, aux acquisitions transfrontalières dans ces secteurs particuliers.

(v) La diversification internationale :

Les opérations internationales apportent une forme de diversification différente de celle qui est généralement recherchée à travers la construction de conglomérats (différents secteurs d'activité) : dans la mesure où les économies ne sont pas parfaitement corrélées, les fusions & acquisitions internationales réduisent la variabilité de revenus qui seraient sinon fortement liés

à la santé économique d'un seul pays. Au contraire de la diversification sectorielle qui n'apporte rien aux investisseurs qui ont la possibilité de constituer par eux-mêmes un portefeuille de titres diversifié, la diversification internationale peut être utile à ces mêmes investisseurs dans la mesure où, pour diverses raisons (manque d'information, difficultés pratiques,...), les portefeuilles de titres risqués comportent une majorité de titres nationaux. En d'autres termes, la diversification internationale serait en mesure de réduire le risque « non systématique » comme le risque « systématique ».

(vi) L'assurance d'un accès à certains facteurs de production:

Quand leur pays d'origine est pauvre en certains facteurs de production, les entreprises peuvent avoir recours à des acquisitions, il s'agit alors d'opérations verticales, afin de s'assurer leur approvisionnement indépendamment des contraintes commerciales.

(vii) Les taux de change :

Les taux de change ont plusieurs effets sur les opérations de F & A. La sous-évaluation momentanée d'une devise peut inciter à des acquisitions ; elle a également un effet direct sur le coût de production et, dans un sens différent, sur la valeur des profits rapatriés par la maison-mère. D'un autre côté, la variation des changes constitue un risque supplémentaire pour les entreprises et serait plutôt de nature à réduire le nombre d'opérations trans-frontalières

Notre objectif est d'apporter un éclairage sur les motifs et conséquences de ces opérations. En effet, l'allocation des ressources est plus efficace dans le cas d'économies d'échelle (taille efficace supérieure au marché national), de synergies (acquisition d'un savoir technologique, mise en commun de ressources en R & D, coût de facteurs plus faible, meilleure coordination des secteurs amont et aval), ou d'amélioration de la gestion des entreprises. En revanche, les opérations n'améliorent pas le bien-être social si leur objectif est la recherche de pouvoir de marché, la pure diversification sectorielle ou la sous-évaluation momentanée en raison d'un change avantageux. L'examen de ces opérations trans-frontalières parmi l'ensemble des F & A concernant les entreprises françaises et de leurs caractéristiques (pays de l'autre entreprise impliquée, secteur, type d'opération) fournit des indications sur objectifs poursuivis. Dans un premier temps, nous rappelons les principaux résultats de la littérature empirique sur le sujet,

et dans un deuxième temps, nous exploitons différentes sources de données pour approcher le cas de la France.

1. Les opérations de Fusions & Acquisitions transfrontalières : les cas américain et européen

Les études récentes sur les opérations de F & A trans-frontalières portent sur des pays et des périodes différents ; cependant, elles mettent en évidence un certain nombre de tendances communes :

- . la proportion d'opérations trans-frontalières augmente régulièrement ;
- . deux pays dominant à la fois en termes de nombre d'opérations et de montants en jeu : les Etats-Unis et le Royaume-Uni;
- . certains secteurs d'activité sont plus particulièrement concernés : le secteur banque/finance et celui de l'assurance, le secteur pétrolier, la distribution, le secteur pharmaceutique, le secteur aéronautique ; simultanément, un grand nombre d'opérations sont lancées par des conglomérats.

1-1. Un nombre croissant d'opérations trans-frontalières

- . Le cas des Etats-Unis

Entre 1981 et 1990, le nombre d'opérations trans-frontalières impliquant une entreprise américaine a augmenté régulièrement pour plus que doubler sur la période³: en 1981, sur 2395 opérations au total, on compte 253 transactions trans-frontalières, en 1990, elles sont au nombre de 532 sur un nombre total d'acquisitions pratiquement inchangé (2258 opérations en 1988), la moyenne étant de 404.7 par an ; avant 1970, les opérations transfrontalières, une vingtaine par an, étaient anecdotiques. Dans le cas des Etats-Unis, on constate que ce sont plutôt les entreprises américaines qui sont acquises par des étrangères que l'inverse (2211 entreprises américaines ont été visées contre 1836 qui ont été acquéreuses sur la période), et le Royaume-Uni à lui seul est à l'origine de 57% des acquisitions aux Etats-Unis. Parmi les 100 plus grosses opérations (en valeur) réalisées aux Etats-Unis, 20 impliquent une entreprise

³ Source : Mergerstat Review.

étrangère (l'acquisition en 1987 de la Standard Oil par British Petroleum, la plus grosse opération de la période, représentait 7.8 milliards de dollars).

. Le cas de l'Europe

L'évolution observée en Europe est similaire. Le nombre de F & A transfrontalières n'a pratiquement pas cessé d'augmenter entre 1985 et 1995 : 1206 opérations en 1994 contre 898 en 1989, (Martinez et Bilbao-Garcia, 1995) en nombre comme en valeur alors même que la tendance était globalement à la baisse pour l'ensemble des transactions à partir de 1990. Les Etats-Unis sont le premier partenaire avec environ un tiers des opérations en nombre et un peu moins en valeur ; vient ensuite le Royaume-Uni avec environ 15% des opérations en volume comme en valeur.

La croissance du nombre des restructurations réalisées à l'intérieur de la Communauté Européenne est encore plus élevée sur cette même décennie, le nombre d'opérations ayant plus que doublé dans l'ensemble du secteur manufacturier. Bon nombre d'entreprises seraient ainsi passées d'une dimension nationale à une dimension européenne (Anderson 1995). A l'intérieur de la Communauté Européenne, les F & A trans-frontalières concernent souvent des entreprises de pays limitrophes : les entreprises espagnoles se tournent naturellement vers le Portugal ou la France, les entreprises françaises vers l'Allemagne ou la Belgique, etc. Globalement, le Royaume-Uni reste le principal acteur à l'intérieur de la Communauté Européenne avec un quart des transactions en nombre et 30% en montant ; les opérations impliquant des entreprises françaises représentent environ 15% de l'ensemble en nombre et 20% en valeur ; les chiffres sont du même ordre pour l'Allemagne et beaucoup plus faibles pour les autres pays.

Ainsi, malgré l'augmentation du nombre d'opérations effectuées entre pays de la Communauté Européenne, il n'en demeure pas moins que les Etats-Unis et le Royaume-Uni restent les principaux acteurs dans les restructurations trans-frontalières en Europe avec respectivement environ un tiers et 15% des achats d'entreprises européennes (Martinez et Bilbao-Garcia 1995).

1.2 Les principaux secteurs concernés

. Le cas des Etats-Unis

Au cours des années 90, ce sont les entreprises américaines du secteur de la grande distribution qui ont été le plus souvent visées par des entreprises étrangères (9% des cibles) ; viennent ensuite les conglomérats (6%), le commerce de gros et le secteur pétrolier (5.6%) et le secteur pharmaceutique (5%).

Du côté des entreprises acquéreuses américaines, viennent en premier des conglomérats (6.5%), suivis d'entreprises du secteur de l'aéronautique, de l'assurance, de SSII (5.6% pour chacun de ces secteurs), et des secteurs bancaire et du commerce de détail (8% globalement) (Gonzalez et al. 1997).

. Le cas de l'Europe

Sur une période de deux années riches en opérations (1986 et 1987) Garella (1989) indique que 73% des opérations concernent le secteur industriel, 13% le secteur de la distribution, 8% le secteur bancaire et 7% celui de l'assurance. Sans avancer des chiffres précis, Anderson cite les mêmes secteurs sur la période 1985-1995 avec en plus celui de l'aviation civile, essentiellement pour des prises de participation minoritaires dans ce dernier cas. Ces deux études confirment la prédominance de certaines branches d'activité dans les opérations de fusions et acquisitions.

2. Les mouvements récents de Fusions & Acquisitions en France

La prise en compte exhaustive en termes statistiques des fusions et acquisitions a souvent posé problème. L'intérêt que portent les instituts publics ou privés de collecte d'informations à cette question est relativement récent. Les études ont souvent porté sur certains secteurs particuliers notamment ceux touchés par la crise et dont une restructuration était nécessaire. D'autres problématiques, comme celle des groupes d'entreprises, ont incité les instituts statistiques à créer des enquêtes spécifiques (comme la LIFI, - liaisons financières -, à l'INSEE). Ce type d'enquêtes est intéressant à plus d'un titre, notamment pour connaître la structure des liaisons financières entre entreprises françaises ou étrangères. Des difficultés apparaissent cependant pour comprendre par exemple comment s'est formée une liaison à

travers le temps⁴ Notre travail s'intéresse ici à une période récente (1994-1998). Les données proviennent de la revue *Fusions & Acquisitions Magazine* qui recense, mois par mois, les différentes opérations de F & A et de joint-ventures. C'est la seule revue qui s'occupe systématiquement de ce recensement pour les entreprises françaises. Toutes les autres sources d'information sont très partielles.

Notre travail consiste dans un premier temps à donner les chiffres globaux sur la période choisie (1994-1998) qui permet de prendre en compte les évolutions récentes les plus significatives. Dans un deuxième temps, nous analysons plus en détail le début et la fin de la période (années 1994 et 1998). Ce choix est dicté par le fait que l'année 1994 a été un creux pour les opérations de F & A en France sur la décennie alors que l'année 1998 a au contraire été le témoin de nombreux mouvements de concentrations. De plus, certaines tendances globales apparaissent et sont communes à l'ensemble des années de la période. Enfin, dans une dernière partie nous donnons quelques éléments d'analyse sur les opérations les plus importantes qui ont eu lieu pendant cette période.

2. 1. Activité globale sur la période

Les secteurs d'activité concernés

Les tableaux 1 à 5 (donnés en annexe) fournissent la répartition en pourcentage des nombres et des montants des transactions de F & A d'entreprises françaises par secteur d'activité pour la période 1994-1998.

On remarque qu'en nombre de transactions, en début de période (année 1994) plusieurs secteurs d'activité sont équivalents. Ainsi, on remarque que le nombre d'opérations est le plus important pour le secteur des services (15.5%), suivi du secteur de l'édition, communication (13.3%), et la distribution (11.4%). Les banques, assurances, holdings, ainsi que l'agro-alimentaire et la chimie-pharmacie représentent des parts sensiblement égales (respectivement 9.9%, 9.8% et 9.2%).

En revanche, en termes de valeur des transactions le secteur des banques, assurances et holdings représente la plus grande partie (24.5%). La chimie-pharmacie vient en deuxième position avec 16.2% et le transport, équipements de transport en troisième position (12.6%).

La part du secteur des banques, assurances et holdings ne cesse d'augmenter, que ce soit en nombre ou en valeur, pour arriver à un pic en 1996 (18.9% des opérations de F & A). En

⁴ Voir Derhy (1995) pour un exposé plus détaillé de ces problèmes.

1997, la part de ce secteur représente même plus de la moitié de la valeur des opérations d'acquisition d'entreprises françaises (52.1%).

Au fil des dernières années, trois secteurs apparaissent dominer les opérations de F & A d'entreprises françaises.

En effet, pour l'ensemble de la période d'étude (1994-1998), les trois secteurs suivants sont dominants :

- . Banques, assurances, holdings : 15.4% du nombre de transactions, 43.1% des montants alloués aux acquisitions d'entreprises françaises.
- . Distribution : 10.5% en nombre et 10.8% en valeur.
- . Services : 17.5% en nombre et 10.5% en valeur.

A côté de ce groupe de secteurs, on trouve un deuxième dont la contribution au nombre d'opérations est relativement importante mais dont la valeur des transactions est moindre. Il s'agit de:

- . Edition, communication : 9.2% en nombre et 7% en valeur.
- . Chimie, pharmacie : 7.4% en nombre et 8.5% en valeur.
- . Transport, équipements du transport : 6.6% en nombre et 4.7% en valeur.
- . Machines-outils, équipements électriques : 8.3% en nombre et 4.2% en valeur.
- . Agro-alimentaire : 8.1% en nombre et 3.1% en valeur.

Remarquons quand même que les secteurs de la Chimie-pharmacie et de l'Edition, communication avaient des contributions plus importantes en début de période (années 1994 et 1995) pour ensuite fléchir les trois dernières années. Ceci pourrait être expliqué par la fin d'une vague de restructurations dans ces deux secteurs.

Enfin, l'ensemble regroupant tous les autres secteurs, constitue un dernier groupe où l'activité de F & A est relativement peu soutenue du début à la fin de la période d'étude. Ceci pourrait s'expliquer pour certains secteurs par le fait qu'ils aient déjà connu d'importantes phases de restructuration dans un passé récent (Textile-Habillement et Métallurgie-Sidérurgie).

Les principaux partenaires de la France dans les fusions trans-frontalières

Il convient de distinguer les acquisitions d'entreprises étrangères par des entreprises françaises et celles d'entreprises françaises par des entreprises d'autres nationalités.

. Acquisitions françaises à l'étranger

Le tableau 6 donne l'ensemble des opérations d'acquisitions françaises à l'étranger (en nombre et en valeur) pour la période d'étude.

En nombre d'opérations, il apparaît que les acquisitions françaises à l'étranger se font essentiellement dans 6 pays : les Etats-Unis (16.8%), le Royaume-Uni (10.8%) ; l'Allemagne (10.7%), l'Italie (9%), l'Espagne (5.5%) et la Belgique (3.3%) (cf. tableau 7).

En revanche, en termes de valeur des acquisitions, la hiérarchie est plus prononcée. On peut ainsi isoler un premier groupe de pays qui représente plus de 75% de la valeur des opérations d'acquisition d'entreprises françaises.

Il s'agit de :

Groupe 1 : Etats-Unis (34.9%), Royaume-Uni (19.4%), Belgique (13.4%).

Suivent ensuite deux groupes de pays avec des parts relativement homogènes (voir tableau 8).

Groupe 2 : Allemagne, Italie, Espagne, Brésil et Pays-Bas.

Groupe 3 : Suède, Argentine, Australie, Suisse.

. Acquisitions étrangères en France

Le tableau 9 fournit le nombre et la valeur des acquisitions étrangères en France par pays pour l'ensemble de la période 1994-1998. Dans ce cas, que se soit en termes de nombre d'opérations ou de montant des acquisitions, on distingue trois groupes de pays (voir tableaux 10 et 11).

Groupe 1 : Etats-Unis (25.1% en nombre, 23.3% en valeur), Royaume-Uni (16.3% et 20.3%) et Allemagne (12.3% et 20%). Représentent à eux trois plus de 50% des acquisitions et plus de 63% de la valeur totale des acquisitions étrangères en France.

Groupe 2 : Italie, Belgique, Suisse, et Pays-Bas. Ce groupe représente 20,2% des acquisitions en nombre et 23.3% du montant total des acquisitions étrangères en France.

Groupe 3 : Canada, Multinationales, Suède, Irlande, Japon et le Brésil. Il représente 7.3% des acquisitions étrangères en nombre et 7.1% en valeur.

2.2 Etude détaillée des années 1994 et 1998

Dans cette partie, nous analysons plus en détail ce qui s'est passé pour les années de début et de fin de période. L'objectif est de répondre aux questions suivantes :

- . Quels types d'opérations sont utilisés par les entreprises ?
- . Les acquisitions se font-elles par des entreprises du même secteur ou au contraire d'autres secteurs ?
- . Quelle est la part des participations sans contrôle relativement à des prises de contrôle ?
- . Quelles sont les grandes tendances des mouvements de F & A en France ?

L'année 1994

. Acquisitions françaises à l'étranger

Dans cette section, nous analysons uniquement les principaux pays (pour l'année 1994) où des entreprises françaises ont acquis des actifs ou la totalité d'une entreprise. Ils sont au nombre de 6 : l'Allemagne, la Belgique, l'Espagne, l'Italie, le Royaume-Uni et les Etats-Unis.

Le tableau 12 montre qu'il y a très peu de fusions à proprement parler, puisque seules 2 fusions ont eu lieu avec des entreprises espagnoles (sur l'ensemble des autres pays on compte une seule avec une entreprise brésilienne). Dans la majorité des cas, les acquisitions françaises à l'étranger sont directement des prises de contrôle (à 100% ou moins du capital) ou des acquisitions d'activités.

Ainsi sur l'ensemble des opérations de F & A françaises à l'étranger (201), on recense 39 qui ont consisté en une participation dans le capital d'une entreprise étrangère et 15 qui ont concerné un relèvement de la participation sans contrôle effectif.

- . *Près de 75% des opérations de F & A initiées par des sociétés françaises à l'étranger ont pour finalité la prise de contrôle effectif, l'acquisition d'une activité, ou encore le relèvement d'un contrôle déjà existant (12 opérations sur les 201).*
- . *Sur les 201 opérations recensées, seules 3 ont pris la forme d'une OPA (Offre Publique d'Achat).*
- . *Dans les secteurs les plus représentatifs des opérations de F & A de l'année 1994, la grande majorité des firmes françaises ont acquis des sociétés du même secteur à*

l'étranger. Globalement, 66% des acquisitions ont lieu à l'intérieur d'un même secteur (voir tableau 13).

. Acquisitions étrangères en France

Dans cette section, nous analysons uniquement les principaux pays (pour l'année 1994) dont des entreprises ont acquis ou fusionné avec des entreprises françaises. Il s'agit des deux groupes de pays mentionnés plus haut (groupes 1 et 2).

Tout comme pour les acquisitions françaises à l'étranger, il s'avère qu'en pratique très peu de fusions sont initiées par des entreprises étrangères en France (voir tableau 14). On en recense 3 (deux avec des entreprises américaines et une avec une entreprise allemande). De plus, dans la majorité des cas, les acquisitions étrangères sont directement des prises de contrôle (à 100% ou moins du capital) d'entreprises françaises ou des acquisitions d'activités en France.

Ainsi sur l'ensemble des opérations de F & A étrangères en France (163), on en recense 29 qui représentent une participation dans le capital d'une entreprise française et 2 qui concernent un relèvement de la participation sans contrôle effectif.

- . Plus de 80% des opérations de F & A entreprises par des sociétés étrangères en France ont pour finalité la prise de contrôle effectif, l'acquisition d'une activité, ou encore le relèvement d'un contrôle déjà existant (5 opérations sur les 163).*
- . Sur les 163 opérations recensées, seules 3 ont pris la forme d'une OPA (Offre Publique d'Achat).*
- . La grande majorité des acquisitions d'entreprises françaises (149 sur 163) le sont par des entreprises étrangères du même secteur d'activité (voir tableau 15).*
- . En 1994, 13 filiales ou participations françaises à l'étranger ont été cédées à d'autres entreprises ou investisseurs français dans les principaux pays où les entreprises françaises ont des intérêts.*
- . 41 filiales ou participations françaises à l'étranger ont été cédées à des firmes étrangères dont 11 aux Etats- Unis (acquises pour 7 d'entre elles par des sociétés américaines).*

. Opérations de Fusions & Acquisitions franco-françaises

Le tableau 16 montre que sur les 537 opérations recensées en 1994, 92 ont concerné uniquement une participation dans le capital et 17 ont donné lieu à un relèvement de la participation sans prise de contrôle.

- . *Plus de 80% des opérations de F & A franco-françaises ont pour finalité la prise de contrôle effectif, l'acquisition d'une activité, ou encore le relèvement d'un contrôle déjà existant (45 opérations sur les 537).*
- . *Il y a eu 20 fusions entre des entreprises françaises.*
- . *Sur les 537 opérations franco-françaises recensées pour l'année 1994, 13 l'ont été sous forme d'OPA, 4 sous forme d'OPE (Offre Publique d'Echange), et 6 sous forme d'OPR (Offre Publique de Retrait)*

L'année 1998

. Acquisitions françaises à l'étranger

Dans ce qui suit, nous donnons quelques informations détaillées sur les acquisitions françaises à l'étranger (plus exactement dans ceux où le nombre d'opérations est supérieur à 10) pour l'année 1998. Nous en avons recensé ainsi dans 8 pays.

Le tableau 17 montre qu'il y a finalement très peu de fusions d'entreprises françaises avec des entreprises étrangères. En tout et pour tout, trois fusions ont eu lieu avec des entreprises américaines. Comme pour l'année 1998, la majeure partie des opérations de F & A initiées par des sociétés françaises à l'étranger sont directement des prises de contrôle (208 opérations sur les 335 recensées), à 100% ou moins, des relèvements de contrôle (13) ou des acquisitions d'activités (23). Par ailleurs, 65 opérations ont consisté en une participation dans le capital d'une entreprise étrangère sans contrôle effectif. 22 opérations ont consisté en un relèvement de participation dont 6 seulement sans prise de contrôle effectif. On constate enfin une participation croisée.

- . *Près de 80% des opérations de F & A initiées par des sociétés françaises à l'étranger ont pour finalité la prise de contrôle effectif, l'acquisition d'une activité, ou encore le relèvement d'un contrôle déjà existant.*
- . *Sur les 335 opérations recensées, 13 ont pris la forme d'une OPA et 4 d'une OPE. Cela représente environ 5% des opérations pour l'année 1998 alors que ce chiffre n'était que de 1.5% pour l'année 1994. On remarque ainsi une forte progression même si les chiffres restent relativement modestes.*

- . *Dans les secteurs où les entreprises françaises ont fait le plus d'acquisitions à l'étranger pour l'année 1998, une forte majorité (plus de 75% des cas) des firmes françaises acquièrent des sociétés du même secteur d'activité (voir le tableau 18).*

. *Acquisitions étrangères en France*

Dans ce qui suit, nous donnons le détail des acquisitions étrangères en France pendant l'année 1998.

A la lecture du tableau 19, il s'avère une fois de plus que très peu de fusions sont initiées par des entreprises étrangères en France durant l'année 1998. Ainsi, nous en avons recensé 1 seule, initiée par une entreprise américaine.

Sur l'ensemble des 255 opérations de F & A étrangères en France, on en recense 48 qui ont pris la forme d'une participation dans le capital d'une entreprise française et 3 qui ont concerné un relèvement de la participation, sans contrôle effectif. Enfin, on trouve également une participation croisée avec une entreprise allemande (Deutsche Telekom et France Télécom 2% de participations croisées).

- . *Près de 80% des opérations de F & A entreprises par des firmes étrangères en France ont pour finalité la prise de contrôle effectif, l'acquisition d'une activité, ou encore le relèvement d'un contrôle déjà existant (19 opérations sur les 255 recensées).*
- . *Sur les 255 opérations recensées, 16 ont pris la forme d'une OPA, soit un pourcentage de 6,3% (à comparer avec les 1.8% de l'année 1994).*
- . *Plus de 70% des opérations de F & A d'entreprises françaises par des entreprises étrangères le sont dans le même secteur d'activité. Quand ce n'est pas le cas, c'est essentiellement des holdings et des banques qui acquièrent des parts du capital des entreprises françaises (voir tableau 20).*
- . *Les principaux partenaires de la France acquièrent des entreprises dans tous les secteurs mentionnés ci-dessus. Le seul fait notable est que les entreprises britanniques et américaines acquièrent plus d'entreprises du secteur « Services pour les entreprises », avec respectivement 10 sur 43, et 14 sur 86.*

. *Opérations de Fusions & Acquisitions franco-françaises pour l'année 1998*

Le tableau 21 montre que sur les 553 opérations recensées, 78 ont consisté uniquement en une participation dans le capital d'une autre entreprise française, et 24 ont donné lieu à un relèvement de la participation sans prise de contrôle.

- . *Plus de 80% des opérations de F & A franco-françaises ont pour finalité la prise de contrôle effectif, l'acquisition d'une activité, ou encore le relèvement d'un contrôle déjà existant.*
- . *16 fusions ont eu lieu entre entreprises françaises.*
- . *Sur les 553 opérations franco-françaises pour l'année 1998, 29 l'ont été sous forme d'OPA, 3 sous forme d'OPE, 1 sous les deux formes, et 3 sous forme d'OPR.*
- . *37 opérations d'acquisition sont le fait d'une mise en redressement judiciaire (entreprises cédées par un administrateur judiciaire, contre 13 pour les acquisitions étrangères en France) et 26 opérations ont donné lieu à une cession de la part des fondateurs ou de familles (contre 10 pour les acquisitions étrangères).*
- . *Plus de 65% des opérations de F & A franco-françaises se font entre des entreprises du même secteur d'activité. Quand ce n'est pas le cas, c'est essentiellement les holdings et les banques qui sont partie prenante de ces acquisitions (voir tableau 22).*
- . *Il apparaît en plus de toutes les opérations analysées ci-dessus, 38 opérations ayant donné lieu à des cessions d'entreprises étrangères sous contrôle français (8) et à des cessions d'actifs d'entreprises françaises à l'étranger (en totalité ou en partie).*

2.3 Principales opérations de F & A pour la période 1994-1998

Dans cette dernière partie, nous nous intéressons aux principales opérations de F & A intervenues durant la période d'étude, 1994-1998. Pour ce faire, nous reprenons les données fournies par *Fusions & Acquisitions Magazine*. Ainsi, pour l'année 1994 nous considérons celles dont la valeur totale dépasse 500 MF. Pour les autres années, nous prenons en considération celles dont la valeur dépasse le milliard de francs.

Presque sans surprise (à part les secteurs de « Produits en Pierre, Argile, Verre et Béton » et celui de l'« Immobilier »), il apparaît que les secteurs qui sont importants en nombre et en valeur des opérations de F & A sur l'ensemble des années, sont aussi ceux où les opérations les plus importantes ont eu lieu.

Ainsi, si on reprend la classification faite auparavant, le secteur agrégé « banques, assurances, holdings » apparaît comme celui où le plus grand nombre de grosses opérations de F & A a eu lieu (soit 113 opérations sur un total de 464). La « distribution » arrive en deuxième position avec 50 opérations. Viennent ensuite les autres secteurs (voir tableau 23).

Les tendances générales que nous avons mentionnées pour l'ensemble des opérations de F & A restent valables.

Ainsi, il s'avère que sur l'ensemble des 284 opérations de F & A initiées par des entreprises françaises (et non leurs filiales à l'étranger), 164 ont concerné d'autres entreprises françaises, 23 des entreprises britanniques et 31 des entreprises américaines. On retrouve par ailleurs l'Allemagne et la Belgique avec 9 opérations dans chacun de ces deux pays.

De l'autre côté, on retrouve toujours les Etats-Unis avec 49 acquisitions d'entreprises françaises (dont 12 filiales américaines et 5 filiales dans d'autres pays) et le Royaume-Uni avec 32 opérations (dont 4 filiales britanniques et 1 filiale aux Pays-Bas). L'Allemagne vient en troisième position avec 18 opérations (dont deux filiales allemandes et une filiale belge).

De plus, dans la quasi totalité des opérations de grande envergure il s'agit de prise de contrôle, de relèvement d'un contrôle déjà existant ou d'acquisition d'une activité.

Enfin, la majorité des fusions les plus importantes sont internes à un même secteur d'activité.

3. Conclusion

L'examen des opérations de F & A trans-frontalières impliquant des entreprises françaises confirme les conclusions d'études antérieures portant sur l'Europe :

- . 80% des opérations ont pour objectif la prise de contrôle
- . Les F & A concernent des entreprises du même secteur dans les deux tiers des cas en 1994 et 75% des cas en 1998
- . les principaux secteurs concernés sont: la banque, l'assurance, la distribution, les services, la chimie-pharmacie, édition-communication.

L'hypothèse de la recherche d'une taille efficace supérieure à celle du pays d'origine semble donc confortée : dans le secteur de la banque et de l'assurance l'intégration européenne donne naissance à des entreprises de taille européenne. Dans le cas de la chimie-pharmacie ou de l'édition-communication, on peut penser que les entreprises visent à mettre en commun leurs

ressources en R & D ou à partager un avantage technologique. Beaucoup d'opérations, mais de moins en moins nombreuses, relèvent de la diversification.

Les données statistiques utilisées ne permettent pas à ce stade de vérifier les hypothèses de la mauvaise gestion, de la sous-évaluation ou de l'acquisition d'une position dominante. L'exploitation à venir de bases de données comptables et boursières ainsi que de données plus macroéconomiques (nombre d'entreprises dans l'industrie, prix et taux de change) devraient fournir des informations plus précises sur l'impact concurrentiel des opérations de F & A d'entreprises françaises.

Bibliographie

Anderson, Donald (1995) : «The European Union and the Outlook for Trade », *Business Economics*; 30(3), Juillet 1995, pages 17-23.

Derhy, Armand (1995) : « Les fusions et acquisitions en France de 1959 à 1992 : Evolution et caractéristiques », *Revue d'Economie Industrielle*; 0(73), 3^{ème} Trimestre, pages 19-44.

Garella, Paolo (1989) : « Fusions et acquisitions dans l'industrie européenne », *Observations et diagnostics économiques*, N° 29, Octobre, pages 185-219.

Gonzalez, Pedro, G.M. Vasconcellos, R.J. Kish and J.K. Kramer (1997): « Cross-Border Mergers and Acquisitions: Maximizing the Value of the Firm », *Applied Financial Economics*; 7(3), Juin, pages 295-305.

Torre-Enciso, Isabel-Martinez; Garcia, Javier-Bilbao (1996): «Mergers and Acquisitions Trends in Europe », *International Advances in Economic Research*; 2(3), Août 1996, pages 279-86.

II. Une évaluation de l'intensité concurrentielle: Unicité du prix et stratégies des entreprises dans l'Union Européenne.

Le rapport "The Economics of 1992"⁵ décrit dans les termes suivants les mécanismes grâce auxquels le Programme de Marché Interne (PMI) devait prendre forme:

"La levée des barrières et la liberté d'offre dont vont bénéficier les entreprises comme une conséquence du Programme de Marché Interne devrait conduire, à travers une pression concurrentielle accrue, à une convergence des prix vers un niveau inférieur, au bénéfice du consommateur. Du point de vue des producteurs, la pression concurrentielle s'exercera tout d'abord sur les marges, tout particulièrement dans les secteurs où ces producteurs auront eu un certain pouvoir de monopole. Les producteurs seront aussi amenés à réduire leurs coûts de production et de distribution pour devenir plus efficaces. La pression accrue qui s'exercera ainsi sur les coûts et les marges, sera un moyen puissant pour induire une convergence des prix vers des niveaux plus cohérents avec l'efficacité technique et économique. "

Or, si le processus d'intégration des marchés européens s'accélère depuis 1993, on est en droit de s'interroger sur les conditions de marché qui en résulteront, et notamment sur la convergence des prix. Cette problématique du prix unique dans un marché intégré prend un relief tout particulier au regard des politiques de la concurrence. La jurisprudence en matière de droit européen de la concurrence laisse en effet apparaître un double objectif de la Commission européenne : veiller au respect des règles de concurrence d'une part, mais aussi, d'autre part, "faire tendre les marchés vers un marché unique ". Ce dernier objectif se traduit notamment par une volonté assez nette de voir se former un prix unique pour un même produit au sein de l'espace européen. Mais l'unicité des prix est-elle une condition nécessaire et suffisante permettant d'affirmer qu'un marché est économiquement intégré ? Sinon, existe-t-il un seuil de dispersion des prix permettant de mesurer l'achèvement de l'intégration⁶ des marchés européens ? Le passage à la monnaie unique conduira-t-il à plus d'homogénéité des

⁵ European Economy -N° 35- Mars 1988

⁶ Si l'on entend par "marché intégré" un espace géographique dans lequel prévaut une seule politique fiscale et où les produits et services s'échangent sans barrières tarifaires par le biais d'une monnaie unique.

prix ? Quelles indications peut-on en déduire pour les politiques de prix des entreprises françaises ?

Nous proposons tout d'abord une étude empirique, mettant d'une part à jour les disparités des prix à la consommation *entre régions françaises en 1996* et comparant, d'autre part, la *dispersion de ces prix (à l'intérieur d'un même pays) avec la dispersion constatée entre pays membres de l'Union Européenne en 1993*.

Puis, nous nous intéressons à la problématique de traitement du *commerce parallèle* par la Commission Européenne. Ce type d'échange conduira-t-il à la convergence des prix ? Nous illustrons cette problématique par le cas de l'industrie automobile, en exposant tout d'abord les principales caractéristiques de ce marché, puis en montrant que les firmes peuvent avoir intérêt à laisser entrer sur le marché des mandataires réimportant les véhicules (notamment en France) depuis les pays européens où les prix sont plus bas. Dans ce cas, le commerce parallèle ne conduit pas à la convergence des prix, et l'arbitrage coexiste avec la segmentation des marchés.

1. Analyse empirique des écarts de prix à la consommation : une comparaison régions françaises vs pays membres de l'U.E.

Si les coûts de transaction liés aux taux de change et à leurs variations disparaissent, les firmes devraient être incitées à uniformiser leurs prix. Cependant, l'étude des disparités de prix entre régions françaises tend à montrer que dans un marché parfaitement intégré, comme celui que forme un pays, des écarts de prix persistent et sont parfois même relativement importants. Une monnaie unique n'aurait alors sur le marché européen qu'un impact incomplet sur la convergence des prix vers un prix unique. Ce mécanisme ne semblerait donc pas être le plus pertinent pour décrire le degré d'unification d'un marché.

L'analyse statistique descriptive nous permet de comparer les disparités de prix à la consommation entre régions françaises avec celles mises à jour entre les Etats Membres de l'Union européenne (U.E.). Nous nous sommes fondés sur un rapport réalisé pour la Commission Européenne⁷. Ce rapport étudie la convergence des prix dans l'U.E. après la mise en place du programme de marché unique. Il apparaît que, malgré une convergence accrue des

⁷ D.R.I, E. de Ghellinck, Horack, Adler & Associates, " Analysis of Changes in Price Disparities in the E.U. following the Launch of the Internal Market Programme ", Février 1996

prix après janvier 1993, des différences de prix persistent pour la plupart des biens et services. Les résultats que nous avons obtenus en comparant les prix à la consommation entre les régions françaises montrent que, pour l'année 1996, sur l'échantillon de produits et services sélectionné⁸, des écarts de prix par catégorie de biens et services existent et sont relativement significatifs pour certains produits.

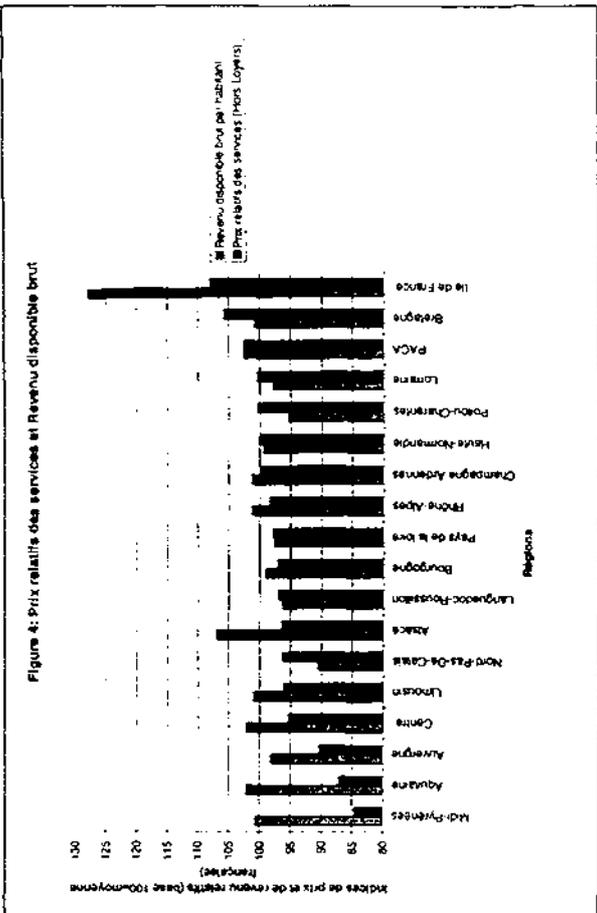
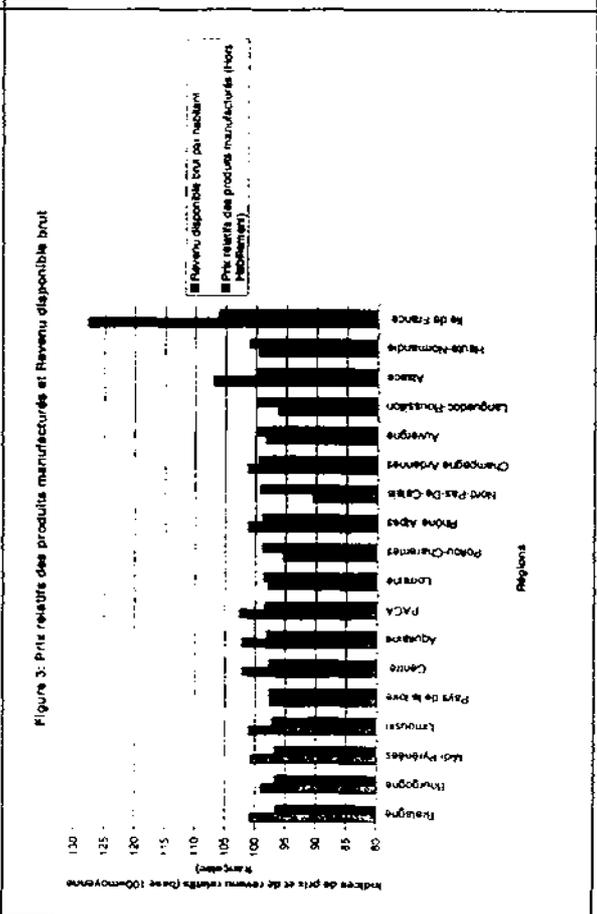
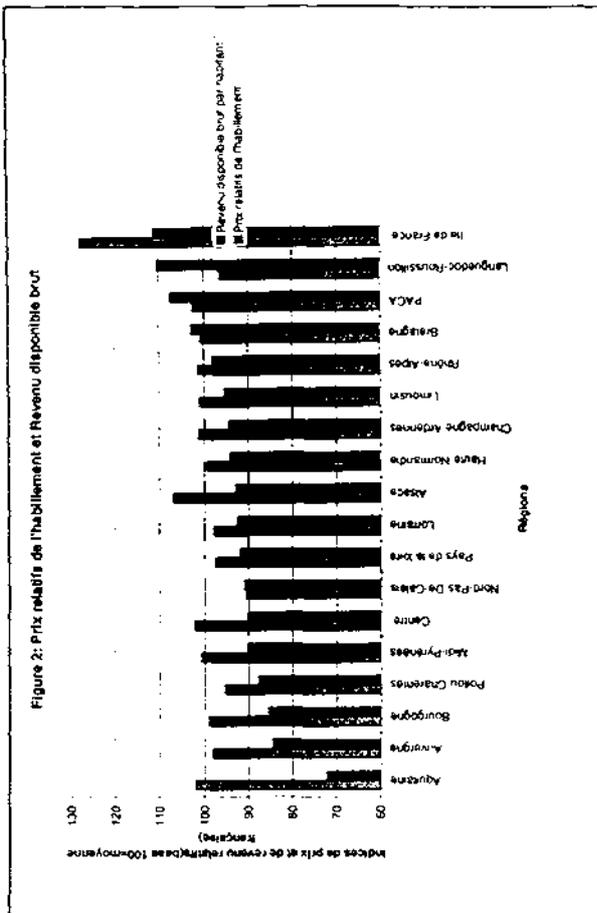
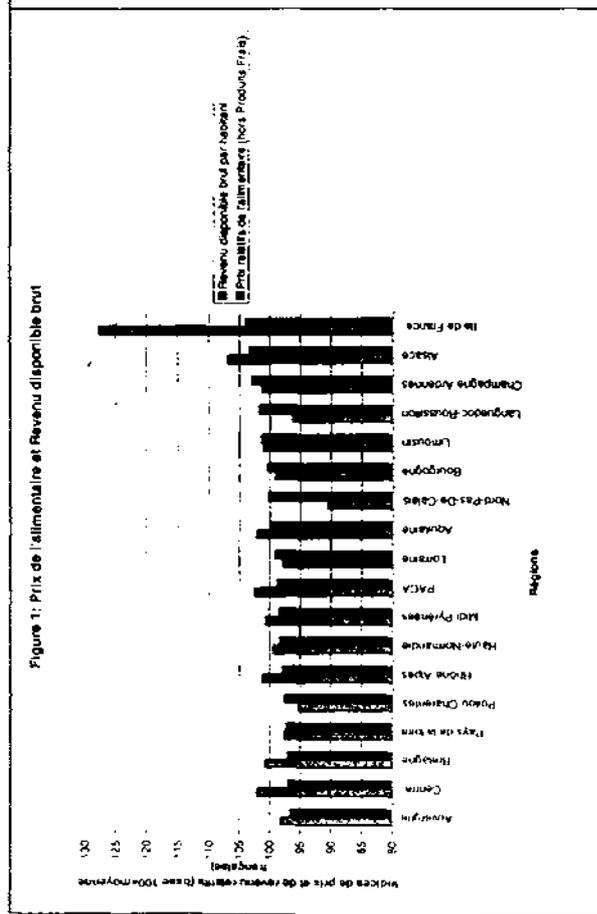
Les figures 1 à 4, donnent les indices de prix relatifs dans les différents secteurs de consommation (alimentaire, biens manufacturés, habillement, services) en utilisant comme base 100 le prix moyen France entière (métropolitaine). Par ailleurs, nous avons relativisé ces indicateurs par le revenu disponible brut des ménages par région. Pour chaque groupe de produits sélectionné (Cf. Encadré 1) on a calculé des prix moyens régionaux, puis des indices de prix relatifs de ces prix par rapport au prix moyen France entière. L'Ile-de-France, tout en ayant le revenu disponible brut le plus élevé, est aussi en tête des régions les plus chères et son indice de prix relatif moyen peut être supérieur à 110 (habillement). Quant aux régions les moins chères en moyenne, on retrouve l'Aquitaine, l'Auvergne ou la Bretagne, avec des indices oscillant entre 85 et 95.

Intuitivement, la dispersion des prix des biens et services *entre les régions d'un pays* de l'U.E. devrait être plus faible que la dispersion des prix *entre les pays* de l'U.E. Le tableau 1 confirme cette intuition, tout en révélant cependant, pour la France, des coefficients de variation des prix non négligeables. Les coefficients de variations entre pays de l'U.E. sont, en moyenne, supérieurs de 13.7 points à ceux calculés entre régions. La différence va d'une fois et demi (pour l'habillement) à trois fois et demi (produits manufacturés). Comme dans l'étude européenne, nous constatons que ce sont les produits les moins "échangeables" qui font apparaître la dispersion la plus forte. Dans notre échantillon il s'agit des services et de l'énergie.

Le résultat selon lequel on trouve en moyenne, pour un même type de produit, moins d'écarts de prix entre deux régions qu'entre deux pays peut signifier que, lorsque les conditions "institutionnelles" sur les marchés sont homogènes -monnaie unique, même politique fiscale- les opportunités qu'ont les firmes d'adapter leur stratégie de prix selon le marché

⁸ Nous avons reconstitué les postes PPA (Parités de Pouvoir d'Achat) des données Eurostat utilisées pour l'étude européenne à partir de la nomenclature Indice des Prix à la Consommation (IPC). La base de données a été constituée à partir des séries mensuelles de l'IPC.

géographique semblent réduites. Toutefois, force est de constater que les prix ne sont pas homogènes dans les régions, ni, *a fortiori*, entre ces régions.



**Tableau 1: Comparaison de la variabilité des prix Etats Membres de l'Union Européenne/
régions françaises**

Secteur	Dispersion des prix entre pays membres de l'Union Européenne ⁹	Dispersion des prix entre régions françaises
Alimentaire	22%	6.7%
Habillement et textile	21.2%	14.5%
Autres produits manufacturés	20.1%	5.7%
Services	28.5%	10.4%

Source: calculs à partir de données I.N.S.E.E et Eurostat

Dispersion des prix mesurée par le coefficient de variation (ratio écart type / moyenne):

$$C_I = \sigma_I / I$$

Encadré 1: Méthodologie

L'année 1996 a été retenue pour des raisons de facilité d'accès aux données. Cependant, les résultats de l'enquête spatiale effectuée par l'INSEE en 1989 montrent que le profil de dispersion des prix entre régions n'a pas changé. Le choix de l'année pour la construction des indices de prix relatifs en France n'est donc pas déterminant.

La dispersion des prix est observée entre régions et non pas entre agglomérations parce que le nombre de relevés par agglomération et par variété ne nous permet pas d'obtenir des prix moyens statistiquement corrects. 96 villes sont enquêtées, soit, en moyenne, entre 4 et 5 agglomérations par région. Afin de reconstituer les postes Parités de Pouvoir d'Achat (P.P.A.) on utilise les pondérations des variétés de produits ou services qui constituent ces postes dans chaque agglomération. Ainsi, un prix moyen annuel par variété de l'indice et par agglomération est calculé. Puis on agrège celui-ci sur toutes les agglomérations d'une même région en utilisant le poids de chaque agglomération dans la consommation nationale de la variété. On construit ensuite des indices de prix relatifs (prix moyens par région par rapport à la moyenne France entière (=100). Enfin, en pondérant chaque variété par le poids qui lui est

⁹ 1993 est l'année retenue pour l'Union Européenne (dernières données disponibles)

accordé dans l'indice, on aura construit un indice de prix relatif du poste, qui correspond en fait à la dispersion des prix d'un panier représentatif du poste.

On calcule la **dispersion** des prix d'un poste de la même manière qu'elle a été calculée au niveau européen, c'est à dire en utilisant le coefficient de variation.

2. La question du commerce parallèle sur le marché automobile européen

L'analyse précédente semble confirmer l'idée selon laquelle les marchés européens restent segmentés, et ce, malgré un processus lent d'intégration économique et monétaire à l'oeuvre depuis 1957. Aujourd'hui, une monnaie unique a remplacé un ensemble de systèmes monétaires ayant donné lieu à des périodes de fluctuations parfois importantes. Les barrières commerciales entre les 15 pays membres ont disparu et la libre circulation des personnes, des biens et des capitaux permettent aux entreprises d'agir dans un vaste marché géographique. Cependant, il subsiste une part non négligeable d'hétérogénéité tant institutionnelle que liée aux caractéristiques des offreurs et des demandeurs, et induisant de la part des entreprises des stratégies adaptées suivant les pays. Ce raisonnement s'applique en particulier à l'analyse du marché automobile.

2.1 Les facteurs d'hétérogénéité macroéconomique

Ainsi, les politiques de taxes indirectes sont encore très différentes suivant le pays européen, comme le montre le tableau 2. L'implication en termes de politique tarifaire pour les firmes françaises comme étrangères est immédiate. Dans des pays comme le Danemark, où les taxes sont très élevées, les constructeurs pratiquent des prix hors taxe très faibles pour conserver leur part de marché. Les consommateurs des autres États membres peuvent effectuer leur achat au Danemark, puisque seule la T.V.A. est acquittée dans le pays d'origine de l'acheteur. Ainsi, les disparités de politiques fiscales sont-elles une des raisons importantes de l'existence du commerce parallèle en Europe.

Par ailleurs, jusqu'en mai 1998, les variations des taux de change des monnaies européennes ont induit de la part des entreprises, et en particulier dans l'automobile, des comportements dits de "*pricing to market*". Lorsque la monnaie locale s'apprécie, la firme exportatrice peut soit laisser son prix tel quel (la compensation de la variation de change sur le prix à l'exportation est nulle), soit compenser cette appréciation en augmentant son prix, mais en risquant de voir diminuer sa part de marché (compensation complète de la variation du taux de change) soit enfin utiliser un compromis entre ces deux stratégies (compensation incomplète). Lorsque la firme adapte son comportement de prix aux variations du taux de change, comme cela a été le cas dans la période 1993/1994, les prix ont chuté, traduisant une stratégie de gain de parts de marché de la part des entreprises sur les marchés britannique et italien.

Enfin, des facteurs liés à la demande, comme les préférences pour les marques domestiques, ou liés à l'offre comme des différences dans les coûts des entreprises ou la présence de marques domestiques sur certains marchés, impliquent que l'uniformisation des prix sur le marché automobile européen est encore loin d'être atteinte.

Le tableau 3 indique, pour 5 pays, les parts de marché par groupe automobile en 1998. Trois de ces marchés ont plusieurs marques domestiques (l'Allemagne, la France et l'Italie) et l'on peut voir que ces marques domestiques ont la plus grande part de marché dans ces pays (45.6% en Allemagne -avec les groupes Volkswagen, BMW et Mercedes, 57.2% en France -groupes P.S.A. et Renault, et 39% pour le groupe Fiat en Italie). Ceci reflète bien la préférence des consommateurs pour les marques domestiques. Par ailleurs, la Belgique comme le Danemark n'ont aucune marque domestique (la Belgique est producteur pour des marques étrangères). Une concurrence plus forte (plus " fluide ") s'exerce et, pour les deux pays, le taux de pénétration des marques japonaises (soumises à des quotas jusqu'en l'an 2000) est beaucoup plus important que pour les trois premiers pays (15.8% pour la Belgique et 25.1% pour le Danemark). On trouve ici une justification des hypothèses posées dans l'analyse du commerce parallèle (Cf. ci-après).

Des conditions de concurrence hétérogènes -en termes de structure de l'offre comme de la demande- règnent dans l'Union Européenne.

Tableau 2 : Taxes d'achat de véhicules dans l'Union Européenne en 1999

Pays	T.V.A.(%)	Taxes d'achat	Frais d'immatriculation (en monnaie locale)
Allemagne	16	Aucune	50 DM
Autriche	20	Basée sur la consommation d'essence. Maximum=16%	842 à 1269 SCH
Belgique	21	Basée sur la cylindrée et l'âge	2500FB
Danemark	25	< 50800 DKR, 105 % > 50800 DKR, 180 %	1070 DKR
Espagne	16	< 1.6 l 7% > 1.6 l 12%	10250 PTA
Finlande	22	100 % - 4600 FM	Aucun
France	20.6	Aucune	Taxe locale 102 à 195

			FF(cartes grise)+charges parafiscales
Grèce	18	16 à 128 %	Aucun
Irlande	21	22.5 à 30 % suivant la cylindrée	Aucun
Italie	20	Provinciale basée sur la puissance fiscale	300000 à 720000 L
Luxembourg	15	Aucune	1128 LFR
Pays-Bas	17.5	Essence : 45.2 % - 3394 G Diesel : 45.2 % - 1278 G	Aucun
Portugal	17	Basée sur la cylindrée	5000 ESC
Royaume-Uni	17.5	Aucune	Aucun
Suède	25	Aucune	Aucun

Source : ACEA.

1 Euro = 13.76 SCH, 38.38 FB, 7.07 DKR, 5.66 FM, 6.56 FF, 1.86 DM, 1842.22 L, 38.38 LFR, 2.10 G, 200.48 ESC, 158.30 PTA

Tableau 3 : Parts de marché dans 5 pays de l'Union Européenne en 1998

Constructeurs/Pays	Allemagne	Belgique	Danemark	France	Italie
P.S.A. Peugeot Citroën	3.5%	15.6%	15.6%	28.2%	6.6%
Peugeot	1.2%	7.2%	7%	11.6%	3%
Citroën	2.3%	8.4%	8.6%	16.6%	3.6%
Renault	6.3%	10.6%	3.7%	29%	7.5%
Groupe Fiat	4.4%	5.2%	6.8%	6.7%	39%
Groupe Volkswagen	28.2%	19.7%	19.9%	11%	12%
Ford	9.8%	8.8%	9.4%	6.9%	8%
General Motors	14.5%	10.3%	10.2%	6.4%	9.1%
BMW Rover	7.3%	4.7%	1.9%	2.9%	3.5%
Mercedes	10.1%	3.7%	1.3%	1.9%	2.6%
Volvo	1.1%	2.6%	1.7%	0.5%	1%
Marques Japonaises	12%	15.8%	25.1%	4.8%	6.4%
Marques Coréennes	1.6%	2.1%	3.9%	1%	3.6%

Source : CCFA

Groupes : P.S.A. = Peugeot + Citroën

Fiat = Alfa Romeo + Fiat + Lancia

General Motors = Opel + Vauxhall + Saab

Volkswagen = Volkswagen + Audi + Seat + Skoda

2.2 Une analyse microéconomique des disparités de prix

Le commerce parallèle consiste, pour une firme, à acheter un bien dans un pays où son prix est bas pour le revendre dans un pays où le prix est élevé en concurrençant les réseaux de distribution exclusive du bien. Dans l'Union Européenne, ce type d'échange a lieu pour toute une catégorie de produits pour lesquels on constate des différentiels de prix importants entre les pays membres. Le marché automobile en est un bon exemple. Les constructeurs français occupent, comme nous avons pu le voir précédemment, une part de marché non négligeable en Europe, tout en pratiquant des prix très différents d'un pays à l'autre. Par exemple, en mai 1999, le prix hors taxes de la même Peugeot 106 pouvait varier, de 38% entre le pays le moins cher (le Danemark) et le pays le plus cher (l'Allemagne). Si la Commission Européenne protège cet arbitrage au nom de la liberté du commerce entre États Membres et de l'unification des marchés, elle autorise aussi, dans le secteur automobile, la distribution exclusive, qui facilite la segmentation des marchés. La Commission semble agir comme si tout arbitrage était socialement efficace et pouvait, *in fine*, conduire à la convergence des écarts de prix vers un prix unique, plus concurrentiel. Cependant, l'offre du bien par un importateur parallèle (mandataire) ne pourrait-elle pas coexister avec un réseau de distribution exclusive ou sélective sans que cela ne coûte au producteur ? Autrement dit, peut-il y avoir de la discrimination par les prix malgré la présence d'un arbitragiste ?

Nous proposons une analyse par un modèle de concurrence en prix où un producteur (un constructeur français) vend son bien dans deux pays¹⁰. Il est en monopole dans un pays et en duopole dans l'autre. Une situation plus concurrentielle dans le deuxième pays l'oblige donc à y fixer un prix plus bas que dans le premier, l'écart de prix laissant alors la possibilité à un importateur parallèle d'entrer sur le premier marché afin de réimporter le bien (le véhicule de marque française). Les deux biens ("autorisé", offert par le réseau et "non autorisé") sont des substituts imparfaits en raison d'un "coût d'achat" à cet arbitragiste, lié à des garanties moindres - services après-vente, temps d'attente plus longs. Dans ce contexte, nous montrons que pour certaines configurations des paramètres, les deux offreurs coexistent à l'équilibre et

¹⁰ Le modèle formel est développé en annexe.

que le producteur a même intérêt à laisser entrer l'arbitragiste, qui lui permet de discriminer parmi les consommateurs du pays à prix élevé (la France). Ainsi, dans ce modèle, nous mettons en évidence des mécanismes qui expliquent pourquoi les différentiels de prix, dans le secteur automobile par exemple, entre deux pays européens pourraient ne pas converger vers la " loi du prix unique " malgré la protection de l'arbitrage par la Commission Européenne. Dans ce contexte, les constructeurs français peuvent continuer à appliquer une politique de prix différenciés par marché, tout en laissant faire les nombreux mandataires agissant pour le compte de consommateurs français.

3. Conclusion

L'harmonisation des prix en Europe a été prédite depuis plusieurs années sans pour autant que l'on ait constaté de grands changements dans la dispersion des prix. Au niveau agrégé, le rapport sur la convergence des prix en Europe (Cf. I.) montre que cette convergence a bien eu lieu depuis 1980 et qu'elle s'est même accélérée après 1993. Cependant, en 1993, il existait encore dans de nombreux secteurs, des différences de prix non négligeables entre les pays membres de l'U.E.

Au niveau d'une industrie comme l'industrie automobile, la convergence des prix a tout lieu de prendre prochainement place, et ce, en raison de la convergence de plusieurs facteurs, comme l'harmonisation fiscale (taxes d'achat). Par ailleurs, les réseaux de distribution actuels (concessions exclusives) sont remis en question par l'arrivée à échéance du dernier règlement d'exemption [...] en 2002. Il se pourrait, là encore, que l'entrée de la grande distribution dans la distribution automobile conduise à se réduire les différences de prix entre les marchés.

DEUXIEME PARTIE

La grande distribution et ses relations avec le secteur productif en France

Introduction

Depuis les années 1970, les entreprises de grande distribution françaises ont connu un développement rapide. Les chiffres d'affaires des plus grands groupes, autour de 200 milliards de francs annuels, montrent bien que la distribution est devenue une activité particulièrement importante dans l'économie française. L'annonce, le 30 août 1999, de la fusion entre Carrefour et Promodès¹¹, a suscité de nombreuses réactions, tant chez les concurrents de ces deux distributeurs, que chez leurs fournisseurs ou chez les autorités en charge de la politique de la concurrence.

Dans cette section, on s'attachera à dresser un tableau de la situation des entreprises de grande distribution françaises, et de leurs rapports avec les producteurs, en particulier dans le secteur agro-alimentaire. On donnera dans un premier temps un aperçu général de la place du commerce et de la grande distribution alimentaire française dans le monde, et plus particulièrement en Europe. Ensuite on s'intéressera à certains points plus précis, en menant un analyse fondée sur des travaux théoriques récents. Ainsi, on développera plus particulièrement les thèmes de la concentration de la distribution en France et du développement des marques de distributeurs, qui semblent des points cruciaux pour comprendre le fonctionnement actuel de ce secteur.

1. Les acteurs et le marché.

Si le libre-service et autres méthodes de ventes caractéristiques des grandes surfaces ont vu le jour aux Etats-Unis et n'ont pas été adoptés en France avant la fin de la seconde guerre mondiale (c'est seulement à la fin des années cinquante que les premiers supermarchés ont été créés), les premiers hypermarchés sont en revanche apparus en France, dans les années soixante. En 1963, le premier hypermarché Carrefour, d'une surface de 2500 m², a été créé en région parisienne. D'autres sociétés ont suivi dans la voie ouverte par Carrefour : le premier hypermarché Auchan s'est ouvert à proximité de Tourcoing en 1967, Promodès a inauguré son

¹¹ Si l'opération est acceptée par les autorités en charge de la politique de la concurrence, le groupe ainsi formé occupera la première place de la grande distribution française avec 26.7% de parts de marché, et la deuxième place mondiale avec 54 milliards d' Euros de chiffre d'affaires.

premier hypermarché en 1970 à Mondeville (Calvados), et Casino a créé son premier hypermarché à Marseille en 1970.

Depuis 1970, l'importance de la grande distribution dans l'activité commerciale française n'a cessé de croître. D'abord cantonnés au secteur alimentaire, les magasins des grands distributeurs ont élargi leur champ d'activité à de nombreux autres domaines, et proposent maintenant un éventail de produits très vaste.

1.1. Structure du marché national.

Le réseau de distribution alimentaire français est dense. Il réunit à la fois de grandes surfaces modernes (supermarchés et hypermarchés essentiellement) et des magasins de type traditionnel, qui sont toutefois en déclin. Depuis 1970, la part de marché des grandes surfaces dans le commerce alimentaire a doublé, au détriment du petit commerce. Actuellement, près des deux tiers des achats alimentaires se font dans les hypermarchés ou les supermarchés. Le tableau suivant illustre cette évolution en donnant la part du marché alimentaire commercialisé dans chaque type de magasin en France:

En %	1970	1980	1996	1997	1998
Hypermarchés (>2500 m ²)	3,6	14,3	36,8	37,7	38,0
Supermarchés (entre 400 et 2500 m ² , y compris hard discount)	9	16,8	31,2	31,2	31,3
Autres grands commerces (grands magasins, ...)	20,7	13,7	1,5	1,5	1,5
Petit commerce (y compris spécialistes)	66,7	55,2	30,5	29,6	29,2

La densité commerciale française est l'une des plus fortes d'Europe: 235 m² de surface commerciale par habitant en 1998. Cette densité montre bien que le marché national semble parvenu à maturité, laissant aux groupes de distribution français des perspectives de croissance du marché national assez faibles. En outre, la nouvelle réglementation a contribué à freiner les nouvelles implantations de grandes surfaces. En particulier, les ouvertures de

grandes surfaces de plus de 300 m² sont réglementées depuis 1996 par la loi Raffarin¹², qui a renforcé la loi Royer de 1973 : toutes les constructions de super- ou d'hypermarchés sont soumises à l'autorisation des Commissions Départementales d'Équipement Commercial, qui ont aussi un droit de regard sur les extensions de grandes surfaces existantes. Depuis 1996, on a constaté une très forte diminution du nombre d'ouvertures de supermarchés : de plus de 300 ouvertures par an entre 1986 et 1995, on est passé à moins de 200 ouvertures annuelles.

Ces facteurs combinés incitent les distributeurs à développer une stratégie de croissance externe sur le marché national, comme nous le verrons dans la section III, et à développer leur activité internationale en s'implantant dans d'autres pays. En outre, les stratégies commerciales se modifient, les distributeurs privilégiant le développement de leurs marques propres pour disposer d'un pouvoir accru sur leurs fournisseurs. Nous analyserons ce phénomène dans la partie IV de cette section.

1.2. Les grands groupes.

L'analyse des parts de marché des distributeurs français laisse apparaître une structure clairement oligopolistique, dominée par six acteurs principaux, de tailles comparables. Le chiffre d'affaires mondial des groupes les plus importants est actuellement proche de 200 milliards de Francs. En comparaison, le plus gros groupe industriel français de l'agro-alimentaire, Danone, réalise un chiffre d'affaires deux fois plus faible : 88 milliards de francs en 1997. Le tableau suivant donne l'évolution de la répartition des parts de marché des grands groupes sur le marché français de la grande distribution généraliste, de 1993 à 1997 :

(en %)	1993	1994	1995	1996	1997
Intermarché	14.6	14.7	15.1	15.1	15.5
Leclerc	14.7	15.0	15.2	15.3	15.2
Auchan	6.9	6.9	7.0	12.5	12.8
Carrefour	12.6	12.8	13.1	12.9	12.6
Promodès	10.2	10.8	11.1	11.0	11.1
Casino	7.6	7.2	7.3	7.5	9.8
Docks de France ¹³	5.6	5.7	5.8	-	-

¹² Loi du 5 juillet 1996 relative au développement et à la promotion du commerce et de l'artisanat.

¹³ Racheté en 1996 par Auchan.

Système U	5.4	5.5	5.6	5.7	5.9
Cora	4.4	4.4	4.4	4.3	4.0
Comptoirs Modernes ¹⁴	2.7	2.8	2.9	3.0	3.2

Le statut juridique du réseau des groupes permet de distinguer deux types d'enseignes. D'une part, le commerce intégré comprend les entreprises exerçant simultanément les activités de grossiste et de détaillant, comme Carrefour, Promodès, Auchan ou Casino. D'autre part, le commerce associé regroupe des entreprises juridiquement distinctes, fédérées par contrat ou par convention autour d'une centrale d'achat, laquelle centralise les activités de grossiste : ceci leur permet en général d'exploiter une enseigne commune. Ainsi, les magasins des groupements Leclerc, Super U ou Intermarché appartiennent à des entrepreneurs indépendants. En revanche, les magasins Carrefour, Casino, ou Auchan sont soit des succursales de ces groupes, c'est-à-dire que les dirigeants des magasins sont des salariés de l'entreprise, soit des franchisés. L'une des principales caractéristiques de la grande distribution française est l'importance des groupements d'indépendants : en 1997, les groupements Leclerc, Système U et Intermarché réalisaient 36.6%¹⁵ des ventes de la grande distribution.

1.3. La place des grands groupes français de distribution sur le marché international,

L'internationalisation des chaînes de grande distribution française est un phénomène déjà relativement ancien, puisqu'il a commencé au début des années soixante-dix. Il se poursuit à présent à un rythme plus soutenu que le développement national des firmes. Les succursales situées à l'étranger, en particulier dans les pays d'Europe du Sud ou de l'Est, sont en général plus rentables que les magasins situés en France. Les grands groupes français se sont développés essentiellement en Europe (surtout en Espagne), mais aussi en Amérique (où Carrefour possède 78 hypermarchés), et ils s'implantent progressivement en Asie (Carrefour y exploitait déjà 28 hypermarchés en 1997). Les entreprises françaises du grand commerce intégré se développent ainsi dans les pays étrangers, alors que peu de groupes étrangers ont pénétré le marché français, hormis les chaînes de maxidiscounte.

¹⁴ Racheté en 1998 par Carrefour.

¹⁵ source Nielsen

Les premiers distributeurs qui ont développé leur activité à l'étranger sont Carrefour, Promodès et Auchan. Ainsi, Carrefour et Promodès se sont installés dès les années 1970 en Espagne. Actuellement, en termes de chiffre d'affaires, Promodès est le premier distributeur espagnol, et Carrefour le second.

Tableau 15 : Chiffre d'affaires des principaux groupes français à l'étranger, 1997

Groupe	CA réalisé à l'étranger (en Mds FF)	% du CA total réalisé à l'étranger
Carrefour	73	43
Promodès	66	39
Auchan	23	19
Cora	11	24
Casino	8	11
Comptoirs Modernes	2	7

Les grands distributeurs français ont concentré leur développement international sur l'Europe, à l'exception de Carrefour qui a massivement investi dans les pays émergents d'Amérique Latine et d'Asie. Actuellement, Promodès réalise 39% de son chiffre d'affaires à l'étranger, mais essentiellement en Europe: 20% en Espagne, 7% au Portugal, 5% en Italie, et 4% en Grèce. En revanche, Carrefour, qui réalise plus de 43 % de son chiffre d'affaires à l'étranger, est largement implanté hors du continent européen : 15% au Brésil, 12% en Espagne, 6% en Argentine, et 4% à Taiwan.

En Europe, la Grèce et l'Italie offrent encore des perspectives de débouchés car la législation y a retardé le développement de la grande distribution. En Espagne, les deux premiers groupes de distribution sont français, et le marché semble atteindre son niveau de saturation. Les pays d'Europe de l'est, qui représentent 70 millions de consommateurs, attirent la plupart des grands distributeurs qui ont commencé à s'y implanter. Ainsi, Casino et Auchan installent des magasins en Pologne, où le pouvoir d'achat croissant des consommateurs et le faible développement des structures commerciales permettent de prévoir un fort potentiel de croissance du secteur commercial. Finalement, la tendance relativement ancienne à l'internationalisation des groupes français semble actuellement se poursuivre et même s'accroître.

Il convient de noter que les indépendants français sont en retard sur les autres groupes, leur développement international se limitant aux pays frontaliers. Ce retard semble dû à leur structure juridique reposant sur un système d'adhésion : un gérant adhérent sur un nouveau marché ne bénéficie pas du soutien financier du groupement et porte seul tous les risques. Intermarché a contourné partiellement ce problème au travers d'une alliance avec un groupement d'indépendants européens Spar Handel AG. Leclerc et Système U sont quant à eux quasi-absents du marché international.

2. Les relations entre producteurs et distributeurs

2.1. Les rapports de force entre producteurs et distributeurs

Lorsque la grande distribution a commencé à se développer, le rapport de forces entre les distributeurs et leurs fournisseurs était clairement en faveur de ces derniers. La réputation des marques des producteurs, encore accentuée par le développement et le succès de la publicité dans les années cinquante, faisait leur force: un distributeur se devait de proposer dans son magasin les marques les plus connues, aux prix les plus bas. Dans ce contexte, les distributeurs, isolés, avaient une marge de manoeuvre restreinte dans leurs négociations avec les producteurs. Les petits commerçants ne pouvaient pas négocier facilement les prix, et étaient soumis aux menaces de rupture d'approvisionnement des producteurs de grandes marques lorsqu'ils se proposaient de distribuer des produits concurrents.

Peu à peu, l'augmentation de la taille des distributeurs et la diminution du nombre d'acteurs dans la distribution, liées au développement de quelques enseignes et à la concentration du secteur, a inversé le rapport de forces, en concentrant les débouchés des producteurs entre un petit nombre de clients. La « puissance d'achat » des distributeurs, c'est-à-dire leur pouvoir de marché en tant qu'acheteurs, s'est considérablement accrue : cette notion recouvre la capacité d'une firme à acheter en grande quantité, et est associée à la possibilité pour cet acheteur d'obtenir des conditions commerciales avantageuses.

D'autres facteurs ont contribué à accroître le pouvoir de négociation des distributeurs face aux producteurs. La crise économique du début des années quatre-vingt a renforcé les exigences en termes de prix des consommateurs, qui se sont reportés sur des produits moins chers, dont l'image de marque était moins développée. Le développement des marques de distributeurs a alors répondu à cette évolution des préférences des consommateurs : l'amoindrissement de l'impact de la publicité sur les comportements d'achat a diminué la puissance des producteurs

de « grandes marques » qui ont dû s'adapter à l'apparition de cette nouvelle forme de concurrence, provenant directement de leurs distributeurs.

Il est bien entendu difficile de mesurer ces changements survenus dans les rapports de force entre producteurs et distributeurs. Différents indices convergents permettent cependant de soutenir la thèse du renversement du rapport de forces. Ainsi, il est notoire que, depuis quelques années, la fidélisation des clients n'est plus l'apanage des marques. D'après une enquête réalisée par l'INSEE en 1997 sur la distribution alimentaire, lorsque leur marque habituelle (pour un produit alimentaire ou non) n'est pas présente dans leur grande surface principale, 56% des consommateurs achètent un produit d'une marque équivalente, 26% reportent leur achat à une visite ultérieure, et seulement 13% vont dans un autre magasin¹⁶. Les reports des consommateurs d'une marque à l'autre semblent donc plus fréquents que leurs reports d'une enseigne à l'autre. Cependant, il faut nuancer ce résultat en remarquant qu'une part non négligeable des consommateurs déclare reporter l'achat à une visite ultérieure, ce qui peut masquer une visite à un autre magasin, car en moyenne, d'après la même étude de l'INSEE, les consommateurs français fréquentent régulièrement plusieurs grandes surfaces (entre deux et trois), vraisemblablement d'enseignes différentes.

D'autres manifestations du renversement des rapports de force entre producteurs et distributeurs se font jour. Il semble que les marges pratiquées par les distributeurs et les producteurs aient connu des évolutions opposées depuis une quinzaine d'années. En outre, la valeur boursière des grands distributeurs croît plus rapidement que les différents indices boursiers français, notamment le CAC 40 ou le SBF 120. Cette évolution indique comment les marchés financiers perçoivent l'évolution des performances des distributeurs, relativement à celles de leurs fournisseurs, dont les cours, pour la plupart, augmentent moins vite ou ont tendance à chuter.

En pratique, l'accroissement du pouvoir des distributeurs se manifeste par des exigences tarifaires accrues. Ils exercent des pressions sur leurs fournisseurs pour réduire le plus possible les prix, mais aussi pour obtenir des rabais et ristournes de fin d'année, ainsi que des droits de référencement¹⁷ et d'autres participations aux frais commerciaux. Les « marges arrières » des distributeurs ont augmenté, c'est-à-dire qu'ils ont fait des profits en négociant des tarifs de

¹⁶ Les 5 % restant ne se prononçant pas dans l'enquête citée.

¹⁷ Un produit est référencé par un distributeur lorsqu'il fait partie de la liste des produits disponibles chez ce distributeur.

gros plus faibles sans changer le niveau des prix de détail. Dans ces négociations, les distributeurs disposent d'un éventail de moyens de pression allant d'un simple ralentissement de l'effort de vente au déréférencement d'un ou plusieurs produits de la marque.

En effet, sur certains marchés, notamment alimentaires, les producteurs ont très peu de débouchés autres que la grande distribution, car le marché de la vente aux professionnels par exemple est assez fermé. Dans ce cas, les grandes chaînes de distribution sont des interlocuteurs très puissants : il n'est pas rare qu'une enseigne représente plus de 20% des ventes d'une marque. On comprend bien alors l'impact d'un déréférencement sur le fournisseur. Cependant, les menaces de déréférencements sont surtout à craindre pour les petits producteurs. En effet, les grands producteurs internationaux, multi-produits, fournissent une part importante des biens vendus par les grands distributeurs. Ces entreprises de taille internationale visent un marché plus vaste, dans lequel les distributeurs d'un pays ont une faible part de marché. Cette avance donne aux grands producteurs internationaux une position de force lors des négociations avec les distributeurs, moins internationalisés.

En revanche, la situation est très différente pour les petits producteurs. Leur pouvoir de négociation face aux grands distributeurs est restreint, au point qu'ils résistent difficilement aux pressions exercées par leurs clients sur les tarifs. On retrouve le problème de la « dépendance économique », recouvrant les situations dans lesquelles l'un des partenaires d'une relation commerciale voit sa survie menacée par la possibilité d'une rupture de cette relation. Les autorités en charge de la politique de la concurrence se sont penchées sur la question de la dépendance économique des fournisseurs envers les distributeurs, afin de dresser un cadre réglementaire permettant de maintenir « l'équilibre et la loyauté des relations commerciales ».

2.2. La loi Galland et son impact

Parmi les modifications majeures apportées par la loi Galland¹⁸ à la réglementation des relations commerciales, on compte l'interdiction de la revente à perte¹⁹ et l'autorisation du

¹⁸ Loi n°96-588 du 01 juillet 1996 sur la loyauté et l'équilibre des relations commerciales.

refus de vente ; en outre, la pratique de « prix abusivement bas » par les distributeurs est maintenant interdite, sans qu'un seuil précis soit clairement défini. De plus, les modifications des règles de facturation interdisent aux distributeurs d'anticiper les rabais de fin d'année lors de la fixation de leurs tarifs : seuls les rabais déjà concédés peuvent être pris en compte dans l'élaboration des prix de détail. Par ailleurs, les primes de référencement versées sans contreparties réelles (c'est-à-dire sans service réellement fournis par le distributeur) sont prohibées, ainsi que les « déréférencements abusifs » : toute rupture d'une relation commerciale durable doit maintenant s'accompagner d'un préavis écrit, permettant aux fournisseurs de trouver d'autres débouchés. Cette clause limite la possibilité pour un distributeur de déréférer subitement les produits d'une marque pour faire pression sur leur producteur lors de négociations.

Ces mesures, qui semblent globalement limiter le pouvoir des distributeurs, n'ont pas encore prouvé leur efficacité. La loi avait deux objectifs : réduire le déséquilibre entre distributeurs et petits fournisseurs, et protéger le petit commerce de la concurrence des grandes surfaces, jugée trop forte. Dans les deux cas, les résultats semblent mitigés, et la loi engendre de nombreux effets pervers, dus essentiellement à l'ambiguïté de sa formulation. En outre, elle a eu un léger effet inflationniste, surtout peu après son application.

3. Le mouvement de concentration dans la grande distribution.

3.1. La grande distribution française est de plus en plus concentrée.

Une comparaison au niveau européen fait ressortir des différences assez marquées entre les pays. Si la distribution est souvent concentrée en Europe du Nord, comme par exemple en Finlande, où les cinq premières firmes se partageaient 96% du marché en 1997, elle l'est en revanche beaucoup moins en Europe du Sud, où les structures de la distribution sont plus traditionnelles et laissent encore une large part au commerce de proximité. Le tableau suivant présente la somme des parts de marché des cinq premiers distributeurs dans les pays européens, en 1997 :

Finlande	96
Suède	87

¹⁹ La loi Galland empêche les distributeurs de pratiquer des marges négatives sur des produits d'appel et de récupérer leurs frais en faisant payer aux fournisseurs des sommes forfaitaires, par exemple des primes de référencement.

Pays-Bas	79
Autriche	79
<i>Danemark</i>	78
Allemagne	75,2
France	67,2
Royaume-Uni	67
<i>Grèce</i>	59
Belgique et Luxembourg	57
Portugal	52
Irlande	50
Espagne	38
Italie	30

Source : LSA, sauf Danemark (« Food Business ») et Grèce (« the retail pocket book 1998 »)

En France, la concentration s'est nettement accrue depuis une dizaine d'années, à la suite d'un important mouvement de fusions et acquisitions. Alors qu'en 1992, les cinq premières entreprises de grande distribution française détenaient 55,7% des parts de marché, elles en totalisaient 67,2% cinq ans après. La première opération de concentration vraiment importante date de 1991, lorsque Carrefour a racheté Euromarché, en déclin, mais dont le chiffre d'affaires annuel représentait environ 27 milliards de francs. L'année suivante, Casino a racheté Rallye et en 1996, Auchan a pris le contrôle de Docks de France. Plus récemment, en 1998, alors que Promodès lançait sur Casino une OPA qui allait échouer, Carrefour rachetait Comptoirs Modernes. Enfin, l'annonce, à la fin de l'été 1999, d'une OPE de Carrefour sur Promodès²⁰, a suscité de nombreuses réactions.

Les groupements d'indépendants ont engagé eux aussi des tentatives de rapprochement, notamment au niveau des centrales d'achat. Ainsi, en juin 1998, Leclerc et Système U ont signé un accord de coopération concernant la négociation de leurs achats, en créant la centrale commune Lucie.

²⁰ Si l'opération est acceptée par les autorités en charge de la politique de la concurrence, le groupe ainsi formé occupera la première place de la grande distribution française avec 26.7% de parts de marché, et la deuxième place mondiale avec 54 milliards d'Euros de chiffre d'affaires.

Depuis la fin des années 1980, la disparition d'un grand nombre de petites enseignes, dont la plupart ont été intégrées à de grands groupes, témoigne de l'importance du processus de concentration dans lequel le secteur distributif est engagé. Sur les 50 principales enseignes de supermarchés existant en 1986, 21 avaient disparu en 1996, et une dizaine d'autres sont en situation précaire et pourraient prochainement disparaître. Cette évolution a été encore plus marquée pour les hypermarchés : 30 enseignes étaient recensées en 1986, dont les deux tiers ont maintenant disparu.

L'hypothèse la plus souvent avancée pour expliquer cette évolution est la saturation progressive des marchés nationaux, dans les pays disposant d'un réseau distributif moderne bien développé. Les grandes chaînes de distribution, ne pouvant plus compter sur la croissance interne de leur activité, cherchent à poursuivre une croissance externe, pour atteindre une « taille critique » leur permettant de bénéficier d'économies d'échelle, et de survivre sur un marché de plus en plus international. Il semble cependant que la croissance des rendements avec la surface de vente, qui était constatée jusqu'à il y a quelques années, soit terminée, les distributeurs semblant actuellement faire face à des rendements constants.

En revanche, un autre facteur explique cette volonté de concentration des distributeurs. L'accroissement de la taille d'un distributeur renforce sa puissance d'achat, et détermine son pouvoir de négociation face à ses fournisseurs, ce qui peut engendrer une autre forme d'économies d'échelle. En effet, plus les distributeurs sont concentrés, plus leur pouvoir de marché en tant qu'acheteurs se développe. Ainsi, lorsque le poids d'une enseigne dans le chiffre d'affaires d'un producteur augmente, le distributeur concentre une part plus importante des débouchés du producteur, et est à même d'exiger des conditions d'approvisionnement plus favorables. L'état de dépendance d'un producteur à l'égard de ses distributeurs risque donc d'augmenter avec leur concentration.

A travers le regroupement de groupes de distribution en centrales d'achat communes, on constate que les grands distributeurs sont encore plus concentrés en tant qu'acheteurs qu'en tant que vendeurs. Les 5 premières centrales d'achat françaises réalisaient, en 1996, 75,1 % des parts de marché à l'achat, alors que les cinq premiers distributeurs ne totalisaient que 66,8% des parts de marché à la vente. Récemment, de nouveaux regroupements sont apparus, donnant naissance à ce que les professionnels du secteur ont appelé des « super-centrales ». En 1999, Casino et Cora ont regroupé leur puissance d'achat au sein de la centrale commune Opéra. En outre, des premiers signes de regroupements internationaux apparaissent. Ainsi, la centrale Agenor est issue du rachat du groupe allemand Spar par Intermarché. Cependant, la

centralisation des achats au niveau européen semble encore assez limitée. En 1997, Leclerc s'est affilié à la centrale d'achat européenne EMD, et Système U l'a rejoint en 1998, mais cette affiliation se limite à la négociation avec les fabricants de produits de base ou de marques de distributeur. Les grands groupes préfèrent en général internationaliser leurs fonctions de vente avant de conclure des accords de regroupements en matière d'achats.

3.2. Concentrations et politique de la concurrence.

Les attitudes des autorités en charge de la politique de la concurrence à l'égard de la concentration dans la grande distribution sont contrastées. En France, en 1996, le ministère des finances a saisi le Conseil de la Concurrence d'une demande d'avis relative à diverses questions portant sur la concentration dans le secteur de la grande distribution. Dans cet avis²¹, le Conseil de la Concurrence mesure le degré de concentration atteint en 1996 dans le secteur de la distribution au niveau national, et évalue les conséquences actuelles de cette concentration tant sur la concurrence sur le marché aval que sur les relations entre fournisseurs et distributeurs. Le rapport conclut que « la concentration de la distribution a peu d'effets sur le marché aval car la concurrence reste vive entre les groupes de distribution ». En revanche, les effets de la concentration des distributeurs sur leurs relations avec les producteurs sont plus sensibles.

En outre, la concentration a des effets sur l'entrée de nouveaux concurrents sur le marché : l'entrée est plus difficile sur les marchés les plus concentrés. Cependant, cet argument n'a jusqu'alors pas été retenu à l'encontre des concentrations dans le secteur de la distribution, car l'arrivée sur le marché français des enseignes allemandes, notamment dans le secteur du hard-discount, a montré qu'il restait des marchés de niches accessibles, et que les barrières à l'entrée n'étaient pas complètes, même si la réaction des enseignes françaises a été fortement défensive.

Finalement, le Conseil de la Concurrence préconise une analyse des fusions entre distributeurs au cas par cas, et conclut qu'« aucune mesure de déconcentration de structures commerciales ne paraît pouvoir être fondée sur une approche générale ». Il y a deux ans, les autorités françaises en charge de la politique de la concurrence ne semblaient donc pas disposées à

²¹ Avis n°97-A-04 en date du 21 janvier 1997 relatif à diverses questions portant sur la concentration de la distribution.

freiner le mouvement de concentration de la grande distribution ou à s'opposer à la poursuite de son développement. Leur position pourrait avoir évolué depuis.

Dans d'autres pays, les autorités en charge de la concurrence sont nettement plus réticentes à accepter les propositions de concentration en provenance du secteur de la grande distribution. Ainsi, les autorités anti-trust américaines sont beaucoup plus sévères à l'égard de la concentration des distributeurs, et utilisent notamment une définition du marché pertinent plus étroite empêchant certains projets de fusion d'être acceptés²². La commission des monopoles et des fusions (MMC) britannique a engagé une enquête sur le secteur de la distribution en 1999. Enfin, la Commission Européenne a généralement adopté une attitude conciliante à l'égard des mouvements de concentration dans le secteur de la distribution, mais son opposition récente à certains projets de fusion dénote un changement d'attitude à l'égard du secteur de la distribution et du problème de la puissance d'achat. On peut citer par exemple le refus opposé par la Commission Européenne au projet de fusion entre deux distributeurs finlandais, Kesko et Tuko²³, mais aussi les conditions imposées au projet de fusion entre Rewe et Reinl, en février 1999²⁴. Les autorités européennes semblent de plus en plus méfiantes à l'égard de la puissance des distributeurs.

3.3. Propositions pour une analyse théorique des fusions entre distributeurs.

Pour les autorités en charge de la politique de la concurrence, la question se pose donc de savoir s'il est opportun de laisser les distributeurs poursuivre le mouvement de concentration observé depuis une dizaine d'années. On peut en effet se demander à la fois pourquoi les distributeurs semblent engagés dans un mouvement de concentration plus vaste et plus rapide que leur fournisseurs, et quelles peuvent en être les effets économiques.

Nous proposons maintenant des éléments de réponse à ces questions, en envisageant, du point de vue de la politique de la concurrence, les déterminants et les effets sur le surplus social de la concentration des distributeurs. Il est important de noter que nous adoptons un point de vue de politique de la concurrence, sans chercher par exemple à trancher la question de savoir s'il

²² On peut ainsi évoquer le cas de Von's (1966, United States vs. Von's grocery compagny, 384 US 270), où la Federal Trade Commission a interdit la fusion de deux distributeurs de Los Angeles, dont les parts de marché locales cumulées ne dépassaient pas 7,5%.

²³ Décision du 20 novembre 1996, relative au cas IV/M.784-Kesko/Tuko, publiée au Journal Officiel des Communautés Européennes n° 110 du 26/04/1997.

est opportun de favoriser le développement d'une entreprise française de taille internationale capable de rivaliser avec des concurrents déjà gigantesques.

Nous présentons ici un aperçu de la modélisation développée et des résultats obtenus. Le modèle complet est présenté en annexe.

Le cadre d'analyse retenu

On considère le marché d'un bien homogène, produit par des producteurs identiques, qui se font concurrence à la Cournot (c'est-à-dire que la variable stratégique de chaque producteur est la quantité qu'il produit). Ce bien est vendu par l'intermédiaire de distributeurs, tous identiques, et eux aussi en concurrence à la Cournot. On fait l'hypothèse que les producteurs ne peuvent pas vendre directement leur production aux consommateurs, et qu'ils doivent passer par un distributeur. Les consommateurs formulent une demande linéaire classique.

Comparaison des incitations à la fusion des distributeurs et des producteurs

Dans un premier temps, nous étudions les incitations à la fusion des firmes aux deux niveaux du marché, en comparant simplement les gains de profits réalisés par un certain nombre de firmes lorsqu'elles fusionnent. Nous envisageons donc des fusions exogènes, en prenant en compte les synergies qui peuvent exister entre des firmes participant à une fusion par l'intermédiaire de leur structure de coûts. Nous paramétrons les résultats en fonction du nombre de firmes présentes en amont et en aval avant la fusion.

Lorsque le nombre de firmes est le même aux deux niveaux du marché, une fusion devient profitable (c'est-à-dire que le profit réalisé par la firme résultant de la fusion est supérieur à la somme des profits des firmes qui y participent) lorsqu'elle regroupe un certain nombre de firmes: le seuil que l'on détermine ainsi est plus élevé pour les producteurs que pour les distributeurs. Ainsi, une fusion entre producteurs doit regrouper au moins un quart des firmes présentes sur le marché, alors qu'une fusion entre distributeurs est toujours profitable dès qu'elle regroupe au moins trois firmes.

Ces premiers résultats nous donnent l'intuition que les incitations à la fusion sont plus élevées pour les distributeurs que pour les producteurs. On peut comprendre ce phénomène en étudiant les profits des firmes: les problèmes de double marginalisation bien connus dans l'étude des relations verticales engendrent une plus grande fragilité des distributeurs dont les

²⁴ Décision du 03 février 1999. M 1221.

marges sont naturellement plus faibles que celles de leurs fournisseurs. Cette fragilité diminue lorsque les distributeurs se regroupent, et l'accroissement de leur pouvoir de monopsonne leur apporte des gains plus élevés que l'accroissement du pouvoir de marché de producteurs qui fusionnent.

On peut cependant noter que l'impact sur le surplus social des fusions entre distributeurs est plus faible que celui des fusions entre producteurs, qui, pour un même nombre de firmes, sont socialement plus nuisibles.

Endogénéisation des décisions de fusion

Cependant les résultats que l'on vient d'évoquer reposent sur une étude de fusions exogènes. Il semble nécessaire d'affiner l'étude en endogénéisant la décision de fusion des firmes, c'est-à-dire en considérant que les entrepreneurs prennent eux mêmes la décision d'acquérir d'autres firmes ou de vendre leur entreprise.

On considère donc un jeu de fusion dans lequel les entrepreneurs ont la possibilité de faire des offres d'achat sur leurs concurrents, et des propositions de vente de leur propre entreprise. On restreint les possibilités de fusions à chaque niveau horizontal : un producteur ne peut se porter acquéreur que d'un autre producteur, et un distributeur d'un autre distributeur, mais aucune possibilité de concentration verticale n'est envisagée.

La détermination des équilibres de ce jeu confirme les résultats précédents : les distributeurs fusionnent beaucoup plus facilement que les producteurs. En particulier, aucun équilibre n'existe dans lequel un producteur rachète d'autres producteurs. Au contraire, chez les distributeurs, la concentration peut aller jusqu'à la monopolisation dès que le nombre de firmes présentes au début du jeu est inférieur à 5, ce qui correspond à de nombreuses situations réelles.

Analyse en termes de surplus social

En termes de surplus social, on a vu précédemment que pour un nombre de firmes donné, la fusion est plus nuisible si elle se produit parmi les producteurs que parmi les distributeurs. Cependant, les incitations des firmes sont telles que les producteurs fusionnent beaucoup moins facilement que les distributeurs. Finalement, les fusions envisageables à l'équilibre du jeu entre les distributeurs sont plus nuisibles en termes de surplus social que les fusions pouvant se produire entre les producteurs.

En conclusion, et malgré les limites évidentes du modèle, nous pensons qu'un processus de concentration important est plus probable au niveau des distributeurs qu'au niveau de leurs fournisseurs sur le marché d'un bien de grande consommation donné. Un contrôle attentif des concentrations semble donc nécessaire sur le marché de la grande distribution, l'émergence de monopoles, même locaux, ne semblant pas exclue.

4. Le développement des marques de distributeurs.

Les marques de distributeurs (MDD) recouvrent des produits vendus sous le nom du distributeur ou sous un nom associé à l'enseigne. Elles ont pris de l'essor dès le début des années 90 et ont connu une accélération forte à la suite de l'adoption de la loi Galland en 1996, qui avait pour but de réduire la concurrence par les prix dans la grande distribution et de renforcer le pouvoir de négociation des producteurs. Les distributeurs ont un taux de marque beaucoup plus fort sur les MDD, ce qui explique leur intérêt croissant pour ce genre de produits. Si la première génération des MDD se contentait de proposer des substituts à bas prix des produits de marque (cette stratégie de « premiers prix » permettait en outre aux distributeurs de faire face à la nouvelle concurrence du *Hard Discount*), désormais, les MDD s'élargissent à une gamme de produits innovants, et rivalisent en qualité avec les « grandes » marques. Toute la difficulté de cette stratégie est de faire comprendre au consommateur cette montée en gamme des MDD.

Les MDD ont atteint en moyenne 20.3% de part de marché de l'alimentaire en 1998, performance qui se situe encore loin derrière celles des autres pays européens²⁵ :

Parts du CA pour les MDD en 1998 (en %)	
Auchan + Mammouth	14.7
E. Leclerc	16.6
Monoprix + Prisunic	17.3
Cora	17.3
Continent	18.9
Champion	19.8
Système U	20.0

²⁵ Les MDD atteignent 50% de part de marché au Royaume Uni par exemple.

Géant	20.5
Carrefour	23.6
Casino	24.3
Intermarché	28.3
total hypermarché et supermarché	20.3

source : secodip / Eurostaf

L'essor des MDD reflète aussi l'évolution des rapports entre distributeurs et producteurs. A travers ces produits, les distributeurs concurrencent leurs propres fournisseurs, ce qui augmente encore la concurrence pour l'accès au linéaire, qui constitue une ressource rare. Cette rareté du linéaire permet ainsi aux distributeurs de choisir leurs fournisseurs, ou de facturer le linéaire plus cher. Par ailleurs les MDD permettent aux distributeurs de négocier des prix plus bas avec leurs fournisseurs, qui n'ont pas à supporter de dépenses de marketing ou de promotion et sont souvent de petits producteurs, voire des antennes de production intégrées comme pour Intermarché et Casino.

Dans un premier temps, on étudie les motivations stratégiques qui déterminent le comportement d'un distributeur lors du choix de la qualité des produits qu'il vend sous sa marque propre. On montre ainsi que la faible différenciation verticale entre marque propre et marque nationale peut relever d'un problème de coordination entre le producteur de marque nationale et le distributeur, nuisant ainsi au profit du producteur et au bien-être social en général. Dans un second temps, on étudie les effets potentiellement pro- ou anti-concurrentiels des structures de vente qui peuvent émerger lorsqu'un producteur de marque nationale s'adresse à un secteur aval oligopolistique disposant de marques propres. On montre ainsi que les motivations stratégiques qui déterminent la politique de distribution d'un producteur de marque nationale sont fondamentalement liées à l'existence de marques propres et au degré de concurrence sur la marché final. D'une part, le fait de proposer la marque nationale à plusieurs enseignes de distribution permet à un producteur de marque nationale de bénéficier d'une prime stratégique aux dépens des enseignes de distribution. L'intensification de la concurrence sur le marché final qui en résulte contribue alors à faire évoluer les rapports de force au sein de l'industrie en faveur du producteur de marque nationale. D'autre part, le fait de restreindre la commercialisation d'une marque nationale à une seule enseigne de distribution peut permettre à la structure verticale composée du producteur de marque nationale et du distributeur servi, de bénéficier d'une prime stratégique aux dépens des distributeurs non servis. Sur certains marchés, en particulier non alimentaires, cet effet peut

l'emporter sur le premier. L'évolution des rapports de force dans le cas d'exclusivité se fait alors aux dépens des distributeurs non servis et en faveur du producteur de marque nationale et du distributeur servi, les rapports de force entre ces derniers n'étant pas affectés. On montre dans ce cadre qu'il convient en particulier de s'assurer que les plus puissants distributeurs ne souscrivent pas d'accord exclusif avec certains fournisseurs de marque nationale, ce qui pourrait diminuer artificiellement la compétitivité de leurs concurrents en leur interdisant l'accès à ces produits.

Troisième Partie :
Concurrence et concentration dans le secteur des
assurances en France

L'assurance est un échange de richesses entre individus qui font face à un aléa (Borch (1962)). Cet échange est défini par la signature d'un contrat, avant que l'aléa ne se réalise. Un contrat efficace vérifie le principe de mutualité: la richesse finale d'un agent n'est pas fonction de son risque mais du *risque agrégé* du groupe. Les agents sont solidaires, les primes des uns sont les remboursements des autres. Ce principe est riche de conséquences. L'assurance, c'est rassembler des agents au sein d'une structure d'échange, et leur faire signer un contrat qui les conduit à se co-assurer. Le coût de l'assurance est donc un coût d'organisation de ce service. Aujourd'hui (sauf pour l'assurance-maladie), l'assurance est mise en oeuvre par des institutions privées, qui proposent des contrats à des agents souhaitant se protéger de risques individuels. On devrait alors s'attendre à ce que ce marché, de par la nature du service qu'il permet d'échanger, soit affecté par des mécanismes qui l'éloignent du fonctionnement bien connu des marchés de concurrence parfaite. Notre travail explore cette idée.

Le marché français de l'assurance s'organise aujourd'hui dans un contexte particulier dans lequel on observe :

- . un *mouvement de concentration* très fort,
- . une *réglementation* qui affecte les échanges,
- . la cohabitation d'acteurs (du côté de l'offre) institutionnellement différents: les *sociétés anonymes* et les *mutuelles*.

Ces traits particuliers posent un certain nombre de questions, pour lesquelles plusieurs types de réponses peuvent être données.

- . La première d'entre elles concerne les fusions. Qu'est ce qui, sur le marché de l'assurance, différencie les incitations et les conséquences des fusions des autres secteurs de l'économie ?
- . La seconde a trait à la réglementation prudentielle. Quels objectifs poursuit-elle, quelle forme prend-elle, quels sont ses effets?
- . Enfin on s'intéresse aux acteurs du côté de l'offre. La cohabitation entre mutuelles et sociétés anonymes est-elle possible et explicable sur le plan théorique ? Existe-t-il une organisation efficace des échanges ? Quel rôle jouent les mutuelles sur ce marché ?

Dans une première partie nous procédons à une analyse descriptive du marché français autour des caractéristiques précédentes. Dans une seconde partie, nous tentons de donner des fondements à l'intensité des fusions observées dans le secteur. Pour cela, nous construisons un modèle théorique qui fonde les effets de taille sur les externalités (de type externalités de

réseau) exercées par un assuré sur les autres clients de sa compagnie (conformément au principe de mutualisation).

1. Une approche descriptive du secteur des assurances en France

1.1 Le bilan des opérations de concentration

Un mouvement de concentration intense

Les pays de l'Union Européenne jouent actuellement un rôle moteur dans les mutations et réorganisations qui s'opèrent sur le marché mondial de l'assurance. Cinq des dix premiers assureurs mondiaux étaient européens en 1997 (contre deux seulement en 1992), et parmi eux, deux groupes français. Une hausse de 20 % par an du nombre de concentrations a été enregistrée sur le marché européen de l'assurance²⁶ ces dernières années. Certains prédisent qu'à terme le paysage européen ne comptera plus que 2000 compagnies (contre 5000 en 1995, année marquant le début du mouvement), relevant d'une dizaine de groupes européens. Comparativement aux autres pays européens, le phénomène de concentration est particulièrement marqué en France.

Le mouvement de concentration s'inscrit avant tout dans une logique de « mondialisation » de l'économie. On peut toutefois s'interroger sur l'ampleur qu'a pris ce phénomène au cours de la dernière décennie. Selon toute vraisemblance, l'environnement concurrentiel accompagnant cette tendance a subi des modifications tangibles. La théorie économique enseigne qu'une structure concentrée du marché peut empêcher, restreindre ou fausser le jeu de la concurrence, ce qui aura inévitablement des retombées négatives sur les assurés (hausse des cotisations et/ou baisse du montant et de la qualité des garanties). Ces effets peuvent néanmoins être compensés par un effet positif (observable à plus long terme) dû quant à lui à la diminution des coûts que deux compagnies qui fusionnent sont susceptibles de réaliser. Dans quelle mesure les facteurs traditionnellement invoqués qui conditionnent l'existence et la puissance de l'un ou l'autre de ces effets se manifestent-ils sur le marché de l'assurance ?

Concentration et pouvoir de marché en assurance

L'existence d'un pouvoir de marché est difficile à cerner tant au niveau théorique que dans la pratique. Une part de marché élevée est une condition nécessaire mais pas suffisante (sauf cas

²⁶ Onzième rapport d'activité (année 1997) de la Commission Européenne de la Concurrence.

extrêmes) pour établir qu'un groupe d'entreprises abuse ou risque d'abuser de sa position. Si l'on mesure le degré de concentration sur un marché par le ratio de concentration (la part des 10 ou 5 premières entreprises sur le marché), les récentes opérations de fusions-acquisitions se sont traduites en France par un accroissement sensible du degré de concentration²⁷ : la part de marché des 5 premières compagnies françaises est passée de 41,4% en 1995 à 51,3% en 1998²⁸.

Le nombre de firmes sur un marché n'est pas plus que le degré de concentration un indicateur fiable de l'exercice d'un pouvoir de marché. Un secteur peut être concentré, et c'est le cas du marché de l'assurance en France, et compter un grand nombre d'entreprises (500 compagnies françaises en 1998). Dans la pratique, Schor (1994) note qu'en France, la formation de groupes de dimension européenne voire mondiale a été suivie par un phénomène d'entrée de nouveaux intervenants de taille plus réduite, particulièrement en assurances de personnes et de dommages. Le marché est-il pour autant devenu plus concurrentiel ?

La possibilité pour un grand groupe d'accroître son pouvoir de marché dépend fortement de l'existence de barrières à l'entrée. Bon nombre d'études empiriques s'accordent à dire que les barrières économiques à l'entrée sont relativement faibles sur le marché de l'assurance²⁹ (coûts fixes et intensité capitalistique faibles, absence de coûts irrécouvrables...), et qu'en Europe, la déréglementation tend à estomper les barrières institutionnelles. La multiplication récente des nouveaux acteurs dans la distribution pourrait par ailleurs conforter ces dires et donner l'illusion d'un accès aisé au marché de l'assurance. C'est oublier que très souvent, les prétendus « nouveaux » opérateurs s'avèrent être des filiales de grands groupes. A titre d'exemple, la première société de vente directe en France (Direct Assurance) est une filiale du groupe AXA. De plus, l'essor de la bancassurance en France n'a été possible qu'à travers la formation d'accords de partenariat entre établissements bancaires et assureurs (par exemple entre les Banques Populaires et la MAAF) ou la création de filiales communes (BNP et AXA

²⁷ Les taux de concentrations nationaux, mesurés en part de marché détenue par les 10 premiers groupes, sont assez disparates sur le marché européen de l'assurance. Ils diffèrent selon les pays et selon la branche d'activité : les taux les plus faibles en assurance non vie sont recensés en Espagne et au Royaume-Uni et les plus forts dans les pays nordiques. Par ailleurs, le marché est plus concentré en assurance vie qu'en non vie (en 1996, les 10 premiers groupes français se partagent 70% du marché d'assurance vie contre 60% du marché non vie) mais la tendance semble s'inverser.

²⁸ Source FFSA.

²⁹ Joskow (1973), Pesticau et Pirard (1989), Schor (1984) ...

ont donné naissance à Natio-Assurances). Il existe d'autres éléments qui corroborent l'existence de barrières à l'entrée.

. Les risques de particuliers sont caractérisés par un phénomène de *proximité*. L'assureur doit être un intermédiaire proche du consommateur. L'entrée sur un marché étranger passe donc le plus souvent par le rachat d'agences locales. A l'évidence, l'argument de proximité joue de manière inégale selon les cultures³⁰.

. Il existe *une certaine inertie des assurés français*. Une augmentation des primes ne fait pas fuir les clients. Le coût de l'assurance n'est donc pas le seul élément expliquant les choix individuels. Des arguments standards peuvent bien sûr être avancés: différenciation des produits d'assurance, différenciation géographique.

La concentration entraîne-t-elle une baisse des coûts ?

La baisse des coûts et plus généralement les gains en efficacité productive constituent l'argument le plus souvent affiché par les entreprises concernées par des opérations de fusion ou d'acquisition. Les entreprises fusionnant dans le secteur de l'assurance mettent en avant de telles perspectives: AXA-UAP de plus d'un milliard de francs et Allianz-AGF prévoit d'économiser 300 Millions de marks à partir de l'an 2000 (Les Echos 23/11/98) . Il est toutefois évident que ces économies ne pourront être réalisées qu'une fois les coûts de restructuration résorbés.

Les économies d'échelle sont la conséquence la plus directe attendue d'une concentration :

. Au niveau de la conception des contrats, la mise en commun des fichiers clients de deux compagnies permet une meilleure catégorisation des risques et donc une tarification plus fine ainsi qu'un meilleur ciblage des nouveaux produits. Pour ce qui est de la gestion des sinistres, une compagnie de taille importante peut créer des unités spécialisées pour les différents types de sinistres, ce qui augmente l'efficacité de cette activité³¹.

³⁰ En Grande-Bretagne, la vente directe a remporté un succès incontestable (30% des contrats d'assurance dommages ont fait l'objet en 1998 d'une souscription par ce mode de distribution). Ceci s'explique en partie par la grande volatilité de la clientèle britannique qui est favorisée par l'absence de tacite reconduction dans les contrats. Au contraire, les assurés Français semblent plus fidèles (ou plus captifs...) à leur assureur.

³¹ A ce propos, on peut évoquer la lutte contre la fraude. Seules les grosses sociétés mettent en place des cellules anti-fraude, les procédures spécialisées qu'elles nécessitent sont trop coûteuses pour des compagnies plus modestes.

que la tradition française combinait, jusqu'en 1986, contrôle des tarifs et de la situation financière des entreprises. Les directives communautaires européennes interdisant désormais tout contrôle préalable et systématique des tarifs d'assurance, le contrôle, au sein de la Communauté, s'est orienté vers la situation financière des entreprises.

L'exigence de solvabilité des sociétés d'assurances constitue l'une des trois composantes du dispositif réglementaire destiné à protéger les assurés contre le non respect des engagements. Elle s'ajoute en effet aux règles relatives aux provisions techniques et aux placements des primes collectées. La réglementation, tant française que communautaire, prévoit que les sociétés d'assurances doivent disposer d'un montant minimal de fonds propres et de plus-values latentes. Elle a pour objectif de compenser une éventuelle insuffisance des provisions techniques ou une baisse de valeur des placements. L'existence de ce type de réglementation repose sur le fait qu'il est difficile, voire impossible, pour les assurés d'avoir accès à des informations relatives à la situation financière de leur compagnie. Même si c'est le cas, il leur est difficile de pouvoir apprécier à partir de ces informations l'éventualité d'un non respect de ces engagements. Il paraît même plausible de supposer, dans l'absolu, que les assurés n'envisagent même pas la possibilité que la compagnie ne respecte pas ses engagements.

Les normes réglementaires

En France, le contrôle des sociétés d'assurances est confié, pour les sociétés du Code des Assurances à la Commission de Contrôle des Assurances (1989) et pour les mutuelles du Code de la Mutualité par la Commission de Contrôle des institutions de prévoyance et des mutuelles (1989).

La réglementation européenne fixe pour les entreprises ayant leur siège dans un des Etats de l'Union, des exigences harmonisées concernant le montant minimal des fonds propres (et de plus values latentes). Ce montant minimal est composé de deux parties, l'une fixe, appelée fonds de garantie minimum et l'autre variant en fonction des engagements de la compagnie, appelée marge de solvabilité réglementaire.

Selon les normes communautaires en vigueur, la marge de solvabilité minimale est :

- . 16 % du chiffre d'affaires ou 23% des sinistres moyens des trois derniers exercices si ce montant est plus élevé pour l'activité non vie ;
- . 4% des provisions mathématiques pour l'activité vie plus 0.1% à 3% des capitaux non couverts par les provisions déjà constituées qui seraient à payer en cas de survenance d'événements garantis aux titres des contrats en portefeuille.

Cette marge est pondérée par le taux de conservation (% des affaires souscrites non cédées aux réassureurs).

On observe en France, depuis 1994 que la marge de solvabilité moyenne du marché se situe à un niveau nettement supérieur aux normes réglementaires (78,8% en 1997 et 95.9% en 1998, soit 6 fois la marge réglementaire pour 1998). Ce résultat peut en partie s'expliquer par la progression du stock de plus-values latentes observé depuis 1994, qui s'est poursuivie en 1998 et 1999 du fait de la forte croissance des marchés d'actions. Mais les fonds propres sont aussi très au-dessus de la marge de solvabilité exigée: à lui seul le ratio fonds propres/chiffre d'affaires est en moyenne de 49.1% pour 1998.

Les effets pervers de la réglementation

Comme nous l'avons souligné, la réglementation prudentielle a été instaurée dans le but de protéger les consommateurs sur le marché de l'assurance et l'on peut se demander si, en ce sens, elle remplit son objectif. La réglementation vise à modifier le comportement des sociétés d'assurance présentes sur le marché. Elle affecte donc le processus concurrentiel. On peut alors tenter s'interroger sur son impact sur ce marché de l'assurance, d'autant plus que ce dernier vit aujourd'hui une période de déréglementation dont l'objectif est le renforcement de la concurrence.

Les économistes s'accordent sur l'idée selon laquelle la réglementation prudentielle a un effet anti-concurrentiel. Notre travail défend cette idée, et explique à la fois la nécessité d'une intervention de l'Etat, les conséquences anti-concurrentielles de cette réglementation.

L'impact des fonds de garantie minimum paraît évident: imposant une contrainte sur les fonds propres nécessaires pour la création d'une compagnie d'assurance, il bloque l'entrée sur le marché de l'assurance aux petites compagnies et ne peut qu'avoir un impact négatif sur la concurrence.

L'impact de la marge de solvabilité est moins évident, elle peut introduire des contraintes de taille maximale pour les compagnies d'assurance opérant sur le marché, quel que soit le critère appliqué (chiffre d'affaires ou charge des sinistres). En effet, pour un montant de fonds propres donné et pour une prime déterminée de manière concurrentielle, le nombre maximal d'individus que la compagnie peut assurer en respectant la réglementation est fixé. Il est alors possible que les compagnies présentes ne soient pas capables d'assurer tout le marché, ce qui encouragerait l'entrée de nouvelles compagnies et dans ce sens aurait un impact positif sur le processus concurrentiel. Mais ceci ne pourrait se réaliser qu'en cas de libre entrée sur le

marché ; le cas échéant, la contrainte de capacité qu'imposent les ratios pourrait s'avérer néfaste pour les consommateurs.

Le cadre réglementaire crée des barrières à la sortie qui favorisent la concentration

Sur un marché, les *barrières à la sortie* agissent, au même titre que les barrières à l'entrée, en faveur du maintien d'une structure oligopolistique. Or, en France, très peu de compagnies font faillite au sens propre du terme: deux en dix ans en France. En effet, les compagnies françaises « en dérive » ou mal gérées ne sortent pas du marché. Ceci est en partie dû au cadre réglementaire qui incite au *rachat des compagnies devenues insolubles*. L'absorption de compagnies « qui vont mal » est ainsi une alternative systématique à leur faillite (BarNiv et Hathorn 1997).

1.3- La cohabitation des mutuelles et des sociétés d'assurances traditionnelles

L'étude de l'évolution de la concurrence sur un marché passe avant tout par la définition des acteurs sur ce marché. Sur le marché français de l'assurance cohabitent principalement deux grands types d'acteurs (les institutions de prévoyance n'entrent pas dans notre champ d'analyse) qui sont les sociétés anonymes d'assurance et les sociétés mutuelles. Ces deux types d'organisations diffèrent par leurs objectifs, par leur régime juridique et par les organisations chargées de leur contrôle. Suivant le secteur d'activité, l'une ou l'autre de ces structures domine. Il peut même exister des secteurs où seule l'une des deux structures est présente.

Il est important de noter que la France se distingue des autres pays européens³² par une très forte implantation des sociétés mutuelles aussi bien sur le marché de l'assurance dommage (la MACIF est le premier assureur automobile français et les sociétés du GEMA avaient en 1997 une part de marché en auto de 32%) que dans le domaine de la protection sociale complémentaire.

Il nous paraît utile, à cette étape, de rappeler les caractéristiques des sociétés mutuelles et leur justification économique. Les sociétés mutuelles sont des sociétés de personnes, dont les membres, le plus souvent d'une même profession, s'assurent les uns les autres contre certaines éventualités. Ces sociétés sont issues des anciennes caisses de solidarité des corporations et

³² Ce type d'organisation est absent au Royaume Uni et très peu représenté dans les autres pays européens où il concerne essentiellement des individus d'une même entreprise ou d'une même catégorie professionnelle.

elles appliquent le principe de mutualité. Les caractéristiques essentielles qui les distinguent des sociétés anonymes d'assurances sont les suivantes. Ce sont des sociétés à but non lucratif, sans capital social, qui sont la propriété de leurs membres, qui peuvent participer à la prise de décisions suivant un principe égalitaire. Les cotisations peuvent faire l'objet d'une révision *ex post* : des ristournes peuvent être versées ou un supplément demandé en fin d'exercice en fonction de la charge des sinistres de l'ensemble des sociétaires. Cette structure est en fait un rassemblement d'individus, mettant en commun leurs risques et leurs cotisations, ce qui les rend individuellement assurés et collectivement assureurs.

Il existe en réalité deux types de sociétés ayant ce mode de fonctionnement, qui se distinguent principalement par leur champ d'intervention et leur régime législatif. Il s'agit des sociétés d'assurance mutuelles, régies par le Code des Assurances et intervenant principalement dans le domaine de l'assurance dommage et les mutuelles proprement dites (appelées mutuelles 45), régies par le Code de la Mutualité et intervenant principalement dans le domaine de la protection sociale complémentaire.

Les stratégies des compagnies d'assurance sont certainement affectées par la présence de sociétés mutuelles dont les objectifs et les règles de fonctionnement sont, comme nous l'avons décrit, fondamentalement différents. De plus, il convient de mentionner que les mutuelles du Code de la Mutualité bénéficient de privilèges fiscaux et réglementaires qui leur sont reprochés par les autres sociétés d'assurances. Le caractère potentiellement anticoncurrentiel de ces privilèges a même fait l'objet de recours déposés auprès de la Commission Européenne par la FFSA concernant notamment l'exonération de la taxe de 7% sur les cotisations d'assurance maladie complémentaire et du régime fiscal de faveur en matière d'impôts directs.

La pertinence de l'appui réglementaire dont bénéficient les mutuelles dépend de l'impact de leur présence sur le marché de l'assurance. Cette présence est-elle suffisamment bénéfique pour les consommateurs ? Peut-on anticiper l'évolution de ces deux types d'organisation : tend-on vers une disparition d'une des deux organisations ou vers une cohabitation ? Quelles seront les conséquences de l'entrée des sociétés anonymes sur le marché de risques jusqu'à présent couverts uniquement par des mutuelles ?

2 - Analyse théorique

L'objectif de cette partie est de comprendre les conséquences de la mutualisation des risques sur le processus concurrentiel que se livrent des sociétés d'assurance, compagnies et mutuelles

sur ce marché. Nous montrons que la mutualisation des risques engendre un effet réseau qui perturbe le jeu de la concurrence.

Nous modélisons la concurrence entre sociétés d'assurance comme un jeu de concurrence en contrats, dans lequel dans une première étape les sociétés présentes sur le marché proposent des contrats, puis dans un second temps les assurés potentiels choisissent le contrat et la société qu'ils préfèrent.

Nous limitons notre analyse aux contrats non discriminants et robustes à la renégociation. Le concept d'équilibre utilisé est celui d'équilibre de Nash parfait en sous-jeu.

Les sociétés étant a priori identiques aux yeux des assurés, elles se font concurrence en attirant les assurés par les contrats qu'elles leur proposent, c'est-à-dire que formellement, le processus concurrentiel est une concurrence à la Bertrand en contrats.

La première étape de ce travail est de déterminer les propriétés des contrats optimaux. Conformément au principe de mutualité, les contrats optimaux sont des contrats d'assurance parfaite, dont les primes dépendent du sinistre moyen du groupe d'agents assurés dans la même société. Dès lors, apparaît un effet taille ou effet réseau, l'espérance d'utilité d'un agent dépend en effet du nombre d'agents appartenant au même groupe. Cet effet réseau a des conséquences sur l'optimalité des fusions.

Si les agents ne sont pas conscients de cet effet taille (on les qualifie de myopes), l'équilibre du jeu est conforme au résultat de Bertrand. L'introduction d'une réglementation prudentielle dans ce modèle affecte les équilibres.

Si les agents sont conscients de cet effet réseau, le jeu admet une multiplicité d'équilibres. Afin d'organiser l'exposé des résultats, nous présentons ces résultats en isolant certains d'entre eux. Nous étudions successivement les cas de monopole pur, monopole concurrencé par une mutuelle, oligopole pur, oligopole concurrencé par une mutuelle.

Dans une deuxième section nous décrivons notre cadre analytique, et étudions les contrats optimaux puis leurs propriétés. La section 3 présente les résultats quand les agents ne perçoivent pas le risque de faillite, dans ce cas de figure on étudie l'impact de la réglementation prudentielle.

La section 4 s'intéresse au jeu dans lequel les assurés comprennent correctement les contrats offerts.

2.1 Le cadre analytique retenu

Les hypothèses

indépendants les uns des autres. Les agents présentent de l'aversion vis-à-vis du risque et souhaitent s'assurer.

L'assurance est possible par le biais de deux organisations.

Sur ce marché co-existent des compagnies d'assurance et des mutuelles. Chacune compagnie d'assurance est constituée d'un groupe d'actionnaires, neutres vis-à-vis du risque. Ces actionnaires ont investi une certaine somme, le capital ou fonds propres. Ils ne sont responsables qu'à hauteur des fonds investis, c'est-à-dire qu'ils peuvent perdre au maximum la totalité des fonds propres. En contrepartie, le modèle étant statique, ils reçoivent à la fin de l'exercice la totalité des résultats de l'activité, c'est-à-dire les fonds propres et le bénéfice net. Ces investisseurs souhaitent maximiser la rentabilité des fonds investis, ce qui conduit à maximiser l'espérance de profit dans le cas d'un stock de capital fixe, ou encore si la société peut accroître ou diminuer son stock de capital, à maximiser le taux de rentabilité des fonds investis.

Une mutuelle est un groupe d'agents qui font face au risque d'accident et qui se co-assurent les uns les autres. Deux individus qui adoptent un accord précisant le partage de leur richesse comme fonction des pertes monétaires qu'ils subissent forment ainsi une mutuelle. La mutuelle n'a donc pas de fonds propres, et organise les échanges entre individus à risque. A ce titre, une mutuelle est une société d'assurance qui ne possède pas d'actionnaires. Notons que le modèle est statique et exclut la réassurance.

Les contrats optimaux

Mutuelle et compagnies proposent aux agents des contrats d'assurance. Ces contrats spécifient les remboursements accordés à un agent en fonction de tous les éléments économiques pertinents, c'est-à-dire en fonction de l'accident de l'agent et des accidents des autres agents appartenant à la même société. Ces contrats doivent être réalisables et donc prévoir des paiements possibles quelles que soient les pertes des membres du groupe. Dans le cas de la mutuelle, cette contrainte impose que la somme des remboursements soit négative ou nulle quel que soit l'état de la nature réalisé. Dans le cas de la compagnie, c'est le montant des fonds propres qui doit être supérieur à la somme des remboursements quelles que soient les charges des sinistres.

Cette contrainte de réalisation impose que la prime d'assurance d'un agent (paiement sans accident) peut être variable si elle dépend des accidents des autres assurés. Cette propriété est bien connue en ce qui concerne les mutuelles. Pour les sociétés d'assurance, elle vient du fait

que les pertes possibles sont infinies alors que les fonds propres de la compagnie d'assurance sont par nature limités. Si les pertes agrégées sont trop importantes, les actionnaires consacreront la totalité de leur fonds propres aux remboursements. Dans ce cas, nous dirons que la compagnie d'assurance fait faillite. Si l'entreprise souhaite continuer son activité l'année suivante (mais ceci n'est pas étudié dans notre modèle statique), elle doit reconstituer son capital.

Nous supposons que seuls des contrats efficaces non discriminants sont proposés à l'équilibre du marché. Ces contrats efficaces sont en effet robustes à la renégociation, et si un contrat non efficace attirait un certain nombre de clients, alors il serait de l'intérêt des parties contractantes, assureurs et assurés, de le modifier en le remplaçant par un contrat efficace.

Plus précisément, un contrat est dit efficace s'il est préféré par l'assuré moyen dans l'ensemble des contrats non discriminants possibles, et compte tenu d'une certaine exigence de profitabilité émanant des actionnaires, s'ils existent, c'est-à-dire dans le cas où la compagnie d'assurance est une société privée.

Ces contrats efficaces vérifient un certain nombre de propriétés.

En premier lieu ces contrats sont tels que les sinistres individuels sont totalement remboursés. Les primes dépendent du sinistre moyen.

Le contrat d'une mutuelle spécifie une prime égale au sinistre moyen. Cette prime est donc aléatoire ex ante, elle équilibre les comptes ex post. Concrètement, un tel contrat peut être réalisé en supposant que les agents payent une prime fixe avant réalisation des sinistres, prime fixe corrigée ex post par un rappel de prime ou un remboursement. Cette caractéristique est conforme au principe de mutualité de Borch, et aux pratiques usuelles des mutuelles, qui calculent les primes de l'année future en fonction du sinistre global de l'année passée.

En ce qui concerne la société d'assurance, le principe de mutualité voudrait que les actionnaires des compagnies, neutres vis-à-vis du risque, assurent totalement les clients de la compagnie. Toutefois, une couverture totale est impossible, le sinistre global peut être infini et les actionnaires ne sont responsables qu'à hauteur des fonds qu'ils ont engagés. Dans ce cas, le principe de mutualité prend la forme suivante. Le contrat d'assurance précise une prime fixe. Si le sinistre global est inférieur à la somme des fonds propres et des primes perçues, l'agent accidenté est totalement remboursé. Si par contre, cette somme est insuffisante pour rembourser tous les accidentés, le principe de mutualité joue comme pour une mutuelle, le contrat proposé est tel que la prime dépend du sinistre moyen réalisé, dans ce cas, les actionnaires consacrent la totalité de leurs fonds propres à rembourser les accidentés et la

prime payée par les assurés est égale au sinistre moyen moins le capital par tête. De nouveau un tel contrat peut être mis en oeuvre par le paiement ex ante d'une prime fixe et un rappel de primes ex post si les avoirs de la compagnie sont insuffisants, rappel qui équilibre alors les comptes. Dans un contexte dynamique, une augmentation des primes de l'année suivante peut s'assimiler à ce rappel de primes.

Le contrat d'assurance d'une compagnie est donc un contrat hybride entre un contrat d'assurance totale avec prime fixe, et un contrat avec prime variable quand le sinistre moyen est trop grand. Remarquons que dans ce cas, la compagnie serait insolvable si elle n'augmentait pas les primes. On dit qu'il y a alors faillite. Cette faillite signifie pour la compagnie d'assurance non pas qu'elle ne peut assurer sa tâche, mais que pour satisfaire ses engagements elle perd ses fonds propres et augmente les primes de ses clients.

Effet taille : propriétés des contrats optimaux

De tels contrats font apparaître un effet taille dans l'activité d'assurance. Le contrat faisant appel au principe de mutualité, l'espérance d'utilité d'un assuré est une fonction du nombre d'assurés qui appartiennent à la même société, qu'elle soit une compagnie ou une mutuelle. Ainsi la prime d'assurance que paye l'agent mutualiste est le sinistre moyen, variable aléatoire qui reflète la somme des risques des membres de la mutuelle. Le même effet joue au sein d'une société d'assurance quand il y a faillite.

La mutualisation des risques (regrouper deux sociétés identiques sans modifier les primes fixes, si cette société est une compagnie) augmente toujours l'espérance d'utilité des agents membres de ces groupes.

Quand le nombre de clients est suffisamment grand, la loi des grands nombres s'applique, éliminant le risque de l'activité d'assurance. L'espérance d'utilité des agents est alors maximale, c'est-à-dire qu'un agent préfère être assuré auprès d'une société qui assure tout le marché plutôt qu'une petite société.

Conformément à l'intuition, *l'espérance d'utilité d'un membre d'une mutuelle d'assurance augmente quand la taille du groupe s'accroît*. Le sinistre moyen est en effet une variable aléatoire d'autant moins risquée que le nombre d'individus concernés augmente.

L'analyse en ce qui concerne les agents qui s'assurent auprès d'une compagnie est plus complexe. En effet, un agent paye une prime variable, constante s'il n'y a pas faillite, et dans le cas inverse, égale à la différence entre le sinistre moyen et le capital par tête. L'impact d'un accroissement de clientèle a alors deux effets. D'une part, être plus nombreux diminue la

probabilité de faillite, et par l'effet mutualisation permet de payer une prime moins variable. Mais être plus nombreux signifie aussi que le capital par tête est plus faible et donc pour un sinistre moyen de même taille, la prime variable payée est plus grande. Ces deux effets jouent différemment selon la taille de l'entreprise. Quand le nombre d'assurés est très grand, la loi des grands nombres s'applique, le sinistre moyen tend vers l'espérance du sinistre et le capital par tête est négligeable. Dans ce cas, comme nous l'avons remarqué précédemment, l'espérance d'utilité de l'agent est la plus grande possible. *Pour des niveaux faibles de clientèle, l'effet capital par tête domine l'effet mutualisation et l'espérance d'utilité d'un assuré diminue quand la taille de l'entreprise augmente. Par contre, quand le nombre d'assurés est suffisamment grand, l'effet capital par tête devient négligeable, et l'espérance d'utilité augmente avec la taille de l'entreprise.* L'effet taille est donc dans ce cas assez subtil. L'effet taille affecte aussi l'espérance de profit de la compagnie. Pour une prime fixe supérieure à la prime actuarielle, l'espérance de profit augmente quand le nombre de clients augmente. Une compagnie d'assurance qui proposerait une prime inférieure à la prime actuarielle réaliserait des pertes (en espérance) dès lors qu'elle assure un nombre suffisamment grand d'agents.

Quand les sinistres sont élevés, la faillite de la compagnie lui garantit un résultat nul. Elle ne subit donc pas les queues de distribution de la perte agrégée. Tout se passe alors comme si l'entreprise appréciait le risque, en particulier la mutualisation des risques diminue son espérance de profit : notre entreprise préfère diviser son capital et ses clients en deux sociétés distinctes (bien sûr ceci se ferait au détriment de l'assuré. En particulier, l'espérance de profit par tête (d'agent assuré) diminue quand la clientèle s'accroît. Tout se passe comme si les rendements d'échelle (dans une vision de l'entreprise d'assurance dans laquelle les facteurs de production sont les assurés et le capital) étaient en fait décroissants.

Un triple conflit d'intérêt existe alors entre la compagnie d'assurance privée et son client : augmenter la prime fixe diminue l'utilité du client mais augmente le profit de la compagnie ; le taux de rentabilité est d'autant plus grand que le capital est faible, mais l'assuré préfère des fonds propres importants ; mutualiser les risques augmente l'espérance d'utilité des agents mais diminue l'espérance de profit des entreprises.

L'optimalité des fusions

Les propriétés engendrées par la mutualisation des risques ont des conséquences pour l'efficacité des structures.

L'optimalité des fusions

Les propriétés engendrées par la mutualisation des risques ont des conséquences pour l'efficacité des structures

En premier lieu, la fusion de deux entreprises identiques (même stock de capital, même nombre d'assurés) est optimale au sens de Pareto c'est-à-dire qu'il existe un contrat après fusion qui augmente à la fois l'espérance d'utilité des agents et l'espérance de profit des entreprises. En assurance, on devrait donc s'attendre à avoir de grosses entreprises.

Cette propriété est identique pour les mutuelles, l'espérance d'utilité des assurés augmente quand les mutuelles se regroupent.

Imaginons que l'Etat souhaite gérer le secteur de l'assurance, et par exemple contrôler les primes et le nombre de compagnies sur le marché (en accordant aux actionnaires privés un taux de rentabilité identique à celui qu'ils obtiendraient sur un autre marché). Ne doit survivre alors qu'une seule entreprise d'assurance, la mutuelle si les exigences des actionnaires sont trop élevées, ou dans le cas contraire une compagnie d'assurance privée. Bien entendu dans ce cas, la réglementation des primes permet d'éviter l'effet anti-concurrentiel de la monopolisation.

2.2 - Concurrence sur le marché de l'assurance et myopie des agents

Une hypothèse souvent évoquée dans le domaine de l'assurance est que les agents ne perçoivent pas les risques auxquels ils font face et en particulier pensent que la probabilité de faillite est négligeable. On dit dans ce cas que les consommateurs sont myopes. L'effet taille engendré par la mutualisation des risques n'affecte pas les choix des agents. Si l'on suppose que les contrats offerts par les compagnies d'assurance sont inchangés, le choix des agents entre deux compagnies d'assurance ne dépend que des primes fixes offertes, et l'agent s'assure auprès de la compagnie qui a offert la prime la plus faible. Remarquons en particulier que cette hypothèse de myopie est incompatible avec l'existence des mutuelles, celles-ci font effectivement « faillite » dès qu'un agent a un accident.

Sous cette hypothèse, la concurrence qui s'établit sur ce marché est très vive et à l'équilibre les compagnies d'assurance proposent une prime actuarielle et se partagent le marché. De plus, dans ce contexte, l'entreprise souhaite réduire autant que possible le montant de ses fonds propres. La concurrence conduit dès lors à une probabilité d'insolvabilité relativement importante (prime et fonds propres sont les plus faibles possibles).

La réglementation prudentielle a un effet néfaste sur la concurrence. En effet, dans les cas où elle présente une contrainte effective pour les compagnies, elle introduit une contrainte de capacité sur le marché, qui conduit alors à une augmentation des primes. Ce mécanisme entre en jeu que le stock de capital soit donné ou résulte d'une décision stratégique. La conclusion est alors très claire : *la réglementation prudentielle atteint son but à l'équilibre, elle conduit à une diminution très nette de la probabilité d'insolvabilité. Mais cet objectif est réalisé aux dépens de l'assuré, la réglementation a un effet pervers anticoncurrentiel, elle augmente la prime d'assurance.*

2.3 - Concurrence sur le marché de l'assurance : mutualisation des risques et effet taille, l'impact sur la concurrence

La modélisation du jeu concurrentiel

Si l'on suppose que les consommateurs sont rationnels et parfaitement informés, ils comprennent le fonctionnement du marché. L'espérance d'utilité d'un assuré dépend ainsi non seulement de la prime d'assurance, mais aussi du stock de capital par tête et du nombre d'individus assurés par la même entreprise que lui. Le processus concurrentiel s'en trouve fondamentalement modifié.

Notre cadre d'étude peut être simplifié par la remarque suivante. On sait que seuls des contrats efficaces seront proposés à l'équilibre. Or on connaît parfaitement ces contrats. Ce sont des contrats d'assurance parfaite, les sinistres individuels sont totalement remboursés, dont la prime dépend du nombre d'agents qui s'assurent dans la société. Par conséquent, les clauses du contrat prévoient des paiements de primes dépendant de ce nombre.

Mais il est possible de remplacer cette écriture théorique par les clauses suivantes. Pour la mutuelle, « le contrat prévoit un remboursement total des sinistres individuels, mais organise un rappel de primes ex post de manière à équilibrer les comptes ». Une telle clause décrit parfaitement le fait que les primes dépendent du nombre d'agents mutualisés.

De la même façon, les clauses du contrat d'une compagnie sont les suivantes. « Le contrat prévoit un remboursement total des sinistres individuels, moyennant le paiement d'une prime fixe. Si toutefois, le montant total des primes et des fonds propres sont insuffisants pour couvrir les pertes, alors la compagnie organise un rappel de primes pour équilibrer ses comptes ». De nouveau cette clause suffit pour décrire un contrat optimal dont les primes dépendent du nombre d'agents assurés par la compagnie.

Le contrat de la mutuelle est alors totalement décrit par la clause proposée. Pour une compagnie, la clause contractuelle ci-dessus et la donnée de la prime décrivent tout autant le contrat optimal. Par conséquent, en première période du jeu, la mutuelle annonce sa clause, et chaque compagnie sa clause et une prime.

Dès lors que l'on limite les annonces possibles aux contrats efficaces, le jeu de concurrence en contrat est en fait un jeu de concurrence en primes, dans lequel chaque compagnie propose une prime dans une première étape. Dans une seconde étape, les agents observent clauses et primes proposées et choisissent (simultanément) une société d'assurance parmi les sociétés présentes ou de ne pas s'assurer. Notons que la mutuelle dans cette perspective a un rôle passif, la prime que payent les agents étant toujours endogène (ou modifiée ex post), la mutuelle n'a pas le contrôle de la prime qu'elle fait payer aux agents.

La seconde étape du jeu décrit comment, suite à des offres des sociétés, les agents se répartissent entre elles. Un équilibre de cette étape est appelée une configuration de marché. Comme nous l'avons déjà remarqué, l'espérance d'utilité d'un agent qui s'assure auprès d'une société dépend du contrat, mais aussi du nombre d'individus s'assurant avec lui. Cet effet réseau engendre une multiplicité d'équilibres en seconde période, donc suite à une offre, plusieurs configurations sont possibles. Or, on sait que parmi ces configurations, il en existe une qui maximise l'espérance d'utilité du marché. Elle est telle que tous les agents s'assurent auprès d'une société unique, soit la mutuelle, soit la compagnie qui a offert la prime la plus faible. Toutes les autres configurations possibles donnent aux agents une espérance d'utilité plus faible. Cette propriété vient de l'application de la loi des grands nombres et est due à la mutualisation des risques. Une question est alors posée, peut-on s'attendre à ce que tous les assurés sélectionnent (ou se coordonnent) sur cet équilibre aux merveilleuses propriétés alors même qu'il existe plusieurs équilibres ? Bien sûr, si les agents obéissaient tous à un chef unique, celui-ci pourrait sélectionner l'équilibre le meilleur pour eux.

Afin de préciser la recherche des équilibres du jeu, nous adopterons la démarche suivante. Dans une première étape de la recherche, nous supposerons que toute offre est spontanément suivie de cette configuration optimale pour les agents. Tout se passe alors dans le jeu comme si les agents pouvaient se coordonner sur l'équilibre de seconde période qu'ils préfèrent. Nous déterminons les équilibres du jeu quand ce comportement est prévu par les entreprises. Cette situation est appelée comportement de type boule de neige. Il implique une monopolisation du marché et une très forte volatilité des assurés, si une compagnie baisse sa prime, elle attire tous les assurés.

Dans un second temps, nous relâchons cette hypothèse et complétons notre analyse en recherchant tous les équilibres du jeu.

Des offres et une configuration de marché forment alors un équilibre du jeu dès lors que chaque compagnie obtient un profit supérieur à celui qu'elle prévoit si elle changeait sa proposition. Mais comme il existe plusieurs configurations d'équilibre qui répondent à l'une de ses déviations, il est possible qu'elle prévoit des situations différentes.

Enfin nous étudions successivement les cas de monopole pur (seule une compagnie est présente), de monopole concurrencé par une mutuelle, d'oligopole pur, puis d'oligopole concurrencé par une mutuelle.

Les équilibres si les agents adoptent toujours le comportement « boule de neige »

Les résultats

Supposons dans un premier temps qu'il existe une seule entreprise. Ce monopole assure tout le marché ou personne et il assure tout le marché tant que l'espérance d'utilité d'un agent assuré (au milieu de tous les autres) est plus grande que son utilité s'il ne s'assure pas. Par conséquent la meilleure offre du monopole est telle que l'agent est indifférent entre s'assurer ou non, il paye alors la prime la plus élevée possible compatible avec le fait qu'il peut refuser l'assurance.

L'introduction d'une mutuelle dans ce modèle modifie très fortement le résultat car désormais le monopole doit offrir une prime telle que l'espérance d'utilité de l'agent assuré (au milieu de tous les autres) est supérieure à celle qu'il obtiendrait si tous les agents (et lui même) s'assuraient dans la mutuelle. La prime offerte tend vers la prime actuarielle. Le taux de profit du monopole est alors très faible

Le phénomène est identique lorsque plusieurs compagnies sont présentes sur le marché. L'une d'entre elles s'accapare le marché et offre la prime actuarielle.

L'introduction d'une mutuelle dans ce processus concurrentiel sauvage n'affecte pas le résultat. Mais, comme précédemment, si le taux de profit est trop important, les compagnies laissent place à la mutuelle. Notons de plus que cette situation ne peut être durable car l'une au moins des entreprises immobilise son capital et ne perçoit aucun profit.

Si les entreprises peuvent entrer et sortir du marché, et si cette entrée est séquentielle au sens où dans un premier temps les entreprises choisissent d'être présentes ou non puis dans un second temps offrent des contrats, et si le taux de profit de l'économie est strictement positif, le marché se monopolise de façon spontanée. S'il n'existe pas de mutuelle, seule une

entreprise privée survit sur ce marché. S'il existe une mutuelle et si le taux de profit est suffisamment grand, seule la mutuelle sera présente sur le marché.

Les conclusions

La mutuelle peut jouer un rôle concurrentiel très important malgré sa passivité stratégique quand le marché souffre d'un défaut de concurrence.

L'équilibre dans le cas de l'oligopole est efficace. Toutefois, il reflète une situation qui n'est pas durable à long terme.

S'il n'existe pas de mutuelle, et de façon paradoxale, parce que les consommateurs se coordonnent ex post spontanément sur la situation qui leur est la plus favorable, la concurrence est détruite ex ante.

Trop de concurrence tue la concurrence dans ce modèle.

Les équilibres du jeu

Remarquons en effet que des propositions et une configuration de marché forment un équilibre si aucun acteur, les assurés, les compagnies, n'ont intérêt à dévier. Or on a vu que l'effet taille était tel qu'il pouvait exister plusieurs configurations en réponse aux offres des compagnies.

Supposons qu'une entreprise dévie. Le profit qu'elle tire de sa déviation dépend de façon fondamentale de la réaction des assurés à cette déviation. Or, si l'on abandonne l'hypothèse selon laquelle les assurés se coordonnent de façon spontanée sur l'équilibre qui est le meilleur pour eux, plusieurs types de configurations peuvent répondre à la nouvelle offre de l'entreprise déviante. On doit donc s'attendre à une multiplicité d'équilibres, et chacun d'entre eux dépend de la configuration prévue par la compagnie en réponse à sa déviation.

Sur le plan théorique, pour déterminer tous les équilibres du jeu, il suffit de considérer le cas dans lequel une entreprise qui dévie prévoit que sa déviation sera suivie de la configuration qui lui est la plus défavorable. En effet celle-ci est possible, mais de plus si l'entreprise n'a pas intérêt à dévier quand elle est optimiste, alors elle ne dévie pas quand elle suppose le pire.

Cette remarque nous permet de comprendre à quel point les entreprises vont perdre leur pouvoir sur ce marché. En effet, désormais, une baisse de prime peut être vue par les

assurés comme une perte d'utilité, il suffit pour cela qu'ils pensent que les autres consommateurs pensent comme eux, et donc que la « ruée » de la section précédente n'ait pas lieu. Dans un certain sens le marché est devenu plus inerte parce que la réaction de chaque agent assuré dépend de celle des autres.

Les résultats

Cette perte de pouvoir concurrentiel est bien illustrée par la situation de monopole.

Le jeu de monopole admet plusieurs équilibres. Parmi ceux-ci, il est possible que le monopole assure tous les agents et facture une prime plus faible que la prime maximale qui était le seul équilibre du jeu de monopole sous l'hypothèse de comportement boule de neige. Plus précisément, il existe un intervalle, dont la borne supérieure est cette prime maximale, tel que toute prime dans cet intervalle est un équilibre du jeu. Très clairement l'espérance de profit de la compagnie a diminué.

Quand le monopole est concurrencé par une mutuelle, celle-ci perd de la même façon son pouvoir concurrentiel. En effet, tous les équilibres de monopole pur tels que tout le marché est assuré restent des équilibres du jeu dans lequel cette compagnie et une mutuelle se livrent concurrence.

Mais, désormais, il est aussi possible qu'à l'équilibre la mutuelle seule assure tous les agents. Enfin, il existe des équilibres dans lesquels les deux structures d'assurance se partagent le marché. Dans ce cas, l'espérance d'utilité de l'agent est plus faible que lorsque seule la mutuelle est active à l'équilibre. Remarquons toutefois que le monopole concurrencé par la mutuelle (quand les deux systèmes coexistent) facture une prime plus importante que le monopole concurrencé sous l'hypothèse « boule de neige ». Apparaît alors une conclusion à priori paradoxale, il est possible que la compagnie profite de la perte de pouvoir concurrentiel. Dans un certain sens, l'effet taille engendre de la concurrence imparfaite.

La nature de ces résultats subsistent dans le cas de l'oligopole. Il semble difficile de déterminer tous les équilibres du jeu, toutefois certains d'entre eux illustrent le défaut de concurrence qui apparaît dans cette analyse.

Si la mutuelle n'existe pas, on peut trouver des équilibres dans lesquels les compagnies d'assurance privées se partagent le marché en facturant une prime supérieure à la prime actuarielle.

Si la mutuelle existe, comme pour le monopole, les équilibres existants avec oligopole seul restent des équilibres. Il est aussi possible que la mutuelle assure seule le marché.

Enfin il existe des équilibres mixtes, dans lesquels les deux structures d'assurance coexistent. Les primes dans ce cas peuvent être assez importantes. Plus précisément dès lors que la mutuelle existe, toutes les offres des compagnies d'assurance telles que les primes sont strictement supérieures à la prime actuarielle et telles qu'il existe une configuration stable de marché sont des équilibres du jeu.

Les conclusions

Quelques faits marquants apparaissent donc quand on abandonne l'hypothèse de comportement en « boule de neige ».

En premier lieu, notons que le jeu admet un très grand nombre d'équilibres, le modèle perd son pouvoir prédictif mais gagne peut-être en réalité. Ces équilibres reflètent une situation de concurrence imparfaite, fondamentalement liée aux effets taille perçus par les assurés qui modifie l'impact qu'ont les diminutions des primes sur l'espérance d'utilité des agents.

De la même façon, la mutuelle perd son pouvoir concurrentiel, et la conséquence est que sa présence peut rendre possible des équilibres très néfastes pour les assurés, toutefois nous n'avons pas montré que le même type d'équilibre n'existe pas sans mutuelle.

Enfin l'équilibre du jeu n'est pas toujours efficace au sens de Pareto.

Conclusion

. Un élément d'explication des mouvements de fusions

Notre modèle fournit un argument supplémentaire en faveur des fusions. Les fusions conduisent à exploiter l'effet de mutualisation des risques et sont donc optimales, si bien sûr on ignore les effets anti-concurrentiels qu'elles peuvent engendrer. La fusion de mutuelles doit toujours être encouragée.

. La réglementation prudentielle, quel impact ?

La réglementation prudentielle engendre un défaut de concurrence. Elle conduit les entreprises à facturer des primes d'assurance très élevées. On pourrait argumenter qu'en protégeant les profits des entreprises, la réglementation protège l'entrée et donc favorise la concurrence. Cette remarque ne change pas notre résultat. Une seule entreprise serait plus efficace sur ce marché.

. *Coexistence des mutuelles et des compagnies, quel concept de concurrence ?*

Notre modèle souligne une différence fondamentale existant entre compagnies et mutuelles. L'absence (théorique dans un modèle statique) de fonds propres des mutuelles affecte le contrat qu'elles proposent et donc la concurrence qu'elles livrent aux autres acteurs. La mutuelle, parce que l'effet réseau y est le plus fort, est la plus fragile face aux comportements du marché quand les agents ne se coordonnent pas spontanément sur le meilleur équilibre. Son pouvoir concurrentiel est maximal quand il existe des défauts de concurrence.

Enfin notons que rien ne garantit que plus de concurrence conduise à plus d'efficacité dans ce modèle. L'effet réseau engendre une imperfection de la concurrence.

Bibliographie

BarNiv R., Hathorn J. (1997) «The Merger or Insolvency Alternative in the Insurance Industry », *Journal of Risk and Insurance*, 64, 89-113.

Joskow, P. L. (1973), « Cartels, Competition and Regulation in the Property-Liability Insurance Industry », *Bell Journal of Economics and Management Science*, 4, 375-427.

Conseil de la Concurrence (1998), *Onzième Rapport d'Activité de la Commission Européenne de la Concurrence (année 1997)*, Journal Officiel de la République Française.

Lorenzi J.- H. (1998) «L'industrie de l'assurance », dans Ewald F. et Lorenzi J.-H. Eds., *Encyclopédie de l'assurance*, Economica.

Pestieau P., Pirard C. (1989) «L'entreprise d'assurance : économies d'échelle et performance », *Revue d'Economie Financière*, 11.

Schor A. D. (1994) « Concentration, contestabilité et concurrence sur le marché français de l'assurance », *Revue d'Economie Industrielle*, 70, 17-31.

QUATRIEME PARTIE

**Comment allouer et financer des missions de service
universel dans un réseau ouvert à la concurrence?**

Une évaluation du système français

La libéralisation des activités de réseau, à l'oeuvre en Europe dans la plupart des activités en réseaux, s'accompagne d'une remise en cause des systèmes traditionnels soutenant les missions de service public³³. Tandis qu'en régime de monopole, ces missions sont assurées par un opérateur unique qui finance par subventions croisées le déficit laissé par les usagers non rentables, l'ouverture à la concurrence incite les firmes rivales à entrer d'abord sur les segments de marché rentables. Du fait de la concurrence par les prix qui s'y exerce, les prix baissent et l'opérateur initialement en charge des missions de service public n'est plus en mesure de financer les pertes liées à ces obligations. Ce phénomène *d'écrémage*, que connaissent à des degrés divers tous les secteurs ouverts à la concurrence, suscite de nombreuses questions pour l'organisation des obligations de service universel.

Tout d'abord, certains pays ont tout simplement renoncé à imposer des obligations de service public ou universel à leurs opérateurs : le cas des télécommunications en Suède est à cet égard particulièrement exemplaire, puisque aucune obligation n'est imposée ni à l'opérateur historique ni à ses concurrents. L'un des résultats de cette situation, unique en Europe à notre connaissance, est le développement extrêmement rapide du téléphone mobile dans ce pays : au lieu d'imposer un raccordement au téléphone fixe socialement coûteux et inefficace pour les résidences isolées (nombreuses en Suède), les autorités n'ont ici imposé aucune contrainte de desserte ni d'uniformité géographique des tarifs, et ont compté sur le développement du marché concurrentiel des mobiles, aux coûts et aux tarifs moins élevés sur ces segments de marché, pour que la couverture téléphonique du pays soit quasi-complète³⁴. Les tarifs du téléphone fixe ne faisant l'objet d'aucune péréquation géographique, la téléphonie mobile a assez naturellement occupé le créneau des résidences secondaires isolées, très nombreuses en Suède. Il semble, intuitivement, bien que nous ne disposions d'aucune étude quantitative sur ce sujet, qu'il s'agisse là, dans le cas de la Suède, d'une solution plus efficace, en termes de coût collectif, que l'imposition d'une contrainte de desserte de territoires relativement déserts, accompagnée d'une distorsion de l'ensemble des tarifs téléphoniques.

³³ Dans tout ce qui suit, nous parlerons indifféremment de service public ou de service universel. Ce que nous entendons précisément par là est défini un peu plus loin.

³⁴ Il faut d'ailleurs noter que le monopole légal n'existait pas en Suède avant la libéralisation du marché européen. Pour plus de détails sur le système suédois, voir C. Henry (1997), *Concurrence et services publics dans l'Union Européenne*, PUF.

On pourrait donc s'interroger, du moins dans certaines configurations (géographiques, démographiques, concurrentielles, sectorielles, etc...) sur le bien-fondé de l'existence même de certaines contraintes de service universel : dans certains cas, et l'exemple précédent le suggère, il pourrait se faire que des évolutions technologiques rentabilisent finalement des services déficitaires lorsqu'ils étaient offerts en monopole et avec les anciennes technologies, rendant caduque la contrainte de desserte géographique universelle par un moyen donné (le téléphone fixe par exemple)³⁵. Par ailleurs, les obligations de service universel imposent souvent la péréquation des tarifs, et certains des développements de notre étude suggèrent qu'il pourrait s'agir là d'une obligation extrêmement coûteuse du point de vue collectif, en raison des distorsions importantes qu'elle entraîne.

Cependant, nous n'envisagerons pas les questions, pourtant fondamentales, du bien-fondé des contraintes de service universel. Il n'est pas douteux par ailleurs qu'en général, les obligations de service universel sont indispensables si l'on poursuit un objectif de taux de couverture de 100% de la population, si bien que les cas où les remarques précédentes sont pertinentes sont à l'évidence très peu nombreux.

Notre recherche vise donc à examiner, parmi une série de possibilités, quelles sont les solutions les plus efficaces pour remplir certaines obligations de service universel, et à situer les options françaises dans cet éventail.

Certains pays européens ont parfois opté pour la *mise aux enchères* de segments de marché relevant du service universel. Tel est, par exemple, le cas de certains pans du service postal en Allemagne. Différents mécanismes d'enchères sont envisageables³⁶, mais quels qu'ils soient, ces mécanismes débouchent tous sur une allocation du segment de marché mis aux enchères à la firme offrant le meilleur prix s'il s'agit d'une enchère au *moins* disant (et donc à la firme la plus efficace), ou le meilleur score dans le cas d'une enchère au *mieux* disant.

³⁵ De la même façon, assurer un maillage fin du territoire par les transports collectifs ne requiert pas nécessairement une desserte *ferroviaire* universelle, un service de transport par bus pouvant atteindre cet objectif à moindre coût.

³⁶ En particulier, une abondante littérature théorique, désormais complétée par quelques travaux empiriques, s'intéresse à la question de l'organisation optimale du mécanisme d'enchères : comment faut-il prendre en compte les problèmes liés à la qualité des biens ou des services ? Faut-il découper le marché vendu aux enchères en lots, et dans l'affirmative, comment ? Faut-il choisir un système au premier ou au second prix ?

Un mécanisme d'allocation par enchères est donc fort différent de celui qui prévaut en France, où les missions de service universel sont en général allouées de façon automatique à l'opérateur historique.

D'autres mécanismes encore sont parfois mis en oeuvre, loin de l'Europe cette fois: ainsi, dans le secteur des télécommunications en Australie, la règle du « pay or play » permet à des entrants d'accomplir, au moins localement, certaines de ces missions plutôt que d'y contribuer financièrement par le biais du paiement d'une taxe. Ce sont donc les opérateurs concurrents eux-mêmes qui peuvent prendre l'initiative de remplir ces obligations.

Les exemples qui précèdent amènent donc à traiter deux questions :

- . une question *d'allocation* des missions de services universel: comment les opérateurs en charge de telles missions devraient-ils être désignés (procédure d'enchères, désignation d'un opérateur par le régulateur, règle du pay or play) ?
- . une question de *financement* du service universel : les déficits laissés par les usagers les moins rentables doivent-ils être couverts par des transferts forfaitaires payés par les contribuables, par une taxation des opérateurs concurrents, par subventions croisées ?

Les réponses à ces différentes questions ont des implications importantes à la fois en termes de redistributions entre les agents (usagers, contribuables, opérateurs) et en termes d'efficacité. C'est ce double point de vue qui est examiné dans notre travail.

1. Les obligations de service universel et les options de régulation :

Suivant les contextes institutionnels, différentes obligations de service universel peuvent être imposées par le régulateur :

- . la *contrainte d'ubiquité* impose que tous les usagers soient connectés au réseau, ce qui suppose que l'opérateur à qui cette contrainte s'impose offre le raccordement à un tarif « supportable » par les agents les plus défavorisés en termes de revenus ou de coût de raccordement.
- . la *contrainte de non-discrimination* impose que tous les usagers se voient offrir le même tarif ou le même menu de tarifs.

Nous intéressons ici aux contraintes *d'ubiquité* et de *non-discrimination géographique*, dans un contexte où la seule hétérogénéité entre les consommateurs est celle qui a trait à leur coût de raccordement au réseau : tous les usagers ont par ailleurs la même fonction de demande. Nous écartons par conséquent la question du traitement des petits consommateurs, qui

pourraient relever aussi du service universel, pour nous concentrer exclusivement sur les aspects géographiques souvent justifiés par des impératifs « d'aménagement du territoire ».

Nous prenons ces contraintes de service universel comme des données, sans revenir sur leur justifications, c'est-à-dire sur la façon dont il faudrait valoriser du point de vue collectif les bénéfices liés au respect de ces contraintes. Nous cherchons plutôt à *évaluer les coûts* liés au choix de telle ou telle organisation institutionnelle. Plus précisément, nous évaluons les avantages et inconvénients de trois systèmes de régulation :

- . le système de « l'entrée restreinte » ce cas correspond au système français. Le régulateur impose à la firme en place (l'opérateur historique) les contraintes d'ubiquité géographique et de non-discrimination. Les pertes supportées par l'opérateur sont compensées (partiellement ou complètement) par taxation: une taxe dont le niveau est choisi par le régulateur est perçue sur chaque unité de bien vendue par l'ensemble des opérateurs, et son montant est reversé à la firme en charge des obligations de service universel. Concrètement, cette taxe vient s'ajouter à la charge d'interconnexion subie par les concurrents de France Telecom lorsqu'ils utilisent le réseau de France Telecom.
- . le système du « pay or play » : ce mécanisme, en vigueur par exemple en Australie, impose également à un opérateur donné de remplir en dernier ressort les contraintes de service universel. Il est également financé par des taxes prélevées sur les quantités vendues par tous les participants au marché. Mais à la différence du mécanisme précédent, chaque opérateur, au vu de la taxe unitaire dont il devra s'acquitter, peut choisir de servir lui-même les consommateurs déficitaires. Il est ainsi dispensé de payer les taxes correspondantes.
- . la mise aux enchères, assortie d'un financement par le biais de transferts forfaitaires. Dans ce cas, nous nous restreignons à une enchère au second prix. Le vainqueur de l'enchère reçoit le montant de subvention immédiatement supérieur au niveau qu'il a réclamé, et ce montant est prélevé sous forme de contribution forfaitaire (n'affectant donc pas la structure de la demande) sur les revenus des contribuables.

Notre référence implicite est celle du secteur des télécommunications (téléphone, câble, Internet) parce que la libéralisation y ayant été amorcée plus tôt, la réflexion théorique ainsi que l'expérimentation pratique sont les plus avancées. Cependant, nous développons une réflexion générale qui s'applique également, dans son essence, à des secteurs assez variés, comme celui de l'énergie ou du transport ferroviaire. L'article peut donc être lu comme une réflexion générale sur les modes d'allocation et de financement d'obligations de service universel dans des secteurs en réseaux variés.

Nous donnons ci-après un aperçu de la modélisation et résumons nos principaux résultats. Le modèle formel est joint en annexe.

2. Le cadre d'analyse adopté :

2.1. Hypothèses de comportement et représentation formelle :

On considère un marché de bien réseau. Le service offert aux consommateurs y est homogène et deux entreprises (I et E) sont potentiellement en concurrence sur ce marché. Ce réseau a un caractère de «service public» au sens où le régulateur, chargé de l'organisation de la concurrence sur le marché, poursuit des objectifs « d'accès égal » de toute la population au réseau.

Du côté de la demande, deux types d'utilisateurs se distinguent par leur coût de raccordement au réseau. Typiquement, si l'infrastructure connaît des rendements croissants, alors les coûts de raccordement sont plus élevés dans les zones de peuplement dense (zones urbaines) que dans les autres zones (zones rurales). Pour simplifier les idées et l'expression, nous désignerons ces deux catégories d'utilisateurs par « urbains » et « ruraux ».

Pour chacune des firmes, nous faisons l'hypothèse qu'il est toujours moins coûteux de raccorder un utilisateur urbain qu'un utilisateur rural au réseau, et ceci quelle que soit la demande exprimée par cet utilisateur. En revanche, nous ne posons aucune hypothèse a priori sur l'efficacité relative des deux opérateurs. Cette efficacité peut même être différente sur les deux segments de marché, l'opérateur le plus efficace pour servir les urbains pouvant être moins compétitif sur le marché des ruraux et réciproquement.

Lorsqu'un consommateur s'abonne chez un opérateur K (K = I ou E), et y consomme une quantité donnée de bien, cette transaction engendre un surplus qui est partagé entre la firme et le consommateur en fonction du tarif (non linéaire) choisi par l'opérateur. Pour mettre l'accent sur les problèmes posés par les obligations de service universel, nous supposons dans la suite que le *service d'un utilisateur urbain engendre un surplus positif* tandis que celui d'un *rural engendre un surplus négatif*. Il n'est donc pas possible de servir profitablement cette catégorie de consommateurs. On suppose que les opérateurs se font concurrence en tarifs ; du point de vue formel, ceci revient à modéliser une concurrence en niveaux d'utilité (ou parts du surplus) laissé(e)s au consommateur.

2.2 La situation de référence : concurrence sans obligation de service universel.

Lorsqu'aucune obligation de service universel n'est imposée aux opérateurs rivaux, ceux-ci se font concurrence séparément sur chaque segment de marché (urbain et rural). Compte tenu des hypothèses ci-dessus, il en résulte que les usagers urbains sont servis par l'entreprise qui est la plus efficace sur ce marché : en effet, les usagers urbains sont rentables et les firmes se battent pour obtenir ces consommateurs. Au contraire, les usagers ruraux ne sont pas servis parce qu'ils ne sont pas rentables. C'est de ce dernier résultat que découle la nécessité, si le régulateur valorise le raccordement de toute la population au réseau, d'imposer au moins la contrainte d'ubiquité.

2.3. Les obligations de service universel :

Nous distinguons deux types d'obligations :

- . *une contrainte d'ubiquité* : le régulateur impose alors à un opérateur de servir les usagers non rentables (les ruraux dans notre représentation de l'économie) ; ceci revient à imposer à l'opérateur en charge de cette contrainte de laisser une part positive ou nulle de surplus aux consommateurs ruraux.
- . *une contrainte de non-discrimination géographique* : le régulateur impose que tous les usagers soient servis au même tarif.

Evidemment, ces deux contraintes font peser sur les opérateurs qui en ont la charge des coûts additionnels par rapport à la situation de référence :

- . la contrainte d'ubiquité engendre des pertes du fait du caractère non rentable des usagers ruraux,
- . la contrainte de non discrimination engendre des pertes du fait qu'elle interdit d'exploiter complètement le surplus des usagers de chaque catégorie.

Par ailleurs, il est clair qu'on peut envisager un traitement séparé de ces deux contraintes, qui peuvent aussi bien être imposées l'une sans l'autre. C'est ce que nous ferons par la suite afin d'identifier les coûts de chacune d'elles.

Les systèmes que nous allons envisager se distinguent par deux caractéristiques:

- . le mode d'attribution des contraintes entre les deux opérateurs ;
- . le mode de financement de ces contraintes.

Comme nous l'avons souligné dans l'introduction, bien des méthodes sont envisageables pour allouer les obligations de service universel entre des opérateurs concurrents. On peut imaginer la mise aux enchères de certains segments de marché, confier ces segments de façon impérative à un opérateur désigné; ou envisager tout un éventail de solutions entre ces deux

solutions «polaires». De la même façon, en ce qui concerne le financement des déficits laissés aux opérateurs en charge des contraintes, des subventions croisées, de la taxation ou des transferts forfaitaires peuvent être mis en place. Nous allons étudier certaines combinaisons de ces différentes approches.

2.4. Le système français : régulation par entrée restreinte

Le système français, que nous qualifierons dans la suite de régulation par «entrée restreinte», combine une attribution par le régulateur des obligations de service universel à un opérateur (l'opérateur installé, soit I dans notre représentation), et un financement par taxation ou par subventions croisées.

Si l'opérateur concurrent (E) évince la firme en place (I) des segments de marché rentables (écrémage des consommateurs urbains), alors les subventions croisées sont impossibles: on est alors dans le cas où l'entrée d'un concurrent produit de l'écrémage et le régulateur prélève sur chaque unité de bien vendue une taxe unitaire, dont le montant sera ensuite versé à l'opérateur I qui subit les pertes. On appellera ce régime un «régime de taxation».

Si au contraire l'opérateur I reste en place sur le segment des urbains, alors l'opérateur E n'est pas actif sur le marché³⁷, et I est en mesure de financer ses pertes par subventions croisées. On appellera ce régime «régime de subventions croisées par I».

L'une ou l'autre de ces configurations se produira à l'équilibre du jeu suivant qui résume la situation envisagée dans cette régulation «à la française»; par entrée restreinte :

- * le régulateur annonce la taxe unitaire qui sera prélevée sur chaque unité de bien dans le régime de taxation.
- * la firme I annonce les surplus (ou de façon équivalente les tarifs) qu'elle laissera à chaque catégorie d'utilisateurs ;
- * la firme E choisit de servir ou non les utilisateurs urbains.

Il en résulte un régime d'équilibre soit de taxation soit de subventions croisées. Il sera intéressant d'évaluer, à l'issue de ce jeu, qui perd et qui gagne à l'imposition des contraintes par rapport à la situation initiale. Clairement le niveau de la taxe, déterminant pour situer

³⁷ Ceci est lié au fait que les opérateurs se font concurrence en tarifs (ou en niveaux d'utilité) : une concurrence en quantités aurait donné des résultats moins tranchés, les opérateurs pouvant alors se partager le marché des utilisateurs rentables.

l'équilibre dans l'un ou l'autre des régimes, a une grande importance pour évaluer les transferts et la répartition entre agents qui en découle.

2.5. La régulation par « pay or play » :

Dans ce mode de régulation, les obligations de service universel sont confiées, en dernier ressort, à l'entreprise historique, I. Toutefois, l'entrant peut choisir de s'en acquitter lui-même.

Dans ce cas, il ne paye pas la taxe. L'arbitrage auquel il est soumis est donc le suivant:

- . soit il choisit de servir (« play ») lui-même les consommateurs non rentables, et dans ce cas, il doit se financer par subventions croisées. Ce cas n'apparaîtra donc que si l'entrant sert également les usagers urbains, et dans le cas où E est « efficace par rapport au niveau de la taxe ».
- . Soit il choisit de ne pas les servir (« pay ») et de se retrouver dans ce cas dans une position comparable à celle de l'entrée restreinte.

Le jeu qui résume cette configuration réglementaire est le suivant:

- . le régulateur annonce le niveau de la taxe;
- . I annonce les niveaux d'utilité qu'il offre aux urbains et aux ruraux ;
- . E choisit les niveaux d'utilités qu'il offre aux urbains et aux ruraux. S'il offre aux ruraux une utilité supérieure à celle laissée par I, il sert ces consommateurs et ne verse aucune taxe. Sinon, c'est I qui sert les usagers non rentables et qui finance alors ses pertes soit par subventions croisées soit par les taxes reçues de E.

A l'équilibre de ce jeu, plusieurs configurations d'équilibre peuvent apparaître : comme dans le cas de l'entrée restreinte, les régimes de « subvention croisées par I » ou de « taxation » restent possibles. Mais apparaît maintenant un régime de « subventions croisées par E » dans le cas où E choisit de servir le marché des ruraux.

2.6 La régulation par enchères et transferts forfaitaires:

Aucune des deux régulations précédentes n'envisage le système, pourtant économiquement attrayant, consistant à confier le service du marché des usagers non rentables à l'opérateur le plus efficace sur ce marché. C'est la solution que mettrait en oeuvre un système d'enchères interdisant les subventions croisées. Nous envisageons donc la possibilité de combiner mécanisme d'enchères (au second prix) et transfert forfaitaire prélevé sur les contribuables et versé à l'entreprise vainqueur de l'enchère.

Plus précisément, le système d'enchères considéré ici consiste à demander aux entreprises de soumettre des offres indiquant la subvention qu'elles réclament pour servir le marché des consommateurs ruraux. Le vainqueur est l'entreprise ayant demandé la plus faible subvention; elle se voit attribuer une subvention égale à la proposition immédiatement supérieure. Le montant de cette subvention est récolté par des prélèvements forfaitaires sur les contribuables. Il existe un coût de collecte des fonds publics, qui rend coûteux le transfert en question.

A l'équilibre, comme les mécanismes d'enchères au second prix sont révélateurs, la firme la plus efficace sur le marché des usagers non rentables sert le marché et reçoit un montant égal au coût demandé par l'autre.

La comparaison de ces différents systèmes (entrée restreinte, pay or play, enchères) aussi bien en termes d'efficacité qu'en termes de redistribution, dépend évidemment du niveau de la taxe fixé par le régulateur dans les deux premiers cas, c'est-à-dire de l'objectif poursuivi par celui-ci.

Nous envisageons certains modes de détermination de la taxe.

2.7. Modes de détermination du taux de taxe

- . Tout d'abord, le régulateur peut simplement poursuivre un objectif de maximisation du surplus global. La taxe qui en résulte, dite taxe de premier rang, peut fort bien laisser subsister un déficit pour l'opérateur qui subit les obligations de service universel, ce qui justifie que d'autres modes de détermination soient adoptés.
- . En second rang, le régulateur choisit la taxe de façon à maximiser le surplus global sous contrainte d'équilibre budgétaire de l'opérateur en charge des obligations de service universel.
- . Enfin, on peut adopter une taxe « concurrentiellement neutre » : c'est le taux de taxe qui égalise les profits de l'opérateur en situation de référence et à l'équilibre du jeu contraint par les obligations de service universel. C'est ce que recommandent les autorités américaines.

2.8 Evaluation des coûts du service universel

Une remarque importante, à ce stade, est la suivante : quelle que soit la méthode de régulation retenue, l'évaluation du coût social des obligations de service universel fait intervenir deux éléments :

- . d'une part, des pertes de bien-être liées à l'inefficience productive : il se peut en effet qu'à l'équilibre, la firme la plus efficace pour servir une catégorie d'utilisateurs ne soit pas celle qui les sert effectivement.
- . d'autre part, une inefficience allocative, liée au mode de financement des obligations. Que le régime soit celui de la taxation, celui des subventions croisées, ou fasse intervenir des transferts forfaitaires, en effet, des pertes de bien-être apparaissent du fait de distorsions dans l'allocation des ressources.

Il en résulte que cette évaluation ne peut être menée que *conditionnellement à une situation de référence* - qui peut être elle-même définie comme une situation concurrentielle ou monopolistique - : évaluer ces coûts dans l'absolu n'a pas de sens. Il s'agit selon nous d'une conclusion importante de ce travail.

3. Les résultats:

Nos principaux résultats peuvent être résumés de la façon suivante :

- 1) Lorsque seule la contrainte d'ubiquité est imposée, les régulations « entrée restreinte » ou « pay or play » permettent toutes deux de décentraliser le choix du régulateur en le déléguant à l'opérateur I. En d'autres termes, celui-ci, qui agit comme leader dans le jeu concurrentiel qui l'oppose à l'entrant, fait des choix stratégiques qui aboutissent à faire coïncider le régime d'équilibre et le régime socialement optimal. Il y a efficacité productive au sens défini plus haut et le coût de la contrainte d'ubiquité est exclusivement allocatif (lié à la présence d'une taxe).
- 2) Lorsque la contrainte de non discrimination est ajoutée à la contrainte d'ubiquité, en revanche, ce résultat ne tient plus. Les deux modes de régulation induisent des sous-optimalités par rapport au régime qu'aurait choisi le régulateur. En outre, du fait des arbitrages des consommateurs entre les deux firmes et de la concurrence sur le marché, même si la contrainte de non discrimination n'est pas imposée à l'entrant, celui-ci la respecte à l'équilibre.
- 3) Pour un niveau donné de la taxe, le « pay or play » domine toujours la régulation d'entrée restreinte lorsque seule la contrainte d'ubiquité est imposée. En effet, l'avantage du « pay or play » est de permettre à l'entrant de remplir l'obligation de service universel si il est

plus efficace. Mais lorsque le régulateur impose aussi la contrainte de non discrimination, l'entrée restreinte, il peut être amené à être actif plus souvent que ne le voudrait l'optimum social. Par ailleurs, le « pay or play » supprime trop souvent aussi le régime de taxation au profit d'un système de subventions croisées.

- 4) Du point de vue redistributif, chacun des modes de régulation induit des transferts différents entre les différentes catégories d'agents (consommateurs urbains et ruraux, entreprise en place et entrante). Un régime de subventions croisées, par exemple, consiste essentiellement en un transfert des consommateurs urbains vers les ruraux. Un système de taxation est plus directement un transfert d'une entreprise à l'autre, même si indirectement la demande et donc le surplus des consommateurs est affecté. Un mécanisme financé par les contribuables implique des transferts de ceux-ci vers les usagers. Il faut donc être conscient du fait que le passage d'un système à l'autre générerait nécessairement des gagnants et des perdants.
- 5) Enfin, contrairement à l'intuition, un mécanisme d'enchères financé par transferts forfaitaires ne domine pas nécessairement les deux autres modes de régulation. L'une des raisons majeures en est qu'il élimine la possibilité de financer les obligations par subventions croisées ; or ce régime peut être le meilleur du point de vue du surplus social.

Conclusion et perspectives de recherche

Cette analyse permet de mettre en relief certaines propriétés, en termes de d'efficacité et de redistribution du surplus, de mécanismes de régulation observés en pratique. En particulier un de nos objectifs était la comparaison du système français avec d'autres solutions réglementaires. Même si cette analyse demeure partielle, elle fait apparaître tout d'abord que si la contrainte d'ubiquité n'est pas distorsive, la contrainte de non-discrimination engendre au contraire des inefficiences importantes. Ensuite, nos résultats font ressortir que le financement par subventions croisées n'est pas nécessairement si inefficace qu'on pourrait le croire au premier abord. Le système de l'entrée restreinte, qui représente bien le cas français, n'est pas systématiquement dominé par des mécanismes plus concurrentiels comme le pay or play. En revanche, les consommateurs gagnent en général à la possibilité d'une concurrence, ne serait-elle que potentielle comme ici. En ce sens, le pay or play leur est généralement plus favorable. Finalement, un élément essentiel dans l'évaluation des coûts du service universel réside dans la manière dont on définit la situation de référence : ici, la référence est celle d'une situation

concurrentielle sans obligations de service universel. On pourrait évidemment imaginer d'autres points de comparaison.

En nous concentrant sur des aspects géographiques, nous avons négligé ici le problème des petits consommateurs. Il s'agit bien sûr d'un autre aspect important du service universel dans un contexte d'ouverture à la concurrence. Nous examinerons cette question ultérieurement.

**CONCURRENCE, INNOVATION ET
INTERNATIONALISATION:
LA DYNAMIQUE DE L'ECONOMIE FRANCAISE**

**CONTRAT CGP
Subvention 5.98**

Vol.1 : LA DYNAMIQUE CONCURRENTIELLE EN FRANCE

Vol. 2 : LES DYNAMIQUES DE L'INNOVATION

Vol 3. Annexes du Vol.1

**CREST-LEI,
URA 2200 (GRECSTA)
Responsable : A. PERROT**

**EUREQua
Université de Paris I, UMR 8594
Responsable : D. ENCAOUA**

**VOLUME 2
*LES DYNAMIQUES DE L'INNOVATION***

COMMISSARIAT GENERAL DU PLAN

Janvier 2000

Table des matières du volume 2

LES DYNAMIQUES DE L'INNOVATION

Première Partie: L'effet de la concurrence sur l'innovation et la croissance économique

- I La problématique
- II Innovations de rattrapage ou/ et innovations de dépassement : un modèle de croissance endogène
- III Caractérisation de l'équilibre stationnaire
- IV Les spécifications de la fonction de coût de R&D et les résultats

Deuxième Partie: Innovation, diffusion des connaissances et croissance

- I L'importance de l'étude des liens entre l'innovation et la croissance au niveau micro-économétrique
- II Un tour d'horizon de la littérature
- III Les modèles théorique et économétrique
- IV Description des données et statistiques descriptives
- V les résultats économétriques

Troisième Partie: Les sources internes et externes d'information dans l'innovation

- I La pertinence de la distinction entre sources internes et externes de l'innovation
- II Les données
- III Le modèle économétrique
- IV Les résultats

Quatrième Partie: Cessation, reprise et survie des entreprises innovantes *versus* non innovantes : un modèle à risques concurrents

- I Pourquoi distinguer les différents types de sortie ?
- II Pourquoi distinguer les entreprises innovantes et non innovantes ?
- III Un modèle de durée: les risques concurrents
- IV Les données
- V Les résultats

Rattrapage ou Dépassement Technologique ? Un modèle de croissance endogène pour étudier l'impact de la concurrence sur l'innovation et la croissance.

David ENCAOUA¹ et David ULPH²

L'objectif principal de cet article est d'examiner l'influence du degré de concurrence sur l'innovation et la croissance économique. L'intensité concurrentielle sur le marché des produits a-t-elle un effet favorable ou défavorable sur la propension à innover ? Avant d'explicitier les prédictions de la théorie économique sur cette question, commençons par préciser le sujet.

Il convient en premier lieu de distinguer la concurrence qui s'exerce sur le marché des produits et celle qui s'exerce pour l'obtention de l'innovation. Concernant cette dernière, plusieurs travaux de base conduisent à une prédiction sans ambiguïté. L'article fondateur de Arrow (1962), suivi de celui de Dasgupta-Stiglitz (1980), puis de l'article de synthèse de Guesneric-Tirole (1985) montrent en effet que l'incitation à innover d'une firme en situation de monopole est plus faible que celle d'une firme sur un marché parfaitement concurrentiel. Ce résultat est une illustration de "l'effet de remplacement" selon lequel un monopole est moins incité à innover qu'une firme concurrentielle, du simple fait qu'un monopole "se remplace lui-même" lorsqu'il innove, alors qu'une firme concurrentielle passe d'une situation de profit nul à une situation de profit de firme dominante en innovant. Le modèle canonique de Lec et Wilde (1980), consacré à l'analyse d'une course à l'innovation où la date aléatoire de succès suit une loi de Poisson ayant pour paramètre une fonction croissante de l'effort de R&D, conduit également à une prédiction analogue. A l'équilibre, la dépense de R&D de chaque participant à la course à l'innovation est une fonction croissante du nombre de participants à cette course. La concurrence pour l'obtention de l'innovation a donc un effet positif sur l'effort de recherche individuel et global. Le même type de résultat apparaît dans l'analyse dynamique de Grossman et Shapiro (1987). Ces résultats, établis le plus souvent dans un cadre d'équilibre partiel, ont été récemment généralisés dans le cadre d'un modèle d'équilibre général de croissance endogène par Aghion et Howitt (1998, section 7.2). L'effet de remplacement d'Arrow subsiste et la concurrence pour l'innovation est favorable à la croissance.

¹ Université de Paris I, Panthéon-Sorbonne, EUREQua

² University College London, ELSE

Cette convergence de résultats s'applique t-elle également à la concurrence sur les marchés de produits ? Autrement dit, quel effet le degré de rivalité des entreprises sur le marché des produits, en matière de concurrence par les prix ou par les quantités, exerce t-il sur l'effort de recherche de chacune d'elles et sur le taux de croissance de l'économie engendré par les innovations correspondantes ? Sur le plan théorique, la prédiction de la plupart des modèles de croissance endogène d'inspiration Schumpeterienne est là aussi sans ambiguïté et elle conduit à un effet inverse du précédent. L'intensité concurrentielle sur le marché des produits est défavorable à l'innovation et à la croissance. La logique du raisonnement conduisant à ce résultat est très simple. Plus l'intensité concurrentielle sur le marché des produits est élevée, plus les rentes anticipées de l'innovation sont faibles, moins les entreprises innovantes sont incitées à fournir d'efforts en matière de R&D et plus le taux de croissance de l'économie généré par l'innovation est faible.

Ce résultat paraît néanmoins en contradiction avec la vision darwinienne ou évolutionniste qui sous tend la plupart des politiques de la concurrence, selon laquelle plus l'intensité concurrentielle est forte plus les entreprises sont incitées à innover pour survivre. Selon cette position, l'existence d'une forte concurrence sur les marchés des produits serait à la source de l'innovation et de la croissance économique. Notons au passage que nous sommes là au coeur d'un débat classique entre deux positions extrêmes: celle qui considère (voir par exemple Jorde et Teece (1990)) que les vertus de la concurrence sur le plan statique de l'allocation des ressources doivent être pondérées par ses effets défavorables sur les plans dynamiques de l'innovation technologique et de la croissance économique, et celle pour qui cette opposition n'est pas fondée (voir par exemple Shapiro et Willig (1990)). Les commentaires suscités par le procès Microsoft aux Etats-Unis illustrent bien la vigueur de ce débat.

Un travail récent de Aghion, Harris et Vickers (1987) a permis cependant une reformulation du problème. Elle consiste à supposer que dans le processus d'innovation, une nouvelle firme se doit d'abord de rattraper la technologie de la firme possédant une avance avant de pouvoir éventuellement la dépasser ultérieurement. Toute nouvelle innovation procéderait ainsi de manière graduelle plutôt que par dépassement de la dernière innovation en date comme le postule l'hypothèse Schumpeterienne de la création destructrice. Dans le cas d'un processus graduel d'innovations, la dynamique industrielle induite conduit à ce qu'à certains moments du temps, les firmes d'une même industrie se retrouvent avoir accès à la même technologie, les amenant alors à être en concurrence frontale sur le marché de leur produit.

On conçoit que dans de telles conditions, plus l'intensité de la concurrence est forte sur le marché, plus chacune des firmes en concurrence frontale est incitée à innover pour échapper précisément à la pression concurrentielle du marché. Cependant, cet effet positif de la concurrence sur l'innovation est contrebalancé par un effet de sens contraire. En effet, plus la pression concurrentielle sur le marché est forte, moins une firme en retard est incitée à innover pour rattraper son retard, dans la mesure où les gains immédiats qu'elle espère obtenir en rattrapant la firme en avance sont limités par l'intensité de la concurrence. L'originalité du modèle de Aghion, Harris et Vickers (1987) est de parvenir à caractériser l'équilibre entre ces deux effets opposés dans le cadre d'un modèle de croissance endogène, en explicitant la proportion à l'équilibre stationnaire des industries qui sont en concurrence technologique frontale. Les situations où à l'équilibre, l'effet positif de la concurrence sur l'effort d'innovation des entreprises en concurrence frontale l'emporte sur l'effet négatif de la concurrence sur l'effort d'innovation d'une entreprise ayant un retard technologique peuvent être ainsi caractérisées.

L'hypothèse d'un processus graduel d'innovation faite dans Aghion, Harris et Vickers (1997) permet donc de restaurer dans certains cas l'influence positive de la concurrence sur l'innovation et la croissance économique. On n'en reste pas moins confrontés au choix de l'hypothèse concernant la caractérisation la plus adéquate du processus d'innovation technologique: l'innovation procède-t-elle par *rattrapage* ou par *dépassement*? Il est clair que dans la réalité, les deux situations sont simultanément présentes. Pour s'en convaincre, on peut se référer par exemple aux enquêtes innovation réalisées aussi bien en Europe (Community Innovation Survey), qu'aux Etats-Unis (Yale Survey). Ces enquêtes font apparaître clairement deux types d'innovation. D'une part, celles qui consistent en la mise au point par une entreprise d'un produit ou d'un procédé qui sont nouveaux pour la firme mais pas pour le marché. D'autre part, celles qui consistent à introduire un produit ou un procédé qui sont nouveaux à la fois pour la firme et pour le marché. Les premières correspondent assez bien à ce qu'on peut appeler des *innovations de rattrapage* tandis que les autres correspondent à des *innovations de dépassement* (pour un traitement économétrique de ces types de données, voir Kleinknecht (1997), Barlet et alii.(1998) et Duguet (1999)).

Comment intégrer simultanément ces deux types d'innovations dans un modèle théorique permettant d'étudier l'effet de la concurrence sur l'innovation et la croissance? L'objectif de cet article est d'apporter une réponse à cette question. Pour expliciter la manière de procéder, revenons à la nature même

du processus d'innovation. Celui-ci est en général entaché de multiples incertitudes, au sein desquelles on voudrait mettre l'accent sur deux d'entre elles. La première, la plus abondamment utilisée dans la littérature se réfère au caractère aléatoire de la date de découverte. Les courses à l'innovation où la date de découverte est modélisée comme une variable aléatoire suivant une loi spécifiée (le plus souvent une loi de Poisson dont le paramètre représente le taux de hasard) en sont une parfaite illustration. La deuxième incertitude, plus rarement prise en compte dans la littérature, se réfère au résultat de l'effort de recherche. Lorsque cet effort est réalisé par une entreprise qui dispose déjà de la technologie la plus avancée, la concrétisation de la recherche par une découverte correspond clairement à un *dépassement* de la technologie précédente. Mais lorsque cet effort de recherche est réalisé par une entreprise retardataire qui n'a pas encore accès à la technologie la plus avancée, une incertitude sur le résultat de son effort demeure. L'entreprise n'est jamais sûre a priori, si la découverte à laquelle elle espère parvenir, sera une simple *innovation de rattrapage* ou une *innovation de dépassement*. Dans ces conditions, on conçoit que si la probabilité d'atteindre une *innovation de dépassement* est strictement positive, la pression concurrentielle sur le marché va inciter l'entreprise ayant un retard technologique à accroître son effort de recherche, dans l'espoir d'échapper précisément à la pression concurrentielle en dépassant son concurrent. L'effet négatif de la concurrence sur l'effort de recherche d'une entreprise retardataire dans le modèle de Aghion, Harris et Vickers (1997) risque ainsi de s'atténuer, voire même de se renverser, renforçant ainsi l'effet positif de la concurrence sur l'effort de recherche des entreprises en concurrence technologique frontale.

Le cas limite où la probabilité de dépassement est égale à 1 correspond à l'hypothèse Schumpeterienne où toute innovation dépasse nécessairement la précédente, à la nuance près toutefois qu'il n'est plus nécessaire de supposer que l'obtention d'une innovation par une firme élimine une firme disposant d'une technologie moins performante. Il y aurait ainsi création sans destruction. La dynamique industrielle induite par ce processus d'innovation conduit à des situations où dans chaque industrie, il y a une lutte permanente pour l'obtention du leadership technologique, sans que la position de leader ne soit assurée une fois pour toutes. Bien entendu, l'autre cas limite où la probabilité de dépassement est nulle correspond à l'hypothèse d'ajustement graduel de Aghion, Harris et Vickers (1997) où une innovation par une entreprise retardataire consiste à rattraper d'abord l'entreprise en avance, avant de pouvoir la dépasser ultérieurement. Les situations intermédiaires entre ces deux cas

limites, où la probabilité de dépassement est comprise entre 0 et 1 permettent ainsi d'étudier l'effet de la concurrence sur l'innovation et la croissance économique en fonction de la probabilité de dépassement.

Le modèle de croissance endogène que nous étudions ici est une généralisation à la fois du modèle de croissance fondé sur l'hypothèse de création destructrice où dans chaque secteur, la dernière entreprise innovante est en situation de monopole jusqu'à l'apparition de la prochaine innovation, et du modèle de croissance fondé sur l'hypothèse d'ajustement graduel où le rattrapage précède nécessairement le dépassement. Il est intéressant ici de revenir sur l'hypothèse de création destructrice, selon laquelle un innovateur en monopole chasse l'autre. Qu'est-ce qui permet de fonder une telle hypothèse ? Le raisonnement habituellement invoqué est que si la concurrence sur le marché du produit est une concurrence à la Bertrand, la firme disposant de la technologie à plus faible coût peut éliminer du marché toute firme utilisant une technologie à coût plus élevé. Il en serait de même si on supposait que la concurrence sur le marché du produit est une concurrence à la Cournot et que l'innovation est drastique. Mais, dans les deux cas, cette argumentation n'est justifiée que dans un cadre statique et non pas dans un cadre dynamique de concurrence intertemporelle. Si la firme disposant de la technologie la plus défavorable a conscience qu'elle peut surmonter son handicap temporaire en consacrant un effort de recherche approprié pour acquérir à son tour ultérieurement une position de leadership technologique, elle ne se laissera pas éliminer du marché, dût-elle supporter des profits nuls pendant la période où elle reste désavantagée sur le plan technologique. Ceci correspond d'ailleurs assez bien à ce qu'on observe dans la réalité. La longue supériorité technologique de Microsoft sur Apple n'a pas empêché cette dernière firme de supporter une traversée du désert avant de réémerger ensuite sur le marché grâce à l'introduction d'innovations appropriées. Le modèle que nous proposons ici et qui généralise le modèle de Aghion, Harris et Vickers (1997) permet de prendre en compte de telles évolutions.

L'article est présenté ci dessous.

Catching-up or leapfrogging? The effects of competition on innovation and growth

David Encaoua* and David Ulph†

January 2000

1 Introduction

2 The basic set-up

2.1 The consumption sector

There exists a continuum of final goods in the economy. Each good is produced in a specific industry indexed by $i \in [0, 1]$. Consumption of output from industry i at time t is denoted by $c_i(t)$. Time is continuous. We suppose that the representative consumer is infinitely lived and has a separable intertemporal utility function given by :

$$U = \int_0^{\infty} \ln C(t) e^{-\rho t} dt \quad (1)$$

where :

$C(t)$ represents an index of overall consumption at date t and is defined as a Cobb-Douglas instantaneous utility function on the continuum of final goods :

$$\ln C(t) = \int_0^1 \ln c_i(t) di \quad (2)$$

*EUREQua Université de Paris I.

†University College London.

$\rho \geq 0$ is the rate of time preference.

We suppose that financial markets are perfect and characterized by an instantaneous interest rate r_t . Since the intertemporal rate of substitution is constant and equal to unity in (1), the utility maximization of the consumer under an intertemporal budget constraint leads to the standard Ramsey equation:

$$\frac{\dot{E}(t)}{E(t)} = \rho - r_t \quad (3)$$

where:

$E(t)$ is the instantaneous global expenditure at date t : $E(t) \equiv P(t)C(t)$,
 $P(t)$ is the general price index at date t defined as: $\ln P(t) = \int_0^1 \ln p_i(t) di$,
 $p_i(t)$ is the price in industry i at date t .

We choose the normalization rule:

$$E(t) = P(t)C(t) = 1 \quad \forall t \in [0, \infty[\quad (4)$$

According to (3) and (4), the interest rate r_t is thus equal to the discounting rate ρ :

$$r_t = \rho \quad \forall t \in [0, \infty[\quad (5)$$

We will denote r the constant rate of interest.

According to (2), each final good has the same weight in the instantaneous utility function. Thus the normalization rule leads to a uniform spending in the different industries:

$$p_i(t)c_i(t) = 1 \quad \forall t \in [0, \infty[, \forall i \in [0, 1] \quad (6)$$

2.2 The productive sector

We suppose that in each industry there are two firms which are involved both in production and R&D. In each industry, the rival firms can be at *different technology levels*. A firm at technology level k can produce φ^k units

of output per unit of labour employed, where the exogenous parameter φ ($\varphi > 1$) represents the labour productivity when $k = 1$. Thus, the cost per unit of output of a firm in technology k is $w\varphi^{-k}$ where w is the wage rate. As our model is written in a partial equilibrium framework, we treat w as an exogenous variable and we choose $w = 1$. The production $q_i(t)$ of industry i ($i \in [0, 1]$) at date t is equal to the consumption of the corresponding good: $q_i(t) = c_i(t)$. The level of production $q_i(t)$ depends on the technology levels of both firms in industry i and on the type of rivalry between the two firms. Global production is equal to global consumption. Let us describe how the technology levels evolve in time.

At each date t , the current state of an industry can be described by the pair of technology levels $(k, k - n)$, where k is the level of the *technological leader* and $k - n$ is the technological level of the *follower*. Thus, n is the *technological gap* between the two firms. This technological gap will be treated as a state variable in what follows. According to (6), a constant proportion of income is spent on the product supplied by each industry. Thus, whatever the nature of product market competition is, firms' equilibrium profit flows derived from competition in the product market depend only upon the *gap* n and not upon the *level* k . We denote by π_n the equilibrium profit flow of a firm which is, from a technological aspect, n steps ahead of its rival (or $-n$ steps behind it if n is negative).

As in A.H.V.(1997), we suppose that the technological gap between firms cannot exceed one step. This assumption, made only in order to simplify the model and to allow analytical solutions, can be justified in three ways. First, it may be too costly (in terms of R&D effort) to a firm to get more than one step ahead of its rival. Second, imitation by the follower might become easy once the leader is more than one step ahead. Third, the length of a patent is in general finite. If the model had been written in discrete time, the duration of the period between two successive dates would correspond to the time length of the patent protection. The consequence of this assumption is that, at any time, each of the two firms in an industry can be in only one of the following three states $n \in \{-1, 0, 1\}$. A firm which is in the state $n = -1$ at some date is the *technological follower* while its rival in the same industry, which is in the state $n = 1$, is the *technological leader*. The corresponding unleveled industry is characterized as being of the *follower-leader* type. When a firm is in the state $n = 0$, both firms are at the same technological level and the corresponding leveled industry is of the *head-to-head* type. An industry can be, at any time, in just one of these two types. But, as time elapses,

the type of an industry changes permanently. We suppose indeed that there exist *process innovations* which lead to an increase of labour productivity by a parameter φ ($\varphi > 1$) (equivalent to a reduction of unit cost).

We are now in position to describe how an industry evolves in time. Consider a period $[t, t + dt]$. Two cases must be considered according to the type of the industry at the beginning of the period.

1. If an industry starts at date t in the *head-to-head* state, neither firm has a gap over the other, and both are at the existing cutting edge of technology. Three situations can occur during the period $[t, t + dt]$. First if both firms innovate during the period, neither will create a gap over the other and the industry will end the period as it began it in the *head-to-head state*. Second, the same is true if neither innovates. Third, if only one firm innovates, it will open a unit gap over its rival and the industry ends the period in the *leader-follower* type.

2. If an industry starts at date t in the *leader-follower* type, then one firm (the leader) has a unit gap over its rival (the follower) and is at the cutting edge of technology. Because the leader is at the cutting edge, if it succeeds in innovating, it can only lower its cost by the specified amount. For the follower, the situation is slightly different. We assume that with probability θ , $0 \leq \theta \leq 1$, the follower, *if it succeeds in innovating*, will be able to acquire an understanding of the technology at the cutting edge, and so will be able to achieve exactly the same technology as the leader would acquire if it innovated. Thus, if its innovation succeeds, the follower can *leapfrog* the preceding leader with probability θ . However, with probability $1 - \theta$, the follower will not acquire the understanding of the technology at the cutting edge, and so, *if it succeeds in innovating*, it will acquire only the technology currently used by the leader. In this case, there is only a *catching-up* by the follower of the current leader's technology. The model captures thus the two polar cases corresponding respectively to *leapfrogging* ($\theta = 1$) and to *step-by-step* innovation ($\theta = 0$). It captures also all the intermediate cases where *leapfrogging* occurs with probability θ and *catching-up* occurs with probability $1 - \theta$. Moreover, note that contrary to the Schumpeterian leapfrogging models where monopoly is typically a universal market structure, our assumptions allow the possibility of a richer market structure. When its innovation succeeds, a technological laggard can advance either by leapfrogging the previous leader or by catching-up it. The possibilities for industry evolution, when starting from a *leader-follower* type, are thus more complex.

Suppose the follower does not succeed in innovating. If the leader innovates, it will open up a gap of 2, but according to our assumption of a maximal gap of 1, the follower gets access to the previous leader's technology and the industry ends the period in the *follower-leader* position. If the leader does not succeed in innovating, the industry ends the period as it began it, namely in a *follower-leader* position.

Suppose the follower succeeds in innovating from the existing cutting edge. If the leader also innovates the gap between them will be reduced to zero during the period and the industry will end the period in the *head-to-head* position. However, if the leader fails to innovate, then the previous follower will have become the new leader and the industry will end the period in the *leader-follower* type (but with the identity of firms reversed).

Suppose the follower succeeds in innovating, but not from the cutting edge. Then if the leader also innovates, it will maintain its gap of 1, and the industry ends the period as it began it in the *leader-follower* type. However, if the leader fails to innovate, then its technological lead will be eliminated, and the industry will end the period in the *head-to-head* position.

Innovative advance and hence economic growth occur at a rate determined by R&D efforts. We suppose that, by employing $\gamma(p_0)$ units of labour in R&D, a firm at the technological frontier which is level with its rival, moves one step ahead with Poisson hazard rate p_0 . Similarly, by employing $\gamma(p_{-1})$ units of labour in R&D, a follower succeeds in innovating with Poisson hazard rate p_{-1} . Conditional to its success, it leapfrogs the leader with Poisson hazard rate $p_{-1}\theta$ and it catches-up with Poisson hazard rate $p_{-1}(1 - \theta)$. Finally, by employing $\gamma(p_1)$ units of labour in R&D, a leader succeeds in innovating with Poisson hazard rate p_1 . The R&D cost function $\gamma(p)$ is supposed to be increasing, continuous and convex. We also assume that $\gamma(0) = 0$.

We can now determine the system of equations whose solutions are the R&D efforts p_{-1} , p_0 , p_1 .

3 The stationary state equilibrium

3.1 The value functions of the markovian game at the steady state

We focus on the determination of equilibrium in Markov strategies at the stationary state of the economy. A Markov strategy specifies a firm's choice of its R&D effort as a function of the current gap¹ in the corresponding industry, leading thus to p_{-1} , p_0 and p_1 . We search for symmetric equilibrium values of p_{-1} , p_0 and p_1 in terms of the parameters π_{-1} , π_0 , π_1 , r , θ , for a given R&D cost function $\gamma(p)$.

Consider the value functions for a firm whose rival's strategy is given by $(\bar{p}_{-1}, \bar{p}_0, \bar{p}_1)$. Let V_n denote the value function of a firm starting from state n . This value gives the firm's expected discounted payoff in the game starting from the state in which it is n steps ahead (or behind if $n < 0$) of its rival ($n = -1, 0, 1$).

In order to obtain the Bellman equations to which obey the value functions, consider for instance the situation of a technological follower which is in the state $n = -1$ at the beginning date of a period $[t, t + dt]$. By spending $\gamma(p_{-1})$ in R&D, it obtains an innovation during this period with Poisson hazard rate $p_{-1}dt$. As it has been explained, we suppose that this innovation may lead the follower either to the cutting edge of the technology, leapfrogging thus the leader with probability $p_{-1}\theta dt$ or to the technology currently used by the leader, catching-up it with the probability $p_{-1}(1-\theta)dt$. With the complementary probability $1 - p_{-1}dt$, the follower does not obtain any innovation during the period $[t, t + dt]$.

During the same period, the leader's innovation occurs with the probability $\bar{p}_1 dt$. Figure 1 describes the corresponding game during the period $[t, t + dt]$. It also gives the discounted payoffs (net of R&D costs) of the follower at the end of this period.

¹Note that a firm which is ahead ($n = 1$) has no incentive to undertake R&D due to our assumption of a maximal gap of 1. Thus, we will find that at equilibrium, $p_1 = 0$.

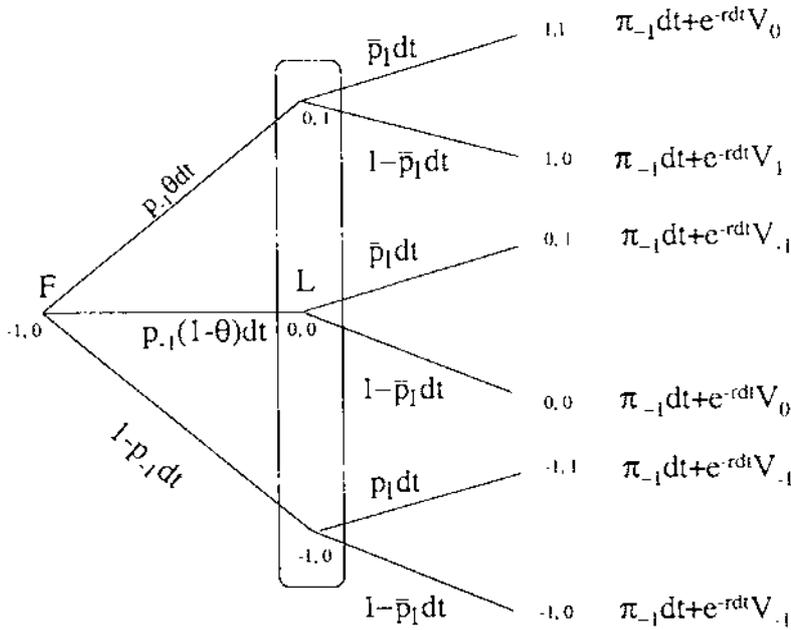


Figure 1: Determination of the follower's value function

The value function V_{-1} of a follower satisfies thus the following Bellman equation :

$$V_{-1} = \underset{p_{-1} \geq 0}{Max} \{ (\pi_{-1} - \gamma(p_{-1}))dt + e^{-rdt} [V_1 p_{-1} \theta dt + V_0 p_{-1} (1 - \theta) dt + V_{-1} \bar{p}_1 dt + V_{-1} - V_{-1} (p_{-1} + \bar{p}_1) dt] \} \quad (7)$$

By using the first order approximation $e^{-\rho dt} \simeq 1 - \rho dt$, and by keeping only the first order terms in dt , one obtains the equivalent Bellman equation:

$$rV_{-1} = \underset{p_{-1} \geq 0}{Max} \{ \pi_{-1} - \gamma(p_{-1}) + p_{-1} \theta V_1 + p_{-1} (1 - \theta) V_0 - p_{-1} V_{-1} \} \quad (8)$$

The corresponding first order condition leads to (for an interior solution):

$$\gamma'(p_{-1}) = \theta(V_1 - V_0) + (V_0 - V_{-1}) \quad (9)$$

The interpretation of condition (9) is straightforward. The LHS is the R&D marginal cost of the follower. The RHS gives the expected marginal revenue which is equal to $\theta(V_1 - V_{-1}) + (1 - \theta)(V_0 - V_{-1})$.

In the same way, one obtains the Bellman equations giving the the values of a firm starting the period $[t, t + dt]$ at the respective states $n=1$ and $n=0$, and the corresponding first order conditions:

$$rV_1 = \underset{p_1 \geq 0}{Max} \{ \pi_1 - \gamma(p_1) - p_{-1}(V_1 - V_0) - p_{-1}\theta(V_0 - V_{-1}) \} \quad (10)$$

$$\gamma'(p_1) \leq 0 \text{ and } p_1\gamma'(p_1) = 0 \quad (11)$$

$$rV_0 = \underset{p_0 > 0}{Max} \{ \pi_0 - \gamma(p_0) - p_0(V_0 - V_1) - \bar{p}_0(V_0 - V_{-1}) \} \quad (12)$$

$$\gamma'(p_0) = V_1 - V_0 \quad (13)$$

The symmetric-equilibrium conditions $\bar{p}_{-1} = p_{-1}$, $\bar{p}_0 = p_0$, $\bar{p}_1 = p_1$ are added to the preceding system of six equations ((8)-(13)). The unknowns of this system are the three value functions V_{-1}, V_0, V_1 and the three intensity R&D efforts p_{-1}, p_0, p_1 . They depend of the following parameters:

r = the interest rate,

θ = the probability of leapfrogging by the follower,

π_{-1} = the current profit of a technological follower,

π_0 = the current profit of a neck by neck technological firm,

π_1 = the current profit of a technological leader,

$\gamma(p)$ = the R&D cost (in units of labor) of a firm moving one technological step ahead with Poisson hazard rate p .

Note that, according to the first order condition (11), if $\gamma'(p_1) \neq 0$, then $p_1 = 0$. This results from the assumption of a maximal gap of 1.

3.2 The determination of industry structures and rate of growth

The steady-state distribution of industry structures is endogenous. Let denote α the proportion of industries that are of the *head-to-head* type in the steady state. In order to determine the value of α , consider a period $[t, t + dt]$. During this time interval, two types of evolution do occur.

In industries which are of the *head-to-head* type (proportion α), succesful innovation made by only one of the two neck by neck firms leads to an evolution towards industries of the *leader-follower* type. This occurs with the probability $2p_0dt(1 - p_0dt) \simeq 2p_0dt$.

In industries which are of the *leader-follower* type (proportion $1 - \alpha$), succesful innovations made either by both the leader and the follower who reaches the cutting edge of technology or by only the follower who just catch-up the leader's technology leads to an evolution towards industries of the *head-to-head* type. This occurs with the probability $(p_{-1}\theta dt p_1 dt) + (p_{-1}(1 - \theta)dt(1 - p_1 dt)) \simeq p_{-1}(1 - \theta)dt$.

Since the distribution of industries remains stationary over time in the steady state, we must have:

$$2p_0\alpha dt = p_{-1}(1 - \theta)(1 - \alpha)dt. \quad (14)$$

This last equation leads to the value of the proportion of industries that are of the *head-to-head* type in the steady state. Note that his value depends directly and indirectly on the parameter θ since the values of p_0 and p_{-1} depend themselves on θ :

$$\alpha(\theta) = \frac{(1 - \theta)p_{-1}(\theta)}{2p_0(\theta) + (1 - \theta)p_{-1}(\theta)} \quad (15)$$

We thus have the following lemma:

Lemma 1 *In the steady state of the economy, the proportion of industries that are head-to-head is given by*

$\alpha(\theta) = \frac{(1-\theta)p_{-1}(\theta)}{2p_0(\theta) + (1-\theta)p_{-1}(\theta)}$, where θ is the probability to leapfrog the leader, conditional to the success of innovation by the follower, $p_0(\theta)$ is the R&D effort by a neck-by neck firm and $p_{-1}(\theta)$ is the R&D effort by a follower. For a given θ in $(0, 1)$ the proportion of industries that are head to head in the steady state is an increasing function of $p_{-1}(\theta)$ and a decreasing function of $p_0(\theta)$.

Note that for $\theta = 0$, we obtain the same result as in A.H.V. (1997). For $\theta = 1$, we have $\alpha(1) = 0$, since, as we will check later $p_0(1) \neq 0$. This means, that in a pure leapfrogging situation where a successful innovation by the follower gives it a leadership position, there are no industries that are of the *head-to-head* type.

Let us now determine the rate of growth of the economy in the steady state. Consider again a period $[t, t + dt]$. The rate of growth g is defined by

$$g = \frac{d}{dt} \ln Q(t) = \frac{d}{dt} \ln C(t). \quad (16)$$

The rate of growth of an industry in a complete cycle is given by $\ln \varphi$, which is the amount by which is increased the log of its output. A complete cycle occurs in two ways.

Firstly, by the evolution from a *head-to head* type towards the next *head-to head* type, where each firm has access to the new cutting edge of the technology. The complete cycle is decomposed in this case into a two-stage cycle. The first stage corresponds to an evolution from a *head-to head* type to a *leader follower* type. Let denote $\ln \varphi_1$ the rate of growth of an industry in this first stage. The probability that an industry of the *head-to head* type (proportion α) moves to an industry of the *leader-follower* type during the period $[t, t + dt]$ (first stage of the complete cycle) is given by $2p_0dt(1 - p_0dt) \simeq 2p_0dt$. The second stage corresponds to an evolution from a *leader follower* type to a *head-to head* type. Let denote $\ln \varphi_2$ the rate of growth of an industry in this second stage. The probability that an industry of the *leader-follower* type (proportion $1 - \alpha$) moves to an industry of the *head-to head* type during the same period (second stage of the complete cycle) is given by $p_{-1}(1 - \theta)dt(1 - p_1dt) \simeq p_{-1}(1 - \theta)dt$. Of course, $\ln \varphi = \ln \varphi_1 + \ln \varphi_2$, since the industry output increases by a factor of φ after a complete cycle.

Secondly, a complete cycle occurs also by the evolution from a *leader-follower* type to the next *leader-follower* type, where the follower succeeds its innovation by leapfrogging the leader, obtaining thus access to the new cutting edge of the technology, and the previous leader does not succeed in innovating. This evolution, which reverses the identity roles of the leader and follower, gives rise to a rate of growth of φ . since one of the two firms (the follower) has increased its technological level by 2 steps. The probability that an industry of the *leader-follower* type (proportion $1 - \alpha$) moves to an industry of the next *follower-leader* type during the same period (complete

cycle) is given by $p_{-1}\theta dt(1 - p_1 dt) \simeq p_{-1}\theta dt$.

The average growth rate of final output is thus given by:

$$g dt = 2\alpha p_0 dt \ln \varphi_1 + p_{-1}(1 - \theta)(1 - \alpha) dt \ln \varphi_2 + p_{-1}\theta(1 - \alpha) dt \ln \varphi \quad (17)$$

Note that g depends again directly and indirectly on the parameter θ . By using the relationship $\ln \varphi = \ln \varphi_1 + \ln \varphi_2$, we obtain finally the following expression of the rate of growth of the economy:

$$g(\theta) = (2\alpha(\theta)p_0(\theta) + \theta p_{-1}(\theta)(1 - \alpha(\theta))) \ln \varphi \quad (18)$$

By substituting the value of $\alpha(\theta)$ given in the previous lemma, one obtains:

$$g(\theta) = \left[\frac{2p_0(\theta)p_{-1}(\theta)}{2p_0(\theta) + (1 - \theta)p_{-1}(\theta)} \right] \ln \varphi \quad (19)$$

We have thus proved the following lemma:

Lemma 2 *For any $\theta \in [0, 1]$ the growth rate at the steady state of the economy $g(\theta)$ is given by (19). It is an increasing function of both the R&D effort $p_{-1}(\theta)$ of a follower in an unleveled industry and the R&D effort $p_0(\theta)$ of a neck-by-neck firm in a leveled industry.*

Note that for $\theta = 0$, which corresponds to the *step by step technological progress* situation where a follower must catch-up the leader before becoming a leader itself, we obtain the same expression for the rate of growth as in A.H.V.(1997) : $g(0) = 2\alpha(0)p_0(0) \ln \varphi = \frac{2p_{-1}(0)}{2p_0(0) + p_{-1}(0)} \ln \varphi$. So, $g(0)$ is an increasing function of $p_{-1}(0)$ and a decreasing function of $p_0(0)$. We will see later that for some specifications of the R&D cost function, the effects of product market competition on $\alpha(0)$ and $p(0)$ are opposite, so that the total effect of competition on the rate of growth $g(0)$ has to be analyzed. For $\theta = 1$, which corresponds to the *pure leapfrogging* situation where a successful innovating follower obtains a technological leadership, we have $\alpha(1) = 0$ and thus $g(1) = p_{-1}(1) \ln \varphi$. In this case, the rate of growth is directly proportional to the follower effort in R&D and this will allow us to examine the effect of competition on the rate of growth in the *pure leapfrogging* situation where an industry is no more monopolized by the last innovator. However,

for a given θ in $(0, 1)$, the variation of g according to p_0 and p_{-1} is a-priori ambiguous since α is a decreasing function of p_0 and an increasing function of p_{-1} .

We have now to compute the solution of the non linear system (8 – 13). Note first that a solution of this system exists if the R&D cost function is continuous and convex. To get a sketch of the proof of this existence, consider the vector $X = (p_{-1}, p_0, p_1, V_{-1}, V_0, V_1)$ and write the system (8 – 13) as $F(X) = 0$, where F is a continuous and convex function from R^6 to R^6 . Choose a convex compact set $B \subset R^6$ sufficiently large to insure that F is defined and have values in B . Now, consider the function $G(X) = F(X) + X$. By the fixed point theorem, there exists a value of X such that $G(X) = X$. Such a value of X is a solution of the system (8 – 13).

In what follows, we try to solve this system for some specifications of the R&D cost function $\gamma(p)$, having in mind the two following questions:

1. How the rate of growth of the economy is affected by the parameter θ which defines the degree of leapfrogging, given that $\theta = 0$ characterizes the step-by step inovative process without any leapfrogging (the situation examined in A.H.V.(1997)) while $\theta = 1$ characterizes a pure leapfrogging model with however an important difference with the usual schumpeterian growth models, namely the fact that there always exist two firms in each industry, competing permanently for the role of technological leader?

2. How the innovative process and the growth of the economy are affected by the degree of rivalry between the two firms, given that rivalry occurs both on the product market and on the innovation side in this model?

4 The solution for two specifications of the R&D cost function

4.1 A linear R&D cost function

Consider the following linear specification of the R&D cost function: $\gamma(p) = dp$, where d is a positive parameter. Let us introduce the parameters $a = \pi_1 - \pi_0$ and $b = \pi_0 - \pi_{-1}$ which play an important role in what follows. These parameters measure the profit flow increments associated, respectively, with gaining the lead and catching up. These parameters depend on φ and on the nature of product market competition (for instance Cournot or Bertrand

competition). Any model of imperfect competition leads to the following relationships: $a > b \geq 0$. Moreover, as functions of the degree of rivalry in the product market, a is increasing, b is decreasing and $a + b$ is increasing.

Suppose that the exogenous parameters b, r , and d satisfy the following condition: $b \geq rd$. The resolution of the system (8 – 13), with the linear R&D cost function $\gamma(p) = dp$, leads to the following results:

$$V_{-1} = \frac{\pi_{-1}}{r}, \quad V_0(\theta) = \frac{\pi_{-1}}{r} + d(1 - \theta), \quad V_1(\theta) = \frac{\pi_{-1}}{r} + d(2 - \theta) \quad (20)$$

$$p_0(\theta) = \frac{b - rd(1 - \theta)}{d(1 - \theta)} \quad (21)$$

$$p_{-1}(\theta) = \frac{a + b - rd(2 - \theta)}{d(1 + \theta(1 - \theta))} \quad (22)$$

By substituting these values in (15), one obtains the following expression of the proportion of industries that are head-to-head in the steady state:

$$\alpha(\theta) = \frac{(1 - \theta)^2 [a + b - rd(2 - \theta)]}{a(1 - \theta)^2 + b(3 - \theta^2) - rd(1 - \theta)(4 - \theta - \theta^2)} \quad (23)$$

Finally, by substituting this value of $\alpha(\theta)$ in the rate of growth given in (19), and replacing $p_0(\theta)$ and $p_{-1}(\theta)$ by their values given by (21) and (22), we obtain the following expression of the steady state economy's rate of growth, which depends only on the exogenous parameters:

$$g(\theta) = \left[\frac{2(a + b - rd(2 - \theta))(b - rd(1 - \theta))}{d[a(1 - \theta)^2 + b(3 - \theta^2) - rd(1 - \theta)(4 - \theta - \theta^2)]} \right] \ln \varphi \quad (24)$$

Recall the two main questions addressed before:

1. How the main variables of the model and mainly the rate of growth vary with the parameter θ ?
2. What is the impact of the degree of rivalry in the product market on R&D efforts and the rate of growth?

Suppose that the rate of interest is very small ($r \simeq 0$). In this case, the following proposition gives an answer to the first question:

Proposition 3 Suppose that the R&D cost function is linear $\gamma(p) = dp$. For very small values of the interest rate, the R&D effort $p_0(\theta)$ by a neck-by-neck firm increases with θ , while the R&D effort $p_{-1}(\theta)$ by a follower in an industry of the leader follower type has a U-form, reaching its minimum around $\theta = 1/2$. The proportion $\alpha(\theta)$ of industries in a head-to-head position decreases with θ , becoming equal to 0 when $\theta = 1$. The rate of growth $g(\theta)$ increases with θ , except for the parameter value $b = 0$, in which case, the rate of growth $g(\theta)$ is equal to 0 for $\theta \in [0, 1[$.

Proof. For $r = 0$, we have $g(\theta) = \left[\frac{2(a+b)b}{d[a(1-\theta)^2 + b(3-\theta^2)]} \right] \ln \varphi$, $\alpha(\theta) = \frac{(1-\theta)^2(a+b)}{a(1-\theta)^2 + b(3-\theta^2)}$, $p_0(\theta) = \frac{b}{d(1-\theta)}$, and $p_{-1}(\theta) = \frac{a+b}{d(1+\theta(1-\theta))}$. These functions are displayed in figures 2 for specific values² of the parameters a, b, d and φ . The curve $p_0(\theta)$ cuts from below the curve $p_{-1}(\theta)$ in the interval $[0, 1]$. The value of $p_0(\theta)$ goes to infinity, but $p_0(\theta)\alpha(\theta)$ converges to 0, as θ goes to the limit 1. The rate of growth increases from $g(0) = \left[\frac{2(a+b)b}{d(a+3b)} \right] \ln \varphi$ to $g(1) = \left[\frac{a+b}{d} \right] \ln \varphi$. However, if $b = 0$, which would happen with Bertrand competition on the product market, the rate of growth of the economy is nil except for $\theta = 1$ (pure leapfrogging case). In this case ($b = 0$), the rate of growth jumps from 0 for any $\theta \in [0, 1[$ to the value $\left[\frac{a}{d} \right] \ln \varphi$ for $\theta = 1$. The threshold of θ above which $p_0(\theta) \geq p_{-1}(\theta)$ is given by $\frac{a+2b-(a^2+4b^2)^{1/2}}{2b}$. One can easily check that the value of this threshold belongs to $]0, 1[$ since $a = \pi_1 - \pi_0 > b = \pi_0 - \pi_{-1}$.

■
²Figures 2 represent respectively the rate of growth $g(\theta)$ (fig. 2a), the proportion of industries head to head $\alpha(\theta)$ (fig. 2b), the R&D effort by a neck-by-neck firm $p_0(\theta)$ (fig. 2c), and the R&D effort by a follower $p_{-1}(\theta)$ (fig. 2d). All these figures have been drawn with the following values of the parameters

$$a = 3, b = 0.1, d = 2, \varphi = e.$$

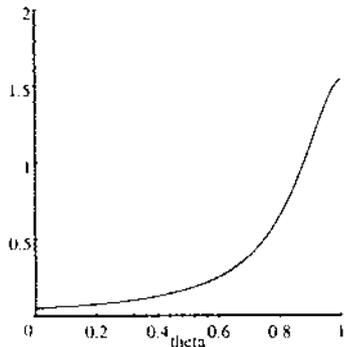


Figure 2a: Rate of growth

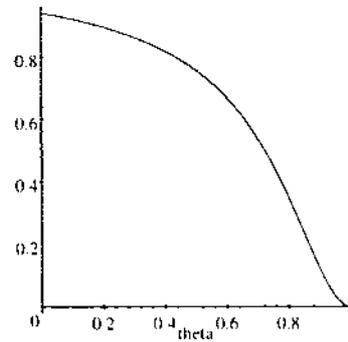


Figure 2b: Proportion of head to head industries

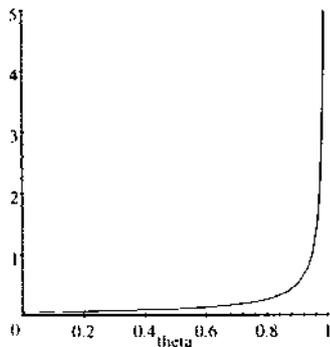


Figure 2c: Neck to neck R&D

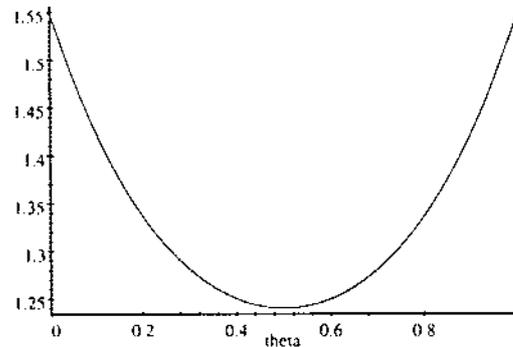


Figure 2d: Follower's R&D

Globally, this proposition asserts that increasing the probability of leapfrogging by the laggard is quite valuable for the innovation process and for the rate of growth of the economy. It also asserts that the R&D effort by neck-by-neck firms becomes higher than the R&D effort provided by a follower, as θ is above some threshold.

Turn now to the second question related to the effects of product market competition. The main findings are summarized in the proposition 2.

Proposition 4 *Suppose that the R&D cost function is linear $\gamma(p) = dp$. For very small values of the interest rate, and for every value of θ in $[0, 1]$, a higher degree of rivalry in the product market lowers the R&D effort $p_0(\theta)$ by a neck-by-neck firm and increases the R&D effort $p_{-1}(\theta)$ by a follower. It also increases the proportion $\alpha(\theta)$ of industries which are head-to-head. Finally, there exists some threshold value $\bar{\theta}$ belonging to $[0, 1[$ such that for $\theta \geq \bar{\theta}$, a*

higher competition in the product market increases the rate of growth of the economy.

Proof. For $r=0$, we have shown that $p_0(\theta) = \frac{b}{d(1-\theta)}$ and we know that in any model of imperfect competition an increase in competition in the product market lowers the parameter $b = \pi_0 - \pi_{-1}$ which measures the profit flow increment associated with gaining the lead. We have also shown that $p_{-1}(\theta) = \frac{a+b}{d(1+\theta(1-\theta))}$ and we know that an increase of competition increases the parameter $a+b = (\pi_1 - \pi_0) + (\pi_0 - \pi_{-1}) = \pi_1 - \pi_{-1}$ which measures the profit gap between the leader and the follower. The proportion of industries which are head to head in the steady state is $\alpha(\theta) = \frac{(1-\theta)^2(a+b)}{a(1-\theta)^2 + b(3-\theta^2)}$ and one can see that $\alpha(0) = \frac{1}{1 + \frac{2b}{a+b}}$ which increases with the degree of competition. Since $\alpha(\theta)$ is a decreasing function of θ , this completes the proof that for each value of θ in $[0, 1[$, more intense is the degree of rivalry in the product market, higher is $\alpha(\theta)$. Finally, concerning the effect of competition on the rate of growth, we observe that in the situation of pure leapfrogging ($\theta = 1$), $g(1) = \left[\frac{a+b}{d}\right] \ln \varphi$ increases with the degree of competition (since $a+b$ is an increasing function of the degree of rivalry). In the situation of pure catching-up ($\theta = 0$), the rate of growth $g(0)$ is proportional to $\frac{a+b}{a/b+3}$. Both the numerator and denominator of this last fraction are increasing functions of the degree of competition. So the fraction itself increases or decreases with competition according to the respective variations of $a+b$ and a/b with the degree of competition. The effect of competition on $g(0)$ is thus indeterminate. However, since $g(\theta)$ is a strictly increasing function of $\theta \in [0, 1]$, except for $b = 0$, the last statement of the proposition is proved when $b \neq 0$. If $b = 0$, the rate of growth is equal to 0. ■

This last proposition is at the heart of our work. It asserts that competition in the product market is good for growth, provided that the possibility given to a laggard to leapfrog it's rival with a succesful innovation is high enough. Thus, in a pure leapfrogging model ($\theta = 1$) where competition between a technological leader and a follower still exist, it appears that competition has positive effects, in a dynamic perspective. This result is to be contrasted with the usual one derived from most of the endogenous growth models based on the Schumpeterian creative destruction hypothesis (e.g. Aghion-Howitt (1982)).

How robust are our results relatively to the chosen linear specification of

the R&D cost? We consider the case of a quadratic R&D cost function in what follows.

4.2 A quadratic R&D cost function

Suppose now that the R&D cost function is given by: $\varphi(p) = \frac{1}{2}dp^2$. In this case, after straightforward substitutions, the system (8–13) leads to a system of two equations whose the variables $p_0(\theta)$ and $p_{-1}(\theta)$ are the solutions:

$$2rp_0 + (1 + 2\theta)(p_0)^2 + 2\theta(p_{-1})^2 - 2\theta^2 p_0 p_{-1} - 2a/d = 0 \quad (25)$$

$$2r(p_{-1} - \theta p_0) - (1 + 2\theta)(p_0)^2 + (p_{-1})^2 + 2p_0 p_{-1} - 2b/d = 0 \quad (26)$$

If, as previously, we consider the case $r = 0$, it is possible to transform this system in order to express p_0 as a function of p_{-1} and p_{-1} as a function of p_0 :

$$p_0 = \frac{2(a + b)/d - (p_{-1})^2(1 + 2\theta)}{2p_{-1}(1 - \theta^2)} \quad (27)$$

$$p_{-1} = \frac{(1 + 2\theta)^2(p_0)^2 - 2(a - 2\theta b)/d}{2\theta(\theta + 2)p_0} \quad (28)$$

By substituting the value of p_{-1} in p_0 , one obtains a fourth order polynomial whose p_0 is the root. The corresponding equation has the form :

$$A(p_0)^4 + B(p_0)^2 + C = 0 \quad (29)$$

where:

$$A = d^2(-1 - 16\theta - 44\theta^2 - 32\theta^3 + 4\theta^4 + 8\theta^5) \quad (30)$$

$$B = d(4a + 32a\theta - 8b\theta + 24a\theta^2 - 32b\theta^2 - 16b\theta^3 + 8b\theta^4) \quad (31)$$

$$C = -4a^2 + 16ab\theta - 16b^2\theta^2 \quad (32)$$

The fourth order equation (29) can be solved in p_0 . We keep only the positive root p_0 for which the value of p_{-1} , given by (28) is positive. Thus, the system (8 – 13) is completely solved in the case of a quadratic R&D cost function.

In order to examine the effects of θ on the different endogenous variables, we take the same specific values of the parameters a, b, d used to draw the curves in the figures 2 of the preceding sub-section. Table 1 gives the values of $p_0(\theta)$, $p_{-1}(\theta)$, $\alpha(\theta)$, and $g(\theta)$ for different values of θ in $[0, 1]$ with a grid of $1/10$.

θ	$p_0(\theta)$	$p_{-1}(\theta)$	$\alpha(\theta)$	$g(\theta)$
0	1.7321	0.7378	0.1756	0.6082
1/10	1.5558	0.7734	0.1828	0.6320
2/10	1.4225	0.8038	0.1844	0.6556
3/10	1.3188	0.8311	0.1807	0.6809
4/10	1.2365	0.8566	0.1721	0.7092
5/10	1.1703	0.8812	0.1584	0.7416
6/10	1.1168	0.9058	0.1396	0.7794
7/10	1.0735	0.9311	0.1151	0.8239
8/10	1.0389	0.9575	0.0844	0.8767
9/10	1.0118	0.9858	0.0464	0.9400
1	0.9915	1.0165	0	1.0165

Table 1: The values of $p_0(\theta)$, $p_{-1}(\theta)$, $\alpha(\theta)$, and $g(\theta)$ for the following values of the parameters:
 $a = 3, b = 0.1, d = 2, \varphi = e, r = 0$.

Other values of the parameters have been tested. They give rise to the same variations of the endogenous variables as obtained in table 1. The results are summarized in the following proposition:

Proposition 5 *Suppose that the R&D cost function is quadratic, $\gamma(p) = \frac{1}{2}dp^2$. For very small values of the interest rate, the R&D effort $p_0(\theta)$ by a neck-by-neck firm decreases with θ , while the R&D effort $p_{-1}(\theta)$ by a follower increases with θ . As θ goes to 1, $p_{-1}(\theta)$ becomes higher to $p_0(\theta)$. The proportion $\alpha(\theta)$ of industries that are head-to head in the steady state has an inverted U-form. It first increases with θ for small values of θ , then it decreases with θ becoming equal to 0 for $\theta = 1$. The rate of growth $g(\theta)$ increases with θ .*

Thus, whatever the R&D cost function is linear or quadratic, we reach the same conclusion. Increasing the probability θ of leapfrogging by the laggard is quite valuable for the rate of growth $g(\theta)$ of the economy. However, in the quadratic case, the R&D effort $p_0(\theta)$ of a neck-by-neck firm decreases with θ , contrary to what occurred in the linear case, and the R&D effort $p_{-1}(\theta)$ of a laggard firm increases with θ in the quadratic case. Note also that the proportion $\alpha(\theta)$ of industries that are head-to head in the steady state has now an inverted U-form, but, as previously $\alpha(\theta)$ is equal to 0 for $\theta = 1$.

What is the effect of competition on the intensity of innovation and on the rate of growth? From (25) and (26), one obtains directly for $\theta = 1$ the values of $p_{-1}(1) = \sqrt[2]{\frac{2(a+b)}{3d}}$ and $p_0(1) = \frac{1}{3} \left[\sqrt[2]{\frac{2(a+b)}{3d}} + \sqrt[2]{\frac{2(a-5b)}{3d}} \right]$. These two R&D efforts increase with the degree of competition on the product market, since $a = \pi_1 - \pi_0$ is an increasing function of competition, $b = \pi_0 - \pi_{-1}$ is a decreasing function of competition and $a + b = \pi_1 - \pi_{-1}$ increases with competition³. Thus the intensity of innovation increases with the degree of rivalry in the product market. Since the rate of growth for $\theta = 1$ is given by $g(1) = p_{-1}(1) \ln \varphi$, (see (19)), it appears that the economy's rate of growth in the pure leapfrogging situation ($\theta = 1$) is an increasing function of the intensity of competition in the product market. The same result was obtained with a linear R&D cost function, and this confirms its robustness. Note however that this result is exactly the reverse of what is derived from usual Schumpeterian growth models (see Aghion-Howitt (1992),

³With Cournot competition in the product market, one obtains $a^{Cournot} = \left(\frac{\varphi}{1+\varphi}\right)^2 \cdot \frac{1}{4}$ and $b^{Cournot} = \frac{1}{4} - \frac{1}{(1+\varphi)^2}$. With Bertrand competition, the corresponding values are $a^{Bertrand} = \frac{\varphi-1}{\varphi}$ and $b^{Bertrand} = 0$. One can verify immediately the following implications: $\varphi > 1 \Rightarrow a^{Cournot} < a^{Bertrand}$, $b^{Cournot} > b^{Bertrand}$, $(a+b)^{Cournot} < (a+b)^{Bertrand}$, $a^{Bertrand} > (a+b)^{Cournot} > a^{Cournot}$.

(1998)). The intuition of the present result is clear. By substituting the usual assumption made in the Schumpeterian growth models, according to which an industry is always monopolized by the last innovator, by a more realistic assumption, according to which both a laggard and a lead-up firm may coexist at any time, given that, at each date, the laggard may leapfrog the leader one opens the way for the analysis of the relationship between competition in the product market on one side and competition in R&D to gain the technological leadership on the other side. In the pure leapfrogging situation, more competition in the product market gives to the follower more incentives to invest in R&D, in order to leapfrog the leader.

What happens for other values of θ ? For $\theta = 0$, we obtain easily from (25) and (26), $p_0(0) = \sqrt[2]{\frac{2a}{d}}$ and $p_{-1}(0) = \sqrt[2]{\frac{2(2a+b)}{d}} - \sqrt[2]{\frac{2a}{d}}$. Thus, the rate of growth $g(0) = \frac{2p_0(0)p_{-1}(0)}{2p_0(0)+p_{-1}(0)} \ln \varphi$ (from (19)) is equal in this case to $g(0) = \frac{2\sqrt[2]{4a(2a+b)-4a}}{\sqrt[2]{d}(\sqrt[2]{2a+b}+\sqrt[2]{2a})} \ln \varphi$. A.H.V.(1997) compare in this case the growth rates obtained with Bertrand and Cournot competition in the product market. In order to do that, it is sufficient to replace a and b with their corresponding values (depending on φ) obtained in these two types of competition (see footnote 3). The finding is that $g^{Bertrand}(0) > g^{Cournot}(0)$ if $\varphi < 7.26$. Thus, unless φ is large, the growth rate for $\theta = 0$ is higher with Bertrand than Cournot, which means that a higher degree of rivalry leads to a higher growth rate. For other values of $\theta \in]0,1[$, the computation of $g(\theta)$ given by (19) is more complex, but by using the analytical solution of equation (29), one can show that for each value of θ , the growth rate is positively affected by the degree of competition if φ is below some threshold $\tilde{\varphi}(\theta)$ which is an increasing function of θ . Note that $\lim_{\theta \rightarrow 1} \tilde{\varphi}(\theta) = \infty$ when θ tends to 1. The following propositions resumes this finding:

Proposition 6 *Suppose that the R&D cost function is quadratic $\gamma(p) = \frac{1}{2}dp^2$. For very small values of the interest rate, and for every $\theta \in [0, 1]$, there exists a threshold function $\tilde{\varphi}(\theta)$ increasing in θ , such that if $\varphi < \tilde{\varphi}(\theta)$, a higher degree of competition in the product market increases the growth rate. For $\theta = 1$, (pure leapfrogging) higher competition always leads to a higher growth rate.*

Conclusion 7

REFERENCES

Aghion, P. and P. Howitt, 1992, A model of growth through creative destruction, *Econometrica* 60, 323-351

Aghion, P. and P. Howitt, 1998, *Endogenous Growth Theory*, Mass.: MIT Press

Aghion, P., C. Harris and J. Vickers, 1997, Competition and growth with step-by-step innovation : an example, *European Economic Review* 41, 771-782

Arrow, K.J., 1962, Economic welfare and the allocation of resources for inventions, in R.R. Nelson, ed. *The Rate and Direction of Technological Change*, Princeton, Princeton University Press

Barlet, C., E. Duguet, D. Encaoua and J. Pradel, 1998, The commercial success of innovations: an econometric analysis at the firm level in French manufacturing, *Annales d'Economie et de Statistique* 49/50, 457-478

Budd, C., C. Harris and J. Vickers, 1993, A model of the evolution of duopoly : does the asymmetry between firms tend to increase or decrease?, *Review of Economic Studies* 60, 543-573

Crépon, B., E. Duguet and I. Kabla, 1996, Schumpeterian conjectures: a moderate support from various innovation measures, in A. Kleinknecht ed. *Determinants of Innovations, The Message from New Indicators*, Macmillan Press

Dasgupta, P. and J. Stiglitz, 1980, Uncertainty, industrial structure and the speed of R&D, *Bell Journal of Economics*, 11, 1-28

Duguet, E., 1999, Innovation, diffusion des connaissances et croissance, in *Innovation et Croissance, Innovation et Performances*, Ministère de l'Economie et des Finances, SESSI

Grossman, G. and C. Shapiro, 1987, Dynamic R&D competition, *Economic Journal* 97, 372-387

Guesnerie, R. and J. Tirole, 1985, L'économie de la Recherche-Développement Introduction à certains travaux théoriques, *Revue Economique* 36, 843-870

Kleinknecht, A., 1996, New indicators and determinants of innovations in A. Kleinknecht, ed. *Determinants of Innovations, The Message from New Indicators*, Macmillan Press

Acknowledgement 8 *A preceding version of this paper has benefited from helpful comments by Jean Pascal Benassy and Mark Machina. The authors gratefully acknowledge the help of Emmanuel Duguet.*

INNOVATION, DIFFUSION DES CONNAISSANCES ET CROISSANCE

Emmanuel DUGUET

Université de Paris I Panthéon-Sorbonne, EUREQua

L'importance des évolutions de la productivité dans la détermination du niveau de vie des nations amène naturellement à rechercher quels sont ses déterminants. Parmi ces déterminants, l'innovation nous intéresse plus particulièrement ici car elle permet de renouveler aussi bien les produits que les procédés de production. Or, la grande majorité des études sur le lien innovation-croissance porte sur des données agrégées par pays ou par industrie et n'utilise pas de mesure directe d'innovation.

Notre étude vise à compléter ce premier ensemble de travaux en étudiant directement le lien innovation-croissance sous l'angle microéconométrique. Pour cela, nous partons des enquêtes sur l'innovation réalisées auprès des entreprises industrielles par le SESSI, en étudiant la relation entre les innovations qui ont été mises en œuvre dans les entreprises et la croissance de leur productivité totale des facteurs. Ceci nous permet d'évaluer dans quelle mesure le résidu de Solow mesure bien l'innovation.

Pour ce faire, nous utilisons un modèle explicatif en deux parties. Dans une première partie, on étudie les déterminants de l'innovation, en insistant sur les sources externes de connaissance. On distingue deux types de sources de connaissances : d'une part, celles qui passent par le système de propriété industrielle, brevets et licences et, d'autre part, celles qui passent par les productions des autres entreprises, biens d'investissements et matériaux nouveaux. Cette décomposition permet d'étudier la question suivante : quelle est la contribution du système de propriété industrielle à la croissance, par rapport aux autres mécanismes de diffusion de la connaissance ?

Dans une seconde partie, on étudie le lien entre le taux de progrès technique, mesuré par la productivité totale des facteurs, et les innovations mises en œuvre par les entreprises industrielles. Dans ce modèle, l'innovation est endogène. La méthode économétrique employée tient donc compte à la fois de la nature qualitative des variables d'innovation employées et de la simultanéité entre les décisions de production (productivité globale des facteurs) et de mise en œuvre des innovations.

L'étude des déterminants de l'innovation fait apparaître que les deux types d'innovation examinés, l'innovation radicale et l'innovation incrémentale, n'ont pas les mêmes déterminants. En fait, deux populations d'innovateurs émergent. D'une part, les entreprises qui réalisent des innovations incrémentales (améliorations de produits et de procédés existants, imitation de produits) utilisent essentiellement des études techniques et de méthode (i.e., de recherche informelle) associées à l'utilisation novatrice de biens d'investissements. On peut interpréter ce résultat comme la mise en évidence d'un comportement d'adoption de nouvelles technologies, qui diffuserait les nouvelles techniques mises au point par d'autres entreprises au sein du tissu productif.

La seconde population d'innovateurs réalise essentiellement des innovations radicales (produits nouveaux pour le marché et premières de procédés technologiques). Les moyens employés pour réaliser ces innovations diffèrent fortement de ceux utilisés pour l'innovation incrémentale. A la base, ces entreprises ont investi dans une activité de recherche et développement interne à l'entreprise (au sens de Frascatti). Mais les sources internes de connaissances ne sont pas utilisées seules ; elles sont combinées à de nombreuses sources externes. Tout d'abord, les brevets dont l'entreprise est titulaire et les licences acquises auprès d'autres organismes jouent un rôle important dans le développement des innovations radicales. Ainsi, le système de propriété industrielle, souvent critiqué dans la littérature empirique, jouerait malgré tout un rôle de premier plan dans l'apparition des innovations radicales. A ces premières sources de connaissances externes, s'ajoutent les matériaux et composants nouveaux, dont l'utilisation comme inputs de l'innovation confirme l'existence d'une dynamique intersectorielle de diffusion de l'innovation allant au delà de l'adoption des nouvelles technologies. Enfin, la recherche externe n'est pas oubliée, puisqu'elle remplace l'effet des licences dans les industries à fortes opportunités technologiques.

Dans l'ensemble nous parvenons aux deux conclusions suivantes sur les déterminants de l'innovation : premièrement, la diffusion des connaissances et des innovations passées contribue fortement au développement des innovations présentes ; deuxièmement, les droits de propriété industriels jouent un rôle clé dans l'émergence des nouvelles technologies, malgré les inconvénients associés à leur mise en œuvre.

La seconde partie de ce travail vise à évaluer la contribution de chaque type d'innovation à la croissance. Pour cela, nous utilisons un modèle dans lequel la productivité globale des facteurs est expliquée par la mise en œuvre des innovations incrémentales et radicales. Dans ce modèle, le gain de croissance que nous évaluons est net. En effet, les innovateurs radicaux sont susceptibles de posséder un plus grand pouvoir de marché que les autres entreprises, de ce fait leurs marges sont automatiquement soustraites de la productivité globale des facteurs des entreprises utilisatrices de leurs innovations. Ce que nous examinons est donc un gain additionnel de croissance que les utilisateurs des innovations seraient en mesure de s'approprier. On en déduit qu'il est vraisemblable que la contribution d'une innovation incrémentale à la croissance d'une entreprise doit être inférieure à la contribution d'une innovation radicale. En effet, l'innovation incrémentale repose sur l'adoption d'une nouvelle technologie, ce qui suppose que la marge du fournisseur de biens d'investissements est retranchée de la croissance de l'entreprise utilisatrice. Au contraire, l'innovateur radical, qui exerce notamment son activité dans les biens d'équipement, est en mesure de prélever une marge sur les autres. On s'attend donc à ce que l'innovation radicale ait une plus forte contribution à la croissance que l'innovation incrémentale. Ce point est renforcé par l'argument suivant : il est vraisemblable que le comportement d'adoption d'une nouvelle technologie est suivi par les concurrents de l'adopteur, de sorte qu'en termes relatifs, la situation concurrentielle de l'adopteur reste inchangée.

Le premier résultat de l'étude économétrique sur la croissance, validé par plusieurs méthodes est que seule l'innovation radicale contribue significativement à la croissance de la productivité globale des facteurs. Ce premier point confirmerait donc la pertinence du résidu de Solow comme une mesure possible de performance des innovations. Le second résultat est que le gain de croissance dû à l'innovation radicale s'accroît avec le degré d'opportunités technologiques de l'activité qu'exerce l'entreprise, de 3% par an pour les faibles opportunités technologiques à près de 9% quand les opportunités technologiques sont fortes. Enfin, notre modèle permet d'agrèger les gains de croissance au niveau de l'industrie manufacturière¹. En tenant compte des

¹ L'agrégation est effectuée parmi les entreprises présentes de manière permanente sur la période 1985-1992.

entreprises qui n'innovent pas, la contribution des innovations radicales à la productivité globale des facteurs agrégée serait de l'ordre de 3% par an sur la période 1986-90.

Nos premières conclusions confirment donc l'importance de la circulation des connaissances dans la mise au point de nouvelles innovations, ainsi que la contribution significative de l'innovation radicale à l'amélioration de la productivité. Mais elles mettent également en lumière l'importance du droit de la propriété industrielle comme soutien de ces mêmes l'innovation malgré les imperfections du système de brevets.



Équipe Universitaire de Recherche en Économie Quantitative
– EURI:Qua
UMR 8594 du CNRS
Maison des Sciences Économiques
106-112, boulevard de l'Hôpital
75647 PARIS Cedex 13

Tél. : (33) 01 44 07 81 98 / Fax : (33) 01 44 07 82 28
e-mail : eduguet@univ-paris1.fr

INNOVATION, DIFFUSION DES CONNAISSANCES ET CROISSANCE

*Une analyse microéconométrique du lien entre la productivité globale
des facteurs et l'innovation sur données d'entreprises*

Emmanuel DUGUET

Cette étude a bénéficié d'un financement du Ministère de l'Industrie (SESSI). Je tiens également à remercier Stéphanie Monjon pour son aide sur la partie empirique de ce travail, ainsi que F. Brousseau, B. Crépon et D. Encaoua et pour leurs commentaires sur une version antérieure de ce travail. Je remercie également les participants au groupe de travail CEMÉ-MAD sur l'innovation, aux séminaires de l'ATOM et du CEMÉ ainsi que les participants aux colloques organisés par le TSER « R&D, Innovation and Productivity » (Paris, Juin 1998 ; Delft, Février 1999) et l'AFSE (Paris, Septembre 1998 ; Nice, Mai 1999).

Résumé

L'importance des évolutions de la productivité dans la détermination du niveau de vie des nations amène naturellement à rechercher quels sont ses déterminants. Un de ces déterminants nous intéresse plus particulièrement ici: l'innovation, qui permet de renouveler aussi bien les produits que les procédés de production. Alors que la grande majorité des études sur le lien innovation-croissance porte sur des données agrégées par pays ou par activité, et n'utilise pas de mesure directe d'innovation, cette étude vise au contraire à utiliser les enquêtes sur l'innovation réalisées auprès des entreprises industrielles par le SESSI, puis agrège, dans un second temps, les contributions individuelles de l'innovation à la croissance, mesurées au niveau des entreprises, pour en déduire son impact agrégé.

Pour ce faire, nous utilisons un modèle explicatif en deux parties. Dans une première partie on relie la mise en œuvre des innovations aux différentes sources de connaissances des entreprises, qu'elles soient formelles ou informelles, internes ou externes. Dans une seconde partie, on relie la croissance de la productivité totale des facteurs sur la période 1986-1990 aux innovations qui ont été mises en œuvre. La méthode employée tient compte à la fois de l'aspect qualitatif des données d'innovation et de la simultanéité entre les décisions de production et de mise en œuvre des innovations.

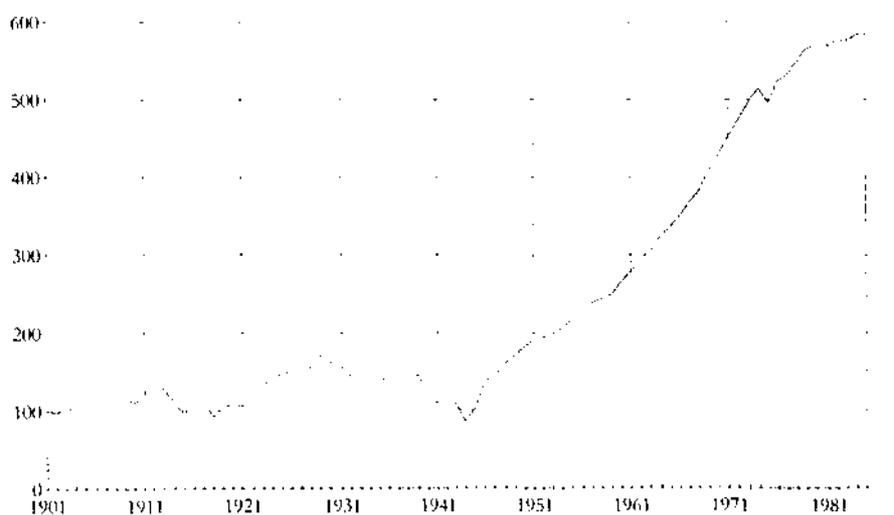
On obtient trois résultats principaux. Premièrement, il existerait une dualité entre d'une part, les entreprises qui ne mettent en œuvre que des innovations incrémentales (améliorations et imitations), qui se reposent essentiellement sur des travaux de recherche informels et l'utilisation novatrice de biens d'équipement et, d'autre part, les entreprises qui mettent en œuvre des innovations radicales (premières technologiques et produits nouveaux pour le marché), qui développent une stratégie basée sur une activité de recherche formelle, accompagnée de l'utilisation de brevets et de matériaux et composants nouveaux. Seules les entreprises qui mettent en œuvre des innovations radicales seraient à même d'améliorer significativement leur productivité globale des facteurs via l'innovation. Deuxièmement, le rendement des innovations, mesuré par la partie de la croissance de la productivité globale des facteurs que l'on peut attribuer à l'innovation, est relativement fort, et croissant avec le degré d'opportunités technologiques de l'activité de l'entreprise. Troisièmement, l'effet agrégé de l'innovation sur la croissance, en prenant en compte les entreprises qui n'innovent pas, serait de 2,8% par an en moyenne dans l'industrie manufacturière.

Introduction

Productivité et niveau de vie

A long terme, l'accroissement du niveau de vie est bien représenté par l'évolution de la production par habitant. Le graphique 1 montre l'évolution de la production intérieure brute réelle par habitant en France de 1901 à 1985. Sur près d'un siècle, la production par habitant a considérablement augmenté, après correction de l'inflation. On constate une relative stagnation de cette quantité du début du siècle jusqu'à la fin de la seconde guerre mondiale. Ensuite, la production par habitant décolle pour atteindre un niveau près de six fois plus élevé qu'au début du siècle. La consommation réelle par habitant est elle-même multipliée par cinq. Cette augmentation coïncide avec la période de plus forte croissance de la population. Mais on constate également un fléchissement de cette croissance dans les années soixante dix et l'engagement de l'économie vers une croissance plus faible au début des années quatre-vingt. Or cet indicateur découle directement de la productivité, égale au produit intérieur brut par actif occupé.

*Graphique 1 – Production intérieure brute réelle par habitant
(base 100 en 1901)*



¹ Calculs effectués d'après le tableau 1 de Villa (1996, p. 116) pour la propension à consommer, Villa (1994) pour la production intérieure brute et Daguet (1995) pour la population.

Cette évolution de la croissance du produit intérieur brut comme celle de la productivité se retrouve dans l'ensemble des pays développés sur la même période (Joly, 1993). Dans le même temps la substitution de capital au travail, qui véhicule le progrès technique, s'est également ralentie dans ces mêmes pays. Le résultat sur la croissance se maintient même si l'on tient compte du taux d'utilisation du capital.

L'innovation à la source des gains de productivité

L'importance de la productivité dans la détermination du niveau de vie amène naturellement à rechercher quels sont ses déterminants. Un de ces déterminants nous intéresse plus particulièrement ici : l'innovation.

L'innovation permet de renouveler l'intégralité du processus productif. Elle peut être réalisée par l'entreprise elle-même ou, au contraire, incorporée dans les biens d'équipement, les nouveaux matériaux et composants, ou la formation du personnel. En modifiant la combinaison productive, l'innovation permet d'accroître la productivité et par là même le niveau de vie à long terme. Mais l'innovation repose sur des connaissances qui sont un bien public et pose de ce fait un problème d'appropriation des bénéfices par les entreprises privées². Les pouvoirs publics cherchent donc à encourager l'innovation aussi bien dans les laboratoires publics que dans les entreprises privées, afin de compenser ce problème d'appropriation (Guillaume, 1998).

Une étude sur données d'entreprises

La grande majorité des études sur la croissance porte sur des pays ou des industries agrégées, qui n'utilisent pas habituellement de mesure de l'effort d'innovation pour évaluer ses effets sur la croissance. Or les enquêtes sur l'innovation, réalisées par le SESSI, permettent aujourd'hui d'identifier les entreprises qui innovent et autorisent ainsi l'évaluation directe de la contribution de l'innovation à la croissance. Nous nous proposons dans cette étude d'évaluer la contribution de l'innovation aux performances des entreprises industrielles, puis d'agréger ces performances pour en déduire celle de l'industrie.

² Pour une synthèse qui incorpore à la fois les aspects stratégiques et les aspects de diffusion du savoir, voir Katz et Ordover (1990).

Dans une première section, nous rappellerons les résultats d'une sélection d'études réalisées sur les liens entre l'innovation et les performances, en insistant sur l'industrie française. Dans une deuxième section, nous présenterons le modèle employé, qui est délibérément proche des modèles antérieurs, afin de faciliter à la fois les comparaisons et son agrégation. Dans une troisième section, les résultats des estimations seront présentés et commentés avant de parvenir, dans une dernière section, aux principales conclusions que l'on peut retirer de cette analyse du lien innovation-croissance. Les aspects plus techniques de l'estimation font l'objet d'une annexe séparée.

Quelques études antérieures

Les études qui nous intéressent ici portent sur la contribution de l'innovation, conçue comme un output de la recherche ou des études techniques, à la croissance de la valeur ajoutée. Dans la littérature, on s'est toutefois souvent orienté vers l'évaluation du taux de rendement des investissements en recherche et développement. On peut distinguer trois grandes familles d'évaluation.

Une première approche, représentée par les travaux de Mansfield (e.g., 1965) consiste à combiner un modèle théorique avec une approche comptable, dans le but d'évaluer le rendement marginal du capital de recherche et développement, c'est-à-dire, de combien aurait augmenté la production d'une entreprise sur la période d'étude si elle avait investi 1 franc de plus en recherche en début de période. On tient donc compte du caractère cumulatif de cet investissement.

Une seconde approche, représentée par l'étude de Minasian (e.g., 1962), consiste à intégrer le capital de recherche dans une fonction de production afin d'évaluer son rendement marginal instantané, c'est-à-dire, de combien augmenterait immédiatement la productivité si on augmentait immédiatement le capital de 1 franc. Cette approche, simple extension de l'estimation de fonctions de production, est celle qui s'est le plus développée dans les études micro-économiques.

Enfin, une troisième approche, suite aux travaux de Solow (1957), évalue le progrès technique comme un terme résiduel. Dans cette approche, on retranche à la croissance de la valeur ajoutée les parts qui sont dues au capital et au travail, ce qui reste est considéré comme le progrès technique, bien que l'auteur original soit prudent sur l'interprétation de ce résidu.

Cette méthode a surtout été appliquée sur données macroéconomiques. Le principal problème posé par cette approche par rapport aux précédentes, est qu'elle mesure le progrès technique sans utiliser la moindre information sur l'innovation ou les dépenses de recherche et développement.

La section suivante présente un ensemble de travaux économétriques représentatifs de ces trois approches, afin de voir quel sont, selon l'industrie, le pays et la période, les différentes évaluations du rendement de la recherche.

Quelques études antérieures

A la suite des travaux de quelques précurseurs (cf. Griliches, 1996), Solow (1957) propose une méthode de décomposition de la croissance qui deviendra un standard. Elle consiste à retrancher à la productivité, définie comme la valeur ajoutée par heure travaillée, la part de cette croissance que l'on peut attribuer à l'emploi et au capital (i.e., aux facteurs de production).³ Le résidu mesure par définition tous les autres changements et est souvent identifié au progrès technique. On appelle également cette quantité la *productivité totale ou globale des facteurs* selon les études. L'auteur applique cette méthode pour étudier la croissance de l'économie américaine entre 1909 et 1949. Il trouve que sur l'ensemble de cette période, le progrès technique est neutre par rapport aux facteurs de production. Il trouve également des gains de productivité significatifs, de l'ordre de 1% par an de 1909 à 1929 et de 2% par an 1930 à 1949. Enfin la décomposition révèle que la productivité par heure de travail, qui a doublé sur la période, serait due à 87% au progrès technique et à seulement 13% au ratio capital-travail⁴. Mais ce dernier résultat traduit plus vraisemblablement des erreurs de mesure des variables et de modélisation de la croissance. En effet, si l'on admet que les nouvelles générations de biens d'équipement sont plus efficaces que les anciennes on peut très bien avoir un accroissement de la productivité associé une baisse de l'intensité capitalistique, puisque cette dernière ne mesure que la *quantité* de capital et non sa qualité. De même, si l'efficacité du travail augmente à capital inchangé grâce à une meilleure formation, le ratio capital-travail peut rester inchangé alors que la productivité horaire augmente. Ceci ouvrira la

³ Il faut également tenir compte des consommations intermédiaires lorsque l'on travaille sur la production au lieu de la valeur ajoutée. Cette mesure a été étendue notamment à la prise en compte des qualifications du travail. Pour une vue d'ensemble sur données agrégées, cf. Accardo et Jlassi (1998).

⁴ Sous l'hypothèse que les rendements sont constants, on montre que le ratio capital-travail résume la totalité des effets du travail et du capital sur la productivité.

voie aux études qui incorporent des générations différentes de biens d'équipement (Jorgenson et Griliches, 1966 ; Mairesse, 1977) et les qualifications de la main d'œuvre (e.g., Crépon et Mairesse, 1993). Ces deux points sont d'ailleurs discutés par Solow lui-même. Mais au delà de ces problèmes de mesure des inputs, se pose celui de la définition retenue du progrès technique.

Progrès technique ou mesure de notre ignorance ?

Solow insiste bien sur le fait qu'il ne mesure que des changements résiduels, c'est-à-dire ce que l'on ne peut pas expliquer à partir des données disponibles. Il s'agit donc aussi, comme le souligne justement Abramovitz (1956), d'une mesure de notre ignorance. En effet, en économétrie, ce qui reste est ce que le modèle ne peut pas expliquer, une simple erreur de prévision comme les autres. Peut-on dès lors l'utiliser pour mesurer le progrès technique ?

Beaucoup d'auteurs se sont, dès cette époque, éloignés de cette voie et ont décidé d'employer des mesures reliées directement au progrès technique, quitte à se limiter aux données disponibles. Ainsi Minasian (1962) , Griliches (1964) ou encore Mansfield (1965) proposent d'évaluer le rendement des dépenses de recherche et développement plutôt qu'un résidu dont l'origine n'est pas identifiable en toute rigueur. En effet, n'importe quelle tendance structurelle de l'économie se retrouve dans ce résidu, et en particulier l'ouverture progressive des économies au commerce international, caractéristique des périodes étudiées. On pourrait également évoquer les forts investissements pendant la reconstruction.

L'article de Minasian (1962) propose de mesurer le progrès technique via les investissements en recherche et développement. Le capital de recherche entre donc dans la fonction de production de manière explicite et le résidu est considéré comme une erreur de prévision habituelle. Bien que l'échantillon soit restreint à 17 entreprises américaines de la chimie observées de 1948 à 1957, les résultats qu'il trouve sont assez proche des résultats qui s'avéreront classique dans ce type d'étude. En tenant compte d'un effet fixe individuel, l'auteur trouve que l'élasticité de la valeur ajoutée au capital de recherche serait de l'ordre de 0,1. Une hausse de 10% du capital de recherche serait associée à une augmentation de 1% de la valeur ajoutée. Ce chiffre est en apparence faible mais il n'en est rien car il implique un rendement marginal élevé du capital de recherche: de l'ordre 54% au lieu de 9% pour le capital physique. Dans l'ensemble, bien que le rendement marginal de la recherche soit élevé, son faible poids

dans les investissements des entreprises (i.e., 5% de la valeur ajoutée dans cette étude) rend son impact global plus modeste. Les raisons pour lesquelles les entreprises n'investiraient pas plus en recherche seraient alors à rechercher dans la littérature sur la propriété industrielle (Levin et al., 1987 ; Katz et Ordover, 1990).

Ce taux de rendement élevé de la recherche se retrouve par des méthodes très différentes. Ainsi, Mansfield (1962) étudie 10 entreprises de la chimie et du pétrole observées de 1941 à 1962, et applique également sa méthode à 10 industrie agrégées. Il évalue le rendement de la recherche comme de l'ordre de 40 à 60% dans l'industrie du pétrole. Dans la chimie, il serait plus faible. Si l'on suppose que le progrès technique est neutre, le rendement serait de 7% il serait de 30% pour le progrès technique incorporé au capital physique. Dans l'ensemble on trouve des ordre de grandeurs comparables aux résultats précédents.

L'étude de Mairesse (1977) propose d'évaluer le taux moyen de progrès technique incorporé au capital sur un échantillon de l'industrie manufacturière française en 1962. Ici, quel est l'impact des biens d'investissement de moins de trois ans sur la valeur ajoutée. L'auteur trouve des taux importants, de l'ordre de 8% par an (i.e., 47% au bout de cinq ans). Toutefois, il ne s'agit pas d'une étude avec dépenses de recherche, car ici le progrès technique doit passer par les biens d'équipement. On n'en déduit pas moins que l'impact du progrès technique incorporé est loin d'être négligeable sur la productivité.

Le rendement de la recherche dans l'industrie française

Pour la France, on peut citer l'étude de Crépon et Mairesse (1993) qui tient compte des qualifications du personnel. Les échantillons utilisés comportent 366 entreprises industrielles qui investissent en recherche et 933 qui n'en déclarent pas⁵. Une fois que l'on tient compte de tous les facteurs explicatifs des différences de productivité des entreprises (i.e., capital, travail et qualifications du personnel), on trouve une élasticité du capital de recherche qui varie entre 0,05 et 0,10 selon le secteur d'activité. Le rendement marginal varie lui entre 22 et 61% selon les secteurs⁶. Les ordres de grandeur sont donc tout à fait comparables à ceux obtenus sur les

⁵ Pour ces dernières entreprises les auteurs disposent de l'importance des ingénieurs et du personnel d'encadrement dans le personnel total. Cette catégorie de personnel fait vraisemblablement de la recherche informelle ou intermittente, combinée à des connaissances externes à l'entreprise.

⁶ Nous avons calculé ces rendements au point moyen de l'échantillon à partir de l'annexe 2 et du tableau 5 de l'article. On constate une propriété intéressante de ces estimations : les rendements marginaux seraient décroissants avec l'intensité technologique du secteur.

entreprises américaines⁷. Cette étude laisse toutefois deux problèmes en suspens: la recherche est une décision de l'entreprise qui doit être incluse dans les estimations, de plus on mesure l'innovation par un de ses inputs, la recherche. Or la recherche n'est pas forcément couronnée de succès⁸. On souhaiterait donc à la fois tenir compte de toutes les entreprises et mesurer la résultante de la recherche, l'innovation, pour expliquer les différences de productivité.

Cette extension a été réalisée par Crépon, Duguet et Mairesse (1998). Un premier ensemble de relations étudie la décision de faire de la recherche ou non, une autre relation étudie le lien entre les dépenses de recherche et l'innovation. Enfin, on étudie le lien entre cette dernière innovation, endogène car décidée par l'entreprise, et la productivité. Ce faisant, on tient également compte du capital physique et des qualifications du personnel. Une première différence avec l'étude précédente est que l'échantillon comprend 6145 entreprises, car la prise en compte des décisions de faire de la recherche évite de se cantonner aux seules entreprises qui en font ou à faire une simple comparaison entre les entreprises qui font de la recherche et les autres. Enfin, deux mesures d'innovation sont mises en concurrence: le nombre de brevets déposés (Crépon et Duguet, 1997) et la part des innovations dans le chiffre d'affaires (Crépon, Duguet et Kabla, 1996 ; Barlet et al. 1997). Les résultats sont encourageants et montrent que l'utilisation directe de données de brevet ou de variables qualitatives, sans les corrections nécessaires⁹, donne des résultats biaisés. Sur la productivité, les dépenses de recherche comme le nombre de brevets donnent des élasticités proches de 0,1.

Le lien innovation-productivité est robuste. Les résultats sont les suivants: une entreprise qui a déposé 2 brevets en moyenne a, toutes choses égales par ailleurs, une productivité de 6% plus élevée qu'une entreprise qui dépose un seul brevet en moyenne. L'écart pour un dépôt de 6 brevets (3^e quartile) est de 17%. Ainsi, malgré ses défauts, le nombre de brevets peut tout de même révéler de forts écarts de performances entre les entreprises.

⁷ Des extensions de la fonction de production de type Cobb-Douglas à une forme Translog ne changent pas les résultats obtenus au point moyen de l'échantillon, alors même qu'elle relâche les hypothèses de rendements constants et d'unité de l'élasticité de substitution capital-travail. Ce type de résultat empirique explique pourquoi on utilise si souvent une fonction de type Cobb-Douglas.

⁸ Le modèle qui suit en tient compte en la traitant comme une variable aléatoire dont on peut influencer le niveau moyen sans pouvoir déterminer toutes les réalisations. Dans le cas des brevets, on montre que l'aléa s'accroît avec le degré moyen d'innovation. Pour une explication plus détaillée, voir Crépon et Duguet (1994, encadré).

⁹ On montre qu'il faut effectuer à la fois des corrections pour la simultanéité et pour la sélection (vid.inf.), si l'on veut obtenir des estimations convergentes.

Mais on peut utiliser également la part des produits de moins de cinq ans dans le chiffre d'affaires ou *ratio d'innovation*. Il explique aussi des différences importantes de productivité. Pour les évaluer, il nous faut un point de référence : nous avons choisi un ratio de 5% car il correspond au centre de la modalité la plus faible de cette variable¹⁰. Par rapport à cette référence, une entreprise qui possède un ratio d'innovation de 10% réalise un gain de productivité de 4,6%. Si l'on prend les centres des autres classes de cette variable, les gains sont respectivement de 9,4%, 16% et 20% pour des ratios de 20%, 50% et 85%. Ces trois derniers cas représentent respectivement 32%, 14% et 3% des entreprises industrielles. Dans l'ensemble, l'innovation permet donc bien d'expliquer des écarts de productivité significatifs et répandus dans l'industrie.

Les études précédentes reposent toutefois sur une approche en coupe transversale, et n'abordent pas directement le problème de la croissance. Elles se focalisent sur des différences de productivité entre les entreprises innovantes et les autres entreprises à une date donnée. Celle qui innove le plus ont effectivement une productivité plus élevée en moyenne, et cette différence de moyenne peut être attribuée aux investissements de recherche¹¹. Les études sur la croissance visent plutôt à examiner comment la productivité d'une entreprise se modifie dans le temps selon qu'elle innove ou non.

Hall et Mairesse (1995) étudient le problème de la croissance de la valeur ajoutée dans les entreprises qui réalisent des investissements en recherche et développement. Le modèle est donc une extension du modèle de Minasian (1962) au cas de la croissance, où la recherche est traitée comme un facteur de production. Ils trouvent que, parmi 197 entreprises qui font de la recherche continuellement sur la période 1980-1987, l'élasticité de la productivité à la recherche est comprise entre 0,1 et 0,2 selon les estimations, ce qui tend à conforter les résultats obtenus en coupe.

Cette étude vise à prolonger l'ensemble de ces travaux d'une part en raisonnant sur l'innovation au lieu de considérer la seule recherche formelle et d'autre part en prenant pour point de départ le modèle de Solow.

¹⁰ Dans l'enquête Innovation du Sessi, le ratio d'innovation comporte quatre modalités : entre 0 et 10%, de 10 à 30%, de 30 à 70% et plus de 70%.

¹¹ La dernière étude tient compte de la recherche informelle à travers la proportion d'ingénieurs dans le personnel. Cette correction réduit l'importance de la recherche formelle dans l'explication de la valeur ajoutée mais nous en avons tenu compte dans le calcul des rendements marginaux données plus haut.

Contribution de cette étude

Le but de ce rapport est de compléter les travaux antérieurs sur les sources extérieures de connaissance, et sur la contribution de l'innovation à la croissance de la productivité globale des facteurs. Plus précisément :

- Nous faisons le lien entre les travaux visant à étudier la fonction d'innovation, qui expliquent le passage des inputs innovants aux innovations, et les travaux sur les performances des entreprises. A ce jour, seul le travail de Crépon, Duguet et Mairesse (1998) propose une approche structurelle de ce type. Ici toutefois nous n'utiliserons pas de données de R&D et de brevet, mais les données de l'enquête innovation. De plus, nous ajouterons à ce modèle des variables mesurant les externalités ou «spillovers» qui ont lieu entre les entreprises. En effet, l'innovation ne trouve pas son origine dans l'entreprise seulement. Les brevets détenus, les licences achetées à d'autres entreprises, les biens d'investissements et les matériaux et composants nouveaux permettent également à une entreprise de développer ses innovations. Ainsi, au lieu de mesurer ces externalités par une moyenne pondérée des dépenses de recherche des concurrents en matière d'innovation, nous mettrons directement les variables précédentes dans l'explication de l'innovation. Ceci présente deux avantages: premièrement, les variables utilisées sont individuelles et non sectorielles, donc ce n'est que dans la mesure où l'entreprise a effectivement utilisé les connaissances extérieures disponibles que nous mesurerons leur impact sur ses innovations. Deuxièmement, il n'est pas nécessaire de calculer une distance technologique entre les entreprises sous des hypothèses spécifiques puisque l'on dispose de mesures directes du phénomène étudié.
- Les études antérieures se concentrent sur les dépenses de recherche et développement et les brevets comme mesures de l'innovation. Ceci pose un certain nombre de problèmes. Dans le cas de la recherche et développement, on aboutira systématiquement à un petit nombre d'entreprises, ce qui implique que l'on omet d'étudier si l'innovation permet aux PME d'améliorer leur productivité. Or, même si l'effet était plus faible dans les PME, l'effet agrégé au niveau de l'industrie pourrait être important en raison de leur grand

Introduire les qualifications n'annule donc pas l'effet de la recherche et elle conserve le rendement élevé constaté dans les études antérieures.

nombre. De plus, il existe d'autres façons d'innover que par un programme de recherche et développement au sens de Frascati.¹² Dans le cas des brevets, on peut simplement remarquer que toutes les innovations ne sont pas brevetées et qu'une innovation peut être protégée par plusieurs brevets. Ce point est confirmé par l'enquête appropriation du SESSI qui montre clairement l'importance de la dimension stratégique du brevet. En effet, une étude récente (Duguet et Kabla, 1998) montre qu'en moyenne un tiers seulement des innovations seraient brevetées, en raison de la divulgation des informations que le brevet implique, et que les principaux facteurs explicatifs du nombre de brevets déposés, outre les investissements en recherche et développement, sont la volonté de se protéger des procès en contrefaçon, l'utilisation des brevets lors de négociations technologique et la part de marché acquise par le passé. Le brevet n'est donc pas seulement une mesure d'innovation.

- Notre but est donc de remplacer les mesures de recherche et développement et de brevets par des mesures directes d'innovation. Aussi, nous devrions pouvoir mieux évaluer l'effet de l'innovation sur la productivité totale des facteurs.

- Nous présentons les résultats agrégés impliqués par notre modèle microéconomique. Un résultat bien connu est que le taux de croissance de l'industrie est simplement une moyenne pondérée des taux de croissance individuels. En agrégeant notre relation de croissance de la même manière nous pourrions mesurer l'impact des innovations individuelles sur la croissance agrégée de l'industrie, en tenant compte de la distribution des productivités entre les entreprises, et également des entreprises qui n'innovent pas.

Dans un premier temps nous rappellerons les fondements du modèle employé. Ce modèle est proche de la comptabilité de la croissance, ce qui facilite son interprétation. Enfin, ce modèle est extensible à des cadres plus généraux que ceux évoqués ici et devrait permettre, à terme, de réintégrer la concurrence imparfaite ou le progrès technique biaisé.

¹² Au moins un chercheur à temps plein (ou l'équivalent).

Le modèle

Introduction

L'approche que nous avons adoptée dans cet article permet d'établir une relation entre la littérature largement macroéconomique sur la croissance et les données d'entreprises, qui permettent de mesurer directement l'innovation.

Nous allons tout d'abord reprendre, en apparence, l'approche de Solow, mais au niveau de chaque entreprise et étudier dans quelle mesure les enquêtes innovation permettent d'évaluer la contribution de l'innovation à l'accroissement de la productivité. A la fin seulement, nous agrégerons les données microéconomiques et confronterons nos résultats avec ceux des études agrégées qui n'utilisent généralement pas de mesure d'innovation. Une façon d'interpréter cette étude est donc de dire qu'elle examine si le «résidu de Solow» est effectivement relié aux innovations des entreprises industrielles.

Une deuxième contribution de cette étude consiste à mesurer la part que prend chaque type d'innovation à la croissance économique. Une amélioration a moins de chance d'accroître la productivité, mais les améliorations sont bien plus nombreuses que les innovations fortes. Lors de l'agrégation des résultats, il se pourrait que les améliorations soient plus productives, tout dépend de leur diffusion. A l'inverse, les innovations radicales, moins nombreuses, sont censées améliorer fortement la productivité. Ces deux types d'innovation contribuent-ils de manière égale à la croissance où faut-il franchir un certain seuil d'innovation pour améliorer ses performances ?

Le progrès technique neutre

Solow (1957) introduit le progrès technique en supposant qu'il est au sens de Hicks. La production est alors donnée par :

$$Q_{it} = \exp(\theta_{it}) \times f(K_{it}, L_{it})$$

où $\exp(\theta_{it})$ représente le progrès technique dont on cherche à mesurer l'évolution. Cette relation permet d'interpréter le progrès technique de la façon suivante : il autorise

l'augmentation de la production à quantités de travail et de capital identiques. A cette première hypothèse on ajoute celles de concurrence pure et parfaite sur les marchés des facteurs de production et du bien produit.

Il est important de noter ici que l'innovation ou la recherche et développement ne sont pas considérés comme des facteurs de production analogues au travail et au capital. On cherche simplement à étudier leurs liens avec le terme de progrès technique θ_t .

La décomposition de la croissance

Nous allons tout d'abord rappeler une décomposition simple de la croissance de la productivité. Nous étudierons ensuite son évolution dans le temps. La croissance de la production peut se décomposer comme:

$$\frac{\dot{q}}{q} = \frac{\dot{A}}{A} + s_k \frac{\dot{k}}{k}$$

où un point désigne une dérivée par rapport au temps. Ainsi \dot{q}/q est le taux de croissance de la valeur ajoutée et s_k la part du capital dans la valeur ajoutée. On peut donc écrire:

$$\frac{\dot{q}}{q} - (\rho - s_L) \frac{\dot{k}}{k} = \frac{\dot{A}}{A}$$

où ρ désigne les rendements d'échelle et s_L est le coût du travail rapporté à la valeur ajoutée. Cette dernière formulation est beaucoup plus pratique sur le plan empirique que la précédente car le coût du travail est facilement mesurable, contrairement au coût d'usage du capital. On écrit généralement cette relation en écart aux rendements d'échelle constants¹³, ce qui permet de tester simplement cette hypothèse :

$$\frac{\dot{q}}{q} - (1 - s_L) \frac{\dot{k}}{k} = \frac{\dot{A}}{A} + (\rho - 1) \frac{\dot{k}}{k}$$

¹³ Notons que les rendements d'échelle peuvent s'écrire : $\rho = s_L + s_K$.

Ainsi, des rendements d'échelle décroissants impliquent une corrélation partielle¹⁴ négative ($\rho-1 < 0$) entre la productivité globale des facteurs et le taux de croissance du capital par heure travaillée.¹⁵ On peut donc étendre le modèle de départ en testant la significativité du taux de croissance de l'intensité capitalistique dans la régression précédente!¹⁶

Le modèle économétrique

Le modèle est constitué de deux parties. Dans une première partie, on explique la mise en œuvre d'une innovation par ses inputs. Ceux-ci incluent les sources internes et externes de connaissances ainsi que la taille de l'entreprise, sa part de marché et la diversification de sa production. Dans une seconde partie, la mise en œuvre des innovations sert à expliquer la croissance. La mesure de performance employée s'apparente à un terme de progrès technique, c'est-à-dire aux gains de valeur ajoutée par heure travaillée une fois que l'on a enlevé les contributions du travail et du capital. Contrairement aux modèles usuels, l'effet de l'innovation n'est pas mesurée par un résidu mais par la partie du résidu que les indicateurs d'innovation peuvent expliquer. Ce qui reste est considéré comme une simple erreur de prévision sans interprétation économique particulière.

La fonction d'innovation

Deux représentations peuvent être envisagées pour l'innovation. Supposer que c'est le simple fait d'innover qui compte ou, au contraire, que c'est le degré d'innovation de l'entreprise qui détermine ses performances. Dans le premier cas, nous utilisons simplement une variable indicatrice¹⁷ comme facteur explicatif de l'innovation. Dans le second cas, nous supposons qu'il existe un degré latent d'innovation dans l'entreprise, inobservable, qui détermine la réalisation d'une innovation. C'est une prévision de cette variable latente qu'il faut

¹⁴ Partielle car on raisonne à taux de progrès technique constant, mesuré par \dot{A}/A , et propre à chaque entreprise.

¹⁵ Il faut que les rendements soient décroissants au sens large, car on obtient la relation précédente sous l'hypothèse de maximisation de la valeur de l'entreprise par rapport au capital et au travail. Dans le cas des rendements constants, on est seulement en mesure de déterminer le capital par heure employée et la valeur ajoutée par heure employée, pas l'emploi. Ceci n'a toutefois aucun impact sur la mesure de performance employée et la relation que nous utilisons reste valable quand les rendements sont constants.

¹⁶ Le taux de croissance est endogène, nous utiliserons donc la méthode des moments généralisés pour corriger les estimations du biais de simultanéité.

¹⁷ Il s'agit d'une variable qualitative qui, par convention, prend la valeur 0 si l'entreprise n'a pas innové, et la valeur 1 sinon (sans perte de généralité).

alors mettre prendre comme variable explicative. Nous effectuerons un test non emboîté pour choisir le bon modèle.

Le modèle structurel est le même pour les deux cas étudiés. Le degré latent d'innovation, noté I^* , est expliquée par les efforts de recherche et d'acquisition des connaissances extérieures de l'entreprise, notées Z , soit :

$$I_i^* = Z_i c + w_i$$

On n'observera la mise en œuvre d'une innovation par l'entreprise i que si ce degré latent d'innovation est suffisamment fort sur la période considérée, ce que l'on peut représenter par

$$I_i = \begin{cases} 1 & \text{si } I_i^* > \kappa \\ 0 & \text{sinon} \end{cases}$$

où κ est un seuil d'innovation fonction de l'activité exercée par l'entreprise. Sous l'hypothèse que la perturbation w_i suit une loi logistique, on obtient un modèle logit, que nous estimons par le maximum de vraisemblance. Notons que cette dernière hypothèse n'est importante que dans la méthode d'estimation en deux étapes (vid.inf.).

Dans le cas où l'effet de l'innovation est constant, c'est la variable indicatrice I qui joue sur la productivité alors que lorsque l'effet est variable, c'est la variable latente I qui explique les performances.

L'équation de croissance

Le modèle que nous utilisons ici décrit l'évolution du logarithme de la productivité globale des facteurs, défini comme :

$$\theta = \ln q - (1 - s_L) \ln k .$$

En différence, on obtient le taux de progrès technique, puisque l'on a :

$$\Delta\theta = \Delta \ln q - (1 - s_L) \Delta \ln k \cong \frac{\dot{q}}{q} - (1 - s_L) \frac{\dot{k}}{k}$$

Ce type de modèle est assez répandu dans l'étude dynamique des performances en général (Geroski, 1998), et implique d'écrire l'évolution de la productivité totale des facteurs, sous hypothèse de rendements constants¹⁸, comme :

$$\Delta\theta_{it} = \beta\theta_{i0} + I_i\gamma + S_i\delta + u_{it}$$

où $\Delta\theta_{it}$ est le taux de croissance de la productivité totale des facteurs sur la période de l'enquête innovation 1985-90, θ_{i0} la valeur initiale de la productivité totale des facteurs, I_i désigne les innovations de l'entreprise et S_i son secteur d'activité. Le terme u est la perturbation du modèle qui représente les chocs aléatoires auxquels l'entreprise a fait face!¹⁹

Ainsi ce modèle élimine d'abord la tendance de croissance de l'industrie à laquelle appartient l'entreprise. Ensuite, il explique l'évolution des performances par le niveau atteint par le passé, dont nous allons voir qu'il représente l'historique d'innovation de l'entreprise, et par les innovations que l'entreprise a réalisé durant la période étudiée. Un terme additionnel représente l'erreur de prévision.²⁰

Les différents coefficients du modèle ont une interprétation intéressante : le premier terme, b , est relié au taux *de dépréciation des innovations*, le deuxième coefficient, γ , mesure l'impact des innovations courantes sur la productivité totale des facteurs et le coefficient δ la tendance générale de croissance de l'industrie à laquelle l'entreprise i appartient.

Pour obtenir l'interprétation de b , il suffit de développer le modèle annuel jusqu'à la date de création de l'entreprise i , notée $t=0$ par convention. On a :

$$\theta_{it} = (1 + b)\theta_{i,t-1} + I_{it}\gamma + S_{it}\delta + u_{it},$$

avec $b < 0$, ce qui implique :

¹⁸ Nous relâcherons cette hypothèse plus loin.

¹⁹ Notons l'analogie entre cette relation microéconomique et le problème de la convergence étudié au niveau macroéconomique. La différence vient de ce que nous ne comparons pas des pays mais des entreprises, de sorte que nous étudions les phénomènes de rattrapage d'une entreprise par une autre. En fait, il s'agit ici d'une convergence conditionnelle à l'industrie d'appartenance et à l'effort d'innovation.

²⁰ On peut également interpréter la perturbation u comme une erreur de mesure sur le taux de croissance de la productivité totale des facteurs.

$$\theta_{it} = (1+b)^t \theta_{i0} + \gamma \sum_{\tau=0}^{t-1} (1+b)^\tau I_{i,t-\tau} + S_i \delta \sum_{\tau=0}^{t-1} (1+b)^\tau + \sum_{\tau=0}^{t-1} (1+b)^\tau u_{i,t-\tau}.$$

La performance de l'entreprise i à la date t dépend donc de la performance de son processus de production à la date initiale, de tous ses efforts d'innovation passés et des aléas qu'elle a rencontré durant son développement.

Lorsque l'on raisonne sur une grande valeur de t , l'influence de la valeur initiale s'atténue et seules les innovations mises en œuvre et les aléas expliquent la performance courante. Ici, on peut faire un parallèle entre l'effet de l'innovation et les modèles avec un capital d'innovations. Ce capital d'innovations, qui résume le niveau technologique de l'entreprise i en début de période, noté T_{it} , est obtenu en appliquant la règle de l'inventaire permanent²¹ :

$$T_{it} = \sum_{\tau=0}^{t-1} (1-\phi)^\tau I_{i,t-\tau},$$

où ϕ est le taux de dépréciation annuel du capital d'innovations. Il existe donc une relation simple entre le taux de dépréciation des innovations et la croissance de la productivité globale des facteurs. Dans notre modèle :

$$\phi = -b.$$

Comme le modèle que nous estimons est en différences longues sur 5 ans, soit 4 périodes de croissance, on définira le taux de dépréciation annuel des innovations comme :

$$b = (1+\beta)^{1/4} - 1.$$

Enfin, remarquons qu'une valeur négative de b implique la convergence des entreprises vers une performance d'équilibre²². En fait, cette convergence est *conditionnelle* puisque la valeur vers laquelle la performance converge dépend à la fois du secteur d'activité et de l'effort d'innovation de l'entreprise. Comme dans notre modèle l'innovation est décidée par

²¹ Soit $T_{it} = (1-\phi)T_{i,t-1} + I_{i,t} = \sum_{\tau=0}^{t-1} (1-\phi)^\tau I_{i,t-\tau}$ où ϕ est le taux de dépréciation du capital d'innovations, supposé constant, de l'entreprise.

²² Un autre cas intéressant est celui de la marche aléatoire (cas $b=0$). Dans cette situation, la productivité totale des facteurs à un instant donné est proportionnelle à la somme de toutes les innovations réalisées par l'entreprise par le passé et de tous les chocs aléatoires qui l'ont affecté à un moment donné. Ceci cas implique

l'entreprise, celle-ci possède toujours la possibilité de changer sa performance d'équilibre en réalisant des investissements immatériels. En innovant, elle peut reprendre une trajectoire de croissance plus élevée, à l'inverse, si elle cesse d'innover, sa croissance s'achemine vers une valeur plus faible.

On voit donc que la modélisation retenue suggère une certaine dualité de croissance entre les innovateurs et les autres, les premiers se dirigeant vers la performance d'équilibre la plus élevée. Il suggère également que l'arrêt de l'innovation par une entreprise la ramènerait vers la performance moyenne de son industrie. Le développement précédent permet de montrer que la performance d'équilibre dépend du taux de dépréciation des innovations. En prenant l'espérance mathématique de la relation précédente et en posant $\Delta\theta=0$, on obtient la productivité totale des facteurs d'équilibre :

$$\theta^* = \frac{\gamma I + \delta S}{-b} = \frac{\gamma I + \delta S}{\phi},$$

où l'on a supposé un effort constant d'innovation, noté I. Plus la dépréciation des innovations est forte, plus la performance d'équilibre est faible. La seule manière de contrebalancer l'effet d'une forte dépréciation des innovations consiste soit à investir de manière permanente en matière d'innovation ou à bénéficier de la diffusion d'une innovation majeure via le terme sectoriel S. Cette relation exprime également la dualité entre les innovateurs et les autres entreprises. Si une partie des entreprises ne réalise jamais d'innovation, $I=0$, leur performance d'équilibre sera durablement plus faible que celle des innovateurs.

Les paramètres structurels du modèle que nous estimons sont donc les suivants :

- Le taux de dépréciation annuel des innovations (ϕ);
- Le gain de croissance moyen associé à l'innovation (γ);
- La tendance de croissance de l'activité de l'entreprise (δ).

en particulier qu'il n'y a pas de dépréciation des innovations et que l'effet des chocs aléatoires ne s'atténue pas au cours du temps. Nos estimations rejettent clairement cette hypothèse restrictive.

Il nous reste à voir comment on peut tester l'hypothèse de rendements constants. Sous l'hypothèse que les rendements sont décroissants, on a :

$$\theta_p = \ln q - (\rho - s_L) \ln k ,$$

en remplaçant θ par θ_p dans la relation précédente, on obtient l'évolution :

$$\Delta\theta_{it} = b\theta_{i,t-1} + (\rho - 1)\Delta \ln k_{it} - b(\rho - 1) \ln k_{i,t-1} + I_i\gamma + S_i\delta + v_{it} ,$$

où v est la perturbation du nouveau modèle. Le coefficient de la croissance du capital permet donc de mesurer l'écart aux rendements constants²³

L'effet de l'innovation est-il variable ou constant?

Les deux modèles explicatifs de la croissance peuvent se résumer sous la forme suivante:

Modèle 1 (effet constant) : $\Delta\theta_i = \beta\theta_{i0} + I_i\gamma + S_i\delta + v_i$ où la différence de performance est directement donnée par le coefficient γ et $I_i = 0$ ou 1 .

Modèle 2 (effet variable) : $\Delta\theta_i = \beta\theta_{i0} + I_i^*\gamma + S_i\delta + v_i$ où la différence de performance entre deux entreprises doit être calculée au cas par cas, et $I_i^* \in \mathbb{R}$.

Pour estimer ces modèles il faut tenir compte du fait que l'innovation est endogène, c'est-à-dire déterminée simultanément avec la productivité totale des facteurs sur l'ensemble de la période. Pour le premier modèle, il faut remplacer I par la probabilité estimée d'innover donnée par le modèle logit, soit $\hat{p} = \exp(Z\hat{c}) / (1 + \exp(Z\hat{c}))$, qui ne dépend que de variables exogènes par définition. Pour le second modèle, il faut simplement remplacer I par sa prévision $Z\hat{c}$, qui ne dépend que des variables exogènes (i.e., les inputs de l'innovation, cf., Maddala, 1983). Bien, entendu, les écarts-types des coefficients ne peuvent plus être calculés par la méthode habituelle, nous utilisons les écarts-types robustes à l'hétéroscédasticité (White,

²³ Cette relation implique également une contrainte entre les paramètres du modèle pour laquelle il faut effectuer un test. Ici, la somme du troisième coefficient et du produit des deux premiers ne doit pas être significativement différente de 0. Notons toutefois que la nullité du coefficient du taux de croissance du capital implique la nullité du coefficient du capital décalé, de sorte que sous l'hypothèse nulle de rendements constants aucune des deux variables ne doit être statistiquement significative, ce que nous avons vérifié.

1980).²⁴ Notons que cette méthode permet d'étudier plusieurs types d'innovations en même temps.

Le modèle à effet variable ne généralise pas à proprement parler le modèle à effet constant car on ne peut pas obtenir le modèle à effet constant en imposant une contrainte sur les paramètres du modèle à effet variable.²⁵ La réciproque étant également vraie, il nous faut recourir à un test non emboîté. Nous reprenons ici l'approche développée par Davidson et Mc Kinnon (1981) et Godfrey (1983) mais avec des écarts-types robustes à l'hétéroscédasticité (cf., Smith, 1992).

L'intuition du test est la suivante. Supposons que le vrai modèle comporte un effet constant, alors, l'ajout d'un effet variable au modèle est inutile. Dans le test, cette propriété se traduit par le fait que la prévision de la performance par le modèle à effet variable ne contient aucune information pertinente pour améliorer le modèle à effet constant. Cette prévision ne doit donc pas être significative dans le modèle à effet constant. Inversement, pour que le modèle à effet variable soit valide, il faut que la prévision de performance obtenue par le modèle à effet constant ne soit pas significative dans le modèle à effet variable, sinon cela signifierait qu'une information pertinente est négligée par le modèle à effet constant.

Il peut sembler étonnant, a priori, qu'un modèle à effet variable puisse fournir de moins bonnes prévisions qu'un modèle à effet constant. Mais ici, cet effet variable est contraint. Il a par définition une forme linéaire qui possède la propriété particulière suivante: les entreprises qui n'innovent pas ont un indicateur d'innovation négatif de sorte que le modèle à effet variable implique une décroissance de la productivité globale des entreprises qui n'innovent pas.²⁶ Or, dans les modèles économiques habituels²⁷, l'absence d'innovation implique souvent l'absence de croissance, de sorte qu'une croissance négative est invraisemblable. Le modèle à effet constant possède l'avantage de se contenter d'imposer un gain de croissance à ceux qui modifient leur produits et leurs procédés.

²⁴ Il serait possible d'améliorer l'efficacité de ces estimateurs soit en tenant compte de la forme spécifique prise par la variance de la perturbation telle qu'elle découle du modèle logit, soit en recourant aux moindres carrés asymptotiques.

²⁵ Le seul point commun entre les deux modèles est obtenu lorsque l'effet de l'innovation est nul sur la croissance.

²⁶ La modélisation retenue implique un indicateur d'innovation négatif dès que la probabilité d'innover est inférieure à 1/2.

²⁷ Sur ces modèles de croissance, voir Barro et Sala-i-Martin (1995)

Dans ce type de test, trois types de résultats sont possibles a priori:

- 1) Les deux modèles ne sont pas rejetés. Il est alors vraisemblable que soit les données manquent de précision, soit que l'hétérogénéité de l'échantillon est insuffisante pour pouvoir conclure en faveur d'un des deux modèles.
- 2) Un des deux modèles est rejeté et l'autre n'est pas rejeté. C'est le cas le plus favorable, car on peut se prononcer sur la représentation adéquate du phénomène étudié.
- 3) Les deux modèles sont rejetés. C'est vraisemblablement un troisième modèle qui explique les deux autres.

Afin d'abrégier la présentation des résultats, nous donnons dès maintenant les résultats des tests non emboîtés. La statistique présentée est le *t* de Student asymptotique robuste²⁸ de la prévision du modèle concurrent. Quand cette statistique est significative, l'hypothèse nulle (i.e., le modèle) indiquée dans le tableau est rejetée. La conclusion est claire: on rejette le modèle à effet variable linéaire et l'on rejette jamais le modèle à effet constant au seuil de 10%.²⁹ Nous garderons donc dans la suite le modèle à effet constant.

Tableau 1 – Modèle à effet constant contre modèle à effet variable

Opportunités d'innovation:	Faibles	Moyennes	Fortes	Ensemble
<i>Hypothèse nulle</i>				
1) Modèle à effet constant	-0.37	-1.61	1.35	0.70
2) Modèle à effet variable	0.58	2.14	0.16	0.35

(*t* de Student robuste de la prévision du modèle concurrent.)

Le modèle à effet constant n'est jamais rejeté au seuil de 10% (valeur critique : 1.645) alors que le modèle à effet variable est rejeté au seuil de 5% (valeur critique : 1.96) dans les activités de moyennes opportunités d'innovation.

²⁸ Par rapport au *t* de Student habituel, il doit être comparés aux fractiles d'une loi normale centrée-réduite.

²⁹ Notons qu'un seuil de 10% est plus restrictif qu'un seuil de 5% car la statistique évalue la significativité du modèle concurrent.

Les données

Les sources internes et externes de l'innovation

Les données proviennent de l'enquête Innovation du Sessi qui porte sur l'ensemble de la période 1986-1990.³⁰ Cette enquête permet de savoir quelles sont les innovations qui ont été mises en œuvre dans les entreprises industrielles, ainsi que les sources de ces innovations. On peut penser que ces données sont plus à même de révéler les différences de croissance que les mesures de brevet et de recherche et développement pour les deux raisons suivantes. La recherche et développement est un input de l'innovation et n'agit donc pas directement sur la productivité, c'est seulement dans la mesure où cette recherche est réussie que l'on peut espérer relever des écarts de productivité entre les entreprises.

Les mesures directes de mise en œuvre des innovations ne sont pas sujettes à ces critiques, même si elles demeurent nécessairement qualitatives. Pour cette étude nous utilisons toutes les sources de l'innovation présentes dans l'enquête. Jusqu'à maintenant seules les dépenses de recherche semble avoir été utilisées. Or ces dépenses ne sont disponibles que pour environ 10% des entreprises industrielles et mettent surtout l'accent sur les connaissances internes à l'entreprise, même si l'inclusion des dépenses de recherche commandées à des organismes extérieurs compense en partie ce problème. Qui plus est, la recherche formelle de l'entreprise, y compris du groupe auquel elle appartient, n'est qu'une des sources de l'innovation. Outre la recherche informelle, on peut citer les sources externes de l'innovation, mises en évidence par le développement des travaux sur les externalités en matière de recherche et développement (Jaffe, Mairesse et Mohnen, 1990). On peut distinguer notamment les licences que l'entreprise se procure mais également les biens d'équipement et la disponibilité de matériaux et composants nouveaux. Ces deux dernières sources sont étroitement reliées aux secteurs des biens d'équipement et des biens intermédiaires. Inclure ces données dans l'explication des succès individuels en matière d'innovation revient donc à introduire l'effet des externalités non pas en construisant une distance technologique entre industries, notamment à partir des données de brevet, mais en mesurant directement l'impact des ces variables sur l'innovation. C'est dans la mesure où les sources externes de connaissances influencent l'innovation qu'elle peuvent avoir un effet sur la productivité.

³⁰ Pour une présentation de l'enquête, voir François (1991).

Plus précisément, l'enquête Innovation propose 8 sources d'innovations:

- [1] La recherche et développement interne à l'entreprise, au sens de Frascati, qui correspond à la mesure que l'on utilise habituellement pour mesurer les inputs de la recherche.
- [2] La recherche et développement interne au groupe auquel appartient éventuellement l'entreprise, qui peut permettre d'expliquer pourquoi des entreprises qui innoveront ont peu de dépenses de recherche internes
- [3] Les activités d'études techniques et de méthodes dans l'entreprise, qui visent à mesurer la recherche informelle, notamment dans les plus petites entreprises.
- [4] La recherche et développement acquise à l'extérieur (y compris les organismes publics). Cette variable est intéressante car elle élargit la notion de recherche et développement externe au secteur public et peut donc permettre de capter des complémentarités avec la recherche privée.
- [5] Les brevets dont l'entreprise est titulaire, permettent de mesurer le poids des outputs passés de l'innovation comme inputs de l'innovation courante et font référence à une certaine dynamique de la connaissance. Les innovations passées génèrent de nouvelles innovations.
- [6] Les droits et licences acquis à l'extérieur de l'entreprise, mesurent le poids des innovations passées des autres comme input de l'innovation actuelle, et complètent la description précédente.
- [7] L'utilisation novatrice de biens d'investissements (installation, processus, matériels), permet de voir les effets des innovations passées provenant de l'industrie des biens d'équipement.
- [8] Les matériaux et composants nouveaux indiquent l'importance de l'industrie des biens intermédiaires dans le processus d'innovation.

Ces informations ne sont disponibles que sous une forme qualitative: « pas du tout », « un peu », « moyennement » et « beaucoup ». Nous verrons que dans l'ensemble elles fournissent des résultats très intéressants.

Pour compléter ces informations qualitatives nous avons ajouté des informations sur la motivation des innovation: le marché («market pull») ou la dynamique propre de la technologie («technology push»). Ces deux informations sont également disponibles au niveau de l'entreprise.

D'autres variables traditionnellement adoptées dans la littérature figurent également dans la relation : la taille et le pouvoir de marché, pour mesurer d'éventuels effets Schumpéteriens, la diversification pour détecter l'éventuelle internalisation des externalités des entreprises diversifiées.

Taille, pouvoir de marché et diversification

Ces variables jouent ici un rôle différent des celui des études antérieures.

Alors qu'elles représentent habituellement une forme réduite de la recherche développement, ici on examine si elles ont un effet additionnel une fois pris en compte les inputs de l'innovation. Ainsi la part de marché ne mesure pas ici l'effet du pouvoir de marché sur l'innovation mais si le pouvoir de marché influence la décision de l'entreprise d'appliquer ou non ses «découvertes». En effet, comme les variables qui représentent la recherche et développement contiennent déjà l'influence du pouvoir de marché sur les investissements immatériels de l'entreprise, l'effet que nous mesurons est plus proche de la décision d'appliquer ces connaissances ou non. Il s'agit donc de tester si, à recherche identique, les entreprises qui ont une plus grande part de marché appliquent plus souvent leurs innovations que les autres entreprises.

On s'attend donc ici à un effet beaucoup plus faible des variables Schumpéteriennes que dans les études antérieures. Si tel était le cas, cela signifierait que l'essentiel de leurs effets concernent les investissements immatériels et non l'utilisation des connaissances par les entreprises. Notons enfin que les effets Schumpéteriens ne sont attendus que sur les variables d'innovations fortes, produit nouveau et première technologique, et dans les secteurs d'activité où l'innovation est effectivement importante.

La taille est mesurée ici par le chiffre d'affaires hors taxe en 1985. Les autres variables nécessitent un traitement de la diversification. Tout d'abord, on calcule le nombre équivalent d'activité industrielles puis la part de marché moyenne des entreprises diversifiées. Ces calculs

ont été réalisés à partir de l'Enquête Annuelle d'Entreprise par fractions (en 1985). Les définitions de ces variables sont identiques à celles de Crépon, Duguet et Kabla (1996).

La productivité totale des facteurs

Afin d'évaluer la productivité totale des facteurs, il faut construire des séries individuelles de capital physique, tenir compte de son taux d'utilisation et du nombre d'heures travaillées. La base de données porte sur les entreprises de 20 salariés et plus mais les immobilisations ne sont disponibles que pour les entreprises de 50 salariés ou plus. Il faut donc recalculer les séries de capital à partir des séries longues d'investissements individuels (1978-1990), afin de garder les entreprises de 20 à 50 salariés.

Nous avons utilisé les sources d'information suivantes: Enquête Annuelle d'Entreprise (de 1978 à 1990). Une base de données macroéconomiques a également été nécessaire¹¹, elle comprend les taux de déclassement du capital au niveau 40 de la NAP, variables chaque année. Pour la première année des entreprises présentes en 1978, nous avons supposé un taux de croissance annuel du capital égal à la moyenne du secteur sur la période et un taux de dépréciation moyen égal à la moyenne de celui du secteur T40 considéré sur la période 1975-1978. Nous utilisons également le prix de la valeur ajoutée (1986 et 1990), le prix de la FBCF (1978 à 1990), le nombre d'heures moyennes travaillées (1986 et 1990) et le taux d'utilisation des capacités de production (1986 et 1990).

Pour calculer le capital, on utilise la formule de l'inventaire permanent avec un taux de déclassement sectoriel et un prix de la FBCF correspondant à l'activité principale de chaque entreprise au niveau 40 de la NAP (notés respectivement δ et p_{it}) :

$$K_t = (1 - \delta) K_{t-1} + I_t / p_{it}$$

Les définitions que nous avons retenues s'efforcent de mesurer au mieux les volumes d'inputs et d'output. Nous avons corrigé les données d'entreprises en les déflatant et en tenant compte du nombre moyen d'heures travaillées et du taux d'utilisation des capacités de production.

¹¹ Cette base de données nous a été transmise par la Division de la Conjoncture de l'INSEE.

L : le nombre d'heures travaillées, obtenues en multipliant le nombre d'employés par le nombre moyen d'heures de travail par employé au niveau 40 de la NAP.

q : valeur ajoutée réelle hors taxe par heure travaillée.

k : le capital précédent multiplié par le taux d'utilisation des capacités de production et divisé par le nombre d'heures travaillées.

s_t : part de la rémunération des salariés (salaires + charges) dans la valeur ajoutée hors taxes. Il est inutile de déflater ici car il s'agit d'une fraction .

L'ensemble de ces données nous permet d'évaluer le taux de croissance de la productivité globale des facteurs sur la période 1986-1990.

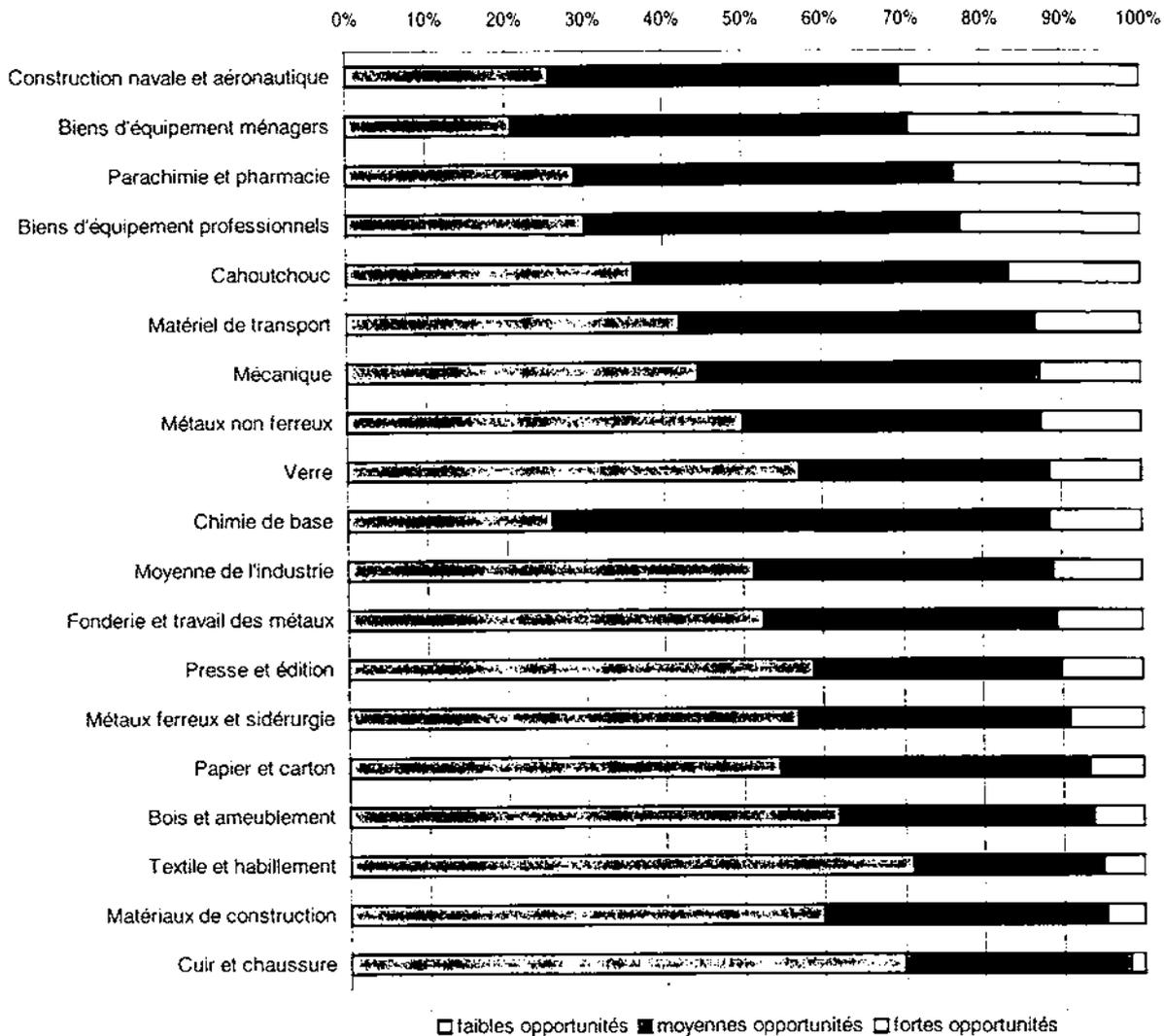
Statistiques descriptives

Les opportunités d'innovation par industrie

Une étude antérieure (Barlet et al., 1998) a clairement montré que le succès commercial d'une innovation dépend des opportunités technologiques de l'activité dans laquelle exerce l'entreprise. Pour cette raison, on distingue, d'après l'enquête innovation, trois types d'activités selon qu'elles sont faiblement, moyennement et fortement innovantes. Ce classement est réalisé par les entreprises elles-mêmes.³² L'intérêt de cette variable est de permettre de distinguer, au sein d'une même industrie agrégée, les activités fortement innovantes et celles qui ne le sont pas. Il existe bien entendu un lien entre ces réponses et le niveau technologique du secteur d'activité, toutefois, une entreprise appartenant à industrie agrégée réputée innovante peut très bien déclarer que son activité, définie au niveau de l'entreprise, n'est pas innovante. Ce point est clairement illustré par le graphique 1 dans lequel nous donnons la décomposition des degrés d'innovation des activités au niveau 40 de la NAP.

³² Il s'agit de la réponse à la question « X1 » de l'enquête Innovation. On a regroupé les modalités « non innovantes » et « faiblement innovantes » dans la première catégorie. Cette définition est identique à celle employée par Barlet et al. (1998) de sorte que les deux études sont comparables.

Graphique 1 - Évaluation par les entreprises du degré d'innovation technologique de leur propre activité



Les industries sont classées par ordre décroissant de fortes opportunités d'innovation

On voit que le pourcentage d'entreprises déclarant une activité fortement innovante est toujours inférieur à 30%. Ce résultat est intéressant si on le rapproche de la remarque selon laquelle la forte hétérogénéité des performances entre les entreprises innovantes compliquerait l'évaluation du rendement de l'innovation, et mènerait ici à une sous-estimation du rendement de l'innovation.

En effet, le graphique 1 montre qu'une part non négligeable des industries «de haute technologie» serait composée d'activité faiblement innovantes et que d'autre part, les industries de faible technologie comporteraient également une part significative d'entreprises fortement innovantes. La résultante serait une sous-estimation de l'écart de performance entre les entreprises innovantes identifiées par leur seule industrie et les autres entreprises. La méthode que nous employons, qui consiste à séparer ces entreprises, doit logiquement permettre de contrôler cette source d'hétérogénéité et contribuer à une meilleure évaluation des effets de l'innovation sur la croissance.

Globalement, si l'on raisonne sur le pourcentage d'entreprises qui considèrent que leur activité est fortement innovante, on retrouve les secteurs de haute technologie : construction navale et aéronautique, biens d'équipement, parachimie et pharmacie, caoutchouc et matériel de transport. L'identification d'un niveau technologique par l'industrie semble donc fonctionner «en moyenne» mais on est loin de l'homogénéité : les industries agrégées regroupent des activités caractérisées par des opportunités d'innovation différentes.

Les innovations mises en œuvre et leurs sources

Dans ce rapport nous retenons deux types d'innovations : les innovations radicales et les innovations incrémentales. Les innovations radicales se définissent comme l'introduction sur le marché d'un produit entièrement nouveau à la fois pour l'entreprise et pour le marché ou de la réalisation d'une première de procédé technologique.³³ L'innovation incrémentale, au contraire concerne l'amélioration d'un produit existant, à l'exclusion du conditionnement, la mise sur le marché d'un produit nouveau pour l'entreprise mais pas pour le marché, ou l'amélioration d'un procédé de production existant.³⁴ On s'attend à ce que l'effet des innovations radicales soit plus fort sur les performances des entreprises dans la mesure où elles imposent des modifications plus importantes de la fonction de production.

Les innovations mises en œuvre et les caractéristiques des entreprises sont décrites dans le tableau 2. On voit apparaître un fait intéressant : les innovations incrémentales et radicales *ne sont pas* parfaitement corrélées. Alors que le pourcentage d'entreprises qui a mis en œuvre une innovation radicale s'accroît avec les opportunités technologiques, le pourcentage d'innovations incrémentales s'accroît d'abord *puis décroît* avec ce degré d'opportunités

³³ Questions « U2 » et « U4 » de l'enquête, respectivement.

technologiques. Ce résultat suggère donc que les déterminants de ces deux innovations ne sont pas les mêmes. Ce point sera confirmé par les régressions logistiques. D'autre part, on voit que la taille et la part de marché moyenne s'accroissent avec les opportunités technologiques, mais ceci peut provenir des différences de coûts fixes entre les différentes industries, comme le suggère le graphique 1.

Tableau 2 – Innovations mises en œuvre par les entreprises, taille diversification et part de marché

Opportunités d'innovation:	Faibles	Moyennes	Fortes	Ensemble
<i>Type d'innovation</i>				
Innovation incrémentale	35.7%	59.0%	45.0%	45.5%
Innovation radicale	18.4%	51.6%	76.7%	37.3%
<i>Moyennes</i>				
Chiffre d'affaires (Millions de francs 1985)	57	99	302	100
Part de marché moyenne	0.8%	1.4%	1.8%	1.1%
Diversification moyenne (nombre équivalent)	1.28	1.30	1.33	1.29

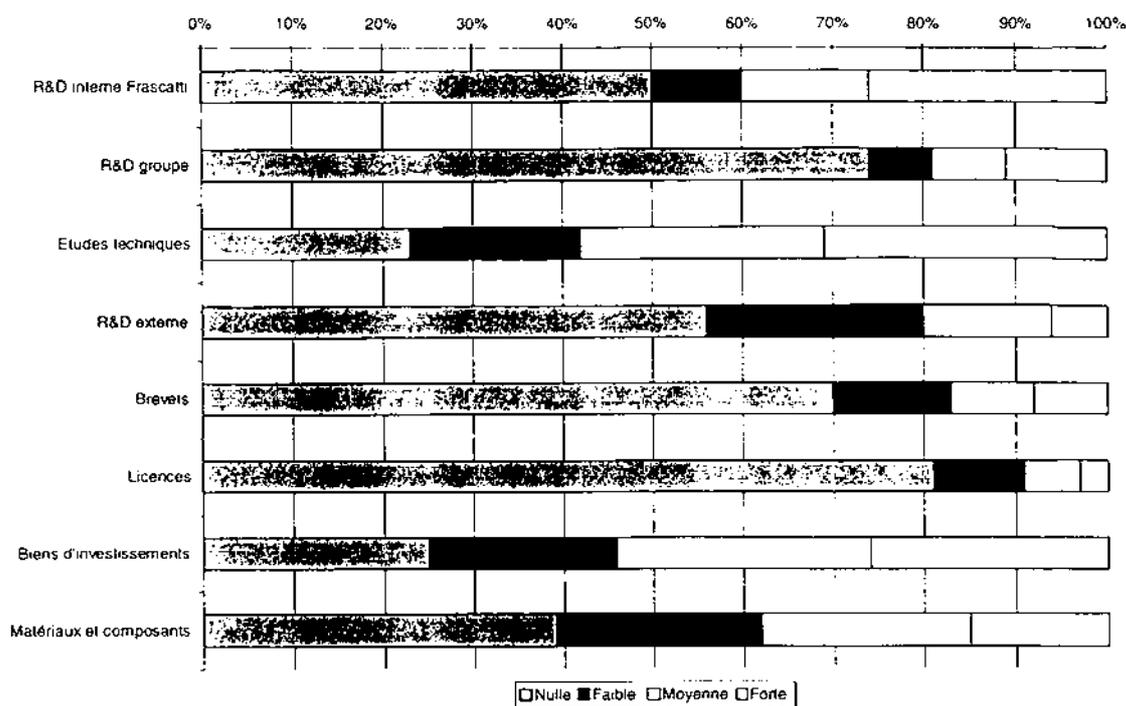
Source : échantillon de 4460 entreprises de l'industrie manufacturière.

Les sources d'innovation sont indiquées dans le graphique 2. On retrouve ici une confirmation que la recherche et développement n'est pas le mode principal d'innovation. Parmi les innovateurs, la moitié déclarent ne pas recourir à une recherche et développement à temps plein. Les sources d'innovations les plus employées sont en fait les études techniques ou de méthodes, identifiées à la recherche informelle, suivies par l'utilisation novatrice de biens d'investissement et les matériaux et composants nouveaux. Les sources d'innovation les plus formalisées, comme la recherche et développement, les brevets et les licences concernent un nombre d'entreprises plus restreint, bien que plus élevé que dans les études antérieures.

Ce résultat ne signifie pas toutefois que la recherche et développement ne serait pas importante pour la croissance, car il se peut que les sources informelles ne mènent qu'à des innovations de faible ampleur (i.e., incrémentales) génératrice de peu de croissance. Pour répondre à cette question nous effectuerons des régressions logistiques qui nous permettront de relier les sources à leurs outputs. Nous verrons que les déterminants des innovations radicales se distinguent nettement des déterminants des innovations incrémentales.

³⁴ Il s'agit respectivement des questions « U1 », « U3 » et « U5 » du questionnaire.

Graphique 2 - Les sources de l'innovation



Répartition des réponses à chaque source selon son importance dans la réalisation des innovations
Échantillon de 3065 entreprises innovantes parmi un total de 4460.

Le problème de sélection

Un problème fréquemment rencontré dans les études sur l'innovation est celui du biais de sélection : un nombre non négligeable d'entreprises n'innovent pas selon certains critères. Ce biais est particulièrement fort sur les données traditionnelles où, selon les études, 10% à 15% seulement des entreprises font de la recherche ou déposent des brevets. Le problème se pose alors de savoir si ce biais de sélection affecte également les enquêtes sur l'innovation.

La réponse dépend de la définition de l'innovation qui est donnée dans l'enquête. On peut en distinguer deux types : l'enquête Innovation de 1990 (limitée à la France), qui définit l'innovation comme un changement significatif de produit, de procédé, de conditionnement (traité séparément), d'organisation ou de pratiques commerciales. Cette définition est très large et est donc peu susceptible de souffrir de biais de sélection.³⁵

³⁵ A l'opposé, les enquêtes communautaires (CIS : « Community Innovation Survey ») effectuées par la suite retiennent une définition plus étroite de l'innovation, qui se réduit souvent aux améliorations significatives de produits et de procédés. En l'état actuel des études empiriques, il est difficile de se prononcer sur les

L'implication de cette sélection sur les estimations est la suivante: seule les entreprises qui ont innové ont répondu aux questions sur les sources de leurs innovations. De ce fait, on tend à surestimer les inputs de la fonction d'innovation quand on se limite aux seuls répondants, donc à estimer une fonction d'innovation qui n'est valable que pour la sous-population qui a répondu à l'enquête, et non pour toute la population.

Avant toute chose, il nous faut donc regarder quel est le pourcentage de sélection dans l'enquête. Les calculs que nous avons effectués montrent que le problème est décroissant avec le degré d'opportunités technologiques de l'activité de l'entreprise: plus l'activité est innovante, moins le problème de sélection est important. Dans cette étude, on trouve qu'il est pratiquement nul pour les activités fortement innovantes (i.e., proche de 1%), de l'ordre de 10% dans les activités moyennement innovantes et atteint son maximum dans les activités faiblement innovantes, à 50%. A première vue, le dernier taux semble élevé, mais il faut garder à l'esprit que le taux correspondant aux données de recherche et de brevets est compris entre 85% et 90%.

Afin de tester l'hypothèse d'absence d'effet significatif de sélection, nous utiliserons la statistique de suridentification des moments généralisés. La raison pour laquelle l'utilisation de cette statistique est justifiée tient à la structure particulière de notre modèle:

- Le critère de sélection des répondants fait partie de notre modèle explicatif il s'agit des fonctions d'innovation, qui donnent la probabilité d'innover, c'est-à-dire de répondre au questionnaire. En fait, le modèle comprend plusieurs innovations et ceci complique un peu sa structure, mais un fait important demeure: en cas de problème de sélection, la perturbation de l'équation de croissance serait nécessairement corrélée aux sources de l'innovation.
- Nous utilisons justement les sources de l'innovation comme instruments de l'équation de croissance, de sorte que la validité de nos instruments dépend de l'absence de biais de sélection.³⁶

différences de résultats que pourraient produire ces deux types d'enquêtes, nous nous contenterons ici d'effectuer un test d'absence de biais de sélection sur l'enquête Innovation de 1990.

³⁶ Dans le cas où les perturbations de l'équation de croissance et des fonction d'innovations seraient sans corrélation, les estimations sont convergentes même en présence de sélection.

- Le test de suridentification teste justement la validité des instruments du modèle. Il est donc adapté, dans notre modèle, au problème de sélection.

Signalons dès maintenant que nous ne rejetons pas la validité des instruments des équations de croissance, de sorte que nous ne trouvons pas d'effet significatif de la sélection. Les statistiques sont données dans les tableaux d'estimations correspondants. Ce résultat était attendu pour les opportunités technologiques moyennes et fortes, en raison du grand nombre d'innovateurs, moins pour les activités à faibles opportunités technologiques. Cependant, nous avons également vu que le critère de sélection de l'enquête Innovation 1990 est le plus faible qui ait été retenu par les différentes enquêtes. Cette large définition de l'innovation explique sans doute le résultat que nous obtenons.³⁷

La croissance de la productivité totale des facteurs

Les tableaux 3 à 5 présentent une décomposition de la productivité totale des facteurs et la comparaison de ces mesures entre les entreprises innovantes et les autres. Tout d'abord, on remarque que les entreprises répondantes aux sources de l'innovation (tableau 4) ont les mêmes caractéristiques moyennes (tableau 3) que les autres entreprises, de ce fait le commentaire ne dépend pas du sous-échantillon considéré.³⁸ Les deux premiers tableaux permettent également de se faire une idée des effets liés à l'agrégation, car ils comparent la moyenne des réponses, qui correspond aux estimations sur données individuelles, à la réponse agrégée.

Sur l'ensemble de la période 1985-1990, la productivité apparente du travail a augmenté de 30%. Toutefois, cette augmentation peut provenir aussi bien d'un accroissement des facteurs de production (capital, travail) que de l'innovation. Dans le même temps, en effet le nombre moyen d'heures employées s'est accrue de 9% et le capital par heure travaillée de 50%. Une bonne part de la croissance vient donc des facteurs habituels. Il reste toutefois un résidu qui est inexplicé. En effet, sous l'hypothèse de rendements constants, dont nous verrons qu'elle n'est pas rejetée par la suite, la contribution du capital à la croissance de la productivité est

³⁷ Ce résultat suggère également une modification des futures enquêtes : ajouter systématiquement une variable d'innovation « faible », comme le conditionnement, afin de mesurer les sources de l'innovation dans les entreprises qui n'ont pas introduit de produit ou de procédé nouveau pendant la période d'enquête, et de réduire ainsi l'impact que pourrait avoir la sélection sur les études économétriques.

³⁸ Les entreprises répondantes aux sources de l'innovation sont celles ayant réalisé une innovation « au sens large » c'est-à-dire de produit, de procédé, de conditionnement, commerciale ou organisationnelle.

pondérée par le complément à 1 de la rémunération du travail dans la valeur ajoutée, soit environ 0.3. De ce fait, il reste une partie inexpliquée malgré la forte croissance de l'intensité capitaliste. Cette partie, la croissance de la productivité totale des facteurs est donnée dans les dernières lignes des tableaux 3 et 4. Pour l'industrie, elle est en moyenne de +26% sur 5 ans, soit 5.9% par an.³⁹ Cette part est suffisamment importante pour justifier une analyse plus détaillée.

Les effets de l'agrégation apparaissent clairement: par définition, elle accorde un poids plus grand aux grandes entreprises, de ce fait les résultats peuvent être différents. Ainsi, la part du coût du travail dans la valeur ajoutée est plus faible au niveau agrégé (i.e., 2/3) qu'au niveau microéconomique (i.e., 3/4) parce que les grandes entreprises ont une intensité capitaliste plus forte que les petites entreprises. De même les taux de croissance sont atténués parce que les grandes entreprises connaissent une croissance plus faible que les petites.⁴⁰

Le modèle que nous employons explique cette croissance de la productivité totale des facteurs par l'innovation. Une première approche consiste simplement à comparer les croissances moyenne et médiane selon que l'entreprise innove ou non. L'avantage de la médiane est qu'elle est moins sensible aux points aberrants qui pourraient éventuellement apparaître.⁴¹

Le test de comparaison des moyennes (tableau 5) met en évidence deux différences significatives dans les secteurs de fortes opportunités d'innovation.⁴² Le premier résultat est que la performance des entreprises qui ont réalisé des innovations radicales est significativement supérieure aux autres entreprises. Le second résultat indique que les entreprises qui ont réalisées des innovations incrémentales auraient une performance moyenne inférieure à celle des autres entreprises.

³⁹ Il y a une différence entre la définition habituelle de la croissance agrégée et celle que nous utilisons ici : nous agrégeons les performances individuelles, ce qui fait intervenir la somme des produits des parts du capital par les taux de croissance des capitaux, alors que dans la définition habituelle on effectue le produit de la part moyenne du capital par la croissance de l'agrégat. Ceci crée un écart.

⁴⁰ Notons que nos résultats vont dans le sens d'une invalidation de la « loi de Gibrat » selon laquelle le taux de croissance d'une entreprise serait indépendant du niveau qu'elle a atteint pour la variable étudiée. Cette étude suggère au contraire que les petites entreprises croissent plus vite que les grandes, et que les grandes entreprises finissent par atteindre un niveau de capital et d'emploi au delà duquel il n'est plus optimal pour elles de croître.

⁴¹ Nous avons examiné les distributions des taux de croissance et éliminé les points isolés en queue de distribution conformément à la pratique habituelle. Les observations éliminées se détachaient nettement de l'ensemble de la distribution.

⁴² Nous commentons ici l'ensemble des entreprises répondantes aux sources de l'innovation, désignées comme innovantes au sens large. Les résultats sont comparables pour l'ensemble des entreprises.

Ces résultats suggèrent donc deux propriétés: la première propriété serait que les gains de croissance dépendent du degré d'opportunités technologiques de l'activité de l'entreprise; la seconde serait que seule l'innovation radicale serait de nature à améliorer la productivité totale des facteurs.

Ces résultats se heurtent toutefois à la critique suivante: le raisonnement en moyenne ne permet pas de distinguer l'effet de l'innovation séparément des effets des innovations passées et de la performance moyenne du secteur d'activité. De plus, on ne peut pas distinguer clairement l'effet de l'innovation incrémentale quand l'innovation radicale est constante⁴³. Pour pouvoir séparer ces trois effets, il faut recourir à des régressions. C'est l'objet de la section suivante.

Tableau 3 - Statistiques descriptives sur l'ensemble des entreprises

Les taux de croissance sont mesurés par les écarts des logarithmes de la période 1985-1990.
Les agrégats sont obtenus en prenant une moyenne pondérée.
En francs de 1985

Opportunités d'innovation	Faibles	Moyennes	Fortes	Ensemble
Capital utilisé par heure travaillée				
- moyenne	93	110	115	102
Nombre d'heures employées (milliers)				
- moyenne	186	273	661	272
Productivité apparente du travail				
- moyenne	113	124	138	120
Salaires + Charges / Valeur ajoutée 1985				
- moyenne	75.4%	73.4%	71.9%	74.2%
- agrégat	71.4%	68.6%	60.8%	66.5%
Croissance du capital utilisé par heure travaillée				
- moyenne	48.8%	51.9%	52.3%	50.4%
- agrégat	22.3%	23.8%	26.4%	23.4%
Croissance du nombre d'heures employées				
- moyenne	6.5%	9.5%	15.5%	8.6%
- agrégat	0.8%	0.0%	-0.5%	0.1%
Croissance de la productivité apparente du travail				
- moyenne	29.9%	30.5%	30.1%	30.1%
- agrégat	28.5%	29.0%	27.0%	28.5%

⁴³ Le problème vient de l'existence d'une corrélation entre les deux dans les activités de faible et moyennes opportunités technologiques.

Croissance de la productivité totale des
facteurs

- moyenne	26.0%	25.7%	22.9%	25.6%
- agrégat	12.1%	9.5%	-12.2%	7.8%
Nombre d'observations	2290	1677	493	4460

*Tableau 4 - Statistiques descriptives sur les entreprises innovantes
au sens large*

Les taux de croissance sont mesurés par les écarts des logarithmes de la période 1985-1990.
Les agrégats sont obtenus en prenant une moyenne pondérée.
En francs de 1985

Opportunités d'innovation	Faibles	Moyennes	Fortes	Ensemble
Capital utilisé par heure travaillée - moyenne	104	109	116	108
Nombre d'heures employées (milliers) - moyenne	237	290	673	332
Productivité apparente du travail - moyenne	116	125	138	124
Salaires + Charges / Valeur ajoutée 1985 - moyenne	74.3%	73.1%	71.9%	73.3%
- agrégat	70.0%	68.6%	60.8%	65.6%
Croissance du capital utilisé par heure travaillée - moyenne	49.8%	51.4%	52.9%	51.1%
- agrégat	22.7%	25.6%	26.4%	24.8%
Croissance du nombre d'heures employées - moyenne	7.3%	9.7%	15.5%	9.8%
- agrégat	2.1%	-0.0%	-0.1%	0.3%
Croissance de la productivité apparente du travail - moyenne	30.3%	30.4%	30.1%	30.3%
- agrégat	28.7%	29.0%	27.0%	7.4%
Croissance de la productivité totale des facteurs - moyenne	26.7%	25.4%	22.7%	25.4%
- agrégat	13.0%	9.0%	-12.8%	6.2%
Nombre d'observations	1081	1501	483	3065

*Tableau 5 – Innovation et croissance de la productivité totale des facteurs
Comparaison des moyennes*

Les taux de croissance sont mesurés par les écarts des logarithmes de la période 1985-1990.
En francs de 1985

Opportunités d'innovation	Faibles	Moyennes	Fortes	Ensemble
<i>Ensemble des entreprises</i>				
- Innovation radicale	26.5%	24.4%	25.5%	25.2%
- Autres entreprises	25.9%	27.0%	14.5%	25.8%
- Test d'égalité des moyennes	0.26	-1.22	2.19*	-0.41
- Test de la médiane	-0.65	-0.61	1.78	-0.53
- Innovation incrémentale	27.6%	24.7%	15.1%	24.8%
- Autres entreprises	25.1%	27.0%	29.4%	26.1%
- Test d'égalité des moyennes	1.32	-1.08	-3.45*	-1.00
- Test d'égalité des médianes	1.17	-1.36	-2.13*	-0.75
<i>Entreprises innovantes au sens large</i>				
- Innovation radicale	26.5%	24.4%	25.5%	25.2%
- Autres entreprises	26.8%	26.7%	12.6%	25.7%
- Test d'égalité des moyennes	-0.13	-1.01	2.59*	-0.32
- Test d'égalité des médianes	-0.72	-0.33	2.07*	-0.23
- Innovation incrémentale	27.6%	24.7%	15.1%	24.8%
- Autres entreprises	23.9%	26.7%	29.2%	26.6%
- Test d'égalité des moyennes	1.23	-0.82	-3.40*	-1.05
- Test d'égalité des médianes	1.82	-1.05	-2.15*	-0.54

Econométrie

Le système suivant fournit une vue d'ensemble du modèle⁴⁴ :

$$\begin{cases} \frac{\dot{q}}{q} - (1 - p_L) \frac{\dot{k}}{k} = \sum_{j=1}^J a_j I_j + u \\ I_j = \mathbb{1}(I_j^* > 0) \\ I_j^* = Zc_j + u_j \quad j = 1, \dots, J \end{cases}$$

On estime ce modèle en deux étapes. Dans une première étape, on estime un modèle *logit* sur l'équation d'innovation, qui explique la probabilité de réaliser une innovation en fonction des variables Schumpéteriennes et des sources de l'innovation. Dans un second temps, la prévision de la probabilité est prise comme variable explicative de l'équation de productivité. Comme on remplace la véritable variable par sa prévision issue du modèle *logit*, on doit calculer des écarts-types robustes à l'hétéroscédasticité⁴⁵

Une seconde méthode d'estimation consiste à estimer directement le modèle de croissance par la méthode des moments généralisés. On instrumente alors l'innovation par ses sources, qui est un bon instrument ici. En effet, il faut trouver des variables instrumentales qui soient corrélées aux outputs de l'innovation, et qui ne soient pas corrélées à la perturbation de l'équation de croissance. Par définition, les innovations sont corrélées à leurs sources, ce qui sera mis en évidence par les estimations du modèle *logit*, et l'effet des sources de l'innovation ne peut passer que la mise en œuvre des innovations dans l'entreprise, ce qui implique qu'elles ne sont pas reliées à la perturbation de l'équation de croissance. Dans la pratique, il se pourrait malgré tout que le résidu reste corrélé aux instruments. Pour s'assurer que les instruments sont valides, on effectuera donc un test de Sargan.

⁴⁴ Nous verrons plus loin que l'on ne rejette pas l'hypothèse de rendements constants, de sorte que cette formulation simplifiée est valable. Le symbole $\mathbb{1}(I_j^* > 0)$ indique une variable égale à 1 si I_j^* est positif, et égal à 0 sinon.

⁴⁵ Une autre méthode consiste à évaluer explicitement la forme de la variance de la perturbation du nouveau modèle et à effectuer des moindres carrés quasi-généralisés. Mais ceci supposerait que l'on fasse des hypothèses supplémentaires sur la perturbation de la première équation. Notre méthode est donc plus robuste.

Résultats

La fonction d'innovation

L'analyse des différents types d'innovations montre que les entreprises n'utilisent pas les mêmes sources de connaissances pour réaliser des innovations incrémentales et radicales. Les résultats de l'estimation du modèle *logit* sont présentés dans le tableau 6.⁴⁶ Nous les présentons par classe d'opportunités technologiques croissante afin de détecter un éventuel passage en continu des innovations incrémentales des activités à fortes opportunités technologiques aux innovations radicales des activités à faibles opportunités technologiques. Un examen du tableau montre que cette continuité ne semble pas avoir lieu : il existe des différences marquées entre les déterminants des innovations incrémentales et radicales, de sorte que les innovations radicales n'apparaissent pas comme le prolongement des innovations incrémentales.

Les innovation incrémentales sont réalisées par les entreprises qui, dans l'ensemble, font le moins de recherche formelle et qui recourent le moins aux brevets et aux licences. Au contraire, ces mêmes entreprises utilisent très fortement les biens d'investissements, et dans une moindre mesure la recherche externe. L'innovation incrémentale apparaît donc comme le résultat d'une diffusion des biens d'équipement vers les différentes industries suppléée par des études techniques et de la recherche externe, ce qui ressemble à une stratégie d'adoption de nouvelles technologies.

L'innovation radicale repose sur des déterminants complètement différents. Les sources les plus importantes sont clairement la recherche et développement formalisée combinée aux brevets et, dans une moindre mesure, aux licences. Ce savoir est complété par l'utilisation des matériaux et composants nouveaux, qui proviennent du secteur des biens intermédiaires. Enfin, ces entreprises recourent également à un appoint de recherche externe (modalité « faible »).

On observe donc une certaine dualité en matière d'innovation entre les entreprises qui réalisent des innovations radicales et les autres. Le premier groupe d'entreprises correspond assez bien à la vision en termes de recherche et de brevets des premières études empiriques où

⁴⁶ L'estimation est réalisée en utilisant la procédure « logistique » de SAS (avec option « descending » pour avoir la probabilité que $I=1$). Sur la définition du R^2 latent employé, voir Barlet et al. (1998). Pour une discussion plus approfondie, voir Windmeijer (1995).

un nombre limité d'entreprises effectuerait l'essentiel des innovations⁴⁷ Le second groupe d'entreprises, qui réalise des innovations incrémentales, repose plus largement sur des connaissances externes car ces innovations apparaissent essentiellement comme des utilisations novatrices de biens d'investissements.

Dans l'ensemble on constate également que les externalités jouent un rôle significatif. Les brevets, dont il n'est pas précisé dans l'enquête s'ils ont été entièrement réalisés par l'entreprise, achetés ou mis au point en collaboration, jouent un rôle de premier plan dans l'élaboration des innovations radicales.⁴⁸ De même, les licences et les matériaux et composants nouveaux jouent un rôle significatif sur ces mêmes innovations. Mais les externalités touchent également les innovations de faible ampleur. Dans ce dernier cas, c'est l'industrie des biens d'équipement qui semble être à l'origine des améliorations de produit et de procédé de l'ensemble de l'industrie.

Ces résultats sur les externalités méritent d'être rapproché du graphique 1, où l'on voit que les activités les plus innovantes se trouvent souvent dans les industries des biens d'équipement. De plus, nous avons vu dans le tableau 2 que les activités à fortes opportunités technologiques réalisent le plus d'innovations radicales. Ainsi, les deux types d'innovations que nous étudions sont reliées.

D'un côté, l'industrie des biens d'équipement réalise des innovations radicales grâce à l'appui de l'industrie des biens intermédiaires et d'un processus d'innovation formalisé, puis elle revend ses innovations radicales aux autres entreprises qui réalisent à leur tour des innovations incrémentales.

Il nous reste maintenant à savoir quelles sont les performances auxquelles ces deux types d'innovations permettent de parvenir.

⁴⁷ Le nombre d'entreprises est toutefois plus élevé que dans les études sur la recherche et les brevets.

⁴⁸ Pour une étude des co-brevets, voir Duguet (1993).

Tableau 6 – Les fonctions d'innovation

Echantillon de 3065 entreprises dans l'industrie manufacturière. Estimation d'un modèle *logit* par le maximum de vraisemblance (les écarts-types asymptotiques sont donnés entre parenthèses).

* : significatif au seuil de 5% ; + : significatif au seuil de 10%

Type d'innovation (variable expliquée)	Incémentale			Radicale		
	Faibles	Moyennes	Fortes	Faibles	Moyennes	Fortes
Opportunités technologiques						
<i>Chiffre d'affaires 1985 (logarithme)</i>	-.03 (.10)	-.21* (.07)	.04 (.12)	.12 (.09)	.23* (.07)	.04 (.15)
<i>Part de marché 1985 (logarithme)</i>	-.18* (.08)	.04 (.06)	-.09 (.10)	0.12+ (.07)	.06 (.06)	.10 (.12)
<i>Diversification 1985 (logarithme)</i>	.40 (.26)	.04 (.19)	-.21 (.34)	.16 (.22)	.22 (.19)	.16 (.45)
<i>Recherche interne (référence = non)</i>						
- un peu	-.70* (.25)	-.64* (.20)	-.13 (.52)	.61* (.22)	.98* (.20)	-.25 (.56)
- moyennement	-.16 (.27)	-.02 (.18)	-.22 (.38)	-.02 (.24)	.41* (.16)	.72 (.47)
- beaucoup	-.15 (.26)	-.50* (.15)	-.25 (.26)	.30 (.23)	.76* (.15)	.64* (.31)
<i>Recherche du groupe (référence = non)</i>						
- un peu	.09 (.30)	.18 (.23)	.07 (.49)	.16 (.26)	-.20 (.24)	.46 (.75)
- moyennement	-.06 (.31)	.01 (.21)	.10 (.40)	-.09 (.28)	.06 (.21)	-.40 (.47)
- beaucoup	.02 (.34)	.25 (.19)	.26 (.30)	.10 (.31)	-.16 (.20)	-.06 (.40)
<i>Etudes techniques (référence = non)</i>						
- un peu	-.02 (.22)	.39+ (.21)	.86+ (.46)	.01 (.19)	-.02 (.19)	-.80 (.52)
- moyennement	.03 (.23)	-.02 (.18)	-.12 (.35)	.25 (.20)	.48* (.18)	-.18 (.42)
- beaucoup	-.01 (.24)	-.12 (.18)	-.06 (.30)	.63* (.20)	.58* (.18)	.12 (.36)
<i>Recherche externe (référence = non)</i>						
- un peu	.27 (.21)	.33* (.15)	-.52+ (.28)	-.13 (.17)	-.20 (.15)	1.22* (.40)
- moyennement	-.35 (.25)	.11 (.17)	1.11* (.33)	-.08 (.23)	-.17 (.17)	-.61+ (.35)
- beaucoup	-.09 (.34)	.47 (.31)	-.32 (.36)	.44 (.29)	-.22 (.28)	.45 (.45)
<i>Brevets détenus (référence = non)</i>						
- un peu	-.32 (.24)	-.26 (.17)	-1.04* (.37)	.75* (.22)	.59* (.18)	.60 (.49)
- moyennement	-.30 (.36)	-.81* (.19)	-1.05* (.36)	.65* (.33)	.68* (.21)	1.74* (.062)
- beaucoup	-1.46* (.43)	-.61* (.22)	-1.49* (.32)	.97* (.44)	1.06* (.26)	1.72* (.49)
<i>Licences acquises à l'extérieur (réf. = non)</i>						
- un peu	-.50+ (.27)	-.17 (.24)	.34 (.33)	.41 (.25)	-.04 (.21)	-.47 (.44)
- moyennement	.07 (.39)	-.21 (.24)	-.21 (.42)	.49 (.33)	.03 (.25)	.57 (.59)
- beaucoup	-.92+ (.48)	-.21 (.32)	-.62 (.48)	.85+ (.48)	.80* (.37)	.68 (.61)

(suite page suivante)

(suite du tableau 6)

Type d'innovation	Incrémentale			Radicale		
	Faibles	Moyenne s	Fortes	Faibles	Moyenne s	Fortes
<i>Opportunités technologiques</i>						
<i>Utilisation novatrice de biens d'investissement (réf. = non)</i>						
- un peu	.72* (.22)	.86* (.18)	.24 (.34)	.16 (.26)	-.46* (.19)	.04 (.42)
- moyennement	.97* (.21)	1.10* (.17)	.68* (.33)	-.09 (.28)	-.56* (.18)	.06 (.42)
- beaucoup	1.22* (.26)	1.12* (.18)	.66* (.31)	.10 (.31)	-.72* (.18)	-.64* (.36)
<i>Matériaux et composants nouveaux (réf. = non)</i>						
- un peu	.32 (.21)	.09 (.17)	-.12 (.36)	.05 (.18)	.17 (.16)	-.50 (.44)
- moyennement	-.33 (.22)	-.25 (.16)	.15 (.30)	.36* (.19)	.42* (.16)	-.52 (.36)
- beaucoup	-.37 (.28)	-.44* (.19)	.36 (.29)	.75* (.25)	.39* (.19)	-.14 (.35)
<i>Dynamique propre de la technologie (réf. = nulle)</i>						
- faible	-.05 (.21)	-.05 (.21)	.72 (.49)	.25 (.19)	.02 (.20)	-.33 (.66)
- moyenne	.14 (.22)	.01 (.19)	.44 (.42)	.18 (.19)	-.13 (.19)	-.63 (.57)
- forte	-.11 (.26)	-.12 (.21)	.13 (.41)	.55* (.23)	-.03 (.20)	-.34 (.57)
<i>Pression de la demande (réf. = nulle)</i>						
- faible	1.29* (.32)	.32 (.33)	.22 (.59)	-.44 (.30)	.32 (.33)	.67 (.72)
- moyenne	.77* (.27)	.32 (.29)	-.11 (.51)	-.04 (.27)	.44 (.29)	.29 (.58)
- forte	.92* (.27)	.20 (.28)	-.05 (.44)	-.15 (.26)	.39 (.28)	.05 (.49)
Terme constant	.05 (.27)	-.05 (.30)	-.66 (.59)	-1.00* (.27)	-.58 (.31)	1.23* (.73)
Biens d'équipement	-.09 (.21)	.20 (.14)	.07 (.25)	.05 (.18)	.10 (.14)	.22 (.32)
Biens de consommation	-.11 (.17)	.19 (.15)	.83* (.27)	.07 (.15)	-.08 (.14)	-.29 (.31)
Log-vraisemblance	-538.52	-880.78	-282.73	-654.74	-892.76	-203.34
R ² du modèle latent	0.408	0.352	0.533	0.392	0.475	0.657
R ² de Mac Fadden	0.104	0.085	0.151	0.099	0.127	0.196
Pourcentage de prévisions concordantes	72%	69%	76%	70%	73%	80%
Nombre d'entreprises	1081	1501	483	1081	1501	483

N.B. : les variables de chiffre d'affaires, de part de marché et de diversification sont centrées.

L'équation de croissance

La diffusion des connaissances s'accompagne-t-elle de gains de croissance significatifs pour les innovations incrémentales ou les entreprises productrices d'innovation radicales arrivent-elles à capter la totalité des gains associés à leurs innovations via la vente des produits qui y sont associés ?

La réponse à cette question revient à s'interroger sur la manière donc les biens d'investissements interviennent sur la productivité globale des facteurs. Ils interviennent d'une part via un accroissement de la productivité et d'autre part comme un accroissement de capital, de sorte que la résultante est un effet net dont le signe est indéterminé. On voit donc que l'étude des fonctions d'innovation amène à étudier l'innovation incrémentale sous un aspect différent : son effet *net* peut très bien être nul si l'accroissement de productivité est compensé par un accroissement de capital. Notons également que cette remarque vaut pour l'innovation radicale dans la mesure où elle implique des investissements. Toutefois, dans ce dernier cas, l'innovation vient plus de l'entreprise elle-même que d'un secteur qui vend ses innovations, de ce fait les innovations radicales pourraient mener à une rémunération supplémentaire propre aux innovateurs.

Les résultats sont donnés dans le tableau 7 qui compare trois méthodes d'estimation. La première méthode, les moindres carrés ordinaires, n'est donnée qu'à titre indicatif car elle est biaisée et non convergente sous les hypothèses de notre modèle. La deuxième méthode est celle dite « en deux étapes » qui tient compte d'une part, du caractère qualitatif de l'innovation et d'autre part de la simultanéité de la croissance et de l'innovation (i.e., l'innovation est endogène).⁴⁹ Enfin, la troisième méthode, celle des moments généralisés, suppose que l'innovation est endogène mais pas qu'elle est de type *logit* : elle impose simplement que l'innovation a un effet constant sur la croissance. Il s'agit de l'estimation la plus robuste⁵⁰

⁴⁹ Dans ce modèle l'innovation est endogène, prise en même temps que les décisions de production.

⁵⁰ L'approche par le modèle *logit* présente l'avantage par rapport aux moments généralisés d'autoriser l'étude des déterminants de l'innovation, qui joue un rôle important dans l'interprétation des résultats. De plus, elle permet de tester la validité du modèle à effet constant (les tests non emboîtés sont présentés dans le tableau 1). Toutefois, ce test est réalisé sous l'hypothèse d'un modèle *logit*, ce qui limite sa portée. L'approche par les moments généralisés est donc plus robuste mais ne propose pas d'estimation des fonctions d'innovation. Ajoutons enfin que les deux méthodes donnent des résultats proches.

Dans l'ensemble, il apparaît que la méthode d'estimation choisie est cruciale pour mettre en évidence les effets de l'innovation sur la croissance. Les moindres carrés ordinaires, biaisés, *sous-estiment* systématiquement l'effet de l'innovation sur la croissance de la productivité totale des facteurs. Le seul effet significatif est trouvé pour les activités à fortes opportunités d'innovation. A l'opposé, les méthodes en deux étapes et des moments généralisés fournissent des résultats cohérents.

Le premier résultat est que seule l'innovation radicale a un effet significatif sur la productivité totale des facteurs, dans toutes les classes d'opportunités technologiques (tableau 7). Ainsi, les améliorations de produits et de procédé ne procureraient pas de gain additionnel, ce qui suggère que les innovateurs des secteurs des biens d'équipement seraient en mesure de capter les bénéfices de leurs innovations, mesuré par les gains de productivité des entreprises clientes.

Le deuxième résultat porte sur l'importance des gains réalisés: elle est croissante avec les opportunités d'innovation de l'activité qu'exerce d'entreprise (tableaux 8 et 9). Pour l'ensemble de l'industrie, le gain de croissance annuel moyen associé à l'innovation radicale est de 4.7%, mais varie de 3% pour les activités de faibles opportunités technologiques à 8.7% pour les activités où les opportunités technologiques sont fortes. Ainsi, on retrouve une extension aux taux de croissance du résultat trouvé par Barlet et al. (1997) sur la part des produits de moins de 5 ans dans le chiffre d'affaires.

On trouve donc bien un effet positif des innovations mises en œuvre, nouveaux produits ou premières de procédé technologique, sur la croissance de la productivité totale des facteurs.

Le modèle qui se dégage de ces estimations séparerait donc d'une part, les entreprises qui produisent des innovations radicales et obtiennent des gains significatifs de croissance et d'autre part, les entreprises qui basent leur croissance sur l'utilisation novatrice de biens d'investissement et qui n'obtiendraient pas de gain additionnel à ceux procurés par les biens d'équipement.

Le troisième résultat porte sur le taux de dépréciation des innovations radicales. Quelque soit le degré d'opportunités technologiques, il resterait autour de 10% par an. Ceci implique également que la performance des entreprises tendrait à converger vers un point d'équilibre qui est fonction du simple fait que l'entreprise a effectué des innovations radicales ou non. Dans tous les cas on rejette le modèle de marche aléatoire. Notons que cela renforce la dualité

précédente, en ce sens que non seulement la performance moyenne des innovations radicales est différente de celle des innovations incrémentales mais également que le *potentiel de croissance* des innovations radicales est supérieur.

Enfin, nous avons également fait les tests de constance des rendements d'échelle (tableau 9). On ne rejette jamais l'hypothèse de rendements constants au seuil de 5%. Par contre, on pourrait rejeter cette hypothèse au seuil de 10% dans les activités de fortes opportunités technologiques. Les rendements seraient alors croissants, toutefois, le coefficient de l'innovation radicale ne semble pas affecté par l'introduction de cette caractéristique, de sorte que nos résultats sont robustes à cette hypothèse.

Globalement, on peut conclure à un effet significatif de l'innovation sur la croissance, qui serait de l'ordre de 5% par an pour les entreprises qui ont introduit des innovations radicales et nul pour les autres entreprises. Le «résidu de Solow» s'expliquerait donc par les innovations radicales sur la période 1986-1990. On peut dès lors s'interroger sur les conséquences agrégées de ce résultat microéconomique.

Tableau 7 - Croissance de la productivité totale des facteurs en fonction de deux types d'innovation

Variable expliquée :

écarts des logarithmes de la productivité totale des facteurs sur la période 1985-1990.

Opportunités d'innovation	Faibles	Moyennes	Fortes	Ensemble
Moindres carrés ordinaires				
- productivité totale des facteurs en 1985 (β)	-.362* (.047)	-.431* (.018)	-.457* (.031)	-.425* (.012)
- innovation radicale	.021 (.028)	.021 (.023)	.092* (.050)	.026 (.016)
- innovation incrémentale	.025 (.031)	-.029 (.024)	-.088* (.041)	-.027 (.017)
- biens d'équipements	.051 (.030)	.011 (.023)	.042 (.040)	.029 (.016)
- biens de consommation	-.061* (.025)	-.047* (.023)	-.002 (.044)	-.050 (.016)
- terme constant	-.362* (.047)	-.352* (.041)	-.436* (.073)	-.355* (.028)
R^2	.262	.275	0.334	.278
Méthode en deux étapes				
- productivité totale des facteurs en 1985 (β)	-.411* (.026)	-.441* (.020)	-.474* (.037)	-.438* (.013)
- innovation radicale	.103 (.081)	.207* (.074)	.287* (.136)	.199* (.057)
- innovation incrémentale	-.092 (.091)	.058 (.094)	-.092 (.127)	-.011 (.073)
- biens d'équipement	.048 (.030)	.001 (.023)	.024 (.043)	.014 (.017)
- biens de consommation	-.062* (.025)	-.043 (.023)	.006 (.044)	-.044 (.016)
- terme constant	-.319* (.104)	-.529* (.106)	-.605* (.160)	-.476* (.079)
R^2	.266	0.279	0.331	.284
Moments généralisés				
Statistique de Sargan (seuil de significativité)	18.68 (0.91)	27.36 (0.49)	32.99 (0.24)	27.33 (0.50)
- productivité totale des facteurs en 1985 (β)	-.403* (.025)	-.451* (.022)	-.452* (.036)	-.437* (.015)
- innovation radicale	.095 (.086)	.259* (.084)	.399* (.176)	.212* (.064)
- innovation incrémentale	-.061 (.098)	.131 (.103)	.003 (.149)	.014 (.078)
- biens d'équipement	.049 (.029)	-.001 (.023)	.007 (.042)	.015 (.017)
- biens de consommation	-.055* (.024)	-.045 (.023)	.020 (.044)	-.042* (.016)
- terme constant	-.329* (.105)	-.622* (.122)	-.713* (.206)	-.501* (.088)

* significatif au seuil de 5%. + significatif au seuil de 10%.

Tableau 8 - Croissance de la productivité totale des facteurs en fonction

de l'innovation radicale

Variable expliquée :

écarts des logarithmes de la productivité totale des facteurs sur la période 1985-1990.

Opportunités d'innovation	Faibles	Moyennes	Fortes	Ensemble
Moindres carrés ordinaires				
- productivité totale des facteurs en 1985 (β)	-.402* (.021)	-.431* (.018)	-.459* (.031)	-.425* (.012)
- innovation radicale	.009 (.023)	.037* (.019)	.149* (.042)	.041* (.013)
- biens d'équipements	.051* (.030)	.011 (.023)	.042 (.041)	.029* (.016)
- biens de consommation	-.061* (.025)	-.048* (.023)	-.015 (.044)	-.051* (.016)
- terme constant	-.339 (.037)	-.381* (.034)	-.520* (.062)	-.380* (.023)
R ²	.262	.275	.328	.277
Méthode en deux étapes				
- productivité totale des facteurs en 1985 (β)	-.409* (.021)	-.441* (.020)	-.475* (.037)	-.438* (.015)
- innovation radicale	.149* (.065)	.172* (.048)	.366* (.089)	.206* (.034)
- biens d'équipement	.048 (.30)	.002 (.023)	.022 (.043)	.014 (.017)
- biens de consommation	-.060* (.025)	-.042* (.023)	-.007 (.039)	-.044* (.015)
- terme constant	-.405* (.054)	-.471* (.048)	-.705* (.090)	-.487* (.030)
R ²	0.266	0.278	0.330	.284
Moments généralisés				
Statistique de Sargan (seuil de significativité)	19.01 (0.92)	29.42 (0.44)	32.93 (0.28)	27.38 (0.55)
- productivité totale des facteurs en 1985 (β)	-.402* (.025)	-.445* (.021)	-.452* (.036)	-.437* (.015)
- innovation radicale	.126* (.069)	.176* (.052)	.396* (.094)	.202* (.036)
- biens d'équipement	.048* (.029)	0.001 (.023)	.007 (.041)	.015 (.017)
- biens de consommation	-.054* (.025)	-.041* (.023)	.020 (.038)	-.041* (.016)
- terme constant	-.387* (.051)	-.483* (.053)	-.710* (.090)	-.486* (.035)

* significatif au seuil de 5%. + significatif au seuil de 10%.

*Tableau 9 – Implications des estimations du modèle à innovation radicale
(à partir des estimations par les moments généralisés du tableau 8)*

Opportunités d'innovation	Faibles	Moyennes	Fortes	Ensemble
- gain de croissance 1985-90	12.6%	17.6%	39.6%	20.2%
- annuel	3.0%	4.1%	8.7%	4.7%
- dépréciation des innovations	44.2%	44.5%	45.2%	43.7%
- annuel	9.6%	9.6%	9.8%	9.5%

*Tableau 10 – Tests de constance de rendements d'échelle
(estimations par les moments généralisés)*

Variable expliquée :

écarts des logarithmes de la productivité totale des facteurs sur la période 1985-1990.

On donne juste les coefficients des innovations ainsi que de la variation du logarithme du capital 1985-90 et du logarithme du capital en 1985. Le logarithme du capital en 1985 a été ajouté aux instruments pour améliorer l'efficacité asymptotique des estimateurs.

Opportunités d'innovation	Faibles	Moyennes	Fortes
Avec innovations radicale			
Statistique de Sargan (seuil de significativité)	19.29 (0.89)	27.62 (0.48)	31.18 (0.31)
- innovation radicale	.125* (.070)	.178* (.054)	.421* (.103)
- $\Delta \ln k$	-.015 (.126)	.129 (.117)	.219* (.129)
- $\ln k$ (1985)	-.009 (.045)	.049 (.041)	.061 (.047)
Avec innovations radicale et incrémentale			
Statistique de Sargan (seuil de significativité)	18.93 (0.87)	26.76 (0.48)	31.34 (0.26)
- innovation radicale	.090 (.089)	.240* (.086)	.399* (.180)
- innovation incrémentale	-.067 (.099)	.102 (.109)	-.020 (.151)
- $\Delta \ln k$	-.005 (.127)	.085 (.126)	.221* (.130)
- $\ln k$ (1985)	-.004 (.045)	.035 (.044)	.062 (.048)

* significatif au seuil de 5%. + significatif au seuil de 10%.

Agrégation des relations individuelles de croissance

L'agrégation des relations de croissance se fait simplement en remarquant que le taux de croissance d'un agrégat est une moyenne pondérée des taux de croissance individuels. Ainsi

$$\sum_{i=1}^n \left[\frac{\theta_{it-1}}{\sum_{j=1}^n \theta_{jt-1}} \right] \times \frac{\theta_{it} - \theta_{it-1}}{\theta_{it-1}} = \frac{\sum_{i=1}^n \theta_{it} - \sum_{i=1}^n \theta_{it-1}}{\sum_{i=1}^n \theta_{it-1}}$$

On peut donc calculer l'impact sur le taux de croissance agrégé des innovations radicales en tenant compte de toutes les entreprises. Pour cela nous remarquons que comme le modèle comporte un effet constant de l'innovation, le coefficient de l'innovation radicale doit être pondéré par la part de la productivité totale des facteurs des entreprises correspondantes en 1985. Cette part est donnée dans le tableau 11 ainsi que la croissance correspondante.

Dans l'ensemble, on voit que la part des innovateurs radicaux est croissante avec le degré d'opportunités technologiques de l'activité exercée : de 39% à près de 84%. Sur l'ensemble de l'échantillon elle est de 58%. Il s'ensuit que l'impact sur la croissance serait de 2.8% par an sur la période 1985-90 pour l'ensemble des entreprises industrielles. Cette croissance est toutefois inégalement répartie selon les activités : de 1.2% par an lorsque les opportunités technologiques sont faibles à 7.2% par an lorsqu'elles sont fortes.

Tableau 11 – Impact de l'innovation radicale sur la croissance agrégée
(à partir des estimations par les moments généralisés)

Opportunités d'innovation	Faibles	Moyennes	Fortes	Ensemble
- pondération des innovateurs radicaux	0.395	0.620	0.814	0.584
- gain de croissance agrégé 1985-90	5.0%	10.9%	32.2%	11.8%
- annuel*	1.2%	2.6%	7.2%	2.8%

* tient compte des entreprises qui n'innovent pas.

Conclusion

Ce rapport permet de dégager un certain nombre de nouveaux résultats sur le lien entre l'innovation et la croissance des entreprises industrielles. Premièrement, les brevets et les licences, souvent critiqués sur la base de leur faible représentativité, jouent en fait un rôle important dans la promotion de la croissance. La principale raison tient à ce que les brevets et les licences font partie des déterminants principaux des innovations radicales, qui sont elles-mêmes les seuls déterminants significatifs de la croissance de la productivité globale des facteurs. De ce fait, nos résultats suggèrent que le droit de la propriété industrielle contribue à diffuser un savoir pertinent pour la croissance dans l'industrie manufacturière.

Le second résultat important est que seules les innovations radicales permettent effectivement d'accroître la productivité globale des facteurs au niveau des entreprises. Ceci ne signifie pas que les innovations incrémentales, basées sur l'utilisation novatrice de biens d'équipement, ne contribuent pas à la croissance, mais que ce sont ceux qui produisent ces biens d'équipement qui en seraient les principaux bénéficiaires. Dans l'ensemble, les innovations incrémentales demeurent importantes pour deux raisons: d'une part, elles contribuent à diffuser de meilleures technologies dans le tissu productif; d'autre part, elles fournissent une clientèle aux entreprises qui innovent le plus, ce qui soutient l'incitation à innover des innovateurs radicaux. Toutefois, ce n'est pas aux simples utilisateurs des nouvelles technologies que les innovations profiteraient le plus.

Dans l'ensemble, ce rapport suggère le fonctionnement suivant de l'industrie manufacturière: les entreprises qui font de la recherche formelle, et recourent aux droits de propriété industriels, produisent un savoir technologique pertinent qui se diffuse dans les entreprises moins innovantes. Ces entreprises faiblement innovantes créent une demande pour les entreprises fortement innovantes, mais ne retirent pas de bénéfice net lié à l'utilisation des nouvelles technologies. Tout ceci favoriserait à la fois la diffusion des nouvelles technologies et la production du savoir pertinent par un ensemble restreint de firmes. Nous trouvons donc une dualité innovation-croissance.

Annexe A : estimation du modèle en deux étapes

On souhaite estimer un modèle comprenant une relation linéaire, l'équation de croissance, et une relation de type logit, la fonction d'innovation. Sans perte de généralité, nous considérons le cas où un seul type d'innovation intervient comme déterminant de la croissance. L'équation de croissance peut s'écrire:

$$y = x b + a I + u$$

où y désigne la croissance de la productivité totale des facteurs, I l'indicateur d'innovation endogène et x désigne les variables explicatives (y compris la valeur décalée de la productivité totale des facteurs). L'erreur du modèle est notée u . La variable I ne peut prendre que deux valeurs : 0 si l'entreprise n'a pas innové, 1 sinon. Si l'innovation était exogène on pourrait estimer le modèle par les moindres carrés ordinaires, mais ici les décisions d'innovation et la performance sont déterminées en même temps. Il faut donc remplacer I par une variable qui soit sans corrélation avec la perturbation, ou l'instrumenter. La solution à ce problème est fournie par la seconde équation du modèle:

$$I^* = zc + v,$$

où I^* désigne le niveau latent d'innovation, z désigne les sources de l'innovation, c leurs coefficients et v la perturbation de la fonction d'innovation. Sous l'hypothèse que v suit une loi logistique on estime ce modèle par le maximum de vraisemblance à partir de la variable dichotomique I . Ceci permet de réécrire l'équation de croissance comme :

$$y = x b + a E(I | z) + u + a (I - E(I | z)) = x b + a E(I | z) + w$$

Selon le modèle logit, l'espérance mathématique de I , qui n'est autre que la probabilité d'innover, est donnée par :

$$E(I | z) = \Pr\{I = 1\} = \frac{\exp(zc)}{1 + \exp(zc)}$$

La méthode en deux étapes peut donc être vue comme un choix particulier d'instrument pour la variable I , ou l'on remplace la variable endogène d'innovation par la probabilité d'innover. Or cette dernière fonction ne dépend que de z , qui sont les instruments du modèle. Il suffit alors d'utiliser le fait que toute fonction des instruments est sans corrélation avec la

perturbation du modèle pour obtenir un estimateur convergent. On remplace donc l'indicateur d'innovation par l'estimation de la probabilité d'innover telle qu'elle résulte du modèle logit et on effectue la régression par les moindres carrés ordinaires.

Un problème se pose alors : l'écart entre la probabilité d'innover et son estimation, qui dépend des valeurs prises par les variables explicatives z , se retrouve alors dans le résidu w de l'équation de croissance modifiée, ce qui implique que la relation devient hétéroscédastique. Il faut donc corriger la matrice de covariance de l'estimateur des moindres carrés ordinaires pour tenir compte de ce problème (i.e., pour pouvoir effectuer les tests de significativité).

Deux approches sont possibles: d'une part effectuer une hypothèse sur la loi jointe des perturbation des relations de croissance et d'innovation (cf. Maddala, 1983), d'autre part utiliser la matrice de covariance robuste à l'hétéroscédasticité de White (1980), qui est valable dans ce cas particulier. Comme cette dernière possibilité implique moins d'hypothèse que la première, c'est celle que nous avons retenue. Les statistiques de significativité de l'équation de croissance sont donc robuste à la distribution de sa perturbation (mais pas à l'hypothèse de loi logistique pour l'équation d'innovation).

Les tests non emboîtés sont réalisés en généralisant les tests de Davidson et Mc Kinnon (1981) et Godfrey (1983). Pour une synthèse sur les tests non emboîtés dans le cadre des moments généralisés, voir Smith (1992).

Annexe B : estimation par les moments généralisés

Soit le modèle linéaire :

$$y = x b + u$$

où la variable y désigne le taux de croissance de la productivité totale des facteurs et la matrice x contient sa valeur décalée, les variables d'innovation et les indicatrices sectorielles. Le problème d'estimation vient de ce que la matrice x contient des variables explicatives endogènes : les innovations, car les innovations mises en œuvre se déterminent simultanément avec la croissance de la productivité totale des facteurs sur la période de cinq ans étudiée. L'essentiel est que la matrice x contienne des variables explicatives endogènes. On admet donc que les régresseurs x sont corrélés avec la perturbation u^1 . On a

$$\text{Plim } x'u / n \neq 0.$$

Dans ce cas, l'estimateur des moindres carrés ordinaires n'est pas convergent. Nous devons donc utiliser une méthode d'estimation par variables instrumentales. Ceci revient à utiliser un ensemble de variables z satisfaisant les trois conditions suivantes: $\text{Plim } z'u / n = 0$, $\text{Plim } z'z / n$ inversible et $\text{Plim } z'x / n$ de même rang que le nombre de variables explicatives du modèle. Dans notre application, la matrice z contient les sources de l'innovation, qui sont à la fois corrélée avec les mises en œuvre d'innovation et sans corrélation avec les performances de l'entreprise une fois que l'on a tenu compte de ses innovations.

La procédure employée, développée par White (1982) consiste, dans un premier temps, à effectuer une estimation par les variables instrumentales (i.e., les doubles moindres carrés), ce qui nous donne un premier estimateur convergent:

$$\hat{\beta} = [x'Px]^{-1} x'Py$$

de covariance asymptotique:

$$\text{Vas}(\hat{\beta}) = [x'Px]^{-1} x'z(z'z)^{-1} \Phi(z'z)^{-1} z'x[x'Px]^{-1}$$

⁵¹ Qui est elle-même autocorrélée dans un modèle en différences premières.

où P est la matrice de projection orthogonale sur $\text{Im}(z)$, $P = z(z'z)^{-1}z'$ et $\Phi = \text{Vas}(z'u)$.

L'estimateur des moments généralisés, noté β^* , qui se ramène ici à l'estimateur 2SIV⁵² de White (1982), fait un meilleur usage de l'information disponible, en tenant compte de la covariance des perturbations. On minimise la quantité suivante :

$$\beta^* = \arg \min_{\beta} u'z\Phi^{-1}z'u$$

Ce qui donne :

$$\beta^* = (x'z\Phi^{-1}z'x)^{-1}x'z\Phi^{-1}z'y$$

$$\text{Vas}(\beta^*) = (x'z\Phi^{-1}z'x)^{-1}.$$

Il nous faut donc un estimateur de Φ pour réaliser l'estimation. Il est facilement obtenu en effectuant d'abord une estimation par les doubles moindres carrés, parfois appelée estimateur des moments généralisés de première étape.

Soit \hat{u}_i le résidu des doubles moindres carrés correspondant à l'individu i et z_i le vecteur des instruments (i.e., les sources de l'innovation) pour ce même individu. Alors, comme les observations sont indépendamment distribuées, on obtient l'estimateur convergent suivant de Φ :

$$\hat{\Phi} = \frac{1}{n} \sum_{i=1}^n z_i' \hat{u}_i \hat{u}_i' z_i$$

A partir de cet estimateur, on recalcule l'estimateur des moments généralisés β^* une seconde fois, ainsi que sa matrice de covariance asymptotique.

Il nous reste alors à tester la validité des instruments utilisés, c'est-à-dire du modèle lui-même. En utilisant l'estimateur de deuxième étape, on effectue un test de suridentification, dont la statistique est donnée par :

$$S = u'^*Z\hat{\Phi}^{-1}Z'u^*$$

⁵² 2SIV : Two-Stage Instrumental Variable estimator.

Sous l'hypothèse nulle $E(z'u) = 0$ elle suit un χ^2 à $p-k$ degrés de liberté où u^* désigne le résidu de deuxième étape.

Dans notre application ce test permet de vérifier les deux points suivants:

1. Les sources de l'innovation n'influencent la croissance que via les innovations mises en œuvre, et non directement.
2. Le biais de sélection sur les sources de l'innovation, inhérent à l'enquête «innovation technologique», n'a pas de conséquence significative sur l'estimation.

Bibliographie

- Barlet C., E. Duguet, D. Encaoua et J. Pradel, 1998, *The commercial success of innovations : an econometric analysis at the firm level in French manufacturing*. Annales d'Economie et de Statistique, n°49-50, pp. 457-478.
- Barro R. et X. Sala-i-Martin, 1995, *Economic Growth*. Mc Graw Hill, New York.
- Crépon B. et E. Duguet, 1997a, *Research and development, competition and innovation : pseudo maximum likelihood methods and simulated maximum likelihood methods applied to count data models with heterogeneity*. Journal of Econometrics, vol. 79, pp. 355-378.
- Crépon B. et E. Duguet, 1997b, *Estimating the innovation function from the patent numbers : GMM on count panel data*. Journal of Applied Econometrics, vol. 12, pp. 243-263.
- Crépon B., E. Duguet et I. Kabla, 1996, *A moderate support to Schumpeterian conjectures from various innovation measures*. In A. Kleinknecht éd., *Determinants of innovation: the message from new indicators*, Mc Millan, Londres.
- Crépon B., E. Duguet et J. Mairesse, 1998, *Research and development, innovation and productivity : an econometric analysis at the firm level*. Document de travail du NBER n° 6696. A paraître dans *Economics of Innovation and New Technology*.
- Crépon B. et J. Mairesse, 1993, *Productivité, recherche-développement et qualifications . .* In Guéllec éd., *Innovation et Compétitivité*. INSEE-Méthodes, n°37-38, *Economica*, 181-221.
- Daguet F., 1995, *Un siècle de démographie française*. Collection INSEE Résultats n°118-119, Décembre.
- Davidson R. et J. Mc Kinnon, 1981, *Several tests for model specification in the presence of alternative hypotheses*. *Econometrica*, vol. 49, 781-793.
- Duguet E., 1996, *Investissement en recherche et développement, innovation et productivité : une analyse économétrique de l'industrie manufacturière française*. Thèse de Doctorat, Université de Paris I-Panthéon-Sorbonne, 22 Novembre.
- Duguet E. et N. Greenan, 1997, *Le biais technologique : une étude économétrique sur données individuelles*. *Revue Economique* 48-(3), 1061-1089.
- Duguet E. et N. Jung, 1997, *An econometric analysis of patent life and value at the firm level*. Document de travail INSEE.
- Duguet E. et I. Kabla, 1998, *Appropriation strategy and the motivations to use the patent system in France : an econometric analysis at the firm level*. Annales d'Economie et de Statistique, n° 49-50, pp. 289-327.
- Godfrey L., 1983, *Testing non-nested models after estimation by instrumental variables or least squares*. *Econometrica*, vol. 51, n° 2, 355-365.
- Gouriéroux C., 1989, *Econométrie des variables qualitatives*. *Economica*, coll. ENSAE.

- Gouriéroux C. et A. Monfort, 1989, *Statistique et modèles économétriques*. Economica, coll. ENSAE.
- Griliches Z. and W. Jorgenson, 1966, *Capital theory: technical progress and capital structure*. American Economic Review, vol. LVI(2), 50-61.
- Griliches Z., 1996, *The discovery of the residual : a historical note*. Journal of Economic Literature, XXXIV, 1324-1330.
- Guillaume H., 1998, *Rapport de mission sur la technologie et l'innovation*.
- Joly P., 1993, *La ralentissement de la productivité : faits et causes*. In Guéllec éd., Innovation et Compétitivité. INSEF-Méthodes, n°37-38, Economica, 39-67.
- Katz M. et J. Ordover, 1990, *R&D cooperation and competition*. Brookings Papers on Economic Activity : Microeconomics, 137-191.
- Levin R., A. Klevorick, R. Nelson et S. Winter, 1987, *Appropriating the returns from industrial research and development*. Brookings Papers on Economic Activity, vol. 3, 783-831.
- Maddala G. S., 1983, *Limited dependent and qualitative variables in econometrics*. Econometric Society Monographs, n°3, Cambridge University Press.
- Mairesse J., 1977, *Deux essais d'estimation du taux moyen de progrès technique incorporé au capital*. Annales de l'INSEF, n°28, 39-75.
- Mairesse J. et P. Mohnen, 1990, *Recherche-développement et productivité : un survol de la littérature économétrique*. Économie et Statistique, n°237-238, 99-108.
- Mairesse J. et M. Sassenou, 1991, *R&D and productivity : a survey of econometric studies at the firm level*. STI Review, OCDE, n°7, 131-147.
- Mansfield E., 1965, *Technological changes: stimuli, constraints, returns : rates of returns from industrial research and développement*. American Economic Review, vol. LV(2), 310-321.
- Minasian J., 1962, *Research and development, production functions and rates of returns*. American Economic Review, vol. LIX(2), 80-85.
- Scherer F., 1982, *Interindustry technology flows and productivity growth*. The Review of Economics and Statistics, vol 64, 627-634.
- Smith R., 1992, *Non-nested tests for competing models estimated by the generalized method of moments*. Econometrica, vol. 60, n° 4; 973-980.
- Solow R., 1957, *Technical change and the aggregate production function*. The Review of Economics and Statistics, 312-320.
- Villa P ; 1994, *Un siècle de données macroéconomiques*. Insee Résultats n°86-87, Avril.
- Villa P., 1996, *La fonction de consommation sur longue période en France*. Revue Economique, 47, n°1.

White H., 1980, *A heteroskedasticity consistent covariance matrix estimator and a direct test of heteroskedasticity*. *Econometrica*, 48, 817-838.

White H., 1982, *Instrumental variables regression with independent observations*. *Econometrica*, 50, 483-499.

Les sources internes et externes d'information dans l'innovation

une analyse économétrique sur données d'entreprises

par

Stéphanie Monjon et Patrick Waelbroeck

La capacité d'une entreprise à innover apparaît aujourd'hui comme un déterminant essentiel de sa performance et de sa compétitivité. Elle constitue par conséquent un enjeu économique essentiel pour la plupart des pays industrialisés en raison de l'intensification de la concurrence internationale. Le succès des entreprises, et des nations, est ainsi devenu plus que jamais dépendant de leur capacité à rassembler et à utiliser les connaissances nécessaires à la mise en place de l'innovation. Cet input de l'innovation n'est néanmoins pas uniforme, et pour innover une entreprise peut être amenée à utiliser des connaissances de natures souvent différentes : il peut s'agir d'un résultat de recherche fondamentale, mais aussi de techniques nécessaires à son application. Le processus de Recherche & Développement (R&D), comme il est coutumier de le considérer, rassemble un groupe d'activités faisant appel à des informations de différentes natures. En premier lieu, la recherche fondamentale a pour objectif la production de connaissances originales. En second lieu, la recherche appliquée correspond à une recherche « tournée » vers le marché, qui a pour but la production de connaissances techniques destinées à être mises en œuvre par des entreprises. Enfin, le développement consiste « en des travaux systématiques basés sur des connaissances existantes obtenues par la recherche et/ou par l'expérience pratique, en vue de lancer la fabrication de nouveaux matériaux produits ou dispositifs, d'établir de nouveaux procédés, systèmes ou services, ou pour améliorer considérablement ceux qui existent déjà » ; cette étape permet donc d'intégrer

l'innovation à l'environnement productif de l'entreprise. Par ailleurs, le changement technologique peut prendre diverses formes, incluant la création d'un nouveau procédé, ou d'un nouveau produit, mais aussi l'adoption d'un nouvel équipement. Chacune de ces formes ne nécessite pas les mêmes connaissances. Ainsi, la façon dont une entreprise a innové est fortement liée aux sources informationnelles qu'elle a utilisées. Par exemple, une entreprise qui utilise de l'information provenant d'un laboratoire de recherche n'innovera certainement pas de la même façon que celle qui aura acquis de l'information auprès de ces fournisseurs. De la même façon, l'information qui aboutit à une innovation de procédé est différente de celle requise pour une innovation de produit.

Cette identification des différentes étapes du processus d'innovation et du mode d'innovation adopté par les entreprises, bien que quelque peu théorique, aide néanmoins à se faire une idée de la diversité des connaissances sollicitées lors du processus d'innovation. Une entreprise peut avoir elle-même mené une activité de recherche fondamentale, ou avoir travaillé à partir de résultats produits dans un organisme extérieur, un laboratoire de recherche publique par exemple. Une entreprise innovante ne se contente pas d'une source informationnelle unique, et fait appel à différentes sources qui ne répondent pas aux mêmes besoins informationnels. Une seule source d'information ne saurait en effet fournir toutes les connaissances utiles, récentes et pertinentes. Les travaux récents ont d'ailleurs montré que le processus d'innovation est hautement complexe et engage généralement plusieurs acteurs externes à une entreprise. S'établissent alors des relations de long terme ou des accords de coopération avec des sources externes de l'innovation. Le processus d'innovation technologique combine ainsi fréquemment des connaissances et des ressources internes d'une entreprise à des connaissances et ressources externes. La tendance actuelle est d'ailleurs à la constitution de véritables réseaux d'information au sein desquels s'associent différents acteurs comme les entreprises, leurs clients et fournisseurs, amis aussi les laboratoires publics de

recherche. Leur objectif est notamment de minimiser les coûts liés à l'acquisition d'informations et de profiter d'actifs complémentaires. Les autorités sont d'ailleurs conscientes de l'importance de réseaux ou de systèmes qui diffusent efficacement les connaissances et l'information et les politiques technologiques de beaucoup de pays industrialisés encouragent aujourd'hui leur création. Mais, le réseau utilisé, et donc les sources informationnelles sollicitées, sont liés à la forme que prend l'innovation.

L'objectif de ce travail est d'analyser de manière empirique la relation existant entre les sources informationnelles utilisées et les différentes façons d'innover, et de montrer que différents réseaux informationnels peuvent mener à des modalités contrastées du changement technologique. En matière de politique économique, cette analyse est donc intéressante car elle permet de mieux appréhender les sources externes d'information qui participent aux différentes façons d'innover. D'ailleurs, la promotion de certaines formes du changement technologique peut se faire en favorisant la communication entre les entreprises et certaines sources informationnelles, l'objectif étant de permettre l'accès à l'information pertinente. Ainsi, le Commissariat Général du Plan (93), quand il conclut « *mais la mise en relation de différents acteurs a un coût et les pouvoirs publics ont pour tâche essentielle de participer au financement de ces mises en réseau, qui pour porter leurs fruits doivent s'inscrire dans la durée [..]* », souligne l'importance des réseaux d'information pour innover, mais n'identifie pas les différents acteurs qui y participent. Une action politique cohérente dans ce sens se doit alors de procéder à cette identification afin de promouvoir le « bon » réseau. Par exemple, si on considère le retard qu'a connu la France concernant les technologies de l'information et de la communication, le problème était très différent de celui d'une création technologique puisque les technologies existaient et étaient disponibles, et renvoyait donc typiquement à un problème d'adoption. Pour accélérer la diffusion de ces technologies, le gouvernement aurait alors pu promouvoir les sources qui offrent l'information pertinente à leur adoption.

L'essentiel des données que nous utilisons provient de l'enquête Innovation Technologique 1997 du SESSI, réalisée à un niveau microéconomique et qui porte sur la période 1994-1996. Cette enquête interroge les entreprises sur la nature et l'origine des innovations qu'elles ont mis en oeuvre durant la période considérée, sur son activité de R&D et les sources informationnelles utilisées par les entreprises pour innover. Toutes ces données sont d'ordre qualitatif. Une première question nous permet d'établir notre première classification. Elle distingue les entreprises qui ont innové en ne participant pas véritablement à la création de l'innovation des entreprises qui ont innové en développant principalement leur innovation en interne. Une deuxième série de question nous permet de différencier les entreprises qui ont introduit de nouveaux produits sur le marché ou dans l'entreprise de celles qui ont introduit un nouveau procédé. Notre échantillon comprend 1680 entreprises françaises de l'industrie manufacturière française, qui ont innové au moins une fois entre 1994 et 1996. Parmi ces 1680 entreprises, 256 ont déclaré avoir innové en adoptant, et 1281 en développant elles-mêmes leur innovation, alors que 1460 ont précisé avoir réalisé une innovation de produit, et 1272 avoir introduit un nouveau processus.

L'enquête interroge par ailleurs les entreprises sur les sources informationnelles qu'elles ont utilisées pour innover. Bien que les réponses soient de nature qualitative, elles offrent un certain degré d'intensité en proposant pour chaque source quatre options : source non utilisée, ou faiblement, moyennement et fortement utilisée. L'enquête propose aux entreprises interrogées douze sources différentes d'information : sources internes à l'entreprise, autres entreprises appartenant au groupe, concurrents, clients, fournisseurs d'équipements et de logiciels, fournisseurs de matières premières et de composants, sociétés de conseils et de recherche marchande, universités ou établissements d'enseignement supérieur, organismes publics de recherche (CNRS, INRA, INSERM) ou institutions privées sans but lucratif, publications de

brevets, conférences, foires et expositions.

Les sources informationnelles proposées sont de natures très différentes : certaines offrent principalement des résultats scientifiques théoriques, alors que d'autres proposent plutôt de l'information technique, portant par exemple sur les caractéristiques d'un nouvel équipement. L'utilisation de ces variables nous permet de rendre compte de la diversité des sources d'informations sollicitées lors du processus d'innovation, qui ne trouve pas son origine seulement dans l'entreprise qui la met en oeuvre : la publication de brevets, les revues professionnelles, ou encore l'information transmise par les fournisseurs contribuent également à ce qu'une entreprise innove. Dans notre échantillon, nous remarquons que plus de cinquante pour-cent des entreprises innovantes ont utilisé plus de six sources informationnelles différentes parmi celles qui leur étaient proposées, et que presque la moitié de l'échantillon a utilisé plus de six sources externes.

Les données d'enquête évitent l'utilisation d'une mesure de brevets ou de R&D pour rendre compte du changement technologique. La R&D est un input de l'innovation, mais c'est seulement dans le cas où cette R&D se concrétise qu'il y a innovation. D'ailleurs, même si la R&D aboutit à des résultats concrets, il se peut que l'entreprise décide de ne pas les utiliser. Mesurer l'innovation par les brevets est également critiquable. Il a en effet été montré, dans une étude récente, que seulement un tiers des innovations était breveté. Une autre originalité de notre étude vient de ce que les variables d'externalités apparaissent en tant qu'inputs de l'innovation, et non en établissant une « distance technologique » entre entreprises et entre secteurs. Elles ne seront par ailleurs prises en compte que dans la mesure où l'entreprise a effectivement utilisé de l'information extérieure pour innover, ce qui évite l'utilisation d'un niveau moyen de R&D dans une industrie. Néanmoins, si les externalités de connaissance sont prises en compte, il nous est difficile de les distinguer des échanges marchands d'information. Les données considérées sont plus larges que les externalités de connaissance, et comprennent

aussi des achats d'information.

Nos résultats confirment que les sources externes jouent globalement un rôle significatif dans le processus d'innovation, et que les sources informationnelles sollicitées diffèrent fortement d'un type d'innovateur à l'autre. Pour les entreprises qui réalisent elles-mêmes le développement de leur innovation, les sources internes jouent un rôle important : il peut s'agir d'information provenant d'employés, dotés de compétences spécifiques, d'activités d'études techniques et de méthodes, présentes souvent dans les petites entreprises, ou d'un véritable service de R&D. Les entreprises de ce type complètent l'information interne avec de l'information externe provenant de leurs concurrents, de leur fournisseurs de matière première et de la publication de brevet ainsi que des foires et expositions. Ces informations sont de nature techniques et principalement en amont du processus de production. Les problèmes rencontrés peuvent être résolus à travers un support technique ou par contacts personnels avec entre ingénieurs lors de foires et expositions.

A l'opposé, les entreprises qui innovent principalement en adoptant des technologies existantes utilisent l'information en provenance de leurs fournisseurs d'équipement ainsi que des revues et conférences. Ces informations sont plus abouties, donc plus facilement assimilable et concerne plus un stage où l'innovateur prend connaissance de l'existence et des caractéristiques de nouvelles technologies. En effet, les entreprises qui adoptent ne mettent pas au point elles-mêmes l'innovation, mais l'utilisent une fois que le produit ou la technologie a été complètement développé. L'innovation repose exclusivement sur de l'information externe.

La seconde typologie envisagée distingue les innovations de produit des innovations de procédé. L'analyse économétrique montre que les sources d'information qui ont participé à ces différents types d'innovations sont différentes. Ainsi, les clients notamment fournissent de l'information pertinente pour les seules innovations de produit, car les clients sont en effet les

utilisateurs du produit proposé par une entreprise et réagiront sur ce dernier. Ils peuvent ainsi avoir inspiré, ou même avoir volontairement initié, une innovation de produit. De plus, l'information provenant des concurrents influence positivement la probabilité d'innover en produit, mais négativement celle d'innover en procédé, car seuls les produits peuvent faire l'objet d'opérations de « reverse engineering ». Finalement, les entreprises qui innover en produit utilisent également de l'information provenant de la consultation des brevets, alors que celle qui innove en procédés font moins appel à cette source d'information. Il faut voir la volonté des entreprises de protéger les innovations de procédé par le secret plutôt que par les brevets.

En conclusion, ce travail offre de premiers résultats intéressants quant à la caractérisation de différents types d'innovateurs. Selon la façon dont une entreprise innove, les sources informationnelles auxquelles elle a recours changent. De plus, les sources externes d'information jouent globalement un rôle important dans le processus d'innovation, mais diffèrent selon le profil de l'innovateur. Les caractéristiques des entreprises diffèrent également selon la façon dont l'entreprise innove. On peut donc avoir l'impression que les différentes façons d'innover renvoient en fait à des entreprises différentes. Or, ce n'est pas forcément le cas. Une même entreprise peut avoir innové de plusieurs façons. Donc, plus qu'un manque de compétences, c'est réellement la stratégie innovante de l'entreprise, et donc les sources informationnelles qu'elles sollicitent, qui mènent à une façon d'innover.

Les sources internes et externes d'information dans l'innovation
une analyse économétrique sur données d'entreprises

Stéphanie Monjon*

Patrick Waelbroeck**

* Université de Paris1-Panthéon-Sorbonne, Equipe Universitaire d'Economie Quantitative (EUREQua),
(monjon@univ-paris1.fr).

** CREST, Laboratoire d'Economie Industrielle (LEI) et Université de Paris1-Panthéon-Sorbonne, Equipe
Universitaire d'Economie Quantitative (EUREQua) (waelbroe@ensae.fr)

1. Introduction

La capacité d'une entreprise à innover apparaît aujourd'hui comme un déterminant essentiel de sa performance et de sa compétitivité. Elle constitue par conséquent un enjeu économique essentiel pour la plupart des pays industrialisés en raison de l'intensification de la concurrence internationale. Les nouvelles théories de la croissance ont permis de mieux comprendre le rôle que jouait le changement technologique sur la croissance, et ont souligné l'importance de l'accumulation de connaissances dans le processus d'innovation. Le succès des entreprises, et des nations, est ainsi devenu plus que jamais dépendant de leur capacité à rassembler et à utiliser les connaissances. Cet input de l'innovation n'est néanmoins pas uniforme, et pour innover une entreprise peut être amenée à utiliser des connaissances de natures souvent différentes : il peut s'agir d'un résultat de recherche fondamentale, mais aussi de techniques nécessaires à son application.

Le processus de Recherche & Développement (R&D), comme il est coutumier de le considérer, rassemble un groupe d'activités faisant appel à des informations de natures variées. En premier lieu, la recherche fondamentale a pour objectif la production de connaissances originales. En second lieu, la recherche appliquée correspond à une recherche « tournée » vers le marché, qui a pour but la production de connaissances techniques destinées à être mises en œuvre par des entreprises¹. Enfin, le développement consiste « en des travaux systématiques basés sur des connaissances existantes obtenues par la recherche et/ou par l'expérience pratique, en vue de lancer la fabrication de nouveaux matériaux produits ou dispositifs, d'établir de nouveaux procédés, systèmes ou services, ou pour améliorer considérablement ceux qui existent déjà » (Frascati, 94) ; cette étape permet donc d'intégrer l'innovation à l'environnement productif de l'entreprise. Cette identification des différentes

¹ Le recours au système de protection par le brevet intervient généralement à ce stade.

étapes du processus d'innovation, bien que quelque peu théorique, aide néanmoins à se faire une idée de la diversité des connaissances sollicitées par les entreprises innovantes.

L'origine de ces connaissances peut être également très variée. Une entreprise peut avoir elle-même mené une activité de recherche fondamentale ou avoir travaillé à partir de résultats produits dans un organisme extérieur, un laboratoire de recherche publique par exemple. Une entreprise innovante ne se contente pas d'une source informationnelle unique, et fait appel à différentes sources qui ne répondent pas aux mêmes besoins informationnels. Une seule source d'information ne saurait en effet fournir toutes les connaissances utiles et pertinentes. Les travaux récents ont d'ailleurs montré que le processus d'innovation est hautement complexe et encourage généralement plusieurs acteurs externes à une entreprise. S'établissent alors des relations de long terme ou des accords de coopération avec des sources externes de l'innovation (Kline et Rosenberg, 86). Le processus d'innovation technologique combine ainsi fréquemment des connaissances et des ressources internes d'une entreprise à des connaissances et ressources externes. La tendance actuelle est d'ailleurs à la constitution de véritables réseaux d'information au sein desquels s'associent différents acteurs comme les entreprises, leurs clients et fournisseurs, mais aussi les laboratoires publics de recherche (OECD, 96). Leur objectif est notamment de minimiser les coûts liés à l'acquisition d'informations et de profiter d'actifs complémentaires. Les autorités sont d'ailleurs conscientes de l'importance de réseaux ou de systèmes qui diffusent efficacement les connaissances et l'information. Ainsi, les politiques technologiques de beaucoup de pays industrialisés encouragent aujourd'hui leur création (Stoneman et Diederer, 94 ; Commissariat Général du Plan, 93). Mais, le réseau utilisé, et donc les sources informationnelles sollicitées, est lié à la forme que prend l'innovation.

Le changement technologique peut prendre diverses formes, incluant la création d'un nouveau procédé ou d'un nouveau produit, mais aussi l'adoption d'un nouvel équipement. Chacune de ces formes ne nécessite pas les mêmes connaissances. Par conséquent la façon dont une entreprise a innové est fortement liée aux sources informationnelles qu'elle a utilisées. Par exemple, une entreprise qui utilise de l'information provenant d'un laboratoire de recherche n'innovera certainement pas de la même façon que celle qui aura acquis de l'information auprès de ces fournisseurs. De la même façon, l'information qui aboutit à une innovation de procédé est différente de celle requise pour une innovation de produit.

L'objectif de ce travail est d'analyser en détail la relation existant entre les sources informationnelles utilisées et les différentes façons d'innover, et de montrer que différents réseaux informationnels peuvent mener à des modalités contrastées du changement technologique. En matière de politique économique, cette analyse est donc intéressante car elle permet de mieux appréhender les sources externes d'information qui participent aux différentes façons d'innover. D'ailleurs, la promotion de certaines formes du changement technologique peut se faire en favorisant la communication entre les entreprises et certaines sources informationnelles, l'objectif étant de permettre l'accès à l'information pertinente. Ainsi, le Commissariat Général du Plan (93), quand il conclut « *mais la mise en relation de différents acteurs a un coût et les pouvoirs publics ont pour tâche essentielle de participer au financement de ces mises en réseau, qui pour porter leurs fruits doivent s'inscrire dans la durée [..]* », souligne l'importance des réseaux d'information pour innover, mais n'identifie pas les différents acteurs qui y participent. Une action politique cohérente dans ce sens se doit alors de procéder à cette identification afin de promouvoir le « bon » réseau. Par exemple, si l'on considère le retard qu'a connu la France concernant les technologies de l'information et de la communication, le problème était très différent de celui d'une création technologique

puisque les technologies existaient et étaient disponibles, et renvoyait donc typiquement à un problème d'adoption. Pour accélérer la diffusion de ces technologies, le gouvernement aurait alors pu promouvoir les sources qui offrent l'information pertinente à leur adoption. L'importance d'une telle politique peut également être illustrée par l'exemple de l'Autriche qui par le passé a comblé son retard technologique en achetant des biens d'équipements à l'étranger, afin d'accroître sa productivité (Léo, 96). Mais, la concurrence de plus en plus aiguë des pays de l'Est à bas salaires l'obligent aujourd'hui à se tourner vers des stratégies de concurrence en qualité plus qu'en prix, ce qui renvoie à des stratégies d'innovation différentes.

L'objectif de ce travail est d'étudier l'origine de l'information qu'utilisent différents types d'innovateurs. Dans la section 2, nous distinguons plusieurs comportements innovants : une première typologie différencie les entreprises en fonction de leur participation dans la création et la mise en forme de l'innovation ; la seconde repose sur le type d'innovations mises en oeuvre. Dans la section 3, nous présentons nos données, avant de faire le point sur la méthodologie utilisée dans la section 4. La section 5 regroupe nos résultats.

2. Différentes modalités du changement technologique

Il existe différentes façons d'innover. Une entreprise peut être présente dès les prémises d'une innovation de produit ou de processus. D'autre part, elle peut avoir assumé elle-même la recherche fondamentale, en développant un laboratoire, ou n'être intervenue qu'en aval du processus en utilisant une nouvelle technologie mise au point ailleurs. La façon d'innover d'une entreprise dépend de sa stratégie d'innovation et des compétences dont elle dispose.

Le potentiel d'une entreprise à innover dépend en effet de sa capacité à produire, ou à assimiler, des connaissances produites à l'extérieur de l'entreprise. L'assimilation d'un résultat produit à partir d'une recherche ou d'un développement externe n'est ni simple, ni immédiate. Ainsi, l'utilisation de cette connaissance peut être compromise en l'absence de personnel capable de traiter cette nouvelle information. En raison d'une mauvaise évaluation des possibilités de la nouvelle technologie, ou de la difficulté de la combiner avec les autres activités de la firme, l'acquisition d'une innovation produite à l'extérieur de l'entreprise peut réduire au final les synergies potentielles entre les différentes activités de la firme, et la rendre peu profitable. L'utilisation de sources externes d'information dépend donc de la propre activité innovante de l'entreprise. C'est ce que Cohen et Levinthal (89, 90) entendent par le concept de *absorptive capacities*. Selon ces auteurs, le stock de connaissances dont dispose déjà une entreprise dans un domaine particulier conditionne son habileté à identifier, assimiler et exploiter de l'information externe. La faculté à exploiter de l'information externe devient une composante primordiale des capacités innovantes d'une entreprise. Au stade élémentaire, il peut s'agir de compétences relativement basiques comme la maîtrise d'un vocabulaire spécialisé ; à un stade plus élaboré, ces capacités peuvent porter sur les derniers développements scientifiques et technologiques. La faculté d'une entreprise à s'approprier telle ou telle information conditionne donc le type d'innovateur qu'elle peut être.

L'information reçue de l'extérieur peut néanmoins prendre des formes très différentes, et nécessiter un besoin plus ou moins grand d'être adaptée, transformée. Selon Julien (94), l'information provenant de brevets, d'organismes de recherche, ou d'autres « canaux informationnels avancés » est coûteuse, au sens où l'information doit subir un traitement poussé avant de pouvoir être appliquée. Le coût de la conversion de l'information en innovation contraint fortement la façon dont une entreprise innove. L'information associée à

la recherche fondamentale par exemple est typiquement moins accessible, moins immédiate à exploiter, que celle associée à une activité de développement. Une entreprise qui ne dispose pas du personnel capable de mener à bien une recherche appliquée à partir de résultats de recherche fondamentale devra se procurer un produit de la R&D plus abouti, donc de l'information qu'elle est capable de s'approprier ; quand l'information est plus facile à assimiler, l'entreprise a alors moins besoin de compétences spécifiques étrangères au métier qu'elle exerce. Une activité de R&D, et plus généralement l'existence de capacités innovantes internes, constituerait alors un préliminaire à l'assimilation de connaissances externes, l'information externe et interne devenant des inputs complémentaires dans le processus d'innovation. Par conséquent, toutes les entreprises ne peuvent pas utiliser tous les types d'information, et certaines peuvent être contraintes sur leur stratégie innovante.

Mais les compétences d'une entreprise ne suffisent pas à expliquer l'ensemble de sa stratégie innovante. En effet, une entreprise peut à la fois choisir de développer un nouveau produit, et adopter une nouvelle technologie, qui participerait à la production de son innovation. Or, pour ces deux opérations, l'information utilisée est différente. En effet, les modalités du changement technologique - développement, adoption, nouveau produit, nouveau procédé - renvoient à des sources informationnelles différentes. Par exemple, l'utilisation de l'information provenant de ses fournisseurs peut conduire à une adoption, plutôt qu'à une création technologique, et à une innovation de processus, plutôt qu'à une innovation de procédé. Au contraire, si une entreprise utilise de l'information provenant d'un laboratoire de recherche, elle innovera certainement en développant elle-même un nouveau produit ou un nouveau processus. Le fait de solliciter certaines sources informationnelles devrait donc fortement conditionner la façon dont une entreprise innove.

3. Les données

L'essentiel des données provient de l'enquête Innovation Technologique 1997 du SESSI², réalisée à un niveau microéconomique et qui porte sur la période 1994-1996. Cette enquête interroge les entreprises sur la nature et l'origine des innovations qu'elles ont mis en oeuvre entre 1994 et 1996, sur son activité de R&D et les sources informationnelles utilisées par les entreprises pour innover. Toutes ces données sont d'ordre qualitatif.

Une première question demande aux entreprises si elles ont réalisé une innovation de produit et/ou de procédé entre 1994 et 1996, ce qui nous permet d'établir notre première typologie reposant sur les types d'innovations mise en oeuvre. L'enquête interroge ensuite les entreprises sur leur participation à la création de l'innovation. Elle permet de distinguer deux types d'entreprises innovantes. La question, « *si vous avez innové, ces produits (respectivement ces procédés) ont-ils été développés principalement par d'autres entreprises ou organismes ?* », nous permet de distinguer les entreprises qui ont innové en ne participant pas véritablement à la création de l'innovation ; cette catégorie essaie de cerner les entreprises qui innovent en adoptant. L'innovation par achat de biens d'équipement se retrouve dans cette catégorie. Le second type isole les entreprises qui ont innové en développant majoritairement leur innovation en interne, et repose sur la question, « *si vous avez innové, ces produits (respectivement ces procédés) ont-ils été développés principalement par votre entreprise ?* ». Ces deux questions nous permettent d'établir notre seconde typologie : entreprises innovantes qui adoptent et entreprises innovantes qui développent en interne³.

² Pour une présentation de l'enquête, voir le 4 pages du SESSI, L'innovation technologique progresse dans l'industrie, n°89, avril 1998.

³ Ce dernier critère reste peu discriminant puisqu'il rassemble des entreprises qui peuvent avoir été présentes

Notre échantillon comprend 1680 entreprises françaises de l'industrie manufacturière, qui ont innové au moins une fois entre 1994 et 1996. Parmi ces 1680 entreprises, 256 ont déclaré avoir innové en adoptant, et 1281 en développant elles-mêmes leur innovation, alors que 1460 ont précisé avoir réalisé une innovation de produit, et 1272 avoir introduit un nouveau processus.⁴

Tableau 1 : Taille des différents échantillons

	Taille des échantillons
Première typologie :	
Adoption	256
Développement en interne	1281
Adoption et développement	101
Seconde typologie :	
Innovation de produit	1460
Innovation de procédé	1272
Produit et procédé	1053

Ce type de données évite d'utiliser une mesure de brevets ou de R&D pour rendre compte du changement technologique. La R&D est un input de l'innovation, mais c'est seulement dans le cas où cette R&D se concrétise qu'il y a innovation. D'ailleurs, même si la R&D aboutit à des résultats concrets, il se peut que l'entreprise décide de ne pas les utiliser. Mesurer l'innovation par les brevets est également critiquable. Il a en effet été montré, dans une étude récente, que seulement un tiers des innovations était breveté (Duguet et Kabla

depuis le début de processus d'innovation, et avoir conduit elles-mêmes leur recherche pure et appliquée, ou simplement avoir développé une invention acquise par ailleurs.

⁴ Nous précisons que certaines entreprises ont déclaré avoir à la fois développé en interne et adopté, et d'autres ont réalisé au cours de la même période des innovations de procédé et de produit. C'est pourquoi les entreprises qui ont adopté ne constituent pas le pendant des entreprises ayant développé, et qu'il est nécessaire de considérer chaque modalité de façon séparée. En outre, en rapport avec notre seconde typologie, l'enquête propose un troisième type, à savoir les entreprises qui ont développé elles-même une innovation avec d'autres entreprises ou organismes. C'est pourquoi, dans la première typologie, la somme des deux activités est inférieure à 1680.

(1998)).

L'enquête interroge par ailleurs les entreprises sur les sources informationnelles qu'elles ont utilisées pour innover. Bien que les réponses soient de nature qualitative, elles offrent un certain degré d'intensité en proposant pour chaque source quatre options : source non utilisée, faible, moyenne et forte. L'enquête propose aux entreprises interrogées douze sources différentes d'information :

- sources internes à l'entreprise
- autres entreprises appartenant au groupe
- concurrents
- clients
- fournisseurs d'équipements et de logiciels
- fournisseurs de matières premières et de composants
- sociétés de conseils et de recherche marchande
- universités ou établissements d'enseignement supérieur
- organismes publics de recherche (CNRS, INRA, INSERM) ou institutions privées sans but lucratif
- publications de brevets⁵
- conférences, réunions ou revues professionnelles, bases de données et réseaux
- foires et expositions

Les sources informationnelles proposées sont de natures très différentes : certaines offrent principalement des résultats scientifiques théoriques, alors que d'autres proposent plutôt de l'information technique, portant par exemple sur les caractéristiques d'un nouvel équipement. L'utilisation de ces variables nous permet de rendre compte de la diversité des sources

d'informations sollicitées lors du processus d'innovation, qui ne trouve pas son origine seulement dans l'entreprise qui la met en oeuvre : la publication de brevets, les revues professionnelles, ou encore l'information transmise par les fournisseurs contribuent également à ce qu'une entreprise innove. Dans notre échantillon, nous remarquons que plus de cinquante pour-cent des entreprises innovantes ont utilisé plus de six sources informationnelles différentes parmi celles qui leur étaient proposées, et que presque la moitié de l'échantillon a utilisé plus de six sources externes.

Tableau 2 : Nombre de sources informationnelles utilisées pour innover -

nombre de sources (tous niveaux d'intensité confondus)	% de l'échantillon		nombre de sources externes (tous niveaux d'intensité confondus)	% de l'échantillon	
	% de	% cumulé		% de	% cumulé
12	10,3	/		/	/
11	9	19,3	11	10,5	/
10	9,3	28,6	10	9	19,5
9	9,1	37,7	9	9,5	29
8	11	48,7	8	9,1	38,1
7	11,1	59,8	7	11,1	49,2
6	9,3	68,3	6	11,5	60,7
5	8,8	78,1	5	9,6	70,3
4	6,9	85	4	9,3	79,6
3	9	94	3	7,8	87,4
2	3,8	97,8	2	8,7	96,1
1	2,1	100	1	3,9	100

Nous utilisons ces sources informationnelles pour caractériser les différents types d'innovateurs : adoption, développement interne, innovation de produit et innovation de

³ Il s'agit de la consultation de brevets, qui peut éventuellement déboucher sur son achat.

procédé. Nous constatons, dans le tableau 3, qu'en moyenne, quelle que soit la façon dont une entreprise innove, elle sollicite un nombre important de partenaires pour obtenir de l'information. Nous attendons néanmoins à ce que l'identité de ces partenaires diffère fortement selon le type d'innovation mis en place.

Tableau 3 : Nombre de sources informationnelles utilisées selon le type d'innovateur

<i>Types d'innovateurs</i>	<i>moyenne</i>	<i>écart-type</i>
<i>Types d'innovations</i>		
Développement interne	7,57	3,03
Développement externe	6,92	3,13
Innovation de produit	7,57	3
Innovation de procédé	7,51	3,09

Une originalité de notre étude vient de ce que les variables d'externalités apparaissent en tant qu'inputs de l'innovation, et non en établissant une « distance technologique » entre entreprises et entre secteurs. Elles ne seront par ailleurs prises en compte que dans la mesure où l'entreprise a effectivement utilisé de l'information extérieure pour innover, ce qui évite l'utilisation d'un niveau moyen de R&D dans une industrie. Néanmoins, si les externalités de connaissance sont prises en compte, il nous est difficile de les distinguer des échanges marchands d'information. Les données considérées sont plus larges que les externalités de connaissance, et comprennent aussi des achats d'information.

Dans cette analyse, seulement une dimension des capacités innovantes est prise en compte : l'activité de R&D. Néanmoins, nos données nous permettent de distinguer les entreprises qui ont mené une activité de R&D permanente et celles qui ont réalisé une activité temporaire. Par ailleurs, nous prenons en compte trois autres variables explicatives, traditionnellement rencontrées dans les travaux économétriques sur l'innovation (Cohen, 95).

La première variable est la taille de l'entreprise, mesurée par le chiffre d'affaires hors taxe. En second lieu, nous prenons en compte un indice de pouvoir de marché, obtenu en calculant une moyenne pondérée des parts de marché des différents secteurs dans lesquels intervient l'entreprise, et un indice de diversification, le nombre équivalent d'activités, obtenu en calculant une moyenne pondérée du nombre de secteurs dans lesquels intervient l'entreprise. Ces indices sont construits à partir de l'Enquête Annuelle d'Entreprises par fractions 1993, qui rend compte du montant des ventes des entreprises dans le secteur où elle exerce son activité principale, et dans les secteurs où elle mène une activité annexe⁶. Concernant ces trois dernières variables, nous ne pourrions pas comparer directement nos résultats avec ceux traditionnellement obtenus dans la littérature, puisque dans cette analyse nous ne cherchons pas à expliquer le fait qu'une entreprise a réussi à innover, ou quelles entreprises s'engagent dans une activité de R&D, mais à savoir si, par exemple, une entreprise diversifiée est incitée à innover d'une certaine façon. Enfin, nous incluons trois variables sectorielles binaires.

4. Le modèle économétrique

Notre objectif est de caractériser les différents types d'innovateurs en fonction de l'intensité avec laquelle les différentes sources informationnelles ont été mobilisées lors de l'innovation. La variable dépendante correspond à la façon dont une entreprise a innové. Nous construisons la variable binaire $y_{i,j}$ pour chaque observation $i=1, \dots, n$ et chaque modalité j telle que :

$$y_{i,j} = \begin{cases} 1 & \text{si l'entreprise } i \text{ a innové de la façon considérée, soit } y_{i,j}^* > 0 \\ 0 & \text{sinon} \end{cases}$$

⁶ L'année 1993 est choisie pour éviter les problèmes de simultanéité avec la variable dépendante. Les définitions de l'indice du nombre équivalent d'activités et du pouvoir de marché moyen se trouvent dans Crépon, Duguet et Kabla (1996).

où $y_{i,j}^* = x_i' \beta_j + \varepsilon_{i,j}$ est une variable continue latente, x_i est le vecteur des variables explicatives, β_j est le vecteur des paramètres à estimer et $\varepsilon_{i,j}$ est le terme d'erreur. L'ensemble des variables explicatives inclut les caractéristiques des entreprises, les sources informationnelles utilisées pour innover et si l'entreprise est engagée dans une activité de R&D permanente ou temporaire. Nous supposons que les termes d'erreur suivent une loi normale et sont distribués indépendamment entre les modalités et entre les observations. Le modèle que nous utilisons correspond donc à un modèle probit standard, que nous estimons par le maximum de vraisemblance.

5. Résultats

5.1.1. Des besoins informationnels contrastés

Nos résultats confirment que les sources externes jouent globalement un rôle significatif dans le processus d'innovation, et que les sources informationnelles sollicitées diffèrent fortement d'un type d'innovateur à l'autre.

5.1.1. L'innovation en développant en interne

Pour les entreprises qui réalisent elles-mêmes le développement de leur innovation, les sources internes jouent un rôle important. L'enquête n'est pas très discriminante concernant ces sources : il peut s'agir d'information provenant d'employés dotés de compétences spécifiques, d'activités d'études techniques et de méthodes présentes souvent dans les petites entreprises, ou d'un véritable service de R&D. Néanmoins, quelle que soit la façon dont se concrétisent ces capacités innovantes internes, elles participent fortement au processus d'innovation.

Les entreprises de ce type complètent l'information interne avec de l'information externe provenant de leurs concurrents. L'information provenant des concurrents tient une place importante : toutes les modalités sont représentées et ce, de façon croissante. Quelle forme prend cette information? Plusieurs cas de figure sont envisageables. Les opérations de « reverse engineering » participent certainement à l'acquisition d'information provenant des concurrents. Il est en effet important pour une entreprise non seulement de mettre au point elle-même des techniques nouvelles, mais aussi de maîtriser celles effectuées par d'autres entreprises. L'information reçue par le biais des opérations de « reverse engineering » peut néanmoins être diversement utilisée. Une entreprise peut se contenter d'imiter exactement une innovation -ce qui pourrait correspondre à la modalité forte-, mais peut aussi chercher à l'améliorer, et contourner alors un éventuel brevet. Dans ce dernier cas, l'information provenant des concurrents aura certainement joué un rôle moins important que dans le cas d'une imitation, ce qui pourrait expliquer que les modalités « faible » et « moyenne » soient également significatives. Dans les deux cas, une activité de « reverse engineering » nécessite de réelles compétences internes afin de s'approprier l'information contenue dans une innovation. Le fait d'analyser les innovations de ses concurrents doit donc constituer un vecteur essentiel de diffusion de l'information liée à une innovation. Les activités de « reverse engineering » ne sont néanmoins pas la seule forme que prend l'information provenant des concurrents. L'information peut également provenir d'une coopération tacite entre concurrents. En effet, utilisant une étude sur les échanges d'information dans l'industrie de l'acier américaine, Von Hippel (87) a constaté que, par le biais de leurs ingénieurs, les entreprises de ce secteur s'échangeaient souvent de l'information, sous réserve d'une réciprocité dans l'échange. Le type d'information concernée est généralement technique, et porte notamment sur le savoir-faire lié à une innovation et qui permet de l'utiliser de façon régulière et efficace. Cette forme particulière de coopération doit certainement se retrouver

dans d'autres industries et pourrait expliquer également la significativité des modalités faible et moyenne. Les informations ainsi transmises n'initient certainement pas l'innovation, mais aident à sa mise en forme.

Les entreprises qui développent en interne recourent également beaucoup à de l'information contenue dans les publications de brevets : toutes les modalités sont significatives. Le rôle de vecteur de diffusion d'information que les autorités cherchent à faire jouer au brevet semble donc fonctionner et l'information qui s'y trouve semble suffisamment précieuse pour que les entreprises acceptent de supporter le coût de son appropriation (Julien, 94). La consultation des brevets participe certainement à la veille technologique que mènent de plus en plus les entreprises innovantes. Alors que cette source publique d'information semble largement utilisée, les autres sources publiques comme les universités ou les organismes de recherche ne sont pas significatives. Pour la modalité « forte », l'information provenant des universités diminue même la probabilité d'innover en développant en interne. Une explication possible serait que les connaissances créées par ces institutions se diffusent du fait de la publication de leurs brevets, mais le lien universités entreprises n'est pas direct. Dans ce cas, une politique qui incite les universités ou les organismes de recherche à breveter leurs inventions faciliterait la diffusion de ce type d'information. Une autre explication pourrait être que le type d'innovateur défini ici ne comprend pas les entreprises qui ont développé elles-mêmes une innovation en collaboration avec d'autres entreprises ou d'autres organismes. Nos résultats suggéreraient alors que les universités pourraient préférer la coopération plutôt que la simple cession des connaissances qu'elles ont créées.

Enfin, l'information provenant des fournisseurs d'équipements et de logiciels réduit la probabilité d'innover en développant en interne. Ce résultat est peu surprenant dans la mesure

où ces fournisseurs proposent fréquemment des produits « clefs-en-main », et offrent donc de l'information qui n'est pas pertinente pour le développement d'une nouvelle technologie. En revanche, le résultat qui suggère que lorsque l'information provient d'autres entreprises, la probabilité de développer soi-même l'innovation diminue est plus surprenant. Nous pouvons en effet nous attendre à ce que, quand des entreprises appartiennent à un même groupe, les échanges informationnelles entre elle jouent un rôle significatif. Nos résultats suggèrent que cela n'est pas le cas.

Un résultat intéressant concerne les deux dernières sources informationnelles : les conférences, réunions, revues professionnelles, bases de données et réseaux d'une part, les foires et les expositions d'autre part. Alors que les entreprises qui développent en interne leur innovation utilisent de l'information provenant des foires et d'expositions, elles n'ont pas recours à l'information qui peut être trouvée aux conférences et dans les revues professionnelles. Les résultats sont inversés pour les entreprises qui ont une stratégie d'adoption. Ces deux sources informationnelles ne doivent donc pas offrir le même type d'information. Les revues et conférences informent de l'existence ou des caractéristiques d'un nouvel équipement disponible, alors que les foires et les expositions sont plus l'occasion de prendre contact, de rencontres entre personnes d'un même métier, voire des lieux d'échanges « de savoir-faire » entre ingénieurs. Les entreprises du premier type n'utilisent en revanche pas la presse ou les conférences pour innover. L'information offerte par ces différents supports doit être relativement générique ou générale. Les foires ou les expositions constituent en revanche des lieux d'échanges d'information plus spécifique, et utile pour innover. L'information provenant d'un support écrit, comme les brevets, se suffit rarement à elle-même. L'information scientifique et technologique exige souvent un dialogue, un échange de savoir-faire (Julien, 94). Les foires et expositions doivent alors constituer les lieux propices à

ce genre d'échanges, et constituent de ce fait une source informationnelle qui complète l'information provenant de la publication de brevets. Bien que fréquemment utilisée, l'importance accordée à l'information qui transite par les foires et expositions n'est pas forte. A l'inverse, les entreprises, qui ont une stratégie d'adoption, ont recours aux revues et conférences pour s'informer, mais n'utilisent pas l'information provenant des foires et expositions.

5.1.2. L'innovation par adoption

Le second type d'innovateur envisagé apparaît très différent du premier, en terme de sources informationnelles sollicitées. L'activité d'adoption a une toute autre figure. Les entreprises ne mettent pas au point elles-mêmes l'innovation, mais l'utilisent une fois que le produit ou la technologie a été complètement développé.

L'adoption est réalisée par les entreprises qui dans l'ensemble font le plus appel à de l'information provenant des fournisseurs d'équipements et de logiciels, et de revues professionnelles, à savoir des agents pourvoyeurs d'informations faciles à assimiler, et le moins à des sources internes. Il semble donc qu'il n'y ait aucune création, même à la marge, de connaissance au sein de l'entreprise. L'innovation repose exclusivement sur de l'information externe. L'information dont a besoin une entreprise qui adopte pour innover portera davantage sur l'existence et les caractéristiques des nouvelles technologies existantes. Elle a alors recours à la presse, à des conférences, et à ses fournisseurs. L'ordre dans lequel sont sollicitées ces différentes sources informationnelles peut varier. Les entreprises peuvent apprendre l'existence de nouvelles technologies par la presse ou lors de conférences, et s'adresser ensuite à des fournisseurs pour se les procurer. Elles peuvent également rechercher, par le biais de la presse, ou de conférences, à vérifier la validité de l'information transmise par

leurs fournisseurs d'équipement et de logiciels.

Un second type d'innovateur/adopteur semble coexister dans cette catégorie : des entreprises, appartenant à des groupes, et qui adoptent des innovations mises au point par une autre entreprise du groupe. L'entreprise en question a bien eu une stratégie d'adoption, mais le développement de l'innovation a néanmoins eu lieu à l'intérieur le groupe.

5.1.3. Innovation de produit versus innovation de procédé

La seconde typologie envisagée distingue les innovations de produit et de procédé. L'analyse économétrique montre que les sources d'information qui ont participé à ces différents types d'innovations sont différentes.

Ainsi, les clients fournissent de l'information pertinente pour les seules innovations de produit. C'était un résultat attendu. Les clients sont en effet les utilisateurs du produit proposé par une entreprise et réagiront sur ce dernier. Ils peuvent ainsi avoir inspiré, ou même avoir volontairement initié, une innovation de produit (Von Hippel (78, 88)). En revanche, les clients, n'ayant généralement pas connaissance du procédé utilisé dans la production, ont peu à apporter à l'entreprise qui innove de cette façon.

Les entreprises qui innovent en produit utilisent également de l'information interne, ce qui ne semble pas le cas des entreprises qui innovent en procédé. Ce résultat était surprenant et nous incité à combiner les deux typologies. Nous trouvons alors le résultat suivante : les sources internes augmentent la probabilité de développer en interne, et diminuent celle d'adopter un nouveau procédé. Ceci est plus conforme à l'intuition. D'ailleurs, nous retrouvons ce résultat pour les innovations de produit.

L'information provenant des concurrents influence positivement la probabilité d'innover en produit, mais négativement celle d'innover en procédé. Ce résultat était également attendu puisque seuls les produits peuvent généralement faire l'objet d'opérations de « reverse engineering ». Nous remarquons néanmoins que seule la modalité « faible » est significative. Ce résultat suggère que les entreprises qui ont innové de cette façon n'ont globalement pas imité, ou que ces mêmes entreprises, même si elles ont copié l'innovation de produit qu'avait réalisée un de leurs concurrents, ont tendance à sous-évaluer le rôle joué par l'information qu'elles se sont appropriées à partir de l'analyse du produit.

Lorsque nous combinons les deux typologies, le résultat précédent devient encore plus fort pour les entreprises qui développent des innovations de produit en interne, mais est également présent pour celles qui se contentent d'en adopter. Sans compétences internes, ces dernières peuvent rarement pratiquer de véritables opérations de « reverse engineering ». Il peut alors s'agir d'informations portant sur l'existence d'une nouvelle technologie, ou de savoir tacite sur la façon d'utiliser un équipement. La stratégie d'innovation de ces dernières entreprises est sans doute une stratégie de suiveur. Elles innoveront non pas pour obtenir un avantage concurrentiel, mais pour rester sur le marché. Le fait que leurs concurrents les plus proches lancent un nouveau produit leur procure des informations essentielles. Elles apprennent notamment si le nouveau produit est bien accueilli sur le marché, si ses concurrents ont rencontré des problèmes pour le produire...

Les entreprises qui innoveront en produit utilisent également de l'information provenant de la consultation des brevets. Toutes les modalités sont significatives, ce qui doit renvoyer à différentes formes d'innovation. La modalité « forte » notamment pourrait renvoyer à un

achat de brevets, alors que la modalité « faible » plutôt à une consultation dans une activité globale de veille technologique. Le résultat concernant le rôle de l'information provenant des brevets est plus ambigu pour les innovations de procédé : cette source informationnelle a un effet négatif lorsqu'elle est utilisée faiblement, mais positive lorsqu'elle est utilisée fortement. Le premier de ces résultats provient de ce que les innovations de procédé sont généralement peu brevetées et sont mieux protégées par le secret (Cohen et Klepper (96) ; Mansfield (85)). L'impact positif renvoie peut-être à des achats de brevets, lorsque les innovations de procédé sont brevetées.

L'information venant des fournisseurs d'équipements et logiciels influence positivement la probabilité de réaliser une innovation de procédé, mais négativement celle de lancer un nouveau produit. Ces résultats sont affinés quand nous combinons les deux typologies. Les entreprises qui développent en interne des innovations de produit n'utilisent pas les fournisseurs d'équipements et de logiciels comme source informationnelle, mais les entreprises qui adoptent des innovations de produit le font, ainsi que celles qui adoptent des innovations de procédé.

Le rôle des fournisseurs de matières premières et de composants comme source informationnelle apparaît en outre quand on combine les deux typologies. Il semble qu'ils jouent un rôle important pour les entreprises qui innovent en produit, que ce soit en adoptant ou en développant en interne, et celles qui développent en interne des innovations de procédé.

5.2. Les caractéristiques des entreprises

L'analyse économétrique montre que certaines caractéristiques conditionnent effectivement la façon dont une entreprise innove. En effet, nous trouvons que les entreprises plus diversifiées

ont une probabilité plus forte d'innover en développant elles-mêmes une nouvelle technologie et que celles qui ont une taille plus importante tendent à innover relativement plus en introduisant de nouveaux procédés. Ce résultat est cohérent avec des travaux précédents (Pavitt et al. (87) ; Scherer (91)). Selon Cohen et Klepper (94), les entreprises les plus grandes sont plus incitées à innover en procédé car, étant capables de produire davantage que des entreprises plus petites, elles peuvent mieux amortir le coût de l'innovation.

Concernant l'activité de R&D que mènent les entreprises de l'échantillon, nous trouvons qu'une activité permanente de R&D augmente la probabilité d'innover en développant en interne, celle d'innover en produit, mais réduit la probabilité d'adopter. Ces résultats sont cohérents avec les portraits d'innovateurs établis jusqu'à présent, et les innovations développées en interne, ou les innovations de produit sont donc davantage le fait d'entreprises qui ont des compétences internes fortes, et notamment qui mènent une activité de R&D. Une activité temporaire de R&D semble en revanche jouer un rôle très marginal, et influence uniquement la probabilité de réaliser une innovation de procédé.

Tableau 4 : Résultats des régressions économétriques

	Types d'innovateurs		Types d'innovations		Innovations de produit		Innovations de procédé	
	Développement	Adoption	Produit	Procédé	Développement	Adoption	Développement	Adoption
<i>Sources internes à l'entreprise (référence=non)</i>								
faible	0.148 (.168)	-0.197 (.175)	0.130 (.185)	0.292 (.178)	0.013 (.168)	0.066 (.217)	0.242 (.165)	-0.257 (.188)
moyenne	0.374* (.144)	-0.446* (.152)	0.197 (.158)	0.110 (.147)	0.303** (.143)	-0.276 (.195)	0.244*** (.141)	-0.396** (.161)
forte	0.693*	-0.733*	0.317**	0.083	0.634*	-0.490**	0.403*	-0.676*

	(.142)	(.150)	(.155)	(.141)	(.140)	(.195)	(.138)	(.160)
Autres entreprises du groupe								
faible	-0.113 (.118)	0.071 (.127)	-0.130 (.134)	0.078 (.109)	-0.207 (.109)	0.023 (.178)	0.003 (.098)	-0.023 (.138)
moyenne	-0.234** (.108)	-0.061 (.121)	0.087 (.130)	0.025 (.101)	-0.168 (.101)	-0.121 (.174)	-0.014 (.091)	-0.078 (.129)
forte	-0.328* (.114)	0.261** (.122)	0.205 (.144)	0.128 (.111)	-0.334* (.107)	0.543* (.154)	0.044 (.098)	0.096 (.133)
Concurrents								
faible	0.215** (.106)	-0.177 (.122)	0.328* (.127)	-0.084 (.104)	0.225** (.100)	0.069 (.163)	0.008 (.094)	-0.251*** (.132)
moyenne	0.282* (.108)	-0.052 (.120)	0.130 (.125)	-0.162 (.105)	0.238** (.102)	0.094 (.163)	-0.131 (.095)	-0.093 (.130)
forte	0.367** (.155)	0.070 (.159)	0.159 (.183)	-0.317** (.140)	0.360** (.143)	0.450** (.206)	-0.107 (.128)	0.042 (.170)
Clients								
faible	-0.210 (.131)	0.061 (.149)	0.233 (.151)	0.044 (.131)	-0.110 (.126)	0.259 (.197)	-0.011 (.118)	0.008 (.160)
moyenne	-0.119 (.121)	0.112 (.136)	0.252*** (.136)	0.002 (.119)	-0.108 (.115)	0.255 (.183)	0.083 (.108)	0.018 (.145)
Fortc	0.005 (.127)	0.135 (.140)	0.243*** (.143)	0.128 (.124)	0.0009 (.120)	0.127 (.191)	0.207*** (.112)	0.034 (.150)
Fournisseurs d'équipements et de logiciels								
Faible	-0.278** (.108)	0.115 (.123)	-0.287** (.135)	0.371* (.097)	-0.311* (.100)	-0.062 (.161)	-0.0009 (.090)	0.174 (.136)
Moyenne	-0.179*** (.106)	0.252** (.116)	-0.505* (.124)	0.572* (.101)	-0.282* (.099)	-0.027 (.155)	0.054 (.091)	0.309** (.127)
Fortc	-0.656* (.121)	0.805* (.127)*	-0.826* (.137)	1.052* (.145)	-0.621* (.118)	0.284*** (.165)	-0.188 (.112)	0.862* (.135)
Fournisseurs de matières premières et de composants								
faible	0.167 (.108)	-0.026 (.120)	0.196 (.126)	0.033 (.101)	0.106 (.100)	0.373** (.161)	0.248* (.092)	-0.114 (.130)
moyenne	0.109	0.051	0.228***	0.079	0.203**	0.360**	0.193**	-0.036

	(.108)	(.118)	(.126)	(.103)	(.102)	(.162)	(.093)	(.127)
forte	-0.016	-0.025	0.210	0.033	-0.025	0.522*	0.104	-0.206
	(.137)	(.150)	(.165)	(.142)	(.131)	(.191)	(.123)	(.164)
<i>Sociétés de conseils et de recherche marchande</i>								
faible	-0.074	0.127	0.127	0.018	-0.015	0.190	0.070	0.141
	(.106)	(.112)	(.135)	(.101)	(.098)	(.150)	(.088)	(.121)
moyenne	-0.146	0.074	-0.022	0.164	-0.158	0.149	0.034	0.172
	(.149)	(.160)	(.183)	(.157)	(.139)	(.212)	(.127)	(.167)
forte	-0.106	0.085	-0.284	-0.195	-0.143	-0.536	0.086	0.166
	(.281)	(.309)	(.320)	(.290)	(.273)	(.523)	(.265)	(.324)
<i>Universités, établissement d'enseignement supérieur</i>								
faible	0.079	-0.031	-0.089	0.093	0.044	-0.098	-0.013	-0.012
	(.113)	(.122)	(.138)	(.103)	(.104)	(.170)	(.092)	(.131)
moyenne	0.083	-0.140	0.232	0.099	0.078	-0.234	0.146	-0.091
	(.143)	(.158)	(.190)	(.134)	(.132)	(.224)	(.116)	(.168)
forte	-0.611*	0.144	0.296	0.218	-0.406**	-0.015	-0.086	0.253
	(.203)	(.227)	(.289)	(.232)	(.198)	(.331)	(.188)	(.234)
<i>Organismes publics de recherche</i>								
faible	-0.042	-0.070	-0.166	-0.035	0.015	-0.134	-0.003	-0.028
	(.121)	(.132)	(.155)	(.112)	(.112)	(.193)	(.098)	(.141)
moyenne	0.215	-0.083	-0.222	0.064	0.222	0.119	0.173	-0.120
	(.163)	(.171)	(.198)	(.147)	(.148)	(.227)	(.127)	(.185)
forte	0.033	0.079	-0.455***	-0.162	-0.048	-0.193	0.076	0.160
	(.233)	(.244)	(.276)	(.236)	(.219)	(.367)	(.202)	(.252)
<i>Publication de brevets</i>								
faible	0.202***	-0.084	0.378*	-0.218**	0.174***	-0.126	-0.021	-0.118
	(.113)	(.124)	(.147)	(.105)	(.104)	(.174)	(.094)	(.134)
moyenne	0.548*	-0.150	0.525*	0.002	0.571*	-0.240	0.145	-0.132
	(.144)	(.147)	(.191)	(.125)	(.131)	(.213)	(.110)	(.156)
forte	0.431**	-0.228	0.367	0.396***	0.468**	0.206	0.201	-0.313
	(.219)	(.235)	(.283)	(.218)	(.200)	(.278)	(.170)	(.256)
<i>Conférences, revues professionnelles</i>								
faible	-0.212***	0.294**	-0.044	-0.075	-0.088	-0.070	-0.201**	0.374*

	(.111)	(.123)	(.130)	(.104)	(.103)	(.163)	(.094)	(.134)
moyenne	-0.329*	0.289**	-0.056	0.017	-0.284*	-0.015	-0.127	0.374*
	(.117)	(.128)	(.139)	(.112)	(.109)	(.171)	(.100)	(.138)
forte	-0.447**	0.298	-0.101	-0.043	-0.452**	0.117	-0.159	0.357
	(.193)	(.204)	(.233)	(.197)	(.184)	(.262)	(.169)	(.219)
Foires et expositions								
faible	0.264**	-0.262**	-0.025	0.006	0.230**	-0.356**	-0.013	-0.225
	(.114)	(.129)	(.133)	(.110)	(.106)	(.167)	(.098)	(.142)
moyenne	0.241**	-0.178	0.037	-0.071	0.284*	-0.357**	-0.029	-0.096
	(.112)	(.124)	(.131)	(.108)	(.105)	(.162)	(.098)	(.135)
forte	0.276***	-0.088	0.171	-0.016	0.341**	-0.191	-0.043	-0.057
	(.159)	(.166)	(.189)	(.152)	(.149)	(.215)	(.136)	(.180)
Activité de R&D :								
<i>permanente</i>	0.395*	-0.350*	0.616*	0.135	0.539*	-0.417**	0.179**	-0.238**
	(.101)	(.112)	(.119)	(.099)	(.095)	(.148)	(.089)	(.121)
<i>temporaire</i>	0.046	-0.073	0.354*	-0.010	0.199**	-0.0007	0.033	-0.027
	(.094)	(.101)	(.108)	(.093)	(.090)	(.132)	(.084)	(.109)
Chiffre d'affaires	-0.027	0.033	0.028	0.085**	-0.023	0.056	-0.006	0.037
1993	(.040)	(.044)	(.048)	(.039)	(.038)	(.058)	(.034)	(.047)
(en log)								
Part de marché 1993	0.039	-0.030	0.025	-0.057***	0.046	-0.056	-0.017	-0.029
(en log)	(.034)	(.037)	(.041)	(.033)	(.032)	(.050)	(.029)	(.040)
Diversification 1993	0.328**	0.002	0.116	-0.054	0.285**	0.026	0.242**	0.104
(en log)	(.145)	(.146)	(.174)	(.127)	(.131)	(.199)	(.113)	(.153)
R² de Mac Fadden	0.338	0.456	0.549	0.263	0.236	0.717	0.048	0.547
Pourcentage de prévisions concordantes	76.7	73.8	79.9	68.3	76.4	77.2	63.6	73.8

Echantillon de 1680 entreprises innovantes de plus de vingt employés dans l'industrie manufacturière

Estimation d'un modèle probit par le maximum de vraisemblance

3 indicatrices sectorielles sont incluses dans chaque régression

* : significativité au seuil de 1 %

** : significativité au seuil de 5 %

*** : significativité au seuil de 10 %

6. Conclusion

Ce travail offre de premiers résultats intéressants quant à la caractérisation de différents types d'innovateurs. Selon la façon dont une entreprise innove, les sources informationnelles auxquelles elle a recours changent. Les sources externes d'information jouent globalement un rôle significatif dans le processus d'innovation, mais différent selon le profil de l'innovateur. Les caractéristiques des entreprises diffèrent également selon la façon dont l'entreprise innove. On peut donc avoir l'impression que les différentes façons d'innover renvoient en fait à des entreprises différentes. Or, ce n'est pas forcément le cas. Une même entreprise peut avoir innové de plusieurs façons. Donc, plus qu'un manque de compétences, c'est réellement la stratégie innovante de l'entreprise, et donc les sources informationnelles qu'elles sollicitent, qui mènent à une façon d'innover.

Bibliographie

Cohen, W. (1995) - Empirical Studies of Innovative Activity, in Stoneman P. (ed.) *Handbook of Innovation and Technical Change*. Oxford, Oxford Blackwell

Cohen, W.M. et Klepper, S. (1996) - A Reprise of Size and R&D, *The Economic Journal*, 106 :925-951

Cohen, W.M. et Klepper, S. (1994) - Firm Size and the Nature of Innovation within Industries : The case of Process and Product R&D, *Review of Economics and Statistics*, 232-243.

Cohen, W.M. et Levinthal, D.A. (1990) - Absorptive Capacity : a new Perspective on Learning and Innovation, *Administrative Quarterly Science*, 35, 128-152

Cohen, W.M. et Levinthal, D.A. (1989) - Innovation and Learning : The two faces of R&D, *The Economic Journal*, 99, 569-596

Commissariat Général du Plan (1993) - *Recherche et Innovation : le temps des réseaux*, la documentation Française.

Crépon B., Duguet, E. et Kabla, I. (1996) - A moderate support to Scumpeterian conjectures from various innovations measures, dans A.Kleinknecht (eds), *Determinants of innovation : the message from new indicators*, McMillan, Londres

Duguet E. et Kabla I. (1998) - Appropriating strategy and the motivations to use the patent system in France : an econometric analysis at the firm level, *Annales d'Economie et de Statistique*, 49-50, 289-327

Florent, F. et François J.-P. (1998) - L'innovation technologique progresse dans l'industrie, *Le 4 pages*, Sessi.

Frascati, M. (1994) - *The Measure of scientific and technological activities*, OECD, Paris

Julien, P.-A. (1994), SME competitiveness, new technology, information and formation, dans Niosi J. (ed.), *New technology policy and social innovations in the firms*, Pinter Publishers, London, 240-259.

Kline, S.J. et Rosenberg, N. (1986) - An Overview of Innovation, in Landau R. and Rosenberg N. (eds), *The Positive Sum Strategy*, National Academy Press : Washington, DC.

Leo, H. (1996) - Determinants of Product and Process Innovation, *Economies et Sociétés*, dynamique technologique et organisation, série W, n°3, 7/1996, 61-77.

Maddala, G.S. (1983) - *Limited-dependant and qualitative variables in econometrics*, Econometric Society Monographs, Cambridge university Press.

Mansfield, E. (1985) - How Rapidly Does New Industrial Technology Leak Out ?, *Journal of Industrial Economics*, 34, 217-23.

Mansfield, E., Schwartz, M. et Wagner S. (1981) - Imitation Costs and Patents : An Empirical Study, *The Economic Journal*, 91, 907-919.

OECD (1996) - *Science, Technology and Industry Outlook*, OECD, Paris.

Pavitt, K., Robson, M. et Townsend, J. (1987) - The size distribution of innovating firms in the UK : 1945-1983, *Journal of Industrial Economics*, 35, 297-316.

Scherer, F.M. (1991) - Changing Perspectives on the Firm Size Problem, dans Acs Z.J. et Audretsch D.B. (eds), *Innovation and Technological Change : An International Comparison*, New York, Harvester Wheatsheaf.

Stoneman, P. and Diederer, P. (1994) - Technology Diffusion and Public Policy, *The Economic Journal*, 104, 918-930.

Schrader S. (1991) - Informal technology transfer between firms ; Cooperation through information trading, *Research policy*, 20, 153-70.

Von Hippel, E. (1988) - *The Sources of Innovation*, New York : Oxford University Press

Von Hippel, E. (1987) - Cooperation between rivals : Informal know-how trading, *Research Policy*, 16, 291-302.

Von Hippel, E. (1978) - *Successful industrial products from customer ideas*, *Journal of Marketing*, 42 :39-49

Cessation, reprise et survie des entreprises innovantes *versus* non innovantes: un modèle à risques concurrents¹

Amel GHARBI-FESSI² et Jacqueline PRADEL³

Janvier 2000

¹Nous tenons à remercier M. MOUCHART et Mme BARDOS pour leurs remarques. Toutes les inexactitudes restant dans ce texte sont de notre fait.

²EUREQua, Université de Paris I Panthéon-Sorbonne.

³EUREQua, Université de Paris I Panthéon-Sorbonne.

L'analyse du phénomène de survie des entreprises a donné lieu à de nombreuses études, qui diffèrent selon les variables explicatives invoquées ainsi que par le phénomène concerné.

La survie de l'entité initiale, définie par son numéro SIREN, a été étudiée en fonction de variables de création (CALLIES [1988]), de caractéristiques socio-démographiques de l'entrepreneur (VIENNET [1991]), ou de variables d'économie industrielle, comme les parts de marché (MATA et PORTUGAL [1994]).

D'autres études faisant intervenir la dimension financière ont concerné les entreprises défaillantes, c'est-à-dire contraintes au dépôt de bilan, dont le traitement est régi selon le droit et conduit à la mise en place d'une procédure de redressement judiciaire dont l'issue peut être la liquidation, la reprise, ou la continuation.

Sur la base de ces travaux, il nous a paru intéressant de conduire l'étude selon trois extensions :

- tenir compte des différents types d'états possibles : cessation totale, cessation pour reprise ou continuation de l'activité.
- utiliser toutes les caractéristiques de l'entreprise : personnelles, financières ou réelles, et relatives au marché.
- examiner comment le fait et la manière d'innover peuvent avoir une influence sur la survie des entreprises.

Cette étude est basée sur un panel d'entreprises françaises du secteur de l'industrie entre 1988 et 1997.

1 Pourquoi distinguer les différents types de sortie?

Lorsque l'entreprise disparaît, elle peut soit cesser totalement son activité, soit la poursuivre sous le nom d'une autre entreprise. Ces deux types de sortie diffèrent non seulement par leurs conséquences économiques et sociales mais aussi par les mécanismes mis en jeu.

Dans la théorie micro-économique, la disparition d'une entreprise paraît une sanction "naturelle", conséquence d'une rentabilité insuffisante ne lui permettant pas de faire face

américaines, mais sans distinguer toutefois les différents types de sortie.

L'activité d'innovation se caractérise par son objet : produit ou procédé. Elle se caractérise également par deux sortes d'intensité :

- une intensité de type endogène définissant la qualité incrémentale ou radicale de l'innovation faite
- une intensité de type exogène définissant l'activité elle-même de l'entreprise, qui peut être technologiquement faiblement ou fortement innovante.

L'impact de l'innovation sur la reprise ou la cessation totale peut être modulé différemment selon ces caractéristiques :

- l'objet de l'innovation, produit ou procédé, a-t-il une influence sur les taux de cessation totale ou de reprise de l'entreprise ?
- le choix que l'entreprise a fait d'une innovation incrémentale ou radicale peut-il diminuer son risque de cessation totale ? L'innovation radicale est l'opération la plus risquée, elle peut aussi bien entraîner la chute que la progression de l'entreprise. Est-ce que la présence d'innovation radicale de la part de l'entreprise augmente ses chances d'être reprise ? L'innovation incrémentale est moins risquée, et on devrait pouvoir observer qu'elle diminue le risque de cessation totale.
- La pression technologique propre à l'activité de l'entreprise représente plus ou moins indirectement le dynamisme innovateur du marché sur lequel intervient l'entreprise. Une activité technologiquement fortement innovante est-elle favorable au mécanisme de croissance externe des entreprises, ou bien au contraire favorise-t-elle les courses individuelles à l'innovation ?

3 Les principaux résultats de l'étude

L'étude du lien entre innovation et survie, ou sortie est rendue possible grâce à trois sources d'informations: les données de l'enquête innovation réalisée en 1991 par le Ministère de

l'Industrie (SFSI), les données comptables du fichier bancaire des entreprises (FIBEN) de la Banque de France et les données du répertoire SIRENE de l'INSEE. Ces données ont été analysées dans le cadre d'un modèle de durée à risques concurrents, permettant de distinguer les différents types de sortie.

Les principaux résultats de l'étude confirment l'influence du profil innovant sur l'évolution des taux de sortie vers deux états "cessation totale" ou "reprise".

En ce qui concerne l'objet de l'innovation (procédés et/ ou produits):

- *L'innovation de procédé n'a pas d'influence significative sur les différents taux de sortie;*
- *L'innovation de produit n'a pas d'influence significative sur le taux de cessation totale, mais multiplie par 3 le taux de reprise.*

Il semble ici que l'on mette en évidence un mécanisme de croissance externe privilégiant l'innovation de produit.

En ce qui concerne les niveaux d'innovation mis en oeuvre par l'entreprise:

- *Innovation radicale n'a pas d'influence significative sur les différents types de sortie;*
- *L'innovation incrémentale n'a pas d'impact significatif sur le taux de sortie pour reprise, mais divise par 3 le taux de cessation totale.*

Il semble ici que ce qui est mis en évidence, c'est le caractère moins risqué de l'innovation incrémentale, et la nécessité pour une entreprise de suivre l'évolution du marché si elle ne veut pas disparaître.

- *Enfin, nous avons constaté que le fait que l'activité de l'entreprise soit technologiquement fortement innovante n'a pas d'impact significatif sur le taux de cessation totale, mais diminue (coef. 0,7) le taux de reprise. C'est donc la course individuelle à l'innovation qui paraît être favorisée dans un marché fortement évolutif.*

Dans le cas d'une cessation pour reprise, les entreprises innovantes sont reprises pour leur rentabilité économique et dans une moindre mesure à cause de leur fragilité financière.

En revanche, les entreprises non innovantes sont davantage reprises à cause de la fragilité de leur structure financière qu'en vertu de leur rentabilité économique.

Dans le cas d'une cessation totale d'une entreprise innovante, les résultats des estimations ont montré que la principale raison de sortie est la fragilité de la structure financière. En revanche, les entreprises non innovantes disparaissent du marché suite à un déclin de leur rentabilité économique combiné à des difficultés financières.

Cessation, reprise et survie des entreprises innovantes *versus* non innovantes : un modèle à risques concurrents*

Amel GHARBI-FESSI[†] et Jacqueline PRADEL[‡]

Janvier 2000

1 Introduction

L'importance du phénomène de survie des entreprises conduit à s'interroger sur ses déterminants. Certaines études relatives à la survie des entreprises retiennent des variables de création (Callies [1988]) et des caractéristiques socio-démographiques de l'entrepreneur (Viennet [1991]). D'autres études retiennent des variables d'économie industrielle (Mata et Portugal [1994]), par exemple les parts de marché. Par ailleurs, les seules études qui ont considéré la dimension financière comme variable d'intérêt n'ont concerné que les entreprises défaillantes. La défaillance d'entreprise est un phénomène généré par un accident financier. L'origine de cet accident peut être économique, financière ou sociale et dont le mode de résolution est régi par le droit ou la mise en place d'une procédure de redressement judiciaire. Certaines résolutions peuvent mener à la continuation de l'activité, d'autres à la reprise, en dernier recours la cessation d'activité (liquidation). Les résultats de ces travaux indiquent que les ratios économiques et financiers sont des déterminants de la défaillance des entreprises. Il nous paraît ici pertinent d'étudier l'importance des variables financières, en plus des variables économiques pour expliquer la pérennité des entreprises qui peut être la cessation totale, la reprise ou la continuité de l'activité.

*Nous tenons à remercier M. MOUCHART et Mme BARDOS pour leurs remarques. Toutes les inexactitudes restant dans ce texte sont de notre fait.

[†]EUREQua, Université de Paris I Panthéon-Sorbonne.

[‡]EUREQua, Université de Paris I Panthéon-Sorbonne.

Sur la base des travaux précédents, trois extensions nous paraissent particulièrement intéressantes afin d'étudier la survie des entreprises : l'introduction d'une variable indiquant le type de sortie (cessation totale, cessation pour reprise et continuité de l'activité), l'introduction de variables qui différencient les entreprises innovantes des entreprises non innovantes et l'introduction de variables financières (rentabilité économique, taux d'endettement, taux d'intérêt apparent, etc).

L'objectif de cet article est ainsi d'étudier les taux de survie et les taux de cessation des entreprises innovantes *versus* non innovantes en tenant compte des différents états de sortie. Les études menées sur les taux de survie des entreprises considèrent souvent de manière globale les sorties des entreprises, sans distinguer précisément le type de sortie alors que les différents états de sortie reflètent des différences de comportements. L'introduction des variables financières nous semble pertinente afin de déterminer si le principal obstacle à la pérennité de l'entreprise est d'ordre économique ou financier. Ces deux extensions seront appliquées à la fois aux entreprises innovantes et non innovantes afin de comparer leurs déterminants de sortie et de survie. Nous procédons à une étude empirique basée sur un panel d'entreprises françaises du secteur de l'industrie entre 1988 et 1997. Ce panel a été réalisé suite à l'appariement de 3 fichiers : le répertoire SIRENE¹ de l'INSEE, l'enquête innovation (1991) du Ministère de l'Industrie (SESST²) et fichier comptable FIBEN³ de la Banque de France. L'outil statistique utilisé pour tester nos hypothèses est le modèle à risques concurrents (*Competing Risks Models*). Ce modèle nous permet d'analyser l'impact des caractéristiques individuelles sur le type de sortie.

L'étude de l'impact de l'innovation sur la survie et sur les différents types de sortie s'avère donc particulièrement attrayante et a d'importantes implications économiques. Dans cette étude, on distingue les différents types de sortie ainsi que leurs déterminants en opposant les entreprises innovantes aux non innovantes. Ces deux catégories d'entreprises se distinguent notamment en terme de rentabilité. Quant au choix des différents types de sortie, il résulte du fait que les objectifs et les conséquences d'une cessation totale sont généralement très différents de ceux d'une cessation pour reprise.

¹Système Informatique pour le Répertoire des ENtreprises et des Etablissemments. Ce répertoire est à la base des études de la démographie des entreprises françaises.

²Service des Etudes et des Statistiques Industrielles.

³Fichier Bancaire des Entreprises.

2 Pourquoi distinguer les différents types de sortie ?

Dans la théorie micro-économique, la disparition d'une entreprise paraît une sanction "naturelle", étant donné que l'entreprise dégage un niveau de rentabilité insuffisant ne lui permettant pas de faire face aux échéances de remboursement de ses engagements financiers. Ainsi, sa disparition permet la réutilisation de ses ressources par des unités plus performantes et économiquement plus rentables. Par contre, les pertes générées par la disparition de l'entreprise sont mesurées en termes de : licenciements des salariés, perturbations des relations inter-entreprises, ruptures des contrats de dettes.

A *contrario*, la cessation pour reprise est un redémarrage prometteur d'un nouveau projet. L'entreprise ne disparaît pas totalement du marché mais elle réapparaît sous une autre unité. Dans ce cas de figure, les contrats sont honorés.

Les reprises sont généralement établies sous condition de garantie d'un minimum de rentabilité. Elle peut avoir deux objectifs différents. D'une part, une entreprise peut avoir des difficultés passagères (par exemple des problèmes financiers temporaires, des difficultés générées par une mauvaise conjoncture...), mais être considérée comme encore rentable. Dans ce cas, elle peut être rachetée par une autre entreprise afin de donner une chance de survie à son activité. D'autre part, une entreprise peut être en pleine croissance et ne connaître aucun problème, ni d'ordre financier, ni d'ordre économique. Elle peut avoir comme objectif, soit de réaliser des économies d'échelle, soit d'augmenter ses parts de marché. Dans ce cas de figure, l'entreprise cherche à prendre le contrôle d'une autre entreprise ou à fusionner.

3 Pourquoi distinguer les entreprises innovantes et non innovantes ?

L'évolution technologique particulièrement rapide de ces dernières années engendre l'apparition fréquente d'innovations. Afin de maintenir leur position sur le marché, voir de survivre, les entreprises peuvent être amenées à réduire leur prix, à imiter de nouveaux produits ou à innover. Or, l'innovation est un projet risqué et rentable à long terme. Les investissements de telles entreprises nécessitent des financements importants. Les contraintes de sélection appliquées à ces dernières risquent d'être un obstacle à leur chance de survie et d'être une explication à leur type de sortie. Cependant, on s'attend à ce que les déterminants des différents types de sortie soient différents selon

le type et l'intensité de l'activité innovante de l'entreprise.

Les travaux empiriques sur l'innovation sont relativement récents. Seul Audretsh en 1995, a essayé d'étudier l'impact de l'innovation sur la survie des entreprises américaines, mais sans distinguer toute fois les différents types de sortie.

4 Le modèle à risques concurrents

Les modèles utilisés pour estimer les taux de survie traitent d'une manière globale les sorties des entreprises, sans distinguer précisément les différents types de sortie (cessation totale, cessation pour reprise, continuité de l'activité) (Audretsh [1994] et Mata et Portugal [1994]). Or, les différents types de sortie (cessation totale et cessation pour reprise) n'ont ni les mêmes objectifs ni les mêmes conséquences. De ce fait, il semble pertinent de tenir compte des différents sorties.

Dans cette étude, les informations disponibles vont nous permettre d'étudier les taux de survie selon trois états de sortie : la cessation totale, la cessation pour reprise et la continuité de l'activité (survie de l'entreprise). On considère ces trois états comme trois états n'ayant pas les mêmes conséquences en terme de croissance économique et financière. De ce fait, nous pouvons vérifier si le taux de sortie évolue de la même façon en fonction de la durée de vie des entreprises selon les différents types de sortie. De même, les caractéristiques individuelles des sortants pourront être comparées afin de déterminer si les cessations totales et les cessations pour reprise concernent des populations d'entreprises différentes.

Afin d'estimer les taux de survie des entreprises innovantes et non innovantes selon les différents types de sortie, nous avons estimé un modèle à risques concurrents dont les intensités de transition vers les différents états de sortie sont spécifiées par un modèle à hasard proportionnel (modèle de Cox), ce modèle permettant à la fois de tenir compte de la "durée de vie" et des déterminants des taux de survie et des taux de sortie.

4.1 La fonction de vraisemblance

En analysant les données nous pouvons constater que l'entreprise peut occuper différents états de sortie. On observe alors pour chaque entreprise la durée de vie à la date de sortie, notée T et l'état de sortie, notée A .

La fonction de vraisemblance pour les variables aléatoires (T, A) est basée sur la spécification de la probabilité jointe du couple⁴

$$\Pr(T > t, A = a) = S(t, a)$$

Cette probabilité peut avoir plusieurs décompositions possibles (Mouchart et Rolin [1995]). On note⁵

$$\begin{aligned} S(t, a) &= \Pr(T > t, A = a) \\ &= \Pr(T > t | A = a) \Pr(A = a) \\ &= E[\Pr(A = a | T) \mathbf{1}_{\{T > t\}}] \\ &= \int_t^{\infty} \Pr(A = a | T = s) dF_T(s) \end{aligned}$$

où $\mathbf{1}_{\{T > t\}}$ est la variable indicatrice qui vaut 1 dès que $T > t$ et 0 sinon.

Dans le cas où $F_T(s)$ est une fonction de répartition différentiable en s , telle que $f_T(s)$ est sa fonction de densité, la contribution à la vraisemblance est de la forme suivante :

$$L_{T,A}(t, a) = \frac{-\frac{d}{dt} S(t, a)}{\Pr(A = a | T = t) f_T(t)} \quad (1)$$

Les contributions à la vraisemblance dépendent de la spécification d'une part de la probabilité de l'état de sortie conditionnelle à la durée observée, et d'autre part de la spécification de la loi marginale de la durée observée. Or généralement, la loi de l'état de sortie conditionnellement à la durée observée est difficile à spécifier. Le modèle à risques concurrents dispense de la spécifier.

⁴Où $S(t, a)$ est appelée une "sous-distribution" car : $S(0, a) = \Pr(A = a) < 1$.

Cependant, la fonction de survie $S_T(t)$ doit satisfaire : $S_T(0) = \Pr(T > 0) = 1$

Dans la suite de notre présentation nous devons admettre que :

$$\Pr(T = 0) = 0 \text{ et } \Pr(T < \infty) = 1 \text{ Dès lors } S_T(t) = \sum_a S(t, a)$$

⁵Les notations adoptées dans la suite de cette section sont similaires à celles de Mouchart et Rolin [1995].

l'enquête. Nous sommes alors en présence d'un problème de sélection endogène, ce qui risque de biaiser l'estimation de la loi de la durée de vie des entreprises.

On note $Y_i = (T_i, A_i)$ la variable jointe et $X_i \in R^p$ les variables explicatives pour $i = 1, \dots, n$ où n est la taille de notre échantillon. Ici le nombre d'entreprises est de 951.

Nous allons spécifier la sous distribution $S(t, a)$ dans un modèle conditionnel en tenant compte du phénomène de sélection endogène, la fonction $S(t, a | t_{dr}, x)$ est notée comme suit :

$$\begin{aligned} S(t, a | t_{dr}, x) &= \frac{\Pr(T > t, A = a | T > t_{dr}, x)}{\Pr(T > t, A = a | x)} \\ &= \frac{\Pr(T > t_{dr})}{S_T(t_{dr})} S(t, a | x) \end{aligned}$$

Afin d'interpréter les observations (T, A) nous allons supposer un mécanisme de risques concurrents (Mouchart [1999]) où T est modélisé comme suit :

$$T = \min \{ \tau_1, \tau_2, \tau_3 \} \in R_+$$

où

τ_j sont les durées latentes associées à l'état de sortie A_j , ($j = 1, 2, 3$)

τ_1 : date de l'événement qui entraîne la cessation totale :

τ_2 : date de l'événement qui entraîne la reprise de l'entreprise :

τ_3 : date de l'événement qui entraîne la fin de l'observation (censure à droite).

A est le premier événement qui s'est réalisé, elle est définie de la manière suivante :

$$A = \min \{ j | \tau_j = T \}$$

On note :

$$a_{i,j} = \begin{cases} 1 & \text{si la période se termine par la sortie de type } j \text{ pour l'entreprise } i \\ 0 & \text{sinon} \end{cases}$$

On considère que la durée latente τ_3 est indépendante de (τ_1, τ_2) , alors que (τ_1, τ_2) risquent d'être corrélées entre elles. Cette corrélation peut être expliquée par le fait qu'une entreprise qui connaît des problèmes financiers peut cesser totalement son activité ou peut être reprise selon que sa situation financière est due à une mauvaise rentabilité (économique ou financière) ou une mauvaise gestion de l'entrepreneur. Or, les données dont nous disposons ne nous permettent pas d'observer ces informations.

La contribution de vraisemblance conditionnellement à la date de réalisation de l'enquête est égale à (cf. annexe 1) :

$$\begin{aligned}
 L_{T,A}(t, a \mid t_{de}, x) &= -\frac{d}{dt} \frac{\Pr(T \geq t, A = a \mid x)}{\Pr(T \geq t_{de})} \\
 &= -\frac{1}{S_T(t_{de})} \frac{d}{dt} \Pr(\tau_j \geq \tau_a \geq t, \forall j \mid x)
 \end{aligned} \tag{3}$$

La vraisemblance de l'ensemble de l'échantillon est donnée par :

$$\begin{aligned}
 L_{T,A}(t, a \mid t_{de}) &= \prod_{1 < i < n} \prod_{1 < j < 3} \left\{ \frac{1}{S(t_{de_i})} S_T(t_i) h_j^c(t_i) \right\}^{a_{i,j}} \\
 &= \prod_{1 < i < n} \frac{S_T(t_i)}{S_T(t_{de_i})} \prod_{1 < j < 3} \left\{ h_j^c(t_i) \right\}^{a_{i,j}}
 \end{aligned} \tag{4}$$

Les fonctions de survie et de hasard vérifient les relations suivantes :

$$S_T(t) = \exp\left(-\int_0^t h_T(u) du\right)$$

et

$$h_T(t) = \sum_{1 < j < 3} h_j^c(t)$$

La log-vraisemblance a pour expression :

$$\log L_{T,A}(t, a \mid t_{de}) = \sum_{1 < i < n} \sum_{1 < j < 3} \left\{ \left[-\int_{t_{de_i}}^{t_i} h_j^c(u) du \right] + a_{i,j} \log h_j^c(t_i) \right\} \tag{5}$$

Dans le cas d'un modèle à risques concurrents, la fonction de vraisemblance ne dépend que de la spécification de la fonction de hasard conditionnel h_u^c .

Afin d'étudier l'impact des caractéristiques individuelles sur la probabilité de sortie, nous avons choisi la spécification de Cox (modèle à hasard proportionnel). L'intérêt de ce type de modèle est de se focaliser sur l'influence des variables explicatives sur les taux de survie et de sortie. La fonction de hasard conditionnel prend dès lors la forme suivante :

$$h_j^c(t \mid t_{de}, x, \theta_j) = h_{0j}^c(t \mid t_{de}, \alpha_j) g(x, \beta_j)$$

où $\theta_j = (\alpha_j, \beta_j)$; $j = 1, 2, 3$

avec α_j paramètre de "nuisance" et β_j paramètre "d'intérêt".

La fonction de hasard conditionnelle peut être décomposée en deux parties :

$h_{0j}^c(t | t_d, \alpha_j)$ est le hasard de base conditionnel. Il peut être considéré comme la fonction de hasard d'un individu dont les variables explicatives sont nulles.

$g(x, \beta_j)$ est l'exponentiel d'une fonction linéaire d'un ensemble de variables explicatives. Cette fonction est positive et telle que :

$$g(x, \beta_j) = \exp\left(\sum_{1 \leq k \leq p} x_k \beta_j\right)$$

La modélisation de la spécification de Cox dépend de deux cas extrêmes :

si la spécification du hasard de base conditionnel est complètement connue alors :

$$h_{0j}^c(t | t_d, \alpha_j) = \alpha_j$$

dans ce cas le hasard de base conditionnel est une loi exponentielle et ne dépend que des paramètres inconnus

si le hasard de base conditionnel dépend de la durée alors :

$$h_{0j}^c(t | t_d, \alpha_j) = \alpha_j(t)$$

dans ce cas α_j est paramètre fonctionnel.

Dès lors, le modèle de Cox est un modèle semi paramétrique. Dans lequel :

α_j prend ses valeurs dans un espace fonctionnel alors que β_j prend ses valeurs dans un espace euclidien.

L'intérêt de l'utilisation de l'approche semi paramétrique est essentiellement pertinente dans les applications économiques où la théorie économique ne peut pas fournir plus d'information sur la structure de la fonction du hasard de base $h_{0j}^c(t | t_d, \alpha_j)$.

Suite à la spécification de la fonction de hasard conditionnel, la fonction de vraisemblance de l'échantillon s'écrit :

$$\log L_{j,A}(t, a | t_d, x, \theta_j) = \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq 3} \left\{ \begin{array}{l} \left[- \int_{t_d}^{t_i} h_{0j}^c(u | t_d, \alpha_j) \exp\left(\sum_{1 \leq k \leq p} x_k \beta_j\right) du \right] \\ + u_{i,j} \log h_{0j}^c(t_i | t_d) + \sum_{1 \leq k \leq p} x_k \beta_j \end{array} \right\} \quad (6)$$

L'estimation des paramètres de la log-vraisemblance dépend du choix de la méthode d'estimation. Dans le cas où on spécifie une distribution au hasard de base conditionnel,

alors, l'estimation est dite paramétrique. En revanche, si aucune spécification n'est attribuée à la fonction de hasard conditionnel, dans ce cas, l'estimation est dite semi-paramétrique.

Avant de décrire l'estimation du modèle, nous souhaitons préciser un problème très rarement développé dans l'estimation des paramètres d'un modèle à risques concurrents. Dans la plupart des travaux, les durées latentes sont considérées implicitement indépendantes. De ce fait, la contribution de vraisemblance est en fonction du hasard marginal $h_j(t)$ (cf. annexe 2) et donnée par :

$$L_{T,A}(t, a) = S_T(t) \prod_{1 \leq j \leq 3} \{h_j(t)\}^{a_j}$$

Si l'on considère que $h_j(t)$ est une fonction de hasard proportionnel, la log-vraisemblance de l'échantillon s'écrit de la façon suivante :

$$\log L_{T,A}(t, a \mid t_{dr}, x, \theta_j) = \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq 3} \left\{ \begin{array}{l} \left[- \int_{t_{dr}}^{t_i} h_{0,j}(u/t_{dr}, \alpha_j) \exp\left(\sum_{1 \leq k \leq p} x'_i \beta_j \right) du \right] \\ + a_{i,j} \log h_{0,j}(t_i \mid t_{dr}) + \sum_{1 \leq k \leq p} x'_i \beta_j \end{array} \right\} \quad (7)$$

En effet, la différence entre les deux fonctions de vraisemblance réside dans le hasard de base. Dans le cas de dépendance des durées latentes, le hasard de base est conditionnel. En revanche, si celles-ci sont indépendantes, le hasard de base est marginal. Les deux formes étant analogues dans les deux cas, cela pose un problème d'identification du modèle. Pour cela nous allons nous référer au théorème (Florens, Fougère et Mouchart [1997]) suivant :

Théorème 1 *Sonnet* :

$S = \{S_*(t_1, t_2, \dots, t_j) : J \text{ -- fonction de survie continûment différentiable}\}$

$\mathbf{S}_J = \left\{ S_* \in S : S_*(t_1, t_2, \dots, t_j) = \prod_j S_j(t_j) \right\}$,

$l_*(t, j)$: la vraisemblance du modèle dans l'ensemble S ,

$l_J(t, j)$: la vraisemblance du modèle dans l'ensemble \mathbf{S}_J ,

Alors $\forall S_* \in S, \exists ! S_J \in \mathbf{S}_J$ sachant que $l_*(t, j) = l_J(t, j)$

et en particulier :

$$h'_{i,*}(t) = h'_{i,J}(t)$$

La démonstration de ce théorème se base d'une part sur la contribution de vraisemblance :

$$l_*(t, j) = h_j^*(t)S_T(t)$$

et d'autre part sur la fonction de hasard de la durée observée :

$$h_T(t) = \sum_j h_j^*(t)$$

De ce fait, la spécification de la fonction de hasard de la durée observée dépend seulement de la somme des fonctions de hasard conditionnel. Or l'égalité $h_{j,*}^*(t) = h_{j,T}(t)$ traduit l'égalité de la contribution de vraisemblance. Par conséquent, que les risques soient considérés dépendants ou indépendants, les résultats des estimations sont identiques. La seule différence réside dans l'interprétation des paramètres estimés.

5 Description des données

Les données utilisées dans cette étude sont issues de trois cohortes de créations d'entreprise qui correspondent aux entreprises créées en 1988, 1989 et 1990. Ces cohortes ont fait l'objet d'un appariement par numéro SIREN des données de l'enquête innovation du Ministère de l'Industrie (SESSI), du répertoire SIRENE (INSEE) et des données FIBEN comptable (Banque de France). L'échantillon est composé de 1181 entreprises. La période d'observation s'arrête à la fin de l'année 1997 et dépend des dates de création des cohortes (10, 9 et 8 ans).

5.1 Les différents ensembles de données

5.1.1 L'enquête innovation 1986-1990

L'enquête innovation a été réalisée pour la première fois en 1991 par le ministère de l'Industrie (SESSI). Elle a permis de révéler l'ampleur inconnue -ou du moins sous-estimée- de l'innovation par rapport à la R&D en France. Son originalité repose sur le fait qu'elle tient compte directement de l'innovation et non plus de *proxie* construite à partir de la R&D. L'innovation a été souvent approximée par le niveau des dépenses en R&D ou par le nombre des dépôts de brevets, alors que ni l'un ni l'autre n'apparaît comme une bonne mesure de l'innovation. Différentes études ont mis l'accent sur l'impact de l'innovation sur la croissance de l'économie française (Barlet, Duguet,

Encaoua et Pradel [1998]). Ainsi, notre étude va permettre de répondre à un certain nombre de questions. Quelle est la proportion d'entreprises innovantes ? Quelles sont leurs probabilités de survie ? Quels sont leurs états de sortie (cessation totale, cessation pour reprise, continuité de l'activité) ?

Les entreprises sélectionnées dans l'enquête sont celles de l'industrie. Parmi ces entreprises, certaines ont innové durant la période 1986-1990, d'autres n'ont pas réalisé d'innovation.

L'information contenue dans cette enquête nous permet d'opposer les entreprises innovantes aux entreprises non innovantes, les innovations de produit et les innovations de procédé, les innovations faibles et les innovations fortes, les innovations radicales et les innovations incrémentales.

En présence d'information provenant d'enquêtes rétrospectives, il convient de prendre certaines précautions dans une approche d'analyse de durée. L'enquête réalisée en 1991 ne concerne que les entreprises survivantes à cette date. Ceci explique l'existence de données "censurées à droite", chose courante dans l'analyse longitudinale. Ainsi, nous pouvons dire que les entreprises observées ont été soumises à un processus de sélection endogène. Ce phénomène de censure peut présenter des risques de biais de sélection.

5.1.2 Les données du répertoire SIRENE

Ce fichier comprend l'intégralité des entreprises du champ ICS⁷ issu du répertoire SIRENE. Les informations fournies par ce dernier concernent la démographie des entreprises : la date de création, la date de cessation d'activité, le type de création (création *ex nihilo* ou création par reprise), le type de sortie (cessation totale, cessation pour reprise, continuation de l'activité), l'effectif salarié, l'activité principale exercée, la forme juridique, la région d'implantation du projet de création. Ce fichier est à la base de l'étude de la survie des entreprises. Mais il ne contient aucune variable macroéconomique ou financière permettant une meilleure explication de la survie des entreprises.

5.1.3 Les données comptables FIBEN

Les données comptables du Fichier Bancaire des Entreprises (FIBEN) de la Banque de France reposent sur les comptes des bilans fiscaux annuels. Ces informations sont constituées en fichier mis à jour mensuellement. On associe à chaque entreprise l'ensemble de ses bilans disponibles. Ces données, en plus de leur caractère individuel, présentent une dimension temporelle.

⁷ Le champ ICS correspond aux entreprises de l'Industrie, du Commerce et des Services.

5.1.4 L'échantillon

L'appariement des trois fichiers repose sur le critère d'appartenance de l'entreprise aux trois fichiers simultanément. Seules les entreprises survivantes au moment de l'enquête innovation ont été retenues dans le fichier correspondant. De plus, dans l'échantillon sélectionné les bilans doivent répondre à des critères de significativité économiques : l'entreprise ne peut avoir un chiffre d'affaires négatif, ni une production négative, ni une valeur ajoutée au coût des facteurs négative, ni des immobilisations corporelles négatives. Les données dont nous disposons sont individuelles et temporelles (données de panel). Elles permettent, d'une part, de tenir compte de l'évolution dans le temps de certaines variables explicatives et, d'autre part, d'avoir une meilleure précision dans les résultats d'estimation. L'appariement des trois fichiers nous a permis de retenir 1181 entreprises parmi lesquelles 684 sont innovantes. Mais, pour cause de données manquantes dans la base FIBEN (entreprises qui n'ont pas déposé leur liasse fiscal bilan une ou plusieurs années durant leur survie), nous n'avons pas pu exploiter l'ensemble des observations. Pour ces raisons, nous ne pouvons pas tenir compte de l'évolution des variables explicatives dans le temps sans restreindre considérablement le panel. Ainsi, pour conserver un nombre d'observations important nous avons procédé en coupe instantanée. Les travaux concernant les faillites d'entreprises, (Blazy et Combiér [1995] et Bardos [1995]) ont établi que les difficultés des entreprises apparaissent dans leur structure productive et financière plusieurs années avant la faillite. L'année de la cessation d'activité ne peut donc être prise comme année de référence pour les variables explicatives déduites d'une déclaration fiscale. En effet, le contenu du bilan et du compte de résultats sont pratiquement inutilisables (Blazy [1996]). Généralement, les symptômes de fragilité reflétés par la structure financière et productive de l'entreprise apparaissent deux ans avant la cessation. Dans notre étude, nous avons retenu les variables explicatives issues de FIBEN deux ans avant la cessation (cessation totale et cessation pour reprise). Le même choix a été appliqué aux entreprises survivantes afin d'avoir une sélection homogène des variables explicatives. En utilisant ces critères de sélection, 951 entreprises sont retenues dont 589 entreprises innovantes, soit environ 62% du total des entreprises de l'échantillon.

Ces données nous permettent d'étudier la survie des entreprises en fonction de variables statiques ou dynamiques (selon l'origine des données) et de variables issues de leur activité économique et financière. Elles permettent de plus de distinguer les différents types de sortie des entreprises. Nous pouvons ainsi chercher si les probabilités de sortie dépendent de la durée de vie des entreprises et d'un certain nombre de variables

explicatives pour les différentes destinations possibles.

5.2 Les variables endogènes

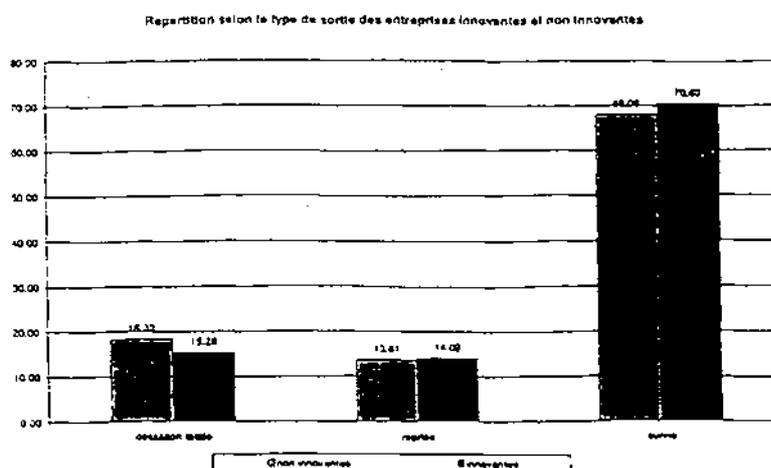
Notre étude consiste à analyser les différents types de sortie selon que l'entreprise est innovante ou non innovante.

5.2.1 Sortie pour cessation totale, sortie pour reprise ou survie ?

Généralement, la cessation d'activité est considérée comme une disparition de l'entreprise de l'activité économique, alors que les faits prouvent que certaines entreprises disparaissent réellement de l'activité économique et d'autres participent encore avec la même activité ou sous une autre forme, à l'occasion d'opération de reprise. Ces différents types de sortie sont définis par la variable "destin" du fichier de cessation du répertoire SIRENE. Cette variable prend trois modalités : cessation totale, cessation pour reprise et continuité de l'activité. Cette variable va nous permettre de distinguer les taux de survie en fonction des différents types de sortie. L'originalité de notre étude est liée à cette information (les travaux antérieurs sur la survie des entreprises ne tiennent pas compte des différents états de sortie). Selon la variable "destin", 17% des entreprises ont cessé totalement leur activité, 14% sont reprises et 69% sont survivantes à la fin de l'observation (31/12/1997) (cf. Figure 1).

5.2.2 Entreprises innovantes *versus* entreprises non innovantes

L'information contenue dans l'enquête nous a permis de stratifier notre échantillon en deux groupes, selon le profil innovant de l'entreprise. L'intérêt d'une telle stratification est de comparer la probabilité de survie des entreprises innovantes et des entreprises non innovantes et de vérifier si leurs déterminants d'une sortie (cessation totale, reprise ou continuité de l'activité) sont différents.



L'innovation technologique dans le Manuel d'OSLO

Les innovations technologiques de produit et de procédé (TPP) couvrent les produits et procédés technologiquement nouveaux ainsi que les améliorations technologiques importantes de produits et de procédés qui ont été accomplies. Une innovation TPP a été accomplie dès lors qu'elle a été introduite sur le marché (innovation de produit) ou utilisée dans un procédé de production (innovation de procédé). Les innovations TPP font intervenir toutes sortes d'activités scientifiques, technologiques, organisationnelles, financières et commerciales. La firme innovante TPP est une firme qui a accompli des produits ou des procédés technologiquement nouveaux ou sensiblement améliorés au cours de la période considérée.

Un produit technologiquement nouveau est un produit dont les caractéristiques technologiques ou les utilisations prévues présentent des différences significatives par rapport à celles produites antérieurement. De telles innovations peuvent faire intervenir des technologies radicalement nouvelles, ou reposer sur l'association de technologies existantes dans de nouvelles applications, ou découler de la mise à profit de nouvelles connaissances.

Un produit technologiquement amélioré est un produit existant dont les performances sont sensiblement augmentées ou améliorées. Un produit simple peut être amélioré (par amélioration des performances ou abaissement du coût) grâce à l'utilisation de composants ou de matériaux plus performants, ou bien un produit complexe qui comprend plusieurs sous-systèmes techniques intégrés peut être amélioré au moyen de modifications partielles apportées à l'un des sous-systèmes.

Il y a innovation technologique de procédé dans le cas de l'adoption de méthodes de production technologiquement nouvelles ou sensiblement améliorées, y compris les méthodes de livraison du produit. Ces méthodes peuvent impliquer des modifications portant sur l'équipement ou l'organisation de la production, ou une combinaison de ces modifications, et peuvent découler de la mise à profit de nouvelles connaissances. Ces méthodes peuvent viser à produire ou à livrer des produits technologiquement nouveaux ou améliorés, qu'il est impossible de produire ou de livrer à l'aide de méthodes classiques, ou essentiellement à augmenter le rendement de production ou l'efficacité de la livraison de produits existants.

5.3 Les variables explicatives

Afin d'expliquer la survie des entreprises tout en tenant compte des différents états de sortie, plusieurs critères ont été retenus : l'origine de création, le secteur d'activité, l'effectif salarié, le type de sortie, la croissance, la rentabilité, la structure productive et la structure financière. Nous pouvons regrouper ces différents critères en deux groupes. D'une part, les variables statiques, et d'autre part, les variables dynamiques.

5.3.1 Les variables statiques

Pour les variables statiques, nous avons retenu d'une part, les variables qui décrivent l'origine de la création, le secteur d'activité, l'effectif salarié et le type de sortie, et d'autre part, les variables relatives à l'innovation.

l'origine de la création : cette variable permet d'indiquer si l'entreprise a été créée *ex nihilo* ou si elle a été reprise (fusion, acquisitions ou absorption...). 59,2% de l'ensemble des entreprises correspondent à des créations par reprises et 40,8% à des créations *ex nihilo*.

le secteur d'activité : nous avons retenu le secteur d'activité au moment du dépouillement de l'enquête innovation au premier semestre de l'année 1991. L'enquête innovation est réalisée auprès des entreprises du secteur de l'industrie. Dans ce secteur, nous avons construit trois classes d'activité : l'industrie des biens d'équipement, l'industrie des biens de consommation et l'industrie des biens intermédiaires. Les proportions des entreprises dans chacun des secteurs sont reportées dans le tableau suivant :

Tableau 1 : répartition des entreprises par secteur d'activité (%)

biens de consommation	26,9
biens intermédiaires	41,7
biens d'équipement	31,4

l'effectif salarié est un critère déterminant dans la survie des entreprises. Plus de 98% des entreprises de notre échantillon ont un effectif inférieur à 500 salariés. Étant donnée la taille de l'échantillon, nous sommes contraints de ne distinguer que trois tranches d'effectifs salariés : inférieur à 40 salariés, entre 40 et 99 salariés et plus de 100 salariés. Les proportions de chacune des tranches d'effectifs sont décrites dans le tableau suivant :

Tableau 2 : répartition des entreprises selon l'effectif salarié (%)

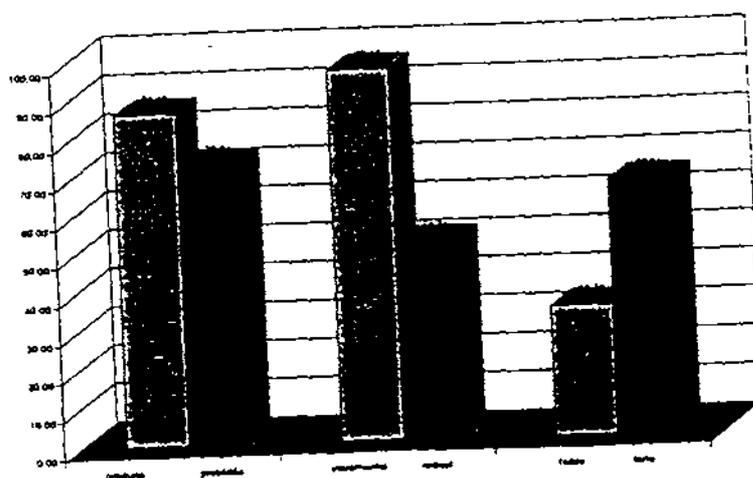
< 40 salariés	35,7
40 à 99 salariés	36,6
≥ 100 salariés	27,7

Les variables relatives à l'innovation : l'enquête innovation comprend des questions sur l'innovation de produit (produits nouveaux ou produits anciens sensiblement améliorés) et l'innovation de procédé (nouvelles techniques de production). Pour les entreprises innovantes, l'enquête nous a permis d'avoir plus d'informations sur le type d'innovation, afin de distinguer :

L'innovation de produit et l'innovation de procédé. L'innovation de produit est le fruit soit d'une amélioration substantielle d'un produit existant, soit de l'introduction d'un nouveau produit sur le marché, éventuellement nouveau pour l'entreprise mais existant déjà sur le marché. L'innovation de procédé est une première réalisation de procédés ou amélioration substantielle des procédés de production. Un certain nombre d'entreprises réalisent à la fois l'innovation de produit et l'innovation de procédé. D'autres ne réalisent que l'une des deux innovations. Parmi les entreprises innovantes, 86,4% ont fait une innovation de produit et 73% ont fait de l'innovation de procédé (*cf.* Figure 2).

L'innovation incrémentale et l'innovation radicale. On appelle innovation incrémentale une innovation qui permet soit une amélioration substantielle d'un produit existant, soit une innovation d'un produit, éventuellement nouveau pour l'entreprise mais existant déjà sur le marché ou soit une amélioration substantielle du point de vue technologique des procédés de production. Une innovation radicale est la création soit d'un produit nouveau sur le marché, soit d'un nouveau procédé de production. Comme dans le cas de l'innovation de produit et de procédé, pour certaines entreprises, l'innovation réalisée est à la fois incrémentale et radicale. Dans d'autres entreprises, l'innovation réalisée n'est que radicale ou incrémentale. Parmi les entreprises innovantes, 96,4% ont fait de l'innovation incrémentale et 51,6% ont fait de l'innovation radicale (*cf.* Figure 2).

- *Le degré d'innovation.* Nous avons stratifié l'activité innovante en deux catégories, selon que l'activité est technologiquement faiblement ou fortement innovante. Parmi les entreprises innovantes, 33,8% sont faiblement innovantes et 66,2% sont fortement innovantes (cf. Figure 2).



Toutes les variables statiques que nous venons de décrire sont introduites dans l'étude sous forme de variables qualitatives (indicatrices).

5.3.2 Les variables dynamiques

Afin d'affiner encore notre étude, nous allons tenir compte d'autres variables décrivant l'évolution de l'entreprise au cours du temps (variables dynamiques). Ces variables permettent d'appréhender la dynamique économique et financière de l'entreprise. Ceci nous permettra de détecter le risque de sortie, qu'il s'agisse d'une cessation totale ou d'une cessation pour reprise (dont l'objectif est soit la croissance externe, soit l'absorption par une autre entreprise suite à des difficultés économiques ou financières). Les variables retenues sont des variables du bilan et du compte de résultat issues du fichier comptable FIBEN. Ces variables ont été regroupées en plusieurs thèmes. Ils concernent les critères de performance (rentabilité), de croissance de l'entreprise, de structure productive, de la structure financière de l'entreprise et de relation inter-entreprises (*cf.* annexe 3 tableau 9). L'analyse des critères est effectuée à partir de certains ratios issus de l'analyse financière de la Banque de France. Ce type de ratio a déjà été utilisé dans des études sur la défaillance des entreprises (Bardos [1996], Blazy et Combier [1995]).

Plusieurs problèmes ont été rencontrés dans à l'utilisation des données du fichier comptable FIBEN. Les résolutions adoptées seront présentées en annexe (*cf.* annexe 3).

Afin d'éviter les informations redondantes fournies par certains ratios et pour garantir une robustesse des résultats des estimations nous avons vérifié les corrélations entre les ratios calculés pour cette étude.

6 Les résultats des estimations

Tout d'abord, il faut noter que nous n'avons pas de terme constant estimé (caractéristique de la vraisemblance partielle) étant donné que ce terme est contenu dans la fonction de hasard de base conditionnel. Les résultats des estimations pour chaque type de sortie ne correspondent pas à la valeur des paramètres estimés (β_j) mais à e^{β_j} appelé le "ratio hasard". Pour les variables qualitatives prenant des valeurs 1 ou 0, nous pouvons interpréter le "ratio hasard" comme l'estimation du hasard pour les variables prenant la valeur 1 par référence à celles prenant la valeur 0. Par exemple, l'estimation du "ratio hasard" pour la variable innovation de produit est de 3.33. Ceci veut dire que dans le cas d'une sortie pour reprise, le hasard (taux de sortie) des en-

entreprises qui font d'innovation de produit est de 333% plus élevé que le hasard des entreprises qui ne font pas d'innovation de produit (toutes choses étant égales par ailleurs). Pour des variables explicatives quantitatives, nous interprétons un pourcentage de variation du hasard estimé suite à une variation d'une unité de la variable explicative. Le pourcentage de variation du taux de sortie pour reprise (hasard) est obtenu par la formule suivante : $100(e^{\beta_j} - 1)$. Si on prend la variable de rentabilité globale pour les entreprises innovantes, le ratio hasard est de 1,037, le résultat reporté dans le tableau 3 est $100(1,037 - 1) = 3,7\%$. Ce résultat peut être interprété comme suit : l'augmentation d'une unité de la rentabilité globale des entreprises innovantes entraîne une augmentation du taux de sortie pour reprise de 3,7%.

Tableau 3: estimation des taux de sorties du modèle à risques concurrents

	Variables	Entreprises innovantes		Entreprises non innovantes	
		Cessation totale	Cessation pour reprise	Cessation totale	Cessation pour reprise
Innovation de produit	innovation de produit	n.s.	3.332 (6.62)		
Qualité de l'innovation	innovation faible	n.s.	1.725 (5.54)		
Innovation incrémentale	innovation incrémentale	0.336 (6.21)	n.s.		
Secteur d'activité	biens d'équipement			<i>référence</i>	
	biens intermédiaire	n.s.	n.s.	0.478 (4.24)	n.s.
	biens de consommation			n.s.	
Tranche d'effectif salarié	moins de 40 salariés			<i>référence</i>	
	40 à 99 salariés	n.s.	0.446 (7.78)		n.s.
	100 salariés ou plus				

	rentabilité d'exploitation	-0.4 (6.79)	n.s.	0.3 (5.20)	n.s.
Rentabilité économique et financière	rentabilité globale	n.s.	3.7 (27.64)*	1.5 (4.63)	1.6 (6.78)*
	rentabilité financière nette	n.s.	-0.1 (5.05)	n.s.	n.s.
	rentabilité financière	n.s.	n.s.	-0.4 (5.73)	n.s.
	taux de valeur ajoutée				
Structure productive	coefficient du capital	n.s.	n.s.	-0.7 (5.33)	n.s.
	taux de la valeur ajoutée	2.7 (15.16)	n.s.	2.5 (17.89)	n.s.
	taux d'accumulation du capital	2 (26.83)*	n.s.	2.5 (17.89)*	n.s.
Fragilité financière	part de l'endettement dans le capital engagé	2.6 (59.38)*	1 (8.67)*	2.8 (25.35)*	1.9 (19.88)*
	taux d'intérêt apparent	0.7 (12.47)	n.s.	n.s.	n.s.
	solvabilité (frais financiers dans l'EBE)	0.2 (7.71)	n.s.	0.7 (34.24)	n.s.
	couverture des capitaux investis	0.4 (11.44)	n.s.	n.s.	-0.5 (3.87)
	délai de rotation des BFR	0.6 (6.68)	n.s.	1.2 (15.58)	n.s.
	délai fournisseur	0.6 (5.21)	n.s.	n.s.	1.1 (12.12)
Crédit interentreprise	délai découvert client	n.s.	0.6 (6.6)	n.s.	n.s.

Les valeurs entre parenthèse sont les statistiques de Wald. -n.s. : non significatif au seuil de 5%. La valeurs critique au seuil de 5% est égale 4.-

*Les valeurs en gras permettent d'opposer les entreprises innovantes et les entreprises non innovantes selon le type de sortie

Le modèle décrit dans la section précédente a été estimé sur deux échantillons : celui des entreprises innovantes et celui des entreprises non innovantes. Le modèle à risques concurrents nous permet d'analyser l'impact des variables explicatives sur le taux de sortie (cessation totale, cessation pour reprise ou continuité de l'activité). Cependant, dans le cas d'une sortie pour une continuité de l'activité (données censurées à droite), elle ne conduit pas à un modèle explicatif pertinent⁸. Ainsi seuls les résultats relatifs aux déterminants des deux types de sortie (pour reprise et pour cessation totale) sont analysés par la suite. Les estimations du modèle sur les deux échantillons font apparaître, comme nous pouvions l'attendre, des différences significatives selon que l'entreprise est innovante ou non innovante et selon le type de sortie (reprise ou cessation totale). En effet, les motivations et les objectifs d'une cessation totale et d'une cessation pour reprise sont différentes.

Afin de tester ces constatations, nous avons établi une comparaison entre la moyenne des ratios pour les entreprises innovantes et celles non innovantes. Les tests d'égalité des moyennes⁹ pour chacune des variables et dans chaque échantillon nous permettent de rejeter l'hypothèse d'égalité des moyennes pour tous les ratios sauf celui de solvabilité. Les résultats de ces tests confirment une différence significative entre les entreprises innovantes et non innovantes, à l'exception du ratio de solvabilité (*cf.* tableau 4).

Afin d'étudier l'impact des caractéristiques économiques et financières sur le type de sortie, nous allons présenter et analyser les résultats des estimations selon que la cessation est totale ou s'il s'agit d'une reprise. Nous vérifierons que le type d'innovation, la taille de l'entreprise, la performance économique et la fragilité financière ont des conséquences différentes selon que l'on considère les sorties pour reprise ou pour cessation totale, aussi bien pour les entreprises innovantes que les entreprises non innovantes.

⁸On note que les estimations ne sont pas significatives. C'est la raison pour laquelle on ne reporte pas de résultats ici.

⁹Le test d'égalité des moyennes permet de tester les hypothèses suivantes :

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_0 : \mu_1 \neq \mu_2 \end{cases}$$

Pour deux populations indépendantes de taille n_1 et n_2 suffisamment grande, où $n_1 = 589$ est la taille de l'échantillon des entreprises innovantes et $n_2 = 360$ et la taille de l'échantillon des entreprises non innovantes ($n_1 + n_2 > 100$ pour des lois quelconques, $n_1 + n_2 > 40$ pour des lois normales), on considère la statistique écart-réduit : $T = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ que l'on suppose gaussienne $N(0, 1)$.

Afin de tester $\mu_1 - \mu_2 = 0$ au niveau α , on acceptera H_0 si : $\frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} < u_{1-\alpha/2}$

avec $u_{1-\alpha/2} = 1.96$, au seuil de $\alpha = 5\%$, si non on refuse l'hypothèse H_0 (Saporta [1990]).

Tableau 4 : comparaison des moyennes des ratios entre les entreprises innovantes et non innovantes

Variables	Entreprise non innovantes	Entreprise innovantes	Statistique calculée*
Rentabilité économique et financière			
rentabilité d'exploitation	21,9731,06 (2,50)	31,06 (5,05)	31,85
rentabilité globale	15,56 (0,87)	18,84 (1,25)	43,76
rentabilité financière nette	5,18 (11,19)	207,09 (8,51)	39,60
rentabilité financière	-239,03 (4,47)	-253,81 (15,21)	22,09
part des dividendes dans valeur ajoutée	2,73 (0,40)	3,75 (0,36)	38,83
Structure productive			
coefficient du capital	108,48 (8,06)	145,82 (6,37)	74,81
Croissance			
taux d'accumulation du capital	18,42 (1,23)	17,55 (0,84)	12,66
Fragilité financière			
part de l'endettement dans capital engagé	42,40 (0,75)	39,39 (1,27)	28,36
taux d'intérêt apparent	22,20 (3,23)	16,36 (1,75)	31,55
Frais financier dans l'EBE	87,09 (6,84)	87,46 (5,38)	0,88
couverture des capitaux investis	128,03 (5,76)	121,24 (3,52)	20,18
délai de rotation des BFR	45,01 (2,80)	64,60 (2,30)	111,68
Credit inter-entreprises			
délai fournisseur	83,03 (2,03)	85,07 (1,50)	16,54
délai découvert client	85,45 (2,09)	93,11 (1,85)	57,187
<p>EBE : excédent brut d'exploitation. BFR : besoins de fonds de roulement. Les valeurs entre parenthèse sont les écarts types. *Statistique d'égalité des moyennes (cf. note de bas de page p211-212). Seule le ratio de solvabilité pour leque on accepte l'égalité des moyennes au seuil de 5%.</p>			

6.1 La cessation pour reprise : un redémarrage porteur d'un nouveau projet

L'expression "cessation pour reprise" ne signifie pas forcément que l'entreprise est en difficulté. Certaines reprises sont d'ordre économique, l'objectif étant d'accroître les capacités de production afin d'atteindre une taille critique pour profiter d'économies d'échelle ou d'étoffer la gamme d'activités et de produits afin de profiter des effets de synergie. D'autres sont d'ordre financier : diversification des portefeuilles d'actifs, réduction du niveau d'endettement ou réalisation d'un "effet de compensation". Il faut signaler que dans le cas des entreprises innovantes, la nature de la reprise peut être encore différente. Nous pouvons notamment penser que la raison essentielle d'une reprise est l'appropriation des innovations réalisées par l'entreprise ou la mise en commun des capacités à innover. Ceci est illustré d'une part par les nombreux rachats de "start-up" observés et, d'autre part par le développement des "joint-ventures" qui ont comme objectif le regroupement des efforts et des moyens en matière d'innovation.

6.1.1 Le type d'innovation

Il apparaît donc que l'objectif d'une majorité d'entreprises innovantes est l'amélioration de la qualité des produits existants ou la mise sur le marché de produits nouveaux (SESSI [1998]). Ce double objectif peut être expliqué par la pression concurrentielle ou par la volonté de l'entreprise d'augmenter sa part de marché en introduisant de nouveaux produits. La concurrence non seulement favorise une allocation efficace des ressources mais incite également les entreprises à élargir la gamme des produits offerts et pousse à la sélection des entreprises compétitives.

D'après les résultats de l'estimation sur notre échantillon d'entreprises françaises (résultats non répertoriés), si l'on ne distingue pas les différents types de sortie, une entreprise qui fait une innovation de produit voit son taux de cessation augmenter de 181%. De même, une entreprise qui fait de l'innovation faible a un taux de cessation qui augmente de 143%. A cette étape, nous pouvons retenir que l'innovation de produit ou l'innovation faible diminue la probabilité de survie des entreprises innovantes. Cependant, le fait de distinguer les différents types de sortie montre qu'il y a un intérêt de l'innovation. Cela n'apparaît pas empiriquement quand on ignore les différents types de sortie. En distinguant les cessations totales des cessations pour reprise, on s'aperçoit que les résultats sont différents et plus pertinents.

Les entreprises qui ont fait une innovation de produit ont un taux de

sortie pour reprise. 3 fois plus élevé que celui des entreprises qui n'en font pas.

Les résultats nous montrent l'impact du type d'innovation sur la destination des entreprises. Il peut être direct dans le cas d'une innovation de produit ou indirect dans le cas d'une innovation de procédé. De plus, on note que le nouveau procédé peut être appliqué à la fois à de nouveaux produits comme à d'anciens produits. Par conséquent, il existe une interaction entre l'innovation de procédé et l'innovation de produit. L'innovation de procédé est presque toujours associée à l'innovation de produit afin de conquérir le marché et de maintenir une chance de survie plus importante.

De même, les entreprises dont l'activité est faiblement innovante, ont un taux de cessation pour reprise 1.7 fois plus grand que celui des entreprises ayant une activité fortement innovante.

Ces résultats peuvent être expliqués par le succès de l'innovation. L'entreprise réalise un profit grâce à son avance technologique lui assurant une position de leader, au moins jusqu'à l'entrée d'un nouveau concurrent ou d'un imitateur. La sortie pour reprise résulte soit d'une volonté de croissance externe, l'entreprise innovante réalisant alors une fusion ou une acquisition, soit des difficultés financières résultant du processus d'innovation, l'entreprise se laissant alors racheter par une autre entreprise.

6.1.2 L'effet taille

Dans notre échantillon, nous n'avons que très peu d'entreprises de plus de 500 salariés. 98% des entreprises sont des PMI. Les résultats empiriques montrent que les PMI sont à l'origine d'un grand nombre d'innovations. L'ensemble des entreprises innovantes représente 61,9% du total de l'échantillon. Parmi ces entreprises 32,4% ont un effectif salarié inférieur à 40, 35,6% ont un effectif salarié compris entre 40 et 99 et enfin 32% ont un effectif salarié supérieur à 100.

Ces résultats de statistique descriptive vont dans le sens de certains travaux théoriques qui ont tenté de résoudre le paradoxe Schumpéterien¹⁰, afin d'expliquer l'existence d'entreprises innovantes de petite taille. Parmi ces modèles figure celui de Jovanovic [1982]. Ce modèle montre que l'existence de petites entreprises innovantes s'explique par un manque d'information des entrepreneurs. L'auteur considère qu'avant la création d'une entreprise, l'entrepreneur ne connaît pas les coûts de sa création et c'est

¹⁰Selon Schumpeter [1983], seules les grandes entreprises sont capables de faire de l'innovation.

seulement *ex-post*, après l'entrée sur le marché, que sa connaissance des coûts s'améliore. L'auteur considère que la taille sous-optimale d'une entreprise est expliquée par le profil du créateur qui peut refléter ses capacités de gestion et d'organisation et ses compétences pour les négociations avec ses bailleurs de fonds. Dans ce modèle, un seuil d'efficience obtenu à partir d'une maximisation de la valeur de l'entreprise constitue le déterminant de sa survie (ou de la sortie) de l'entreprise. Si l'entreprise accroît son efficience, elle augmente ses chances de survie, sinon elle peut être amenée à sortir du marché. Cette analyse va dans le sens d'une préférence pour la grande taille.

On peut se demander pourquoi, des grandes entreprises laissent inexploitées certaines opportunités, tandis que des entreprises de petite taille à la création réussissent et peuvent accroître leur efficience ? En 1994, Audretsch a montré que la petite entreprise accroît son efficience et augmente ses chances de survie parce qu'elle est "porteuse de changement". De plus, on peut considérer qu'il existe une relation de complémentarité entre les entreprises de petites et de grandes tailles (Arrow [1983], Arora et Gambardella [1994] et Larue de Tournemine [1994]). Ceci s'explique par le fait que la présence des entreprises de grandes tailles peut inciter les petites entreprises à adopter une stratégie d'innovation plus offensive de manière à compenser le désavantage de taille.

En effet, les résultats obtenus sur notre échantillon confirment l'impact de la variable tranche d'effectif salarié sur la sortie pour reprise pour les entreprises innovantes.

La taille de l'entreprise n'est significative, que pour les entreprises innovantes. Le taux de sortie pour reprise des entreprises qui ont un effectif salarié compris entre 40 et 99 est environ la moitié de celui des entreprises de taille inférieure.

La variable de taille n'a pas d'effet significatif sur les entreprises qui ont un effectif salarié supérieur ou égal à 100. La taille de l'entreprise n'est pas significative pour les entreprises non innovantes.

Nom du ratio	Expression
Rentabilité économique et financière	
Rentabilité d'exploitation	$100 * (\text{EBE} / \text{capital d'exploitation})$
Rentabilité globale	$100 * (\text{EBG}^1 / \text{capital engagé})$
Rentabilité financière nette	$100 * (\text{CAFN}^2 / \text{fonds propres nets})$
Rentabilité financière	$100 * (\text{CAF}^3 / \text{financement propres nets})$
Part des dividendes dans la valeur ajoutée	$100 * (\text{Dividendes} / \text{Valeur ajoutée})$
Capital d'exploitation = immobilisations corporelles + BFR ⁴	
EBG = Valeur ajoutée globale - charges de personnel	
Capital engagé = actif immobilisé d'exploitation + BFR	
CAFN = CAF - charges de maintien du potentiel de production	
Structure productive et structure de bilan	
Coefficient de capital	$100 * (\text{capital d'exploitation} / \text{valeur ajoutée})$
Taux de valeur ajoutée	$100 * (\text{Valeur ajoutée} / \text{Chiffre d'affaires})$
Croissance	
Taux d'accumulation du capital	$100 * (\text{investissements} / \text{immobilisations corporelles})$
Fragilité financière	
Part de l'endettement dans le capital engagé	$100 * (\text{Endettement} / \text{capital engagé})$
Taux d'intérêt apparent	$100 * (\text{Frais financiers} / \text{endettement})$
Part des frais financier dans l'EBE	$100 * (\text{frais financier} / \text{EBE})$
Couverture des capitaux investis	$100 * (\text{Ressources stables} / \text{capitaux investi})$
Délai de rotation des BFR	$360 * (\text{BFR} / \text{Chiffre d'affaires})$
Crédit interentreprises	
Délai découvert client	$360 * (\text{découvert clients} / \text{production globale})$
Délai fournisseur	$360 * (\text{crédit fournisseurs - avances et acomptes}) / \text{D}^*$
Découvert clients = stocks de produits et travaux en cours - avances et acomptes reçus + créances clients et effets escomptés non échus	
D* = achats + sous-traitance + TVA déductible + volume d'affaires commission en principal	

¹ EBG : excédent brut global.

² CAFN : capacité d'autofinancement nette.

³ CAF : capacité d'autofinancement.

⁴ BFR : besoins de fonds de roulement.

6.1.3 Les rentabilités économique et financière

Afin de mesurer la performance économique de l'entreprise et d'étudier son impact sur le taux de sortie, nous nous référerons à la notion de rentabilité. Cette notion, qui mesure l'efficacité économique du processus de production, constitue l'un des meilleurs indicateurs de la santé de l'entreprise. Pour retracer les différents aspects de l'activité économique, l'Observatoire des Entreprises de la Banque de France adopte une description précise de la rentabilité du point de vue de l'activité d'exploitation et du point de vue de l'activité globale de l'entreprise.

La valeur de la rentabilité d'exploitation (*cf.* encadré) dépend à la fois du secteur d'exploitation, de la longueur du cycle d'exploitation, et des relations inter-entreprises¹¹. En revanche, la rentabilité globale (*cf.* encadré) est moins dépendante du secteur d'activité de l'entreprise¹² (Bardos [1995]).

Dans le cas d'une cessation pour reprise, la rentabilité d'exploitation n'a d'impact significatif ni sur les entreprises innovantes ni sur les entreprises non innovantes, alors que la rentabilité globale a un effet significatif, que les entreprises soient innovantes ou non.

L'augmentation d'une unité de la rentabilité globale entraîne une augmentation du taux de sortie de 3,7% pour les entreprises innovantes et de 1,6% pour les entreprises non innovantes.

De plus, pour les entreprises innovantes, la rentabilité financière (*cf.* encadré) a un impact négatif significatif. A contrario, ce ratio n'est pas significatif pour les entreprises non innovantes.

L'augmentation d'une unité de la rentabilité financière réduit le taux de sortie pour reprise des entreprises non innovantes de 0,1%.

Les résultats obtenus ne vont pas à l'encontre de l'intuition. On peut en effet penser que les reprises sont motivées par la rentabilité de l'activité. Mueller [1986] a constaté que lorsqu'une entreprise fusionne avec une autre entreprise de rentabilité nettement supérieure, le taux de rentabilité de la nouvelle entreprise est supérieur à celui des deux entreprises d'origine. Ces résultats sont cohérents avec le fait que l'objectif d'une reprise est d'acquérir un savoir faire et d'accroître les capacités de production, ainsi,

¹¹Le capital d'exploitation est défini comme suit :

capital d'exploitation = immobilisations + besoins de fonds de roulement (BFR).

Il dépend donc du crédit inter-entreprise à travers le BFR.

¹²L'approche globale de l'activité concerne toutes les opérations d'exploitation et hors exploitation.

en cas de concurrence, elle peut augmenter la production et faire baisser les prix afin de faire face à ses concurrents.

Il faut noter que, l'augmentation d'une unité de la rentabilité globale a un impact sur le taux de sortie pour reprise de 2,3 fois plus grand¹³ pour les entreprises innovantes que celui des entreprises non innovantes. De ce fait, nous pouvons considérer que l'augmentation du taux de cessation pour reprise peut être un signe de bonne performance économique des entreprises innovantes par rapport à celles non innovantes.

6.1.4 La fragilité financière

De nombreux travaux théoriques et empiriques confirment l'impact de la structure financière sur le risque de défaillance (Bardos [1995] et Rapinel [1994]...). De notre côté, nous avons testé l'impact de la structure financière des entreprises innovantes *versus* non innovantes sur leur taux de sortie. Si la hausse de l'endettement augmente à terme le risque de cessation d'activité, l'apport de fonds externes soulage immédiatement le niveau de la trésorerie de l'entreprise, ce qui peut retarder, même à très court terme, le risque de cessation d'activité. Pour détecter les difficultés financières d'une entreprise, il faut observer aussi bien l'endettement que le détail du crédit inter-entreprises.

Les entreprises qui n'ont pas recours à un financement externe pour leurs projets de création sont très rares. Les contraintes de financement constituent un handicap qui diffère selon que l'entreprise est innovante ou non innovante. Le ratio qui nous permet d'étudier l'effet de la fragilité financière sur la sortie est la part de l'endettement dans le capital engagé (*cf.* encadré). Ce ratio traduit l'impact de la structure financière sur le comportement des entreprises. Il est significatif pour les deux profils d'entreprise.

L'augmentation d'une unité de la part de l'endettement dans le capital engagé fait augmenter le taux de sortie pour reprise de 1% pour les entreprises innovantes et de 1,9% pour les entreprises non innovantes.

L'impact d'une augmentation de la part de l'endettement dans le capital engagé sur le taux de sortie des entreprises non innovantes est 2 fois plus important¹⁴ que celui des entreprises innovantes. Cette différence peut être expliquée par la variété du mode de financement des projets innovants : le financement d'un projet innovant fait souvent appel à des fonds propres, étant donné qu'un tel projet butte souvent sur la

¹³écart-réduit observé 47, 89.

¹⁴écart-réduit = 22, 39.

contrainte de sélection utilisée par les banques. Les projets non innovants, en revanche, font d'avantage appel à un financement bancaire.

Pour les entreprises innovantes, le délai découvert client (*cf.* encadré) est significatif et positivement corrélé avec le taux de sortie. Ce ratio n'est pas significatif pour les entreprises non innovantes.

L'augmentation d'une unité du délai découvert client entraîne une augmentation du taux de sortie des entreprises non innovantes de 0,6%.

Ce résultat montre bien l'impact du délai découvert client sur la sortie pour reprise des entreprises non innovantes. Ce résultat confirme la dépendance de ces entreprises *vis-à-vis* du crédit inter-entreprises sur le type de sortie, en plus des difficultés rencontrées au niveaux du financement bancaire.

En revanche, la couverture des capitaux investis (*cf.* encadré) n'a un effet négatif sur la sortie pour reprise que pour les entreprises non innovantes.

Une augmentation d'une unité de la couverture des capitaux investis entraîne une diminution de 0,5% du taux de sortie des entreprises non innovantes.

Afin de limiter tout risque de rupture de financement, les capitaux investis doivent être financés par des ressources stables. Ainsi, une détérioration de l'équilibre du haut du bilan envoie un signal négatif aux repreneurs.

Dans le cas d'une cessation pour reprise, on a montré que : la rentabilité économique et la fragilité financière ont des conséquences différentes sur les cessations pour reprises selon le caractère innovant ou non innovant de l'entreprise. Pour les entreprises innovantes, la variation de la rentabilité globale joue plus fortement sur la sortie par reprise que la variation de la part de l'endettement dans le capital engagé. Pour les entreprises non innovantes nous observons le résultat inverse.

Ainsi, les résultats de cette section mettent en évidence le fait que la reprise d'entreprises innovantes semble être plutôt motivée par des raisons de performances économiques (rentabilité économique plus importante), tandis que la reprise d'entreprises non innovantes s'explique davantage par une logique financière (part de l'endettement dans le capital engagé plus élevée).

6.2 La cessation totale : une croissance mal maîtrisée

On distingue trois motifs principaux de cessation totale d'activité : la cessation peut être volontaire (par exemple il s'agit d'un départ en retraite) ou involontaire (si elle survient suite à un accident, comme un incendie par exemple), ou bien la conséquence d'un déclin des performances économiques et financières.

Pour comprendre les causes d'une cessation totale, nous analysons les effets de l'innovation, du secteur d'activité, de la croissance, de la rentabilité et de la fragilité financière sur les taux de sortie.

6.2.1 Le type d'innovation

Le fait de faire de l'innovation incrémentale a un impact positif que ce soit sur la croissance interne de l'entreprise (continuité de l'activité) ou sur la croissance externe de l'entreprise (reprise).

Le taux de sortie pour une cessation totale pour l'entreprise faisant de l'innovation incrémentale est égal à 33,6% du taux de sortie pour cessation totale pour une entreprise qui ne fait pas d'innovation incrémentale.

Duguet [1996] montre que les déterminants d'une innovation incrémentale (amélioration, imitation) sont, en premier lieu, l'utilisation de biens d'équipement et en second lieu les relations avec la clientèle. En effet, l'innovation s'explique par l'exigence de modernisation des biens d'équipement, d'une part, et de satisfaction de la demande si l'innovation a conquis le marché avec succès, d'autre part. Ce résultat confirme l'intuition : l'entreprise qui améliore ses moyens de production ou ses produits a comme objectif de maintenir sa part de marché afin de faire face à la concurrence et de conserver ses chances de survie.

6.2.2 L'effet secteur d'activité

Le secteur d'activité n'a un impact significatif que pour les entreprises non innovantes.

Le taux de sortie pour cessation totale des entreprises du secteur des biens intermédiaires est égale à 47,8% du taux de sortie pour cessation totale d'une entreprise du secteur des biens d'équipement.

En revanche, les entreprises non innovantes du secteur des biens de consommation n'ont pas un profil de sortie pour cessation totale significativement différent des autres.

6.2.3 La croissance

Généralement, la croissance de l'entreprise est un indicateur favorable de sa santé. Cependant, une forte croissance peut être mal maîtrisée et conduire à la disparition de l'entreprise. Ce phénomène a été mis en évidence par un certain nombre de travaux sur les défaillances d'entreprises et principalement ceux de la Banque de France (Bardos [1995]). La croissance de l'entreprise peut être appréhendée par le taux d'accumulation du capital (cf. encadré).

L'augmentation d'une unité du taux d'accumulation du capital induit un accroissement du taux de sortie de 2% pour les entreprises innovantes et de 2.5% pour les entreprises non innovantes.

L'investissement est une décision qui engendre un coût immédiat et dont le rendement est incertain : c'est en un sens "un pari sur l'avenir". De ce fait, l'investissement est un engagement durable de l'entreprise dans un processus de production qui peut exercer une contrainte financière plus ou moins forte dans le futur.

"L'investissement n'implique pas seulement l'acquisition de capital sur le marché, mais aussi la transformation de ce capital "indéterminé" en capital spécifique et propre à l'usage de chaque entreprise. C'est principalement ce deuxième coût de spécification du capital qui produit un effet d'irréversibilité dans la décision d'investissement, c'est-à-dire qui rend plus coûteux de faire, puis de défaire, que de ne rien faire" (cf. Bourdier, Coeur et Sedillot [1997])

La décision d'investissement est liée à la demande et à la fluctuation de l'environnement de l'entreprise. Par conséquent, l'entreprise peut se trouver avec des capacités de production excessives suite à un retournement de la demande. Pour certaines entreprises, l'accroissement du niveau de l'investissement semble plus contraint par un besoin de renouvellement des machines que d'être le résultat d'une pure volonté d'augmenter les capacités de production. Ainsi, l'objectif de l'investissement est souvent de moderniser l'appareil productif dans le but de faire face à la concurrence et préserver la position sur le marché.

Afin de financer leurs investissements, les entreprises sont obligées de s'endetter, ce qui augmente leurs frais financiers et diminue leur chance de survie.

Par ailleurs, certains travaux récents, comme ceux de François et Favre [1998], ont permis de constater que pour les entreprises innovantes, l'investissement immatériel

correspondant aux dépenses d'innovation est devenu plus important dans le secteur industriel et a même dépassé le niveau des investissements corporels. On peut considérer que la spécificité d'un investissement en innovation réside dans sa rentabilité plus risquée et plus "lointaine" que celle d'un investissement en capital physique. Ainsi, ce type d'investissement risque tout particulièrement de fragiliser l'entreprise.

6.2.4 La structure productive

La valeur ajoutée traduit l'apport spécifique de l'entreprise dans le processus de production et de distribution d'un bien. Cet apport dépend étroitement de la qualification du personnel et des équipements utilisés. Le chiffre d'affaires correspond à la valeur hors taxe des ventes de la production et des marchandises en l'état. De ce fait, le taux de valeur ajoutée (*cf.* encadré) permet d'évaluer la consistance du surplus économique de l'entreprise par rapport à la valeur de ses produits finis.

L'augmentation d'une unité du taux de valeur ajoutée fait croître le taux de sortie pour une cessation totale de 2,7% pour les entreprises innovantes et de seulement 1,6% pour celles non innovantes.

Ce résultat s'explique par un apport de l'entreprise dans le processus de production et de sa façon d'utiliser les moyens de production (qualité du personnel et de l'équipement). L'entreprise qui gère mal son apport risque d'accroître ses risques de sortie du marché. Pour les entreprises innovantes, le taux de sortie pour une cessation totale est 1,7 fois plus grand¹⁵ que celui des entreprises non innovantes. Ceci peut être expliqué par un développement plus rapide et mal équilibré de l'activité des entreprises innovantes (en terme d'apport de bien d'équipement et de personnel qualifié), par rapport aux entreprises non innovantes.

Le coefficient de capital (*cf.* encadrée) est un ratio qui mesure le montant de capital nécessaire à la réalisation d'une unité de la valeur ajoutée. L'effet du coefficient de capital n'est significatif que dans le cas des entreprises non innovantes.

L'augmentation d'une unité du coefficient du capital entraîne une diminution du taux de sortie des entreprises non innovantes pour une cessation totale de 0,7%.

Ce résultat traduit le fait que plus l'entreprise est capitalistique plus elle a des chances de survie ou d'être reprise.

¹⁵écart-réduit = 9,61.

Le poids des besoins en fonds de roulement dans le chiffre d'affaires a une influence directe sur les performances et la structure financière des entreprises :

L'augmentation du poids des besoins en fonds de roulement dans le chiffre d'affaires entraîne une augmentation du taux de sortie de 0,6% pour les entreprises innovantes et de 1,2% pour les entreprises non innovantes.

Ce ratio peut traduire un alourdissement des besoins et entraîner des difficultés de paiement et une fragilité financière.

6.2.5 La rentabilité économique

La différence entre le taux de sortie pour cessation totale des entreprises innovantes et celui des entreprises non innovantes peut être expliquée par la rentabilité d'exploitation¹⁶.

L'augmentation d'une unité de la rentabilité d'exploitation diminue le taux de sortie des entreprises innovantes de 0,4%, alors que pour les entreprises non innovantes l'augmentation d'une unité de ce ratio entraîne une augmentation du taux de sortie de 0,3%.

Ceci est confirmé par l'impact de la rentabilité globale¹⁷ sur le taux de sortie qui n'est significatif que pour les entreprises non innovantes.

L'augmentation d'une unité de la rentabilité globale induit un accroissement du taux de sortie des entreprises non innovantes de 1,5%.

Par conséquent, l'accroissement de la rentabilité économique de l'entreprise non innovante ne suffit pas à maintenir ses chances de survie alors que cela semble être le cas pour une entreprise innovante.

De plus, la part des dividendes dans la valeur ajoutée risque d'aggraver la situation des entreprises qui sont les plus fragiles et même de diminuer leur chance de survie.

Pour les entreprises innovantes ou non innovantes une augmentation d'une unité de la rémunération des actionnaires augmente les taux de sortie respectivement de 3% et de 3,8%.

¹⁶ La rentabilité d'exploitation est le rapport entre l'excédent brut d'exploitation et le capital d'exploitation.

¹⁷ La rentabilité globale est le rapport entre l'excédent brut global et le capital engagé.

Ces résultats peuvent s'expliquer par la volonté des actionnaires de récupérer un maximum de bénéfice avant la cessation ou par une mauvaise stratégie et une mauvaise gestion du profit dégagé par l'entreprise. L'entreprise choisit de ne pas distribuer les dividendes afin de financer son projet d'investissement. Or, les résultats obtenus dans le cas des entreprises innovantes ou non innovantes nous permettent d'en déduire que les entreprises qui sortent ont plutôt fait le choix de rémunérer leurs actionnaires afin de maintenir leur confiance. Ceci peut traduire un dysfonctionnement dans l'organisation interne de l'entreprise. Il existe un conflit entre, d'une part, les dirigeants qui cherchent à développer l'activité de l'entreprise en augmentant l'investissement et, d'autre part, les actionnaires qui sont davantage intéressés par le versement de dividendes. Cette divergence de comportement entre les dirigeants et les actionnaires de l'entreprise risque d'aggraver sa situation économique et financière (Khanna [1995] et Bourdier et Sedillot [1993]).

6.2.6 La fragilité financière

En 1967, Baxter est l'un des premiers à avoir établi une relation entre le niveau d'endettement et le risque de défaillance de l'entreprise. Cette constatation a été confirmée par les travaux empiriques plus récents. De plus, il y a un impact de la taille sur le coût de l'endettement ce qui peut entraîner un risque de non paiement parmi lesquelles 90% ont été liquidées (sortie pour cessation totale) (Bardos [1990]).

Comme dans le cas des cessations pour reprise, la part de l'endettement dans le capital engagé explique de manière significative le taux de sortie. Ceci confirme les résultats de la Banque de France sur les défaillances.

Pour les entreprises innovantes et non innovantes, l'augmentation d'une unité de la part de l'endettement dans le capital engagé augmente leur taux de sortie respectivement de 2,6% et de 2,8%.

Ces résultats confirment et expliquent les difficultés financières liées à l'activité économique des entreprises. Ce résultat corrobore celui annoncé dans la contribution de Bardos (1990).

"[...] à taille égale, l'endettement est d'autant plus coûteux que le risque de défaillance est fort. L'effet du risque et de la taille sont ainsi en partie indépendants, et donc cumulatifs". (Bardos [1990])

C'est seulement dans le cas des entreprises innovantes que le taux d'intérêt apparent (*cf.* encadré) est significatif.

Une augmentation d'une unité du taux d'intérêt apparent fait augmenter la sortie pour cessation totale des entreprises innovantes de 0,7%.

Dans le cas des entreprises innovantes le taux d'intérêt apparent est influencé simultanément par l'effet taille (PMI) et par l'ampleur et la structure de l'endettement (Bardos [1990]).

Ces résultats sont confirmés par l'effet de la contrainte de solvabilité (*cf.* encadré) mesurée par la part des charges financières dans l'excédent brut d'exploitation. Ce ratio est très sensible à l'évolution de la situation de l'entreprise.

L'augmentation d'une unité de la part des charges financières dans l'excédent brut d'exploitation fait accroître le taux de cessation totale de 0,2% pour les entreprises innovantes et de 0,7% pour les non innovantes.

Ce résultat s'explique par la part des ressources que l'entreprise doit consacrer à la rémunération des capitaux empruntés une fois retranchés les coûts d'exploitation, ainsi que les charges salariales. L'augmentation des charges financières peut se traduire par un déséquilibre du bilan qui, à son tour, crée un risque d'insolvabilité. Dans les travaux sur la défaillance d'entreprises (Cheveau et Medjaoui [1987], Longueville [1992], Combier et Blazy [1998] et Bardos [1995]), ce ratio est considéré comme un indicateur expliquant significativement la défaillance d'entreprises (Bordes et Metzitz [1992]). Les petites entreprises sont moins armées pour négocier des taux d'intérêt favorables avec leurs banquiers. La faiblesse de leur fonds propres les rend plus dépendantes de l'accès au crédit.

Pour le crédit inter-entreprises, seul le délai de paiement des dettes fournisseurs a un effet significatif sur le taux de sortie pour cessation totale.

Une augmentation d'une unité du délai fournisseur accroît de 0,6% le taux de sortie des entreprises innovantes et de 1,1% celui des entreprises non innovantes.

Ce ratio traduit la fragilité et la dépendance de l'entreprise vis à vis de son fournisseur. L'augmentation du délai fournisseur peut s'interpréter comme étant la conséquence d'une détérioration de l'activité économique. Ce résultat a été mis en évidence dans plusieurs études (Dietsch et Kremp [1996 et 1998], Bardos [1996]). On remarque une différence significative de l'impact du délai fournisseur sur le taux de sortie pour

une cessation totale. Pour les entreprises non innovantes, le taux de sortie pour une cessation totale est presque 2 fois plus grand¹⁸ que celui des entreprises non innovantes.

Pour les entreprises innovantes, l'augmentation d'une unité de la couverture des capitaux investis induit un accroissement du taux de sortie pour cessation totale de 0.4%.

Etant donnée la vulnérabilité de certaines entreprises innovantes, le déséquilibre du financement peut entraîner la disparition totale de l'entreprise.

Le résultat principal de l'analyse des cessations totales d'activité dans le cas des entreprises innovantes et non innovantes est l'influence de l'effort d'un investissement intensif associé à la fragilité financière. Toutefois, cette influence est plus forte pour les entreprises non innovantes. Une croissance rapide expliquée par un taux d'investissement élevé suppose la mobilisation d'importantes ressources financières, ce qui n'est possible le plus souvent que par un recours à l'endettement. Or, le recours de l'entreprise au financement externe dépend du coût de l'endettement qui est généralement plus élevé pour les PMI. De ce fait, la fragilité financière peut provoquer l'insolvabilité de l'entreprise et sa disparition du marché. La rentabilité d'exploitation différencie les entreprises innovantes et des entreprises non innovantes. L'augmentation de la rentabilité est le signe d'un investissement important effectué dans le passé que supporte l'entreprise innovante, ce qui n'est pas le cas pour l'entreprise non innovante.

Ces résultats montrent l'importance de l'impact de l'endettement sur le taux de cessation totale par rapport à une cessation pour reprise.

Le tableau suivant (tableau 5) met en évidence les principales variables expliquant les deux types de sortie pour les entreprises innovantes et non innovantes.

¹⁸écart-réduit = 23.45.

**Tableau de synthèse: les variables explicatives pour chaque type de sortie dans le cas
des entreprises innovantes et non innovantes**

	Cessation pour reprise	Cessation totale
Entreprises innovantes	<ul style="list-style-type: none"> - innovation de produit - innovation faible - rentabilité globale - part de l'endettement dans le capital engagé 	<ul style="list-style-type: none"> - taux d'accumulation du capital - taux de la valeur ajoutée - part des dividendes dans la valeur ajoutée - part de l'endettement dans le capital engagé - délai fournisseur
Entreprises non innovantes	<ul style="list-style-type: none"> - rentabilité globale - part de l'endettement dans le capital engagé 	<ul style="list-style-type: none"> - rentabilité d'exploitation - rentabilité globale - taux d'accumulation du capital - taux de la valeur ajoutée - part des dividendes dans la valeur ajoutée - part de l'endettement dans le capital engagé - délai fournisseur

Conclusion

L'originalité de notre étude est double. À notre connaissance, la majorité des études portant sur la survie des entreprises ne distinguent pas les différents types de sortie (cessation pour reprise et cessation totale). Or, les motivations d'une cessation pour reprise sont totalement différentes de celles d'une sortie pour cessation totale. La cessation pour reprise est principalement motivée par la rentabilité économique et financière, en dépit de la fragilité financière de l'entreprise dans certains cas. À contrario, une cessation totale est généralement expliquée par la détérioration des performances économiques de l'entreprise qui entraîne sa disparition totale du marché économique. Par ailleurs, les déterminants qui ont été pris en compte dans l'étude de la survie des entreprises sont des variables de secteur, de taille, etc. Seules les études portant sur la défaillance d'entreprises ont tenu compte des variables financières. Le choix des variables issues de la liasse fiscal, nous a permis de mesurer l'impact des caractéristiques individuelles, économiques et financières, sur le type de sortie pour des entreprises innovantes et non innovantes.

Les résultats marquants de cette étude empirique sont l'existence d'une différence significative dans l'explication des taux de sortie (donc des taux de survie) selon les différents états de sortie et selon que l'entreprise est innovante ou non innovante. Dans le cas d'une cessation pour reprise, les résultats permettent de différencier les causes de sortie selon que l'entreprise est innovante ou non innovante. La cessation pour reprise des entreprises innovantes est plus fortement motivée par leurs performances économiques que par la fragilité de leur situation financière. En revanche, la cessation pour reprise d'une entreprise non innovante s'explique davantage par sa fragilité financière que par sa rentabilité économique. Cette différence peut être expliquée, d'une part, par le fait que les projets d'innovations sont des projets risqués et généralement bien étudiés (meilleure rentabilité économique) par rapport à certains projets de créations d'entreprises non innovantes. D'autre part, la fragilité financière plus marquée des entreprises non innovantes s'explique par leur mode de financement. Les entreprises non innovantes font davantage appel au crédit bancaire en cas d'insuffisance d'apport personnel, alors que les entreprises innovantes font plutôt appel aux organismes de capital-risque.

La cessation totale d'un projet d'innovation peut provenir d'une croissance rapide qui génère des coûts de financement difficiles à surmonter malgré une bonne rentabilité économique. Pour les entreprises non innovantes, la croissance rapide et la fragilité financière sont associées aux problèmes d'insuffisance de rentabilité et de pression

concurrentielle conduisant l'entreprise à disparaître. Dans ce type de sortie, les différentes variables expliquant la sortie pour cessation totale sont compatibles avec celles obtenues dans les travaux de la Banque de France et de l'INSEE sur la défaillance d'entreprises.

Annexe 1

Dans le cas d'un modèle à risques concurrents, la fonction $S(t, j)$ est complètement déterminée par les fonctions de survie jointes des durées latentes. Afin de faciliter la présentation du modèle, nous allons prendre par convention les notations suivantes¹⁹ :

$$S_{\bullet}(t_1, \dots, t_J) = \Pr(\tau_1 > t_1, \dots, \tau_J > t_J)$$

Quelque soit $(t_1, \dots, t_J) \in \mathbb{R}_+^J$, la fonction de survie des durées observées $T = \min\{\tau_1, \dots, \tau_J\} \in \mathbb{R}_+$ satisfait la relation suivante :

$$S_T(t) = S_{\bullet}(t, \dots, t)$$

La contribution de vraisemblance conditionnellement à la date de réalisation de l'enquête est égale à :

$$\begin{aligned} L_{T,A}(t, a \mid t_a, x) &= -\frac{d \Pr(T \geq t, A = a \mid x)}{dt \Pr(T \geq t_a)} \\ &= -\frac{1}{S_T(t_a)} \frac{d}{dt} \Pr(\tau_j \geq \tau_a \geq t, \forall j \mid x) \end{aligned} \quad (8)$$

Afin d'alléger les notations nous faisons abstraction du conditionnement par rapport aux variables explicatives.

Dans le cadre de notre modèle, on considère que les fonctions S_{\bullet} , et par conséquent S_T sont continûment différentiables et telles que :

$$S_{\bullet}(t_1, \dots, t_J) = \int_{t_j}^{\infty} -\frac{\partial}{\partial t_j} S_{\bullet}(t_1, \dots, t_J) dt_j$$

$S_{\bullet}(t_1, \dots, t_J)$ peut être écrite en fonction d'une probabilité conditionnelle :

$$S_{\bullet}(t_1, \dots, t_J) = \int_{t_j}^{\infty} S_{j,j}(t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_J \mid \tau_j = t_j) * f_j(t_j) dt_j$$

avec

$$S_{j,j}(t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_J \mid t_j) = \Pr(\tau_1 > t_1, \dots, \tau_{j-1} > t_{j-1}, \tau_{j+1} > t_{j+1}, \dots, \tau_J > t_J \mid \tau_j = t_j)$$

¹⁹ Les notations adoptées dans la suite de cette section sont similaires à celles de Mouchart et Rolin [1995].

où $\bar{j} = \{1, 2, \dots, J\} \setminus \{j\}$

A partir de ces deux équations nous pouvons en déduire la relation suivante :

$$-\frac{\partial}{\partial t_j} S_*(t_1, \dots, t_J) = S_{j/\bar{j}}(t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_J | t_j) * f_j(t_j)$$

On note pour $t_1 = t_2 = \dots = t_J = t$,

$$S_{j/\bar{j}}(t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_J | t_j) = S_{j/\bar{j}}(t)$$

Alors la contribution à la vraisemblance est :

$$\begin{aligned} L_{T,a}(t, a \mid t_d) &= -\frac{1}{S_T(t_d)} \frac{d}{dt} \int_t^{\infty} S_{a/a}(u) f_a(u) du \\ &= -\frac{1}{S_T(t_d)} S_{a/a}(t) * f_a(t) \\ &= -\frac{\partial}{\partial t_j} S_*(t_1, \dots, t_J) \end{aligned} \quad (9)$$

Or $\frac{\partial}{\partial t_j} S_*(t_1, \dots, t_J)$ peut être noté comme suit :

$$\frac{\partial}{\partial t_j} S_*(t_1, \dots, t_J) |_{t_1 = t_2 = \dots = t_J = t} = S_*(t, \dots, t) \frac{\partial}{\partial t_j} \ln[S_*(t, \dots, t)]$$

Soit la relation entre la fonction de hasard et la fonction de survie dans le cas univarié :

$$h_T(t) = -\frac{d}{dt} \ln S_T(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} \Pr(t \leq T < t + dt | T \geq t)$$

De plus, en notant $\{T = \min\{\tau_1, \dots, \tau_J\} > t\} \subset \{\tau_j > t\}$

dès lors $\{T > t\} = \{\tau_j > t, T > t\}$.

On définit alors la fonction de hasard conditionnel des durées latentes $\tau_j (1 \leq j \leq J)$ noté $h_j^*(t)$

$$\begin{aligned} h_j^*(t) &= \lim_{dt \rightarrow 0} \frac{1}{dt} \Pr(t \leq \tau_j < t + dt | \tau_j \geq t, T \geq t) \\ &= \lim_{dt \rightarrow 0} \frac{1}{dt} \Pr(t \leq \tau_j < t + dt | T \geq t) \\ &= -\frac{\partial}{\partial t_j} \ln S_*(t_1, \dots, t_J) |_{t_1 = t_2 = \dots = t_J = t} \end{aligned}$$

La fonction de vraisemblance est le produit des probabilités de sortie conditionnelles, pour les observations non censurées et de la survie à la fin de l'observation, pour les données censurées sachant la date de début de l'enquête innovation.

$$\begin{aligned}
L_{T,j}(t, j \mid t_{dt}) &= -\frac{1}{S_T(t_{dt})} \frac{\partial}{\partial t_j} \ln S_*(t, \dots, t) \\
&= \frac{1}{S_T(t_{dt})} S_T(t) h'_a(t)
\end{aligned} \tag{10}$$

La vraisemblance de l'ensemble de l'échantillon est donnée par :

$$\begin{aligned}
L_{T,\Lambda}(t, a \mid t_{dt}) &= \prod_{1 \leq i \leq n} \prod_{1 \leq j \leq 3} \left\{ \frac{1}{S(t_{dt_i})} S_T(t_i) h'_j(t_i) \right\}^{a_{i,j}} \\
&= \prod_{1 \leq i \leq n} \frac{S_T(t_i)}{S_T(t_{dt_i})} \prod_{1 \leq j \leq 3} \{h'_j(t_i)\}^{a_{i,j}}
\end{aligned} \tag{11}$$

Annexe 2

Si on considère que les risques sont indépendants. Ceci se traduit par l'indépendance des probabilité des durées latentes (τ_j). La fonction de survie est le produit des probabilités jointes :

$$S_{\bullet}(t_1, t_2, t_3) = \prod_{1 \leq a \leq 3} S_a(t_a) \quad (12)$$

avec $S_a(t_a) = \Pr(\tau_a > t_a)$

et $S_a(t) = \Pr(\tau_a > t)$

$$S_T(t) = \prod_{1 \leq a \leq 3} S_a(t) \quad (13)$$

La fonction $S(t, a)$ peut être écrite de la manière suivante :

$$\begin{aligned} S(t, a) &= \int_t^{\infty} \Pr(\cap_{i \neq a} \{\tau_i > s\} / \tau_a = s) * dF_a(s) \\ &= \int_t^{\infty} \prod_{i \neq a} S_i(s) * dF_a(s) \\ &= \int_t^{\infty} \frac{1}{S_a(s)} S_T(s) * dF_a(s) \end{aligned} \quad (14)$$

Or la fonction de hasard intégré est de la forme suivante :

$$H_a(t) = \int_0^t \frac{f_a(s)}{S_a(s)} ds = \int_0^t \frac{1}{S_a(s)} dF_a(s) \Rightarrow dH_a(t) = \frac{1}{S_a(s)} dF_a(s)$$

D'où

$$S(t, a) = \int_t^{\infty} S_T(s) * dH_a(s) \quad (15)$$

$$\text{Or } H_a(t) = -\ln S_a(t) \Rightarrow \frac{d}{dt} H_a(t) = -\frac{d}{dt} \ln S_a(t) = h_a(t)$$

$$\text{Or } H_T(t) = -\ln S_T(t) = -\ln \prod_{1 \leq a \leq 3} S_a(t) = \sum_{1 \leq a \leq 3} -\ln S_a(t) = \sum_{1 \leq a \leq 3} H_a(t)$$

$$dH_T(t) = d \sum_{1 \leq a \leq 3} H_a(t) dt = \sum_{1 \leq a \leq 3} dH_a(t) dt = \sum_{1 \leq a \leq 3} h_a(t) dt$$

$$h_T(t) = \sum_{1 \leq a \leq 3} h_a(t) \quad (16)$$

Or $h_T(t) = \sum_{1 \leq a \leq 3} h_a^c(t)$. On peut en déduire que dans le cas où les risques sont indépendants, la fonction de hasard marginale est égale à la fonction de hasard conditionnelle.

$$h_a(t) = h_a^c(t) \quad (17)$$

D'où la spécification de la fonction de vraisemblance est déterminée par la spécification du hasard marginal :

$$L_{T,A}(t, a) = S_T(t) * h_a(t) \quad (18)$$

Dans le cas où les risques sont indépendants, la vraisemblance de l'échantillon est égale à

$$L_{T,A}(t, a) = S_T(t) * \prod_{1 \leq j \leq 3} \{h_j(t)\}^{a_j} \quad (19)$$

Si nous retenons la même spécification pour la fonction de hasard, hasard proportionnel, la fonction de la log-vraisemblance de l'échantillon se présente comme suit :

$$\log L_{T,A}(t, a / t_d, x, \theta_j) = \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq 3} \left\{ \left[- \int_{t_d}^{t_i} h_{0j}(u/\alpha_j) \exp\left(\sum_{1 \leq k \leq p} x'_k \beta_j \right) du \right] \right. \\ \left. + a_{i,j} \log h_{0j}(t_i) + \sum_{1 \leq k \leq p} x'_k \beta_j \right\} \quad (20)$$

Annexe 3

Les variables économiques et financières retenues dans notre modélisation sont issues des données comptables du fichier FIBEN. L'utilisation de ce dernier a nécessité plusieurs opérations :

1) **Nettoyage des données comptables** : les variables comptables sont à la base du calcul d'un certain nombre d'agrégats. Elles sont aussi à la base de la détermination des ratios utilisés dans le modèle. Avant de calculer les ratios nous avons sélectionné les entreprises ayant des bilans cohérents et significatifs d'une activité normale²⁰. Les seules entreprises retenues ont :

- un chiffre d'affaires net positif ;
- une production positive ;
- une valeur ajoutée au coût des facteurs positive ;
- une valeur ajoutée globale positive ;
- un capital engagé positif ;
- une immobilisation corporelle positive.

Les entreprises ne répondant pas à ces critères sont considérées comme des cas particuliers qui risquent d'avoir un impact sur la robustesse des résultats de l'estimation. De ce fait, nous avons choisi de les éliminer de notre échantillon. Le nombre de cas²¹ éliminés représente 4,5% du total.

²⁰ A titre illustratif, dans le cas où la valeur ajoutée est négative, l'entreprise peut avoir soit une structure juridique spécifique (exemple coopérative), ou bien une situation très exceptionnelle. De ce fait, il faut l'éliminer de l'échantillon.

²¹ Il faut noter que le nombre de cas ne correspond pas au nombre d'entreprise, mais au nombre de fois qu'on trouve dans le fichier un agrégat négatif (chaque agrégat en plus de leur caractère individuel, présentent une dimension temporelle).

Tableau 6 : nombre d'agrégats négatifs dans l'échantillon issu de FIBEN

Agrégats négatifs ou nuls	nombre de cas
chiffre d'affaires	17
production	84
valeur ajoutée au coût des facteurs	204
valeur ajoutée globale	221
capital engagé	264
immobilisation corporelle	94

2) Le retraitement des ratios : les travaux menés par la Banque de France sur les défaillances d'entreprises ont montré que les ratios qui décrivent les risques de défaillance sont utilisés comme des indicateurs à jugement monotone. Ils permettent de dire si l'augmentation d'un ratio entraîne une augmentation ou bien une diminution du risque de défaillance. Afin d'avoir des jugements homogènes sur l'ensemble des entreprises et ne pas affecter les résultats de l'estimation, il paraît souhaitable de retraiter certains ratios.

Le retraitement concerne le cas où le numérateur et le dénominateur sont simultanément négatifs. Le ratio obtenu est positif. Ce qui donne une signification contradictoire à la réalité de l'entreprise. A titre d'exemple : la rentabilité financière est le quotient de la capacité d'autofinancement et du financement propre. La valeur positive de ce ratio est un bon indicateur de la situation financière de l'entreprise. Si cette valeur positive est obtenue suite à une capacité d'autofinancement négative et d'un financement propre négatif, notre commentaire est totalement contradictoire avec la réalité de l'entreprise. Afin d'éviter ce type de confusion, nous n'éliminons pas l'entreprise de l'échantillon, mais nous affectons à cette dernière la valeur médiane du ratio calculée à partir de l'ensemble des entreprises dont le dénominateur est positif.

Le retraitement concerne le cas où le dénominateur est nul, la valeur du ratio est indéfinie. Pour ne pas éliminer l'entreprise, nous affectons à ce ratio la valeur médiane calculée à partir des dénominateurs non nuls.

Dans le tableau suivant nous présentons le dénombrement des agrégats à valeurs négatives ou nulles. Ces agrégats sont le numérateur ou le dénominateur des ratios qui nécessitent un retraitement.

Tableau 7 : les variables à valeur négatives ou nulles des ratios qui nécessitent un retraitement dans l'échantillon.

Agrégats	valeur négative	valeur nulle
Excédent brut d'exploitation	954	1
Capital exploité	169	0
Capacité d'autofinancement	995	1
Financement propre	268	0
Capacité d'autofinancement nette	1766	4
Fonds propres nets	670	3
Capital investi	114	0
Ressource stable	39	0
Dividendes	0	4962
Concour courant bancaire	0	1564
besoin de fonds de roulement	847	1
Frais financiers	0	64

Enfin, le retraitement concerne le cas où l'entreprise a un ratio à valeur extrême, alors que les autres ratios de l'entreprise ont un niveau standard. Dans ce cas nous ne pouvons éliminer l'entreprise étant donné qu'elle est économiquement significative. Dans ce cas nous procédons un traitement par "bornage". Il consiste à ramener à une borne supérieure (percentil 99) les valeurs des ratios excédant cette borne, mais ramener aussi à une borne inférieure (percentil 1) les valeurs qui sont en dessous. Par conséquent, l'entreprise n'est pas éliminée de l'échantillon et la valeur extrême de ce ratio est neutralisée.

3) Le redressement des flux : les entreprises peuvent faire leurs déclarations plusieurs fois dans l'année et à durées d'exercices différentes. Afin d'homogénéiser les durées d'exercices des variables du compte de résultat (variables déterminées à partir de l'exercice de l'année) pour l'ensemble des entreprises de l'échantillon, nous procédons à un redressement par la formule suivante :

Tableau 8 : redressement des agrégats flux

Ratio redressé	formule de redressement
production	production* 12 mois/d5 _n ²²
valeur ajoutée	valeurs ajoutée*12 mois /d5 _n
valeur ajoutée au coût des facteurs	valeur ajoutée au coût des facteurs*12 mois/d5 _n
charges du personnel	charges du personnel*12 mois/d5 _n
excédent brut d'exploitation	excédent brut d'exploitation*12 mois/d5 _n
produits financiers	produits financiers*12 mois/d5 _n
produits et charges d'exploitation	produits et charges d'exploitation*12 mois/d5 _n
valeur ajoutée globale	valeur ajoutée globale*12 mois /d5 _n
capacité d'autofinancement	capacité d'autofinancement*12 mois/d5 _n
autofinancement	autofinancement* 12 mois/d5 _n
capacité d'autofinancement nettes	capacité d'autofinancement nettes*d5 _n /12 mois
dividende	dividende*12 mois/d5 _{n-1}

Les ratios qui ont été sélectionnés pour l'étude sont représentés dans le tableau 9.

²²d5_n : durée de l'exercice de l'année n.

Croissance	
Taux d'investissement productif	100*(investissement / valeur ajoutée)
Taux d'accumulation du capital	100*(investissements/ immobilisations corporelles)
Fragilité financière	
Part de l'endettement dans le capital engagé	100*(Endettement / capital engagé)
Part des fonds propres nets dans le capital engagé	100*(fonds propres nettes / le capital engagé)
Part des CBC dans l'endettement	100*(CBC / l'endettement)
Taux d'intérêt apparent	100*(Frais financiers / endettement)
Part des dividendes dans les fonds propres nets	100*(dividendes / fonds propres nets)
Part des dividendes dans le chiffre d'affaires	100*(dividendes / chiffres d'affaires)
Part des dividendes dans la valeur ajoutée	100*(dividendes / la valeur ajoutée)
Capacité de remboursement	100*(CAF /endettement)
Part des frais financier dans l'EBE	100*(frais financier / EBE)
Crédit interentreprises	
Part des créances douteuses dans les créances clients	100*(créances douteuses / créances clients)
Part des CDL ³ dans la valeur ajoutée	100*(CDL / la valeur ajoutée)
Délai découvert client	360*(découvert clients /production globale)
Délai fournisseur	360*(crédit fournisseurs -avances et acomptes)/D*
Délai de rotation des BFR	360*(BFR / Chiffre d'affaires)
Découvert clients =stocks de produits et travaux en cours – avances et acomptes reçus + créances clients et effets escomptés non échus	
D* =achats + sous-traitance + TVA déductible + volume d'affaires commission en principal	

³ CDL : clients douteux et litigieux.

Tableau 9 : les ratios sélectionnés pour l'étude

Nom du ratio	Expression
Rentabilité économique et financière	
Taux de marge	$100 * (EBE^1 / VACF^2)$
Rentabilité d'exploitation	$100 * (EBE / \text{capital d'exploitation})$
Rentabilité globale	$100 * (EBG^3 / \text{capital engagé})$
Rentabilité financière nette	$100 * (CAFN^4 / \text{fonds propres nets})$
Rentabilité financière	$100 * (CAF^5 / \text{financement propres nets})$
Part des dividendes dans la valeur ajoutée	$100 * (\text{Dividendes} / \text{Valeur ajoutée})$
<p>VACF = Valeur ajoutée + Subventions d'exploitation + loyers de crédit-bail + Personnel extérieur - Impôts et taxes</p> <p>Capital d'exploitation = immobilisations corporelles + BFR⁶</p> <p>EBG = Valeur ajoutée globale - charges de personnel</p> <p>Capital engagé = actif immobilisé d'exploitation + BFR</p> <p>CAFN = CAF - charges de maintien du potentiel de production</p>	
Structure productive et structure de bilan	
Intensité capitalistique	$100 * (\text{immobilisation corporelle} / \text{effectif salarié moyen})$
Coefficient de capital	$100 * (\text{capital d'exploitation} / \text{valeur ajoutée})$
Taux de valeur ajoutée	$100 * (\text{Valeur ajoutée} / \text{Chiffre d'affaires})$
VACF par effectif salarié moyen	$100 * (VACF / \text{effectif salarié moyen})$
Part de la VACF dans les charges du personnels	$100 * (VACF / \text{charges du personnels})$
VACF rapportés aux immobilisations corporelles	$100 * (VACF / \text{immobilisations corporelles})$
Couverture des capitaux investis	$100 * (\text{Ressources stables} / \text{capitaux investi})$
Part des CBC ⁷ dans les BFR	$100 * (CBC / BFR)$
Part des dettes fiscales et sociales dans la VACF	$100 * (\text{dettes fiscale} / VACF)$

¹ EBE : excédent brut d'exploitation.

² VACF : valeur ajoutée au coût des facteurs

³ EBG : excédent brut global.

⁴ CAFN : capacité d'autofinancement nette.

⁵ CAF : capacité d'autofinancement.

⁶ BFR : besoins de fonds de roulement.

⁷ CBC : concours bancaires courants.

Annexe 4

L'estimation du modèle

La méthode d'estimation choisie est donc le maximum de vraisemblance partiel. Une des propriétés intéressantes de cet estimation est qu'il dépend uniquement du rang de la durée, et non de sa valeur numérique.

L'appellation de hasard proportionnel traduit la proportionnalité du hasard de n'importe quel individu de l'échantillon par rapport à n'importe quel autre individu.

$$\begin{aligned} h_{ij}^c(t_i \mid t_{d_{i,j}}, x, \theta_{ij}) &= h_{0j}^c(t \mid t_{d_{i,j}}, \alpha_j) g(x, \beta_{ij}) \\ h_{i'j}^c(t_{i'} \mid t_{d_{i',j}}, x, \theta_{i'j}) &= h_{0j}^c(t \mid t_{d_{i',j}}, \alpha_j) g(x, \beta_{i'j}) \end{aligned}$$

avec $j = 1, 2, 3$

on a donc

$$\frac{h_{ij}^c(t_i \mid t_{d_{i,j}}, x, \theta_{ij})}{h_{i'j}^c(t_{i'} \mid t_{d_{i',j}}, x, \theta_{i'j})} = g(x, (\beta_{ij} - \beta_{i'j}))$$

avec $j = 1, 2, 3; i \neq i'$

$$h_{ij}^c(t_i \mid t_{d_{i,j}}, x, \theta_{ij}) = g(x, (\beta_{ij} - \beta_{i'j})) * h_{i'j}^c(t_{i'} \mid t_{d_{i',j}}, x, \theta_{i'j})$$

La fonction de vraisemblance est le produit des vraisemblances des individus de l'échantillon. Dans le cas de vraisemblance partielle (PL), elle est le produit de tous les événements observés. Si on considère L le nombre d'événements on peut noter :

$PL = \prod_{l=1}^L L_{lj}$ où L_{lj} est la vraisemblance pour le l ième événement de la j -ième sortie.

$$L_{lj} = \frac{h_{0j}^c(t \mid t_{d_{i,j}}, \alpha_j) \exp(x_l \beta_j)}{h_{0j}^c(t \mid t_{d_{i,j}}, \alpha_j) \exp(x_l \beta_j) + \dots + h_{0j}^c(t \mid t_{d_{i,j}}, \alpha_j) \exp(x_n \beta_j)} \quad (21)$$

où x_l est la valeur de x pour la l -ième entreprise. Suite à la simplification par $h_{0j}^c(t \mid t_{d_{i,j}}, \alpha_j)$:

- [27] Geroski P. A. (1991) : "Domestic and foreign entry in the United Kingdom : 1983-1984", in GEROSKI SCHWALBACH "Entry and market contestability, An international comparison", Basil Blackwell, p.63-88 ;
- [28] Geroski P. A. (1995) : "What do we know about entry?", International Journal of Industrial Organization, n°13, p.421-40 ;
- [29] Gharbi A. (1996) : "Les modèles de durée : application à une cohorte d'entreprises", Actes des Journées de Méthodologie Statistique 18-19 octobre 1995, Insee Méthodes, p.343-366 ;
- [30] Jovanovic B. (1982) : "Selection and the evolution of industry", Econometrica, Vol 50, n 3, May, p.649-70 ;
- [31] Khanna N. et Poulsen A. (1995) : "Managers of financially distressed firms : Villains or scapegoats?", Journal of Finance : vol.50, n°3, p.919-940 ;
- [32] Larue-de-Tournemine R. (1980) : "Progrès technique et dépenses dans les techniques avancées : le cas des dépenses spatiales européennes", Revue d'Economie Industrielle : 0 (13), 3ièmes trimestre 1980, Pages 38-53 ;
- [33] Lollivier S. (1997) : "Modèles univariés et modèles de durée sur données individuelles" Méthodologie statistique document de travail n°9702 INSEE ;
- [34] Longueville G. (1992) : "La multiplication des défaillances d'entreprises : contexte permissif ou fragilité financière, Lettre de conjoncture de la BNP, Juillet-Aout ;
- [35] Mata J., Portugal P. (1994) : "Life duration of new firms", The Journal of Industrial Economics, Vol, XLII n°3, p.227-45 ;
- [36] Mouchart M. et Rolin J.-M. (1995) : "Competing risks models : problems of modeling and of identification", Discussion paper, Institut de statistique, Université Catholique de Louvain, Louvain-la-Neuve Belgique ;
- [37] Mouchart M. (1999) : "The econometrics of duration data and of point process", Cours, Université Catholique de Louvain, Belgique ;
- [38] Mueller D.C. (1986) : "Profits in the Long Run", Cambridge, Cambridge University Press ;
- [39] Rapinel L. (1994) : "Caractéristiques financières des PMI ayant bénéficié d'un plan de redressement judiciaire",
- [40] SESSI (1998) : "Les compétences pour innover dans l'industrie, chiffres clé, références", Ministère de l'économie, des finances et de l'industrie, Secrétariat d'État à l'industrie ;

- [41] Shumpeter J. (1983) : "Histoire de l'analyse économique", III, p.60. 1983 ;
- [42] Viennet H. (1991) : "Créer ou reprendre une entreprise", Insee Première, n°159;

**CONCURRENCE, INNOVATION ET
INTERNATIONALISATION:
LA DYNAMIQUE DE L'ECONOMIE FRANCAISE**

**CONTRAT CGP
Subvention 5.98**

Vol.1 : LA DYNAMIQUE CONCURRENTIELLE EN FRANCE

Vol. 2 : LES DYNAMIQUES DE L'INNOVATION

Vol 3. Annexes du Vol.1

**CREST-LEI,
URA 2200 (GRECSTA)
Responsable : A. PERROT**

**EUREQua
Université de Paris I, UMR 8594
Responsable : D. ENCAOUA**

**VOLUME 3
*Annexes du Vol.1***

COMMISSARIAT GENERAL DU PLAN

Janvier 2000

LA DYNAMIQUE CONCURRENTIELLE EN FRANCE

- Annexe 1** Les Fusions-Acquisitions
- Annexe 2** Importations parallèles sur le marché automobile européen : une analyse en termes de différenciation verticale
- Annexe 3** Horizontal Mergers and Vertical Relationships
- Annexe 4** Intégration verticale avec source alternative d'approvisionnement
- Annexe 5** Mutualisation des risques et concurrence sur le marché de l'assurance
- Annexe 6** Allocating and Funding Universal Service Obligations in a Competitive Network Market

ANNEXE 1

Les Fusions-Acquisitions

Secteur d'activité	Nombre	Pourcentage	Valeur	Pourcentage
Banques, assurances, holdings	79	9.9	58.1	24.5
Distribution	91	11.4	19.5	8.2
Services	123	15.5	26.4	11.1
Edition, communication	106	13.3	21.8	9.2
Chimie, pharmacie	73	9.2	38.4	16.2
Transport, équipement de transport	66	8.3	29.9	12.6
Machines-outils, équipements électriques	61	7.7	5.1	2.2
Textile, habillement	27	3.4	0.6	0.3
Agro-alimentaire	78	9.8	18.5	7.8
Métallurgie, sidérurgie	34	4.3	1.4	0.6
Bâtiments et travaux publics	18	2.3	2.9	1.2
Bois, papier, mobilier	17	2.1	6.6	2.8
Restauration	0	0	0	0
Autres	22	2.8	7.9	3.3
Total	795		237.2	

Tableau 1 : Répartition des fusions & acquisitions d'entreprises françaises par secteur d'activité en nombre et en valeur (Mds de francs) pour l'année 1994. Source : *Fusions & Acquisitions Magazine*.

Secteur d'activité	Nombre	Pourcentage	Valeur	Pourcentage
Banques, assurances, holdings	152	15.1	50.5	31.6
Distribution	107	10.6	9.2	5.7
Services	147	14.6	13.3	8.3
Edition, communication	95	9.4	14	8.8
Chimie, pharmacie	78	7.8	30.9	19.3
Transport, équipement de transport	64	6.4	8.3	5.2
Machines-outils, équipements électriques	90	8.9	7.6	4.8
Textile, habillement	41	4.1	2.6	1.6
Agro-alimentaire	86	8.5	5.4	3.4
Métallurgie, sidérurgie	43	4.3	5.7	3.6
Bâtiments et travaux publics	29	2.9	5.6	3.5
Bois, papier, mobilier	43	4.3	1.4	0.9
Restauration	0	0	0	0
Autres	31	3.1	5.3	3.3
Total	1006		237.2	

Tableau 2 : Répartition des fusions & acquisitions d'entreprises françaises par secteur d'activité en nombre et en valeur (Mds de francs) pour l'année 1995. Source : *Fusions & Acquisitions Magazine*.

Secteur d'activité	Nombre	Pourcentage	Valeur	Pourcentage
Banques, assurances, holdings	191	18.9	221.8	46.3
Distribution	97	9.6	54.6	11.4
Services	183	18.1	50	10.5
Edition, communication	65	6.4	19.2	4
Chimie, pharmacie	72	7.1	32.5	6.8
Transport, équipement de transport	62	6.1	17.6	3.7
Machines-outils, équipements électriques	91	9	35.4	7.4
Textile, habillement	44	4.4	14	2.9
Agro-alimentaire	77	7.6	14.2	3
Métallurgie, sidérurgie	38	3.8	10.2	2.1
Bâtiments et travaux publics	18	1.8	3.5	0.7
Bois, papier, mobilier	28	2.8	2.6	0.5
Restauration	0	0	0	0
Autres	43	4.3	3	0.6
Total	1009		478.7	

Tableau 3 : Répartition des fusions & acquisitions d'entreprises françaises par secteur d'activité en nombre et en valeur (Mds de francs) pour l'année 1996. Source : *Fusions & Acquisitions Magazine*.

Secteur d'activité	Nombre	Pourcentage	Valeur	Pourcentage
Banques, assurances, holdings	176	16.7	247.5	52.1
Distribution	109	10.3	24.6	5.2
Services	198	18.7	33	7
Edition, communication	98	9.3	47	9.9
Chimie, pharmacie	70	6.6	40.8	8.6
Transport, équipement de transport	67	6.3	14.7	3.1
Machines-outils, équipements électriques	84	7.9	14.1	3
Textile, habillement	37	3.5	0.6	0.1
Agro-alimentaire	76	7.2	14.7	3.1
Métallurgie, sidérurgie	47	4.4	9.1	1.9
Bâtiments et travaux publics	17	1.6	3	0.6
Bois, papier, mobilier	26	2.5	0.6	0.1
Restauration	10	0.9	10.5	2.2
Autres	42	4	14.6	3.1
Total	1057		474.9	

Tableau 4 : Répartition des fusions & acquisitions d'entreprises françaises par secteur d'activité en nombre et en valeur (Mds de francs) pour l'année 1997. Source : *Fusions & Acquisitions Magazine*.

Secteur d'activité	Nombre	Pourcentage	Valeur	Pourcentage
Banques, assurances, holdings	140	15.3	238.5	43.7
Distribution	97	10.6	96.2	17.6
Services	184	20.1	77.1	14.1
Edition, communication	75	8.2	31.3	5.7
Chimie, pharmacie	62	6.8	19.2	3.5
Transport, équipement de transport	58	6.3	18.6	3.4
Machines-outils, équipements électriques	70	7.6	17.9	3.3
Textile, habillement	45	4.9	5.7	1
Agro-alimentaire	69	7.5	5.2	1
Métallurgie, sidérurgie	32	3.5	4.4	0.8
Bâtiments et travaux publics	10	1.1	3.9	0.7
Bois, papier, mobilier	21	2.3	3.7	0.7
Restauration	13	1.4	1.8	0.3
Autres	41	4.5	21.9	4
Total	917		545.6	

Tableau 5 : Répartition des fusions & acquisitions d'entreprises françaises par secteur d'activité en nombre et en valeur (Mds de francs) pour l'année 1998. Source : *Fusions & Acquisitions Magazine*.

Pays cible	1994		1995		1996		1997		1998		1994-1998	
	N	V	N	V	N	V	N	V	N	V	N	V
E-U	21	21.3	37	21.7	31	28.4	52	54.7	75	109.8	216	236
R-U	24	12.5	26	37	24	6.4	37	45.6	28	29.6	139	131.3
Belgique	15	2.1	0	0	9	25.5	7	1	11	61.4	42	90.4
Allemagne	35	3.2	24	4.5	23	3.4	29	9.7	27	15.7	138	36.7
Italie	19	3.2	15	2.8	22	1.7	22	8.8	37	10	115	26.9
Espagne	21	4.2	0	0	14	1.6	23	6.8	12	9.8	70	22.6
Brésil	0	0	0	0	10	3.1	13	1.6	12	12.8	35	17.3
Pays-Bas	4	4.1	0	0	0	0	9	5	10	5.1	23	14.3
Suède	3	2.3	3	1.4	4	2.4	2	2.6	0	0	12	8.7
Argentine	0	0	0	0	8	0.8	9	1.3	8	5.9	25	8
Australie	0	0	4	4.2	0	0	0	0	7	1.9	11	6.2
Suisse	4	2	4	0.9	5	1.7	0	0	0	0	13	4.7
Autres	56	6.3	102	9.4	84	7.7	96	20.2	107	28.5	445	72.4
Total	202	61.5	215	82.3	234	83.2	299	157.8	334	291.2	1284	676.3

Tableau 6 : Acquisitions françaises à l'étranger par pays pour la période 1994-1998 en nombre et en valeur (Mds de francs). Source : *Fusions & Acquisitions Magazine*.

Pays cible	1994		1995		1996		1997		1998		1994-1998	
	N	%	N	%	N	%	N	%	N	%	N	%
E-U	21	10.4	37	17.2	31	13.2	52	17.4	75	22.5	216	16.8
R-U	24	11.9	26	12.1	24	10.3	37	12.4	28	8.4	139	10.8
Belgique	15	7.4	0	0	9	3.8	7	2.3	11	3.3	42	3.3
Allemagne	35	17.3	24	11.2	23	9.8	29	9.7	27	8.1	138	10.7
Italie	19	9.4	15	7	22	9.4	22	7.4	37	11.1	115	9
Espagne	21	10.4	0	0	14	6	23	7.7	12	3.6	70	5.5
Brésil	0	0	0	0	10	4.3	13	4.3	12	3.6	35	2.7
Pays-Bas	4	2	0	0	0	0	9	3	10	3	23	1.8
Suède	3	1.5	3	1.4	4	1.7	2	0.7	0	0	12	0.9
Argentine	0	0	0	0	8	3.4	9	3	8	2.4	25	1.9
Australie	0	0	4	1.9	0	0	0	0	7	2.1	11	0.9
Suisse	4	2	4	1.9	5	2.1	0	0	0	0	13	1
Autres	56	27.7	102	47.4	84	35.9	96	32.1	107	32	445	34.7
Total	202		215		234		299		334		1284	

Tableau 7 : Pourcentages des acquisitions françaises à l'étranger par pays pour la période 1994-1998 en nombre. Source : *Fusions & Acquisitions Magazine*.

Pays cible	1994		1995		1996		1997		1998		1994-1998	
	V	%	V	%	V	%	V	%	V	%	V	%
E-U	21.3	34.6	21.7	26.4	28.4	34.2	54.7	34.6	109.8	37.7	236	34.9
R-U	12.5	20.4	37	45	6.4	7.7	45.6	28.9	29.6	10.2	131.3	19.4
Belgique	2.1	3.5	0	0	25.5	30.7	1	0.7	61.4	21.2	90.4	13.4
Allemagne	3.2	5.3	4.5	5.6	3.4	4.1	9.7	6.2	15.7	5.4	36.7	5.4
Italie	3.2	5.3	2.8	3.5	1.7	2.1	8.8	5.6	10	3.5	26.9	4
Espagne	4.2	6.9	0	0	1.6	2	6.8	4.3	9.8	3.4	22.6	3.3
Brésil	0	0	0	0	3.1	3.8	1.6	1.1	12.8	4.4	17.3	2.6
Pays-Bas	4.1	6.7	0	0	0	0	5	3.2	5.1	1.8	14.3	2.1
Suède	2.3	3.7	1.4	1.7	2.4	2.9	2.6	1.7	0	0	8.7	1.3
Argentine	0	0	0	0	0.8	1	1.3	0.8	5.9	2	8	1.2
Australie	0	0	4.2	5.2	0	0	0	0	1.9	0.7	6.2	0.9
Suisse	2	3.2	0.9	1.1	1.7	2.1	0	0	0	0	4.7	0.7
Autres	6.3	10.4	9.4	11.5	7.7	9.3	20.2	12.8	28.5	9.8	72.4	10.7
Total	61.5		82.3		83.2		157.8		291.2		676.3	

Tableau 8 : Pourcentages des acquisitions françaises à l'étranger par pays pour la période 1994-1998 en valeur (Mds de francs). Source : *Fusions & Acquisitions Magazine*.

Pays acquéreur	1994		1995		1996		1997		1998		94-98	
	N	V	N	V	N	V	N	V	N	V	N	V
E-U	49	16.9	51	46.9	63	30.4	88	39.6	94	42	345	176
R-U	42	36	40	11	41	28.8	56	30.7	45	47	224	153.8
Allemagne	26	15.8	37	4.8	34	14.3	33	38.2	39	78	169	151.2
Italie	9	2.1	22	1.6	20	5.9	20	12	15	26.3	86	48
Belgique	14	5.8	15	4.9	18	22.5	23	11.3	10	1.1	80	45.8
Suisse	14	6.6	0	0	15	24.3	11	9.5	7	0.5	47	41
Pays-Bas	11	2.1	0	0	22	4.2	18	4.4	15	30	66	40.8
Canada	0	0	7	6	14	3.4	8	5.8	7	2	36	17.3
Multi-pays	0	0	9	1.3	15	10.2	0	0	0	0	24	11.6
Suède	4	5.6	0	0	0	0	9	1.3	4	2	17	8.9
Irlande	3	5.7	0	0	0	0	2	1.8	0	0	5	7.6
Japon	0	0	0	0	9	3.3	6	2.2	3	0.4	18	5.9
Brésil	0	0	1	1.6	0	0	0	0	1	0.6	2	2.2
Autres	43	9.9	65	6.6	52	8.5	52	13.6	46	7	258	45.7
Total	215	106.9	247	84.8	303	156.2	326	170.9	286	237.3	1377	756.3

Tableau 9 : Acquisitions étrangères en France pour la période 1994-1998 par pays en nombre et en valeur (Mds de francs). Source : *Fusions & Acquisitions Magazine*.

Pays acquéreur	N	%	N	%	N	%	N	%	N	%	N	%
	E-U	49	22.8	51	20.6	63	20.8	88	27	94	32.9	345
R-U	42	19.5	40	16.2	41	13.5	56	17.2	45	15.7	224	16.3
Allemagne	26	12.1	37	15	34	11.2	33	10.1	39	13.6	169	12.3
Italie	9	4.2	22	8.9	20	6.6	20	6.1	15	5.2	86	6.2
Belgique	14	6.5	15	6.1	18	5.9	23	7.1	10	3.5	80	5.8
Suisse	14	6.5	0	0	15	5	11	3.4	7	2.4	47	3.4
Pays-Bas	11	5.1	0	0	22	7.3	18	5.5	15	5.2	66	4.8
Canada	0	0	7	2.8	14	4.6	8	2.5	7	2.4	36	2.6
Multi-pays	0	0	9	3.6	15	5	0	0	0	0	24	1.7
Suède	4	1.9	0	0	0	0	9	2.8	4	1.4	17	1.2
Irlande	3	1.4	0	0	0	0	2	0.6	0	0	5	0.4
Japon	0	0	0	0	9	3	6	1.8	3	1	18	1.3
Brésil	0	0	1	0.4	0	0	0	0	1	0.3	2	0.1
Autres	43	20	65	26.3	52	17.2	52	16	46	16.1	258	18.7
Total	215		247		303		326		286		1377	

Tableau 10 : Pourcentages des acquisitions étrangères en France pour la période 1994-1998 par pays en nombre. Source : *Fusions & Acquisitions Magazine*.

Pays acquéreur	1994		1995		1996		1997		1998		94-98	
	V	%	V	%	V	%	V	%	V	%	V	%
E-U	16.9	15.9	46.9	55.2	30.4	19.5	39.6	23.2	42	17.7	176	23.3
R-U	36	33.8	11	13	28.8	18.5	30.7	18	47	19.9	153.8	20.3
Allemagne	15.8	14.9	4.8	5.6	14.3	9.2	38.2	22.4	78	32.9	151.2	20
Italie	2.1	2	1.6	1.9	5.9	3.8	12	7.1	26.3	11.1	48	6.4
Belgique	5.8	5.5	4.9	5.9	22.5	14.4	11.3	6.6	1.1	0.5	45.8	6.1
Suisse	6.6	6.2	0	0	24.3	15.6	9.5	5.6	0.5	0.2	41	5.4
Pays-Bas	2.1	2	0	0	4.2	2.7	4.4	2.6	30	12.7	40.8	5.4
Canada	0	0	6	7.1	3.4	2.2	5.8	3.4	2	0.9	17.3	2.3
Multi-pays	0	0	1.3	1.5	10.2	6.6	0	0	0	0	11.6	1.5
Suède	5.6	5.2	0	0	0	0	1.3	0.8	2	0.8	8.9	1.2
Irlande	5.7	5.4	0	0	0	0	1.8	1.1	0	0	7.6	1
Japon	0	0	0	0	3.3	2.1	2.2	1.3	0.4	0.2	5.9	0.8
Brésil	0	0	1.6	2	0	0	0	0	0.6	0.2	2.2	0.3
Autres	9.9	9.3	6.6	7.8	8.5	5.5	13.6	8	7	3	45.7	6
Total	106.9		84.8		156.2		170.9		237.3		756.3	

Tableau 11 : Pourcentages des acquisitions étrangères en France pour la période 1994-1998 par pays en valeur (Mds de francs). Source : *Fusions & Acquisitions Magazine*.

	Allemagne	Belgique	Espagne	R-U	Italie	E-U
Acquisition d'une activité	4	1	0	4	1	0
Contrôle relevé (CR)	2	2	2	0	0	0
CR à 100%	0	0	1	0	0	1
Fusion	0	0	2	0	0	0
Participation	3	1	4	4	4	4
Prise de Contrôle (PC)	6	5	5	8	4	3
PC à 100%	16	2	3	7	8	10
Participation relevée (PR)	3	2	2	2	2	1
PR pour PC	1	2	2	0	0	1
PR pour PC à 100%	0	0	1	0	0	0
Total	35	15	22	25	19	20

Tableau 12 : Ventilation pour les 6 pays selon le résultat final de l'opération pour l'année 1994. Source : traitement des auteurs à partir de données de *Fusions & Acquisitions Magazine*.

Secteur d'activité de la cible	Nombre d'entreprises cibles	Acquises par des firmes du même secteur
Agro-alimentaire	19	17
Chimie, Pharmacie	25	21
Machines, Equipements électroniques et électriques	14	6
Communications	14	10
Electricité, Gaz et Services sanitaires	9	6
Distribution	14	10
Banques	9	9
Assurances	9	7
Sociétés à Portefeuille et d'Investissements	7	2
Services pour les entreprises	24	18
Total	201	134

Tableau 13 : Nombre d'opérations de F & A françaises à l'étranger, pour les principaux secteurs d'activité (année 1994).

	Allemagne	Belgique	R-U	Italie	Pays-Bas	Suisse	E-U
Acquisition d'une activité	3	0	3	0	0	1	4
Contrôle relevé (CR)	0	0	0	0	1	1	1
CR à 100%	0	0	1	0	0	0	0
Fusion	1	0	0	0	0	0	2
Participation	2	4	6	3	0	1	5
Prise de Contrôle (PC)	4	3	10	2	3	6	4
PC à 100%	7	2	15	2	4	3	17
Participation relevée (PR)	1	0	0	1	0	0	0
PR pour PC	0	1	0	1	0	0	0
PR pour PC à 100%	0	0	0	0	1	0	1
Total	18	10	35	9	9	12	34

Tableau 14 : Ventilation pour les 7 pays selon le résultat final de l'opération pour l'année 1994. Source : traitement des auteurs à partir de données de *Fusions & Acquisitions Magazine*.

Secteur d'activité de la cible	Nombre d'entreprises cibles	Acquises par des firmes du même secteur
Agro-alimentaire	15	11
Chimie, Pharmacie	18	17
Machines-Outils et Equipements de bureau	11	8
Machines, Equipements électroniques et électriques	8	4
Instruments de mesure, d'analyse et de contrôle	8	5
Communications	7	7
Distribution	15	11
Banques	7	4
Assurances	1	1
Sociétés à Portefeuille et d'Investissements	4	2
Services pour les entreprises	15	10
Total pour l'année 1994	163	149

Tableau 15 : Acquisitions étrangères en France, pour les principaux secteurs d'activité (année 1994).

	Nombre d'opérations
Acquisition d'une activité	36
Contrôle relevé (CR)	38
CR à 100%	7
Fusion	20
Participation	92
Prise de Contrôle (PC)	112
PC à 100%	190
Participation relevée (PR)	17
PR pour PC	17
PR pour PC à 100%	4
Recapitalisation	3
Echange d'actifs	1
Total	537

Tableau 16 : Répartition des opérations de F & A franco-françaises selon la nature du contrôle final pour l'année 1994.

	Allemagne	Belgique	Brésil	Espagne	R-U	Italie	Pays-Bas	E-U
Acquisition d'une activité	3	0	0	1	4	1	0	6
Contrôle relevé (CR)	0	1	0	0	0	0	1	0
CR à 100%	0	1	0	0	0	1	0	5
Fusion	0	0	0	0	0	0	0	3
Participation	4	2	3	3	1	11	1	6
Prise de Contrôle (PC)	12	6	6	4	16	10	5	20
PC à 100%	4	0	2	4	9	8	3	36
Participation relevée (PR)	2	0	0	0	0	1	0	0
PR pour PC	0	1	1	0	0	3	0	0
PR pour PC à 100%	0	0	0	0	0	1	0	0
Participation Croisée	1	0	0	0	0	0	0	0
Total	26	11	12	12	30	36	10	76

Tableau 17 : Ventilation pour les 8 pays selon le résultat final de l'opération pour l'année 1998. Source : traitement des auteurs à partir de données de *Fusions et Acquisitions Magazine*.

Secteur d'activité de la cible	Nombre d'entreprises cibles	Acquises par des firmes du même secteur
Agro-alimentaire	21	17
Edition, Presse, Imprimerie	12	8
Chimie, Pharmacie	29	22
Produits en Pierre, Argile, Verre et Béton	27	19
Industrie Métallurgique Primaire	10	9
Machines-Outils et Equipements de bureau	10	8
Machines, Equipements électroniques et électriques	23	18
Communications	13	10
Electricité, Gaz et Services Sanitaires	14	12
Distribution	24	18
Banques	17	16
Assurances	8	7
Sociétés à Portefeuille et d'Investissements	2	1
Services pour les entreprises	41	33
Service Variés	14	5
Total pour l'année 1994	335	252

Tableau 18 : Acquisitions françaises à l'étranger pour les principaux secteurs d'activité (année 1998).

	Allemagne	Belgique	R-U	Italie	Pays-Bas	E-U
Acquisition d'une activité	4	0	0	2	1	9
Contrôle relevé (CR)	3	1	1	0	2	1
CR à 100%	2	0	1	0	2	3
Fusion	0	0	0	0	0	1
Participation	1	1	16	1	2	12
Prise de Contrôle (PC)	17	3	9	3	4	27
PC à 100%	6	4	15	4	3	32
Participation relevée (PR)	1	0	0	1	1	0
PR pour PC	0	0	1	2	0	0
PR pour PC à 100%	0	0	0	0	0	1
Participation Croisée	1	0	0	0	0	0
Total	35	9	43	13	15	86

Tableau 19 : Acquisitions étrangères en France pour les 6 principaux pays selon le résultat final de l'opération pour l'année 1998. Source : traitement des auteurs à partir de données de *Fusions et Acquisitions Magazine*.

Secteur d'activité de la cible	Nombre d'entreprises cibles	Acquises par des firmes du même secteur
Agro-alimentaire	12	10
Chimie, Pharmacie	15	15
Machines, Equipements électroniques et électriques	17	11
Communications	11	8
Distribution	22	14
Banques	8	8
Assurances	4	4
Sociétés à Portefeuille et d'Investissements	4	2
Services pour les entreprises	39	27
Total pour l'année 1994	255	180

Tableau 20 : Principaux secteurs d'activité pour les acquisitions étrangères en France (année 1998).

	Nombre d'opérations
Acquisition d'une activité	38
Contrôle relevé (CR)	15
CR à 100%	20
Fusion	16
Participation	78
Prise de Contrôle (PC)	190
PC à 100%	151
Participation relevée (PR)	24
PR pour PC	19
Privatisation	1
Total	553

Tableau 21 : Répartition des opérations de F & A franco-françaises selon la nature du contrôle final (année 1998).

Secteur d'activité de la cible	Nombre d'entreprises cibles	Acquises par des firmes du même secteur
Agro-alimentaire	48	37
Industrie Textile	16	9
Confection	12	4
Edition, Presse, Imprimerie	28	15
Chimie, Pharmacie	21	15
Machines-Outils et Equipements de bureau	19	15
Machines, Equipements électroniques et électriques	22	15
Matériel de Transport	17	10
Communications	12	6
Distribution	69	39
Banques	13	12
Assurances	19	8
Immobilier	14	9
Services pour les entreprises	78	58
Sociétés à Portefeuille et d'Investissements	25	16
Total	553	360

Tableau 22 : Principaux secteurs d'activité des opérations de F & A franco-françaises (année 1998).

	1994	1995	1996	1997	1998	1994-1998
Agro-alimentaire	9	1	7	4	0	21
Chimie, pharmacie	8	8	5	4	5	30
Produits en pierre, Argile, Verre et Béton	8	1	2	2	4	17
Machines, Equipements électroniques et électriques	3	4	2	5	8	22
Télécommunications	10	1	8	4	8	31
Electricité, Gaz et Services sanitaires	1	3	5	2	7	18
Distribution	7	3	7	19	14	50
Banques	5	4	5	12	13	39
Assurances	12	4	5	8	10	39
Immobilier	4	2	2	6	10	24
Société à portefeuille et d'investissement	5	4	10	11	5	35
Services pour les entreprises	3	1	5	4	7	20
Total	103	49	82	109	121	464

Tableau 23 : Principales opérations de F & A par secteurs d'activité (plus de 500 MF pour l'année 1994 et plus d'un Md pour les autres années). Source : *Fusions & Acquisitions Magazine*.

ANNEXE 2

Importations parallèles sur le marché automobile européen : une analyse en termes de différenciation verticale

Importations parallèles sur le marché automobile européen : une analyse en termes de différenciation verticale

Séverine HALLER* Marie-Hélène JEANNERET†
(Version Préliminaire)

Octobre 1999‡

*E.S.S.E.C et EUREQua, Université Paris I, email : haller@univ-paris1.fr

†Université René-Descartes Paris V et EUREQua, Université Paris I.

‡ Nous tenons à remercier Simon Anderson et Victor Ginsburgh pour leurs commentaires et leurs encouragements, ainsi que tous les participants au groupe de travail d'organisation industrielle et du séminaire "comportements" de EUREQua à l'université Paris I et du THEMMA à l'université de Cergy-Pontoise

Résumé

Des écarts de prix importants sur les marchés automobile européens donnent lieu à des importations parallèles entre les pays de l'Union Européenne. Ce commerce "gris" va-t-il faire converger les prix vers un prix unique, plus concurrentiel ? Cet article a pour objet d'analyser les stratégies de prix d'une firme en monopole sur son marché domestique et en duopole sur un autre marché européen. Le produit est vendu par le biais d'un réseau de distribution exclusive dans les deux pays. La firme peut être concurrencée sur son propre marché par un ou plusieurs arbitragistes tirant partie d'un écart de prix. Nous montrons que lorsque le bien réimporté a une qualité perçue différente du "bien autorisé" et selon l'intensité de la concurrence dans le pays étranger la firme peut avoir intérêt à laisser l'arbitrage (les importations parallèles) se faire et continuer à pratiquer des prix différents entre les deux pays.

Mots clés : Automobile, Importations parallèles, Différenciation Verticale, Politique de la Concurrence

Codes JEL : L62, L13, L15, L4

Abstract

Important price gaps on european car markets give rise to parallel imports between countries of the European Union. Will this "gray market" make prices to converge towards a unique and lower price ? This paper analyses price strategies of a car producer which sells its product in two countries. In its own country its situation is monopoly whereas in the other country he is in competition with another producer. The product is sold by an exclusive dealing network in both country. One or several arbitragists can enter the domestic market to supply the good at a better price. We show that when the imported good has a lower perceived quality than the "authorized" product, and according to the degree of competition in the foreign country, it can be in the interest of the producer to let the arbitragist enter the market. Thus, the domestic firm can go on make different prices in the two countries.

Key Words : Automobile, Parallel Imports, Vertical Differentiation, Competition Policy

JEL Codes : L62, L13, L15, L4

1 Introduction

Le commerce parallèle consiste, pour une firme, à acheter un bien dans un pays où son prix est bas, pour le revendre dans un pays à prix élevé en concurrençant les réseaux de distribution (exclusive dans le cas de l'automobile) du bien. Dans l'Union Européenne ce type d'échanges a lieu pour toute une catégorie de produits pour lesquels on constate des différences de prix entre les pays membres. Le marché automobile s'y prête tout particulièrement.

En Europe en 1997, la part de marché estimée du commerce parallèle d'automobiles se situe entre 3% et 10%. Si l'on se base sur une part de marché moyenne de 6%, ce commerce représente environ 600000 véhicules. Jusqu'alors, la source la plus importante du *marché gris* (du terme anglais *grey market*) était l'Italie et la destination principale était l'Allemagne (pour l'année 1997, 330000 véhicules). La raison de ces échanges particuliers : les dévaluations importantes de la lire italienne, surtout par rapport au Deutsche Mark. En effet, les producteurs automobiles comme Volkswagen n'augmentaient pas leurs prix en Italie pour compenser la chute de la lire. Plutôt que de perdre des clients au profit de la grande marque nationale (Fiat), le producteur allemand maintenait des prix compétitifs sur le marché italien, subissant des pertes jusqu'au réalignement des monnaies. C'est ce que l'on appelle le *pricing to market*¹.

Aujourd'hui, avec les parités fixes irrévocables entre les monnaies européennes, ce type de stratégies disparaît. Mais pas le commerce parallèle. Celui-ci est maintenu par des conditions de marché hétérogènes entre les Etats membres. Les pays à bas prix comme la Belgique, le Luxembourg, l'Espagne ou le Portugal, sont des sources d'importations parallèles vers des pays comme la France, le Royaume-Uni ou l'Allemagne. Un taux de pénétration des importations plus important lié à l'absence d'un producteur domestique (comme en Belgique et au Luxembourg), des revenus moyens plus bas (comme en Espagne et au Portugal) sont autant de raisons qui amènent à un "morcellement" des conditions de concurrence entre les pays de l'Union. Par ailleurs, la distribution exclusive dans le secteur automobile permet aux constructeurs de maintenir leurs différences de prix. L'arbitrage prend alors place parallèlement aux réseaux de distribution exclusive. Or, paradoxalement, cet

¹ voir Goldberg et Verboven [1998] pour un état des lieux des explications empiriques de la dispersion des prix dans l'automobile

arbitrage, souvent l'oeuvre de mandataires² ou de garagistes indépendants, est désormais protégé par la Commission Européenne, alors même que celle-ci autorise la distribution exclusive dans le secteur automobile³. En effet, dans sa décision à l'encontre de Peugeot dans l'affaire IV/33.157_ "ECO System contre Peugeot", l'appréciation juridique portée par la Commission est que "les importations parallèles sont de nature à contribuer de façon significative à réduire les écarts de prix existant à l'intérieur de la Communauté, alors que [...] Peugeot [...] concourt à perpétuer le cloisonnement des marchés nationaux". La Commission Européenne annonce clairement sa politique en souhaitant, parallèlement à des règles de concurrence loyale, l'intégration économique des marchés.

L'analyse de la jurisprudence en matière de droit européen de la concurrence laisse, en effet, apparaître depuis le début des années 90, des sanctions systématiques ("*per se*") envers les firmes ayant empêché les importations parallèles de leurs automobiles vendues dans d'autres Etats membres. Il y a deux types d'importations parallèles, celles provenant des consommateurs eux-mêmes, et les importations d'entreprises (garagistes indépendants ou mandataires). C'est cette deuxième "catégorie" qui a plus particulièrement attiré notre attention. En France, relativement récemment⁴, un garagiste indépendant importait des véhicules neufs en concurrençant les concessionnaires exclusifs locaux. L'avis de la Cour de Justice fut favorable à l'activité de revendeur indépendant, les contrats signés entre les constructeurs et les concessionnaires exclusifs n'étant pas opposables à des tiers. Mais les constructeurs doivent aussi faire face à la présence sur le marché européen de mandataires, qui ne sont pas nécessairement des garagistes. Ainsi, à la fin des années 80, un mandataire français, Eco-System⁵, se vit-il refuser la vente de véhicules Peugeot en Belgique et au Luxembourg. Eco-System faisait bénéficier les consommateurs français de prix très attractifs dans ces pays. Peugeot se défendit auprès de la Commission en insistant sur le bénéfice du règlement d'exemption qu'elle avait obtenu. Elle fut néanmoins sanctionnée

²Un mandataire est une firme qui importe des véhicules automobiles sur l'ordre -le mandat- de consommateurs

³Règlement (CE) N°1475/95 du 28 juin 1995 "concernant l'application de l'article 85 paragraphe 3 du traité de Rome à des catégories d'accords de distribution et de services de vente et d'après-vente de véhicules automobiles". J.O. du 29.6.95, N° L145/25

⁴Arrêt de la C.J.C.E. Affaire du Grand Garage albigeois, C-226/94

⁵Voir aussi, Ginsburgh[1998]

pour restriction du commerce parallèle et donc du commerce entre États membres⁶.

Ces exemples, parmi d'autres⁷, laissent une interrogation. La Commission Européenne semble agir comme si tout arbitrage pouvait, *in fine*, conduire à la convergence des écarts de prix vers un prix unique, plus concurrentiel. Cependant, l'offre de l'importateur parallèle ne pourrait-elle pas coexister avec un réseau de distribution exclusive ou sélective? Autrement dit, peut-il y avoir de la discrimination par les prix malgré la présence d'un arbitragiste?

Dans la littérature, le commerce parallèle résulte d'un détournement des réseaux de distribution exclusive des producteurs opéré par des agents opportunistes (*free riding*). La cause de ce commerce est la discrimination par les prix à laquelle se livrent ces producteurs pour segmenter les marchés mondiaux, et notamment européens. Malueg et Schwartz [1994] font une analyse en terme de bien-être de la discrimination par les prix et des importations parallèles. Ils supposent que le commerce parallèle conduit nécessairement à une tarification uniforme de la part des firmes participant au commerce international. Dans ce contexte, ils montrent que les importations parallèles peuvent conduire à des situations Pareto dominantes pour la collectivité dans la mesure où le prix uniforme empêche certains marchés d'être servis. En effet, lorsque la demande est très dispersée (en termes de différences dans l'élasticité prix) entre les différents marchés, la discrimination du troisième degré par les prix⁸ permet d'atteindre de nouveaux consommateurs. Le bien-être mondial s'en trouve alors amélioré. Pour Anderson et Ginsburgh [1999], l'arbitrage affecte la discrimination par les prix entre les marchés. Dans le cadre d'un monopole vendant son bien dans deux pays, ils montrent que si tous les consommateurs ont le même coût d'arbitrage (qui peut être, par exemple, interprété comme un coût de transport), la firme gagne à segmenter les marchés mais alors le bien-être global de l'économie diminue. En revanche, si les consommateurs peuvent s'"auto-sélectionner" par des différences dans leurs coûts d'arbitrage (par exemple, ceux qui habitent près de la frontière ont un

⁶"Sont incompatibles avec le marché commun et interdits tous accords entre entreprises, toutes décisions d'association d'entreprises et toutes pratiques concertées, qui sont susceptibles d'affecter le commerce entre États membres...", Traité de Rome, art.85

⁷Le marché automobile illustre un phénomène que l'on retrouve pour d'autres produits (médicaments, alcool, matériel Tv-Hifi, téléphones mobiles)

⁸Discrimination parmi les consommateurs suivant des caractéristiques exogènes (Pigou, 1920)

coût d'arbitrage plus faible que les autres), le monopole peut servir l'autre marché (pays) même s'il n'y a aucune demande, afin de discriminer parmi les consommateurs du premier pays. L'augmentation du coût d'arbitrage est alors profitable tant pour le producteur que pour la collectivité. Nous reprenons l'analyse de la discrimination par les prix inspirée par la situation du marché européen de l'automobile en réutilisant l'hypothèse d'arbitrage coûteux pour les consommateurs. Ce coût reflète un risque de défaut de l'importateur parallèle, mais également un coût non monétaire lié à des temps d'attente plus longs entre la commande et la livraison et les lenteurs de la procédure d'immatriculation du véhicule. Dans notre modèle, nous supposons que ce coût est perçu par tous les consommateurs de façon objective comme une réduction de la qualité perçue du bien considéré, ce qui entraîne une disponibilité à payer le bien auprès de l'importateur parallèle plus faible qu'auprès du concessionnaire. C'est pourquoi on peut considérer que le bien "autorisé" et le bien offert par l'importateur parallèle sont des substituts imparfaits. Nous montrons que lorsque les consommateurs n'ont pas tous la même qualité perçue (ou encore différent suivant leur coût d'arbitrage), le constructeur peut avoir intérêt à ce que l'arbitrage ait lieu, autrement dit, à laisser entrer un ou plusieurs mandataires. Tout d'abord, en effet, le coût de l'arbitrage induit une situation dans laquelle toute la demande ne se reporte pas sur l'offre de prix plus avantageuse de l'arbitragiste. Ensuite, une partie des consommateurs qui, auparavant, n'achetaient pas le bien parce qu'il était trop cher, peuvent l'acheter par l'intermédiaire de cet importateur "hors circuit". Le constructeur sert donc une partie plus importante du marché dans le pays à prix élevé tout en continuant de pratiquer des prix discriminatoires entre les deux pays.

Cet article est construit de la manière suivante. Nous présentons le modèle de base dans la section 2. Puis nous caractérisons au cours de la section 3 l'incitation qu'il y a à empêcher l'installation d'un mandataire unique selon l'intensité de la concurrence et le coût de l'arbitrage. La section 4 propose une analyse en statique comparative des différents équilibres trouvés ainsi que des indications pour les politiques de concurrence. Enfin nous présentons nos conclusions dans la section 5.

2 Le modèle

Les hypothèses du modèle proviennent de l'observation des marchés automobile dans l'Union Européenne. Ainsi, on considère 2 pays E et F et 2 firmes, la firme 1 et la firme 2 produisant un bien différencié de qualité q_1 pour la firme 1 et q_2 pour la firme 2, avec $q_1 > q_2$. La firme 1 vend aux consommateurs des deux pays alors que la firme 2 ne vend qu'aux consommateurs du pays E. On reprend le modèle développé par Anderson et Ginsburgh [1999] mais on suppose que les conditions de concurrence sur les deux marchés considérés sont hétérogènes. On cherche ainsi à modéliser une situation où le pouvoir de marché des producteurs n'est pas le même suivant le pays⁹. On veut, en effet, montrer l'influence de l'intensité de la concurrence dans un pays étranger sur le comportement d'une firme domestique présente sur son marché et sur le marché étranger, vis à vis de l'entrée d'un arbitragiste qui réimporterait son produit. On se placera dans un cadre où les qualités des produits sont supposées exogènes : il ne semble en effet pas réaliste d'imaginer que l'entrée potentielle d'un arbitragiste puisse changer les décisions d'une firme en matière de qualité de son produit. L'existence d'un importateur parallèle est alors modélisée comme une situation où un agent A achète le bien 1 de qualité intrinsèque q_1 dans le pays E pour le revendre dans le pays F où la qualité perçue de ce produit sera inférieure à la qualité intrinsèque. On admet que cette diminution est totalement déterminée du côté des consommateurs, compte tenu du risque de défaut de l'importateur et des lenteurs de la procédure. On désignera par $q_1=1$ la qualité intrinsèque du produit 1, $q_A=1-s$ la qualité perçue par les consommateurs en F du produit 1 lorsqu'il est vendu par le mandataire et $q_2=1-s_2$ la qualité intrinsèque du produit 2.

Les firmes 1 et 2 et le mandataire ont comme objectif la maximisation de leurs profits respectifs. La firme 1 vend aux consommateurs du pays F à un prix p_1^F , aux consommateurs du pays E et au mandataire à un prix p_1^E . La firme 2 vend aux consommateurs du pays E à un prix p_2^E . Les coûts marginaux de production et de distribution des firmes 1 et 2 sont égaux et constants. Par souci de simplification nous les prendrons pour nuls dans ce modèle.

Le mandataire vend le bien 1 aux consommateurs du pays F à un prix p_A . Il supporte, en plus du coût d'achat du produit dans le pays E (p_1^E), un

⁹Le marché automobile belge, par exemple, est marqué par l'absence d'un constructeur domestique et, de ce fait, plus ouvert à la concurrence japonaise que le marché français.

coût de transport et de distribution unitaire $t \geq 0$.

Les coûts liés à l'arbitrage sont donc, d'une part, le coût de transport supporté par cette firme, et d'autre part un coût non monétaire subi par les consommateurs.

Le choix du contexte théorique de différenciation verticale est lié au fait que les véhicules automobiles sont des biens différenciés. En réalité, la différenciation est double. Verticale, par le choix parmi différentes qualités d'automobiles correspondant à différentes gammes de véhicules ; mais aussi horizontale, puisque le choix d'un véhicule dans un même gamme se fait suivant la préférence pour la marque, les options, les couleurs. Il nous a semblé préférable, dans un premier temps, de concentrer notre attention sur la dimension verticale de cette différenciation.

Les deux pays, E et F, sont supposés strictement identiques du point de vue des consommateurs. Cette hypothèse est certes restrictive mais elle est aussi raisonnable pour des domaines comme celui de l'automobile et pour des pays de niveaux de vie comparables. Chaque consommateur est caractérisé par un paramètre θ représentant sa préférence pour la qualité. On suppose que θ est uniformément distribué sur $[0, 1]$ pour chacun des pays considérés. La disponibilité $q_i\theta$ du consommateur θ à payer le produit i ($i=1, 2, A$) de qualité perçue q_i est croissante avec cette même qualité. Nous utilisons le cadre théorique de Mussa et Rosen [1978]. Le surplus du consommateur θ du pays j achetant le bien i est défini par $S_j(\theta, p_i) = q_i\theta - p_i$. Chaque consommateur achète une unité du bien qui lui donne le surplus maximum pourvu qu'il soit positif.

$$\text{En F : } S_F(\theta, p_i) = \begin{cases} \theta - p_1^F & \text{si achat du bien directement à la firme 1} \\ & (i = 1) \\ \theta(1 - s) - p^A & \text{si achat du bien à l'arbitragiste} \\ & (i = A) \\ 0 & \text{si pas d'achat} \end{cases}$$

$s \in]0, 1[$ représente la réduction qualité d'un achat du bien à A par comparaison à un achat à un revendeur exclusif. s est constaté par tous les consommateurs mais chacun le valorise différemment suivant sa disponibilité à payer pour la qualité.

En E, les consommateurs désireux d'acheter le bien ont le choix entre deux modèles : celui proposé également en F par le producteur 1 (de qualité

normalisée à 1) et celui proposé uniquement en E par la firme 2. La firme 2 propose un bien de qualité $(1-s_2)$. $s_2 \in]0, 1[$ est la réduction de qualité du bien 2 par rapport au bien 1. Le surplus net d'un consommateur en E est :

$$\text{En E : } S_E(\theta, p_i) = \begin{cases} \theta - p_1^E & \text{si achat du bien à la firme 1 (} i = 1 \text{)} \\ \theta(1 - s_2) - p_2^E & \text{si achat du bien à la firme 2 (} i = 2 \text{)} \\ 0 & \text{si pas d'achat} \end{cases}$$

3 Le jeu

On cherche l'équilibre de Stackelberg d'un jeu non coopératif à deux étapes où :

- 1) les firmes 1 et 2 déterminent leurs prix simultanément. La firme 1 maximise son profit global sur les marchés E et F et la firme 2 maximise son profit en E.
- 2) l'arbitragiste (ou mandataire) décide d'entrer dans le pays F et de proposer le produit de la firme 1 à un certain prix ou de ne pas entrer.

Il s'agit donc de résoudre ce jeu par récurrence vers l'amont. On s'intéresse dans un premier temps aux décisions de l'arbitragiste, puis à celles des firmes 1 et 2 compte tenu de ces décisions. Dans cette première étape les firmes 1 et 2 se font concurrence en prix en E, et la firme 1 choisit son prix de vente en F.

Dans un premier temps, on cherche donc les équilibres de Nash parfaits de ce jeu, puis on identifie les différents équilibres selon les valeurs des paramètres. Enfin, on étudie l'impact d'une variation des différents paramètres sur ces équilibres.

3.1 La décision du mandataire

Les prix p_1^F et p_1^E du bien 1 respectivement dans le pays F et dans le pays E, et le prix p_2^E du bien 2 en E sont considérés comme donnés. L'arbitragiste, A, cherche à maximiser son profit par rapport à son prix de vente. Les calculs des demandes et du prix de l'arbitragiste sont reportés en annexes A et B. Le programme de A est :

$$\begin{aligned} \underset{p_A}{Max} \Pi_A &= (p_A - (p_1^E + t)) D_A(p_A, p_1^E) \\ \text{avec } D_A(p_A, p_1^E) &= \frac{(1-s)p_1^E}{s(1-s)} \frac{p_A}{p_1^E} \text{ si } p_A < (1-s)p_1^E \\ &= 0 \text{ sinon} \end{aligned}$$

On obtient la décision suivante :

Si

$$p_1^E(1-s) > p_1^E + t \quad (1)$$

l'arbitragiste entre sur le marché F et propose :

$$p_A(p_1^E, p_1^E) = \frac{(1-s)p_1^E + p_1^E + t}{2} \quad (2)$$

Son profit d'équilibre est alors :

$$\Pi_A = \frac{[(1-s)p_1^E - (p_1^E + t)]^2}{4s(1-s)}$$

Sinon, A n'entre pas.

La condition (1) peut s'analyser de la manière suivante. Si l'arbitragiste veut avoir la possibilité d'être sur le marché, le consommateur marginal du produit 1 en F si A n'est pas là ($\theta = p_1^E$) doit avoir une disponibilité à payer le produit réimporté ($p_1^E(1-s)$) plus élevée que son coût ($p_1^E + t$).

3.2 Concurrence entre les firmes 1 et 2

Dans cette première étape les firmes 1 et 2 se font concurrence en prix en E, et, simultanément, la firme 1 choisit aussi son prix de vente en F. Dans un premier temps, on cherche donc les équilibres de Nash de ce jeu, puis on identifie les différents équilibres selon les valeurs des paramètres. Enfin, on étudie l'impact d'une variation des différents paramètres sur ces équilibres.

3.2.1 Résolution du jeu

Dans la suite, on calcule les différents candidats à l'équilibre possibles. On note que la firme 2 n'a jamais la possibilité de dévier (ie : tarifier un prix

qui exclut la firme 1 du marché) puisqu'elle produit une qualité inférieure à un coût de production identique à la firme 1. Puis on compare les profits réalisés par la firme 1 dans chacun de ces cas, l'équilibre étant le cas où ce profit est maximum.

On commence par se demander quels sont les prix (p_1^F, p_1^E) qui maximisent le profit de 1 dans l'ensemble des prix qui sont tels que A n'entre pas.

On sait qu'à l'équilibre la demande s'adressant à la firme 2 est non nulle de telle sorte que $p_2^E < (1 - s_2)p_1^E$. En effet, la firme 2 n'a jamais intérêt à tarifier $p_2^E > (1 - s_2)p_1^E$.

On peut montrer que Π_1 est concave en p_1^F et en p_1^E . De même que Π_2 est concave en p_2^E . Les prix de l'équilibre non contraint - ie : sans imposer au préalable que A n'entre pas - sont (Cf annexe C1) :

$$\begin{cases} p_1^{F*} = \frac{1}{2} \\ p_1^{E*} = \frac{2s_2}{3 + s_2} \\ p_2^{E*} = \frac{s_2(1 - s_2)}{(3 + s_2)} \end{cases} \quad (3)$$

Ces prix sont les prix d'équilibre du jeu auquel on s'intéresse et ils sont tels que effectivement A n'entre pas si et seulement si $s_2 > s_2^*(s, t) = \frac{3(1 - s) - 6t}{(3 + s + 2t)}$.

Lemma 1 Si $t > \frac{1-s}{2}$ alors A n'a jamais intérêt à entrer sur le marché.

Preuve. Annexe C1 ■

Lorsque le coût de transport supporté par A est important, il n'est pas intéressant pour lui d'entrer sur le marché. De même, plus la qualité perçue de l'achat à A est faible (s élevé) plus le coût de transport doit être faible pour permettre à A d'exercer son activité de mandataire.

On peut remarquer que dans ce cas, la firme 1 maximise ses profits séparément sur les deux marchés. Elle réalise ainsi un profit de monopole en F. Les profits sont alors les profits de "monopole naturel", Π_1^{M*} et Π_2^{M*} décrits en annexe C1.

Si $s_2 > s_2^*(s, t)$ la firme 1 est en "monopole naturel" sur le marché F. En revanche, dès que $s_2 \leq s_2^*(s, t)$, la firme 1 est menacée par la concurrence potentielle de A et a la possibilité d'empêcher l'entrée de A en pratiquant en F un prix limite $p_1^F = \frac{p_1^E - t}{(1 - s)}$. Les prix de l'équilibre contraint sont alors (Cf annexe C1) :

$$\begin{cases} p_1^{F*} = \frac{3(1-s)t + s_2(4+t) - s(2+t)}{4+5s_2 - 2s(3+5s_2) + s^2(3+5s_2)} \\ p_1^{E*} = \frac{2s_2(t - 3s + s^2 - 2t)}{3+5s_2 - 2s(3+5s_2) + s^2(3+5s_2)} \\ p_2^{E*} = \frac{(1-s_2)s(2-3s+s^2-2t)}{3+5s_2 - 2s(3+5s_2) + s^2(3+5s_2)} \end{cases} \quad (4)$$

Dans ce cas, les profits $\Pi_1^{\text{lim}^*}$ et $\Pi_2^{\text{lim}^*}$ sont les profits de l'équilibre "prix limite" (Cf. annexe C1)

On cherche maintenant les prix qui maximisent le profit de la firme 1 dans l'ensemble des prix qui entraînent une entrée de A sur le marché F. Dans ce cas, la firme 1 est en duopole avec A sur le marché F et en duopole avec la firme 2 sur le marché E. On obtient les prix et les profits suivants :

$$\begin{cases} p_1^{F/A*} = \frac{2s^2(3+5s_2) - 3t - s_2(8+t) + s(3(t-2) + s_2(2+t))}{2(3(1-s^2) - s_2(5-s^2))} \\ p_1^{E/A*} = \frac{2s_2(s - 2 + s^2 + t)}{(3(1-s^2) - s_2(5-s^2))} \\ p_2^{E/A*} = \frac{s_2(1-s_2)(s^2 + s + t - 2)}{(3(1-s^2) - s_2(5-s^2))} \end{cases} \quad (5)$$

Ces prix sont effectivement compatibles avec l'entrée de A sur le marché à l'étape suivante si $p_1^E(1-s) \geq p_1^E - t$ c'est à dire si et seulement si $s_2 < s_2^{**}(s, t)$, avec $s_2^{**}(s, t) = \frac{3(1-s)|2s^2 + t + s(3t-2)| - 10}{2s^2 - 5t - 2s(3+t) + s^2(4+3t)}$

Lemma 2 Une condition nécessaire à l'entrée de A sur le marché est que $t < \frac{2(1-s)s}{3s+1} < \frac{1-s}{2}$

Preuve. Annexe C2 ■

Lorsque le coût de transport est trop élevé, le mandataire A n'a jamais la possibilité d'entrer sur le marché.

Les profits des firmes 1 et 2, $\Pi_1^{D,A*}$ et $\Pi_2^{D,A*}$, dans cet équilibre de "duopole avec A" sur le marché F sont décrits dans l'annexe C2.

Le prix de A est alors :

$$p_A^* = \frac{2s(3(t-1) + s_2(t+7)) - s_2(16+7t) + s^2(3+5s_2)(4+t) - 2s^3(3+5s_2) - 9t}{4(s^2(3+5s_2) - 3 - 5s_2)}$$

et son profit :

$$\Pi_A^{D,A*} = \frac{\{2s^3(3+5s_2) - (3+5s_2)t + s^2(12+4s_2+9t+3s_2t) - 2s[3(1+t) + s_2(3+t)]\}^2}{16(1-s)(3+5s_2 - s^2(3+5s_2))^2}$$

¹⁰ Remarque : $s_2^{**}(s, t) < s_2^*(s, t)$. Cf. annexe C3

3.2.2 Les équilibres

Grâce à l'étude effectuée en 3.2.1, nous sommes maintenant en mesure d'identifier les différentes configurations d'équilibre. Les équilibres du jeu en deux étapes sont résumés dans les propositions 3 et 4.

Proposition 3 Soit $s_2^*(s, t) = \frac{3(1-s)ct}{(3+st+2t)}$. Si $s_2^*(s, t) < s_2 < 1$, le système de prix de l'équilibre ne contraint aucune des firmes à entrer : la firme 1 est en "monopole naturel" en F. Les prix sont ceux du système (3)

Preuve. Annexe D ■

Quand la concurrence en E est relativement faible (lorsque le paramètre s_2 est relativement élevé), alors l'écart de prix entre E et F n'est pas suffisant pour que A puisse entrer sur le marché. Dans ce cas, la firme 1 peut maximiser séparément son profit sur chacun des marchés : son comportement n'est, en fait, pas contraint par la menace d'entrée de l'arbitragiste.

De la même manière, on peut dire que cet équilibre a d'autant plus de chances d'émerger que le risque perçu par les consommateurs lors de l'achat à A est élevé (s élevé). Dans ce cas en effet la concurrence potentielle de A est peu agressive, et $s_2^*(s, t)$ relativement plus faible : A a donc moins intérêt à entrer sur le marché. Le même raisonnement tient pour le rôle du coût de transport, t .

Proposition 4 Il existe une valeur $s_2(s, t) \in]0, s_2^*(s, t)[$ telle que :

(i) Si $s_2(s, t) < s_2 < s_2^*(s, t)$ alors à l'équilibre la firme 1 bloque l'entrée de l'arbitragiste en pratiquant un prix limité. Le système des prix est (4).

(ii) Si $0 < s_2 < s_2(s, t)$ alors l'arbitragiste est présent à l'équilibre. Le système des prix est alors (5).

Preuve. Annexe D ■

Cette proposition nous permet de dire que, sous les hypothèses de ce modèle, lorsque la concurrence en E est assez forte (s_2 faible), la firme 1 a intérêt à laisser se réaliser l'arbitrage entre les marchés tout en continuant à fixer deux prix différents (cas (ii)). En effet, le profit plus faible réalisé dans le pays E du fait de l'intensité de la concurrence est ainsi compensé par les achats du mandataire. Dans la situation intermédiaire (cas (i)), la concurrence est telle que la firme 1 n'est plus en "monopole naturel", en ce sens que, pour les prix prévalant à cet équilibre, le mandataire a intérêt à entrer. Mais pour autant la firme 1 n'a tout de même pas intérêt à laisser entrer le mandataire,

elle pratique donc un prix limité.

Nous distinguons dans la suite deux cas. D'une part une situation de référence dans laquelle le coût de transport est nul : cela permet d'isoler les effets purement concurrentiels entre la firme 1, la firme 2 et l'arbitragiste. D'autre part le cas où t est non nul permet cette fois-ci d'isoler l'effet du coût de transport sur les équilibres, toutes choses égales par ailleurs.

Lorsque $t = 0$ le seul coût de l'arbitragiste est le prix p_1^B d'acquisition du bien en E. On a alors $s_2^*(s, t) = s_2^{**}(s, t) = \frac{3(1-s)}{3+2s} = \bar{s}_2$. On remarque que, logiquement, si s augmente ou si la disponibilité à payer le produit chez le mandataire diminue (risque plus élevé, délais plus long), alors l'équilibre de "monopole non contraint" a plus de chances d'émerger ($s_2^{**}(s, t)$ diminue). La concurrence potentielle du mandataire est en effet moins inquiétante pour la firme 1 si s augmente. Dans le même ordre d'idées, on remarque que les possibilités d'entrée du mandataire sont plus importantes. Les figures de l'annexe F montrent l'incitation de la firme 1 à laisser entrer le mandataire en fonction de l'intensité de la concurrence en E ($1-s_2$). Les figures 1 à 3 montrent, lorsque le coût de transport est nul, comment évolue cette incitation suivant le degré de concurrence de l'arbitragiste (que l'on peut mesurer par le niveau de $1-s$). On constate que plus la concurrence de A est forte (s faible) plus ses possibilités d'entrée sont importantes¹¹.

Pour $t \neq 0$ la figure 4 comparée à la figure 2 montre que, pour un s moyen, l'intervalle de s_2 pour lequel la firme 1 a intérêt à laisser entrer A est plus faible que lorsque $t = 0$. Les cas où A n'entre pas (s_2 relativement élevé) sont plus nombreux. Le coût de transport supporté par le mandataire a donc comme influence, toutes choses égales par ailleurs, de faire se déplacer les équilibres vers des valeurs plus faibles de s_2 , c'est à dire de diminuer les chances de voir émerger un équilibre avec présence de mandataire.

3.2.3 Analyse des prix dans chaque configuration d'équilibre

Nous allons maintenant étudier les effets d'une variation des paramètres de notre modèle (intensité de la concurrence - s_2 -, coût de transport - t - et risque de l'achat au mandataire - s -) sur les prix dans les différents équilibres que nous venons de trouver.

¹¹Ayant prouvé l'existence de s_2 mais n'ayant pas calculé sa valeur en fonction des paramètres, nous ne pouvons pour l'instant raisonner que sur la base de simulations.

Tout d'abord, une hausse de l'intensité de la concurrence à l'étranger (diminution de s_2) oblige la firme 1 à tarifier un prix plus faible dans ce pays ($\frac{\partial p_1^F}{\partial s_2} > 0$ dans les différents régimes d'équilibre). Toutes choses égales par ailleurs, cela améliore donc la situation du mandataire qui voit ses coûts diminuer puisqu'il achète moins cher dans le pays E. On distingue alors trois cas :

- soit les produits 1 et 2 restent malgré tout suffisamment différenciés et dans ce cas le prix du produit 1 en F ne change pas. On reste dans le cas du "monopole naturel".

- soit les produits 1 et 2 restent très peu différenciés et dans ce cas la baisse des coûts du mandataire (qui achète moins cher à l'étranger) entraîne une hausse de l'intensité de la concurrence dans le pays F où la firme 1 doit baisser son prix ($\frac{\partial p_1^F}{\partial s_2} > 0$).

- soit les produits restent moyennement différenciés, et dans ce cas, pour empêcher le mandataire d'entrer sur le marché, la firme 1 est obligée de diminuer l'écart de ses prix entre les deux pays ($\frac{\partial p_1^F}{\partial s_2} > 0$ et $\frac{\partial p_1^F}{\partial s_2} - \frac{\partial p_1^E}{\partial s_2} > 0$).

Il nous reste maintenant à étudier l'impact d'une variation des coûts d'arbitrage (monétaire, t ou non monétaire, s) sur les différents équilibres.

En premier lieu, une hausse des coûts de transport entraînera la hausse du prix de la firme 1 en F ($\frac{\partial p_1^F}{\partial t} > 0$) et ce que l'on soit dans le cas de l'équilibre avec mandataire ou dans celui du prix limite. En effet, lorsque t augmente, la concurrence (effective ou potentielle) du mandataire est moins dangereuse et la firme 1 peut augmenter son prix en F et le diminuer en E ($\frac{\partial p_1^E}{\partial t} < 0$). La baisse du prix de la firme 1 à l'étranger ne compense alors pas l'augmentation du coût de transport ($\left| \frac{\partial p_1^E}{\partial t} \right| < 1$) : elle permet donc à la firme 1 d'intensifier la concurrence dans le pays E tout en bénéficiant d'une hausse du coût du mandataire. On retrouve là un résultat de Anderson et Ginsburgh[1999] (à la différence près que dans leur modèle, ce coût était supporté par les consommateurs eux-mêmes). Lorsque t augmente, en effet, la concurrence du mandataire est moins rude et la firme 1 peut se permettre de tarifier son produit à un prix plus élevé dans le pays F, où elle était en monopole. De plus elle peut aussi se permettre de répondre de manière plus intensive à la concurrence de la firme 2 dans le pays E en diminuant son prix ($\frac{\partial p_1^E}{\partial t} < 0$). L'écart des prix du bien 1 entre les deux marchés se creuse donc avec t , ie : lorsque l'entrée est moins facile pour le mandataire.

Enfin, $\frac{\partial p_1^F}{\partial s} < 0$ confirme que lorsque l'entrée sur le marché est plus difficile pour A (s élevé indiquant un coût élevé pour les consommateurs en F achetant le bien à A), le prix en F diminue, puisque la menace de l'arbitrage est plus faible. La firme 1 peut donc "privilégier" l'intensité de la concurrence en E et diminuer son prix face à la firme 2, sans pour autant craindre de voir le mandataire lui nuire sur le marché F.

Après avoir analysé la statique comparative des prix par rapport aux paramètres du modèle dans chaque équilibre, nous proposons une analyse comparative de ces équilibres entre eux en termes d'écart de prix mais aussi de couverture du marché.

4 La Commission doit-elle autoriser le commerce parallèle ?

Dans cette section nous proposons une analyse en statique comparative des équilibres. L'analyse en termes de surplus social des deux pays s'avérant complexe, nous pouvons cependant, dans un premier temps, décomposer les effets prix et couverture de marché. On compare deux à deux tout d'abord la situation où le commerce parallèle dans le marché automobile est interdit à celle dans laquelle celui-ci est autorisé et effectif, puis, avec la situation où les mandataires sont permis mais absents du marché. Nous verrons les différences en termes de prix et de couverture de marché dans ces situations. Cette analyse devrait permettre d'apporter des éléments d'appréciation à la Direction Générale de la concurrence de la Commission Européenne sur le traitement du commerce parallèle dans le secteur de l'automobile mais également dans d'autres secteurs (téléphones mobiles...).

4.1 Si les mandataires étaient interdits

Cette situation fictive peut être analysée comme la situation déjà rencontrée dans la section 3 de "monopole naturel" de la firme 1 en F. Dans ce cas on aurait les couvertures de marché suivantes :

$$\begin{aligned} \text{En F : } D_1^F &= \frac{1}{2} \\ \text{En E : } D_1^E &= \frac{2}{3+s_2} \text{ et } D_2^E = \frac{1}{1+s_2} \end{aligned}$$

On trouve dans ce cas le système de prix (3) de la section 3.

4.2 Si les mandataires sont autorisés et un seul est sur le marché

Dans ce cas, la firme 1, qui joue la première, a le choix d'empêcher l'entrée en fixant un écart de prix limite entre les marchés E et F. Les couvertures de marché lorsque A est présent sur le marché F sont :

La demande de la firme 1 :

$$D_1^F = \frac{s+t}{2s}$$

La demande s'adressant au mandataire :

$$D_A = \frac{s^3(3+s_2) + s^2(3(t-2) + s_2(t-2)) - 3s(s_2-1) - (3+5s_2)t}{2s(s-1)(3-5s_2 + s(3+s_2))}$$

Sur le marché E on a la configuration suivante :

La demande en bien 1 par les consommateurs E :

$$D_1^E = \frac{2s^2 - (s+t)(1+s_2) - 1 - 3s_2}{s^2(3+s_2) - 3 - 5s_2}$$

Et la demande de la firme 2 :

$$D_2^E = \frac{s^2 + s + t - 2}{s^2(3+s_2) - 3 - 5s_2}$$

Le système des prix qui prévaut dans l'équilibre avec un mandataire en F est le système (5).

Lorsque le commerce parallèle des mandataires est autorisé, l'entrée d'un mandataire sur le marché provoque une baisse de la demande des deux firmes en E. La firme 2 tarifie un prix plus élevé lorsque le mandataire est présent. La demande qui s'adresse à la firme 1 en F augmente lorsque le mandataire est présent.

Nous avons vu que dans le cas d'une concurrence forte en E la firme 1 a intérêt à laisser entrer le mandataire. Qu'en est-il de l'écart des prix entre les marchés E et F par rapport à une situation dans laquelle le commerce parallèle par les mandataires serait interdit par la réglementation ? Dans ce cas fictif, les prix d'équilibre sont ceux de la situation dans laquelle la firme 1 maximise son profit global sans être "inquiétée" de l'entrée du mandataire. Dans les figures 5 et 6 nous comparons les écarts de prix dans les situations d'équilibre deux à deux : présence du mandataire et absence d'arbitrage, absence de mandataire et prix limite, prix limite et arbitrage. Quelles que soient les valeurs de s et de t, l'écart des prix avec arbitrage est plus faible que dans la situation de monopole en F pour la firme 1. La diminution des écarts en

présence du mandataire est d'autant plus forte que s_2 est faible, autrement dit, que la concurrence en E est forte. (Cf. Fig. 5.).

Nous trouvons donc que la présence sur le marché F d'un mandataire conduit bien à la réduction des différences de prix entre deux pays européens, sans toutefois que cette réduction amène à la convergence complète des prix vers le prix uniforme. Les raisons de ce résultat sont que les consommateurs ont un coût lié au risque de l'achat au mandataire, d'une part, et que, d'autre part, les conditions de l'offre (nombre de firmes présentes sur le marché étranger ainsi que la nature de la concurrence) sont différentes entre les deux marchés.

4.3 Si les mandataires sont autorisés mais absents du marché

Cette situation résulte d'une stratégie de prix limite de la part de la firme 1. Dans ce cas :

La demande en F qui s'adresse à la firme 1 s'écrit :

$$D_1^F = \frac{3(1-s)(1-s-t) + s_2(s^2 + s^t + (1-t))}{3(1-s)^2 + s_2(s^2 + 2s(1-s))}$$

En E, la demande à la firme 1 est :

$$D_1^E = \frac{2s^2 - s(3-s_2) + (1+3s_2+2s)(1-s_2)}{3+5s_2 - 2s(3+s_2) + s^2(3-s_2)}$$

Et la demande à la firme 2 :

$$D_2^E = \frac{2-3s+s^2-2t}{3+5s_2 - 2s(3+s_2) + s^2(3-s_2)}$$

Les prix à l'équilibre "prix limite" sont ceux décrits par le système (4)

Proposition 5 *La levée de l'interdiction du commerce parallèle par des mandataires sur le marché automobile implique :*

- i) *la demande qui s'adresse à la firme 1 en F augmente (et le prix baisse), que le mandataire entre effectivement ou que la firme 1 soit en stratégie de prix limite. La demande en F pour la firme 1 diminue (et le prix augmente).*
- ii) *La demande de la firme 2 diminue (et son prix augmente) sur le marché E, que le mandataire soit présent ou que la firme 1 soit en prix limite.*

Preuve. Elle résulte de la comparaison des différentes valeurs d'équilibre des prix et demandes. ■

Enfin, le plébiscite du commerce parallèle dans l'automobile par la Commission conduira inéluctablement vers une situation de concurrence forte entre les mandataires. Or nous pouvons montrer que la concurrence pure et parfaite entre les mandataires ne fait qu'accentuer les résultats trouvés lorsqu'un seul mandataire est présent sur le marché F.

5 Conclusion

Nous venons de montrer que le processus d'arbitrage entre les différents marchés géographiques européens ne conduit pas à la convergence des prix vers le prix unique. En raison des structures et comportements différents d'un pays à l'autre de l'Union européenne, les prix pour un même bien sont parfois loin de vérifier la "loi du prix unique". Dans ces conditions il peut être intéressant pour une firme vendant un bien dans deux pays européens à des prix différents de laisser l'arbitrage avoir lieu tout en continuant à "discriminer" en prix. La firme en place (le constructeur du pays F) peut alors profiter des coûts d'arbitrage différents selon les consommateurs afin de segmenter son propre marché. Le mandataire (l'arbitragiste de notre modèle) sert une partie du marché domestique que le constructeur n'atteignait pas en pratiquant son prix de monopole. Nous retrouvons ici le résultat de Anderson et Ginsburgh [1999].

Nous pouvons conclure que lorsqu'une firme a une position dominante (forte image de marque, préférence pour la marque domestique dans le cas de l'automobile) sur son marché domestique et que, par ailleurs elle est contrainte par les conditions de concurrence dans un autre marché, l'écart de prix qu'elle pratique est réduit par la présence d'importations parallèles sans toutefois être annulé. Cet écart est d'autant plus réduit que la concurrence en F est forte (Cf. Fig 5).

Par ailleurs, nous souhaiterions apporter à ce modèle quelques extensions. Il nous semble en particulier que la présence de la firme 2 également sur le marché F devrait restreindre le degré de liberté de la firme 1 dans ses stratégies de prix. On peut supposer qu'en F cette présence de la firme 2 se résume à la frange concurrentielle d'un marché dans lequel les firmes domestiques (deux constructeurs en France) ont une forte part de marché. Par ailleurs, une modélisation en différenciation horizontale de la concurrence en F serait plus proche de la réalité du marché automobile. En effet, on peut

supposer qu'un modèle de voiture est en concurrence avec les autres modèles d'une même gamme, c'est-à-dire à qualité égale. Dans ce contexte, il serait intéressant de rajouter cette dimension horizontale des préférences.

Références

- [1] Arrêt de la Cour de Justice des Communautés Européennes, 15 février 1996, Affaire C-226/94. "Grand Garage Albigeois"
- [2] Anderson S.P., V.Ginsburgh [1999], "International Pricing with Costly Consumer Arbitrage", *Review of International Economics* 7(1), pp.126-139
- [3] Décision de la Commission du 4 décembre 1991 Affaire IV/33.157_ "ECO System contre Peugeot"
- [4] Ginsburgh V.[1998], "Pourquoi y a-t-il des différences de prix dans l'Union Européenne? Le cas de l'automobile", Mimco U.L.B.,C.O.R.E
- [5] Goldberg P.K. et F. Verbeven [1998], "The Evolution of Price Discrimination in the European Car Market". N.B.E.R. Working Paper 6818
- [6] Livre Vert sur la Politique de Concurrence Communautaire et les Restrictions Verticales. Bruxelles. Janvier 1997. COM(96)721 final
- [7] Malueg D.A., M. Schwartz [1994], "Parallel Imports, Demand Dispersion, and International Price Discrimination", *Journal of International Economics* 37 pp.167-195
- [8] Mussa M. et Rosen S.[1978], "Monopoly and Product Quality", *Journal of Economic Theory* N° 18 pp.301-317
- [9] Pigon , A.C. [1920] "The Economics of Welfare", London :MacMillan
- [10] Règlement (CE) N° 117/95 de la Commission "concernant l'application de l'article 85 paragraphe 3 du traité à des catégories d'accord de distribution et de service après-vente de véhicules automobile"

ANNEXES

ANNEXE A : Fonctions de demande

Il s'agit de trouver les fonctions de demande s'adressant aux différentes firmes. Chaque consommateur va chercher à maximiser sa satisfaction, étant donné les conditions de l'offre. Deux cas sont possibles. Soit A n'est pas présent sur le marché et alors les consommateurs décident d'acheter ou non suivant la valeur de leur surplus net. Soit A est présent et propose p_A en élargissant l'ensemble des choix des consommateurs.

A.1. A n'est pas présent sur le marché Si A ne fait pas partie des offreurs, les consommateurs en F ont le choix entre consommer une unité du bien 1 et ne pas consommer. Ceci se traduit par les fonctions d'utilité suivantes :

Si $S_F(\theta, p_1) = \theta - p_1^F > 0$ achat d'une unité de bien 1

$S_F = 0$ si pas d'achat

Le bien sera effectivement acheté si $\theta > p_1^F$

On en déduit la demande s'adressant à la firme 1 :

$$D_1^F(p_1^F) = 1 - p_1^F$$

Si A n'entre pas, la demande s'adressant à la firme 1 en E sera composée uniquement de la demande totale pour le bien 1 en E. Un consommateur E a en effet le choix entre consommer une unité de ce bien ou une unité de bien 2. Étant donné les fonctions d'utilité décrites précédemment, ceci se traduit par les demandes aux firmes 1 et 2 respectivement :

$$\begin{aligned} D_1^E(p_1^E, p_2^E) &= s_1 \frac{p_1^E + p_2^E}{s_2} \\ D_2^E(p_1^E, p_2^E) &= \begin{cases} \frac{(1-s_2)p_1^E - p_2^E}{s_1(1-s_2)} & \text{si } p_2^E < (1-s_2)p_1^E \\ 0 & \text{sinon} \end{cases} \end{aligned} \quad (6)$$

A.2. A est présent et propose un prix p_A Lorsque A entre sur le marché, les consommateurs en F vont voir le choix de leurs "distributeurs" du produit s'élargir. Par ailleurs, ils ont une réduction de la qualité perçue s lorsqu'ils choisissent d'acheter le produit à A plutôt qu'à la firme 1.s, comme nous l'avons vu précédemment, est connu par tous mais valorisé différemment suivant les consommateurs. L'arbitrage en F est donc le suivant : Acheter le produit à la firme 1 si $\theta - p_1^F > 0$ et si $\theta - p_1^F > \theta(1-s) - p_A$. Tous les consommateurs de F qui sont tels que $\theta > p_1^F$ et $p_1^F - p_A < \theta s$ achètent le produit à la firme 1. Dans le contexte choisi pour cette section, les fonctions de demande s'adressant aux deux firmes, 1 et A, sur le marché F sont les suivantes :

$$D_1^{F/A}(p_1^F, p_A) = \frac{s - p_1^F + p_A}{s}$$

$$D_A(p_A, p_1^F) = \begin{cases} \frac{(1-s)p_1^F - p_A}{s(1-s)} & \text{si } p_A < (1-s)p_1^F \\ 0 & \text{sinon} \end{cases} \quad (7)$$

Sur le marché E, la demande s'adressant à la firme 2 ne change pas. Elle est toujours constituée de la somme des demandes unitaires individuelles des consommateurs préférant acheter la marque 2 au prix p_2^E plutôt que la marque 1 au prix p_1^E . Soit :

$$D_2^E(p_1^E, p_2^E) = \begin{cases} \frac{(1-s_2)p_1^E - p_2^E}{s_2(1-s_2)} & \text{si } p_2^E < (1-s_2)p_1^E \\ 0 & \text{sinon} \end{cases}$$

La firme 1, elle, fait face à une demande supplémentaire provenant de l'arbitragiste. Soit :

$$D_1^{E/A}(p_1^E, p_2^E) = s_2 \frac{p_1^E + p_2^E}{s_2} + \frac{(1-s_2)p_1^E - p_A}{s(1-s)}$$

ANNEXE B : décision de l'arbitragiste

Le profit de A est :

$$\Pi_A = (p_A - (p_1^F + t))D_1^{F/A} - (p_A - (p_1^E + t)) \left(\frac{(1-s_2)p_1^E - p_A}{s(1-s)} \right)$$

On montre assez trivialement que Π_A est strictement concave en p_A ($\frac{\partial^2 \Pi_A}{\partial p_A^2} < 0$) et l'on trouve la fonction de réaction de A grâce à la condition de premier ordre de maximisation de Π_1 ($\frac{\partial \Pi_1}{\partial p_A} = 0$)

ANNEXE C : Stratégies des firmes 1 et 2

C.1. $\{p_1^F, p_1^E, p_2^E\}$ est tel que A n'entre pas Les fonctions de profit de 1 et de 2 sont :

$$\Pi_1 = (1 - p_1^F)p_1^F + \frac{s_2 - p_1^E + p_2^E}{s_2} p_1^E$$

$$\Pi_2 = \frac{(1 - s_2)p_1^E - p_2^E}{s_2(1 - s_2)} p_2^E$$

A nouveau on montre que Π_1 et Π_2 sont strictement concaves en p_1^F et p_1^E et en p_2^E . La maximisation du profit sans contrainte de prix limite par la firme 1 a conduit au système de prix d'équilibre (3). Or ce système de prix ne vérifie la contrainte qui empêche A d'entrer sur le marché que si (1) est vérifiée, soit, si et seulement si on a la relation suivante entre les prix d'équilibre $p_1^{F*}(1 - s) < p_1^{E*} + t$, soit si et seulement si

$$\frac{1}{2}(3 - s) > \frac{2s_2}{3 + s_2} + t$$

$$s_2 > \frac{3(1 - s) - 6t}{(3 + s) + 2t}$$

Le système de prix d'équilibre (3) n'est possible que si $s_2 > s_2^*(s, t) = \frac{3(1 - s) - 6t}{(3 + s) + 2t}$

Par ailleurs, les profits dans non-contraint sont :

$$\begin{cases} \Pi_1^{N*} = \frac{6s_2}{(3 + s_2)^2} + 1 \\ \Pi_2^{N*} = \frac{(1 - s_2)}{(3 + s_2)^2} \end{cases}$$

Lorsque la stratégie d'équilibre de la firme 1 est le prix limite, les profits des firmes sont :

$$\left\{ \begin{array}{l} \Pi_1^{\text{lim}^*} = \frac{9(1-s)^2 t (s(1-t) + s_2^2) + 6s_2 t (s^2 t + 9s_2 - s(16+11t-2t^2)) + s^2 (t^2 - 3t + 4)}{(3+s_2 - 2s(3+2s_2) + s^2(3+s_2))^2} \\ \frac{s_2^2 [2s_2(8+2s^2 t - 2t^2) + 4s^2(1-t) + s^2(26-t-3t^2) + 3s(2t^2 + t - 8)]}{(3+s_2 - 2s(3+s_2) + s^2(3+s_2))^2} \\ \Pi_2^{\text{lim}^*} = \frac{(1-s)s_2(21s^2 - 3s - 2t)^2}{(3+s_2 - 2s(3+2s_2) + s^2(3+s_2))^2} \end{array} \right.$$

Démonstration du Lemme 1 : On observe que $s_2^*(s, t)$ est positif si et seulement si $t > \frac{1-s}{2}$. Dans le cas contraire la condition (1) est toujours vraie, ie : A n'est jamais sur le marché. Si (1) est vérifié alors on est sûr que le système des prix d'équilibre est une solution intérieure des programmes de maximisation des profits des firmes 1 et 2

C.2. $\{p_1^F, p_1^E, p_2^E\}$ est tel que A entre sur le marché En E les profits des firmes sont :

$$\begin{aligned} \Pi_1^{DA} &= p_1^{F/A} D_1^{F/A}(p_1^{F/A}, p_1^{E/A}) + p_1^{E/A} D_1^{E/A}(p_1^{F/A}, p_2^{E/A}) \\ &= \frac{s - p_1^F + p_A}{s} p_1^{F/A} + \frac{s_2 - p_1^E + p_2^E}{s_2} + \frac{(1-s)p_1^E - p_A}{s(1-s)} p_1^{E/A} \end{aligned}$$

$$\Pi_2^{DA} = p_2^{E/A} D_2^{E/A}(p_1^{F/A}, p_2^{E/A}) = \frac{(1-s_2)p_1^E - p_2^E}{s_2(1-s_2)} p_2^{E/A}$$

La maximisation des profits par les producteurs 1 et 2, avec la prise en compte par le producteur 1 de la réaction de A (2) dans sa fonction de profit (jeu de Stackelberg) conduit au système de prix (5) et aux profits d'équilibre correspondant

$$\left\{ \begin{array}{l} \Pi_1^{DA*} = \frac{4s^6(9+14s_2+s_2^2) - (3+s_2)^2 t^2 + 4s^3 [(- 2+t)(9+s_2^2) + s_2(6t-4)] + 2s(t-2)[9t+2s_2(16+7t) + s_2^2(32+9t)]}{8(1+s)s(3+5s_2-s^2(3+s_2))^2} \\ \frac{4s^2(9-18t-2s_2(17+6t-t^2) + s_2(2t^2-2t-39)) + s_2^2[9(-4+t^2) + s_2^2(-20+16t+t^2) + 2s_2(-52+24t+3t^2)]}{8(-1+s)s(3+5s_2-s^2(3+2s_2))^2} \\ \frac{4s_2^2(9(-s+t) + s_2^2(-61-8t+t^2) + s_2(-256+16t+16t^2))}{8(-1+s)s(3+5s_2-s^2(3+2s_2))^2} \\ \Pi_2^{DA*} = \frac{(1-s_2)s_2(s_2^2+t-2)^2}{(3+s_2-s^2(3+2s_2))^2} \end{array} \right.$$

La contrainte sur les prix $\frac{p_1^{1-t}}{1-s} < p_1^{F/A}$, qui permet à A d'entrer sur le marché est alors vérifiée si et seulement si $s_2 < s_2^*(s, t) = \frac{3(1-s)[2s^2+t+s(3t-2)]}{2s^2-5t-2s(3+t)+s^2(4+3t)}$.

Démonstration du Lemme 2 : On peut à nouveau montrer que $s_2^*(s, t)$ est positif si et seulement si $t < \frac{2(1-s)^2}{3+4s}$. Dans le cas contraire la condition (1) est toujours vraie, i.e. A n'est jamais sur le marché. Si $\frac{p_1^{F/A}+t}{1-s} < p_1^{F/A}$ est vérifié alors on est sûr que le système des prix d'équilibre est une solution intérieure des programmes de maximisation des profits des firmes 1 et 2.

C.3. On a $s_2^*(s, t) \geq s_2^{}(s, t) \Leftrightarrow \frac{3(1-s)-6t}{(3+s+2t)} \geq \frac{3(1-s)[2s^2+t+s(3t-2)]}{2s^2-5t-2s(3+t)+s^2(4+3t)} \Leftrightarrow \frac{12t[-2+13s-s^2+2t]}{(3+s+2t)(2s^2-5t-2s(3+t)+s^2(4+3t))} \geq 0$** Or on peut montrer que, d'une part, le numérateur de cette expression est négatif ou nul $\forall t \in [0, \frac{1-s}{2}]$ et que, pour ces mêmes valeurs de t , le dénominateur est strictement négatif.

ANNEXE D : Les équilibres

Démonstration de la proposition 3 : Les prix d'équilibre qui résultent de la maximisation non contrainte du profit par la firme 1 vérifient la condition nécessaire et suffisante pour que A n'entre pas sur le marché, i.e. :

$p_1^F(1-s) < p_1^F + t$ si et seulement si : $\frac{1}{2}(1-s) < \frac{2s_2}{3+4s_2} + t \Leftrightarrow s_2 > s_2^*(s, t) = \frac{3(1-s)-6t}{(3+s+2t)}$ (Cf. annexe C)

Démonstration de la proposition 4 :

Nous recherchons les valeurs de s_2 qui annulent la différence de profit entre les stratégies de prix limite et d'accomodation. Soit $s_2 / \Delta\Pi(s_2) = \Pi_1^{*DA}(s_2) - \Pi_1^{*lim}(s_2) = 0$

$\Delta\Pi(s_2) = 0$ est équivalent à une équation polynomiale de degré 4 en s_2 et on peut montrer qu'elle admet deux solutions complexes et deux solutions réelles. Parmi les deux solutions réelles, on trouve $s_2^*(s, t) = \frac{3(1-s)[2s^2+t+s(3t-2)]}{2s^2-5t-2s(3+t)+s^2(4+3t)}$ et une autre racine, $s_2(s, t)$. On peut montrer que cette deuxième racine réelle est comprise entre 0 et $s_2^*(s, t)$. En effet, $\Delta\Pi = \Pi_1^{*DA} - \Pi_1^{*lim}$ et chacun des profits Π_1^{*DA} et Π_1^{*lim} sont des fonctions strictement concaves en s_2 sur $[0, 1]$. Par ailleurs, $\Pi_1^{*DA}(0) = \frac{(2s-1)}{s(1-s)} > 0$ et $\Pi_1^{*lim}(0) = 0$. De plus, la dérivée de la

fonction $\Delta\Pi$ en $s_2^*(s, t)$ est strictement positive. On en déduit que $\Delta\Pi$ est d'abord positive et décroissante sur l'intervalle $]0, 1[$ puis croissante à partir d'une certaine valeur négative de s_2 . Par conséquent il existe une et une seule valeur $\underline{s}_2(s, t)$ solution de $\Delta\Pi(s_2) = 0$ et telle que $\underline{s}_2(s, t) \in]0, s_2^*(s, t)[$

ANNEXE E : Incitation de la firme 1 à laisser entrer l'arbitrage ($\Delta\Pi(s_2) = \Pi_1^{D^A}(s_2) - \Pi_1^{lim}(s_2)$ en ordonnées) **suivant l'intensité de la concurrence en E** (mesurée par s_2 en abscisses)

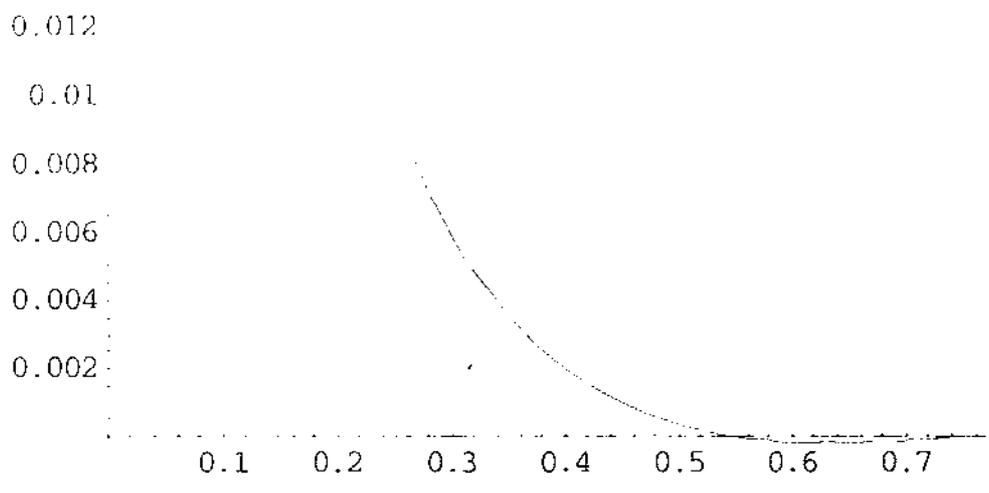


Fig 1 : $t=0, s=0.2, \bar{s}_2 = 0.75$

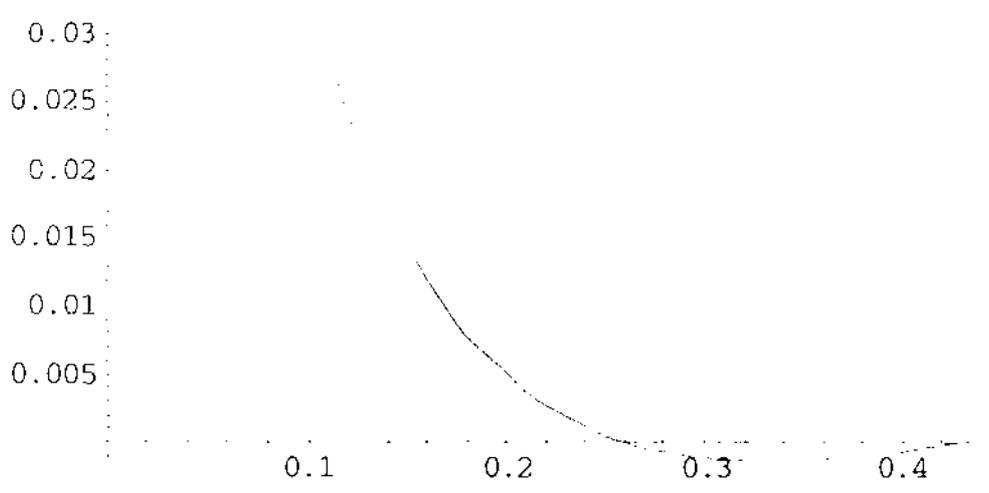


Fig 2 : $t=0, s=0.5, \bar{s}_2 = 0.428571$

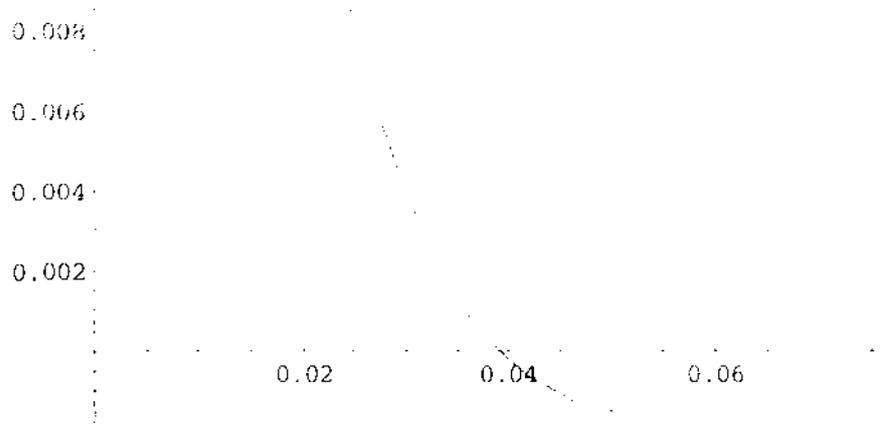


Fig 3 : $t=0$ et $s=0.9$, $\bar{s}_2 = 0.0769231$

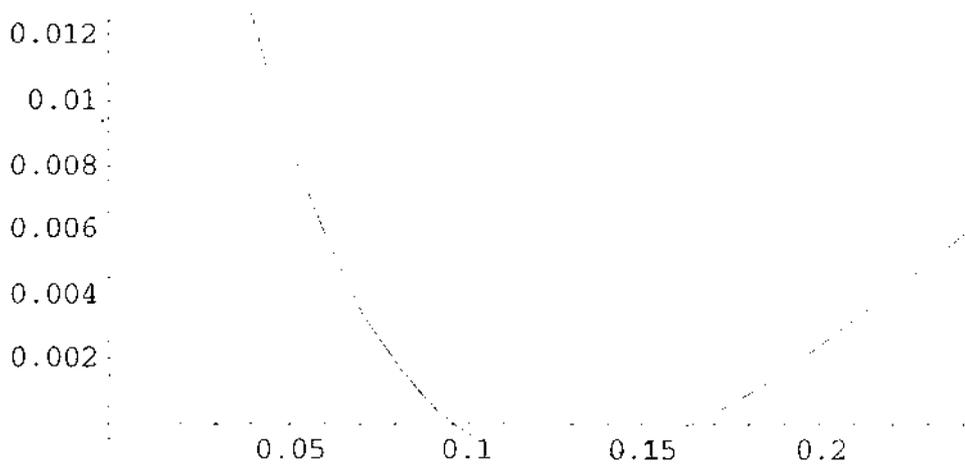


Fig 4 : $t=0.1$, $s=0.5$ et $s_2^*=0.243243$, $s_2^{**}=0.164835$

ANNEXE F : Les différences dans les écarts de prix d'équilibre selon s_2 , le degré de différenciation en E ($t=0$ et $s=0.5$).

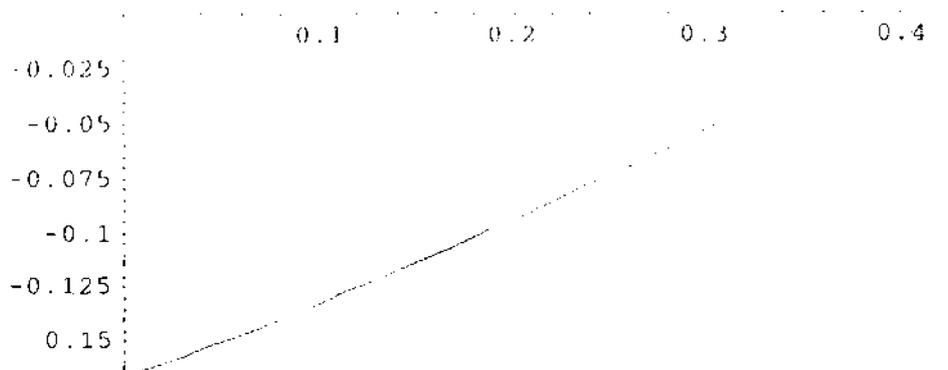


Fig 5 : Ecart de prix avec arbitrage-Ecart de prix monopole en F en fonction de s_2

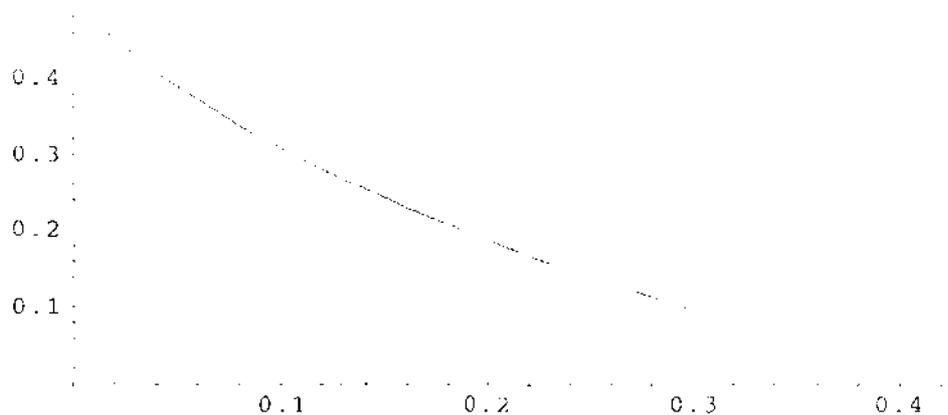


Fig 6 : Ecart de prix monopole en F-Ecart de prix limite en fonction de s_2

ANNEXE 3

Horizontal Mergers and Vertical Relationships

Horizontal Mergers and Vertical Relationships

Marie-Laure Allain* and Saïd Souam†

February 4, 2000

Abstract

In this paper, we propose a model that gives an interpretation of the high concentration of the retailing industry in the convenience goods market by comparing the incentives to merge for producers and retailers. In a two-level model where producers and retailers compete à la Cournot, we do a comparative static analysis and show that retailers face stronger incentives to merge than producers. Our analysis also reveals some surprising results: producers' incentives to merge increase with the number of retailers, and *vice versa*. Moreover, mergers involving a reasonable number of firms induce greater profit gains for retailers than for producers. Finally, the analysis of a simple merger game supports the previous results: we show that there is no equilibrium with merger at the upstream level, whereas mergers can occur in equilibrium at the downstream level.

1 Introduction

The retailing industry is becoming increasingly concentrated. The number of mergers and acquisitions has been particularly high among retailers in Europe during the last ten years. Furthermore, the concentration level in the food retailing industry has increased faster than in most of the other convenience goods industries: even if producers are already quite concentrated in some markets, retailers are merging faster

*CREST-LEI, and ENSAE, Timbre J 120, 3 avenue Pierre Larousse, 92 245 Malakoff Cedex, France (Tel : 33 1 41 17 65 30, E-Mail: allain@ensae.fr)

†CRESE, University of Besançon, EUREQua, and CREST-LEI, ENPC, 28 rue des Saints-Pères, 75007 Paris, France (Tel: 33 1 44 58 27 46; E-Mail: souam@ensae.fr).

and at a larger scale than their suppliers.

The size of firms influences their bargaining power in vertical relationships. The growing concentration of the retailing industry goes hand in hand with an increase in retailers' buying power. Of course, other factors have fostered retailers' strength as well. The development of own brands¹, for instance, and the increasing differentiation of retailers, have shifted the balance of power between producers and retailers. But the increasing concentration of the retailing sector also has a determining influence on the relationships between producers and retailers, in particular in the convenience goods markets. Some papers on vertical restraints² have stressed the market power of large retailers and their ability to determine market outcomes.

Retailers' buying power involved by the concentration of the retailing industry is an important issue in Europe. The subject was raised in the European Commission's Green Paper on Vertical Restraints, in January 1997. Investigations into mergers proposals have led to different conclusions among European countries. For example, in 1997, the European Commission examined and rejected the proposal of concentration between the two major Finnish retailers, Kesko (whose market share was about 40% of the Finnish market) and Tuko (20 % of the Finnish market). At the national level, retail power is also seen to be of concern, and is subject to enquiries³. For instance, the United Kingdom's competition authority led an investigation on this subject in 1999. Moreover, in 1997, the French Competition Authority analysed the concentration of the French retailing sector. It concluded that, even if retailers concentration has no negative impact on downstream competition in France, as retailers remain very competitive, the increasing concentration in the retailing sector significantly fosters retailers buying power.

Given the observed increase of concentration in the retailing sector, it seems of interest to study from a theoretical point of view the impact of a merger between retailers on their bargaining strategy with producers, and *vice-versa*. In particular, we may wonder whether retailers face stronger incentives to merge than producers or not. We also wonder whether a monopoly may emerge at one level of the market, as

¹See Mills (1995).

²See for example Comanor and Rey (1995) and Steiner (1985).

³See Dobson and Waterson (1999).

the outcome of a competitive merger game: that question is an important issue in terms of competition policy.

Several hypotheses have been put forward to explain the high concentration of the retailers. In particular, the existence of increasing returns in the retailing sector has been mentioned. In this paper, we give another interpretation for this phenomenon, relying on the comparison of the incentives faced by producers and retailers when there are strategic interactions between upstream and downstream firms. We study the firms' incentives to merge in a static setting, following Salant, Switzer and Reynolds (1983), Farrell and Shapiro (1990) and Gaudet and Salant (1991). At first, our analysis consists in studying the profitability of an exogenous merger (upstream or downstream) by doing a comparative static analysis between the initial Cournot-Nash equilibrium and the *post-merger* equilibrium, after s firms have merged. A group of s firms have an incentive to merge when the profit of the merged entity is greater than the sum of the profits of the s firms before the merger. In this framework, we compare upstream and downstream incentives to merge. Subsequently, we introduce a merger game, to endogenize the merger decision and the concentration process.

In section 2, we present a simple model with Cournot competition upstream and downstream, and where all firms face constant marginal costs. This first model gives the following intuition: retailers suffer more from a merger between producers than producers do from a merger between retailers. Section 3 presents a more general model where the firms' costs structure is modified to integrate the synergies due to a merger, and the increase of productivity of the merged firm. We show that, in this context, retailers' profits increase more than producers' profits after a merger, and that mergers among retailers are profitable for a smaller number of firms involved than among producers. Consequently, retailers have greater incentives to merge than producers. Section 4 introduces a static merger game, thereby endogenising the firms' decision to merge at one stage of the market. We show that, given the number of firms at both stages, a retailer is able to acquire a greater number of competitors than a producer. Moreover, we show that monopolization of the industry can be an equilibrium of the game for retailers when the number of downstream firms is sufficiently small, but such a monopolization cannot be an equilibrium of the merger game among producers. The last section concludes.

2 A Simple Model

Let us first study a very simple model which provides us with preliminary intuitions. We consider m producers competing à la Cournot in a homogeneous good market. They all face the same marginal cost denoted c . They sell their production on an intermediate market where n retailers stock up. All retailers face the same marginal cost γ . Moreover, they compete à la Cournot to sell this product to consumers.

Let us denote q_l^P the quantity produced by producer l and q_k^D the quantity bought by retailer k . In equilibrium in the intermediate market, the total produced quantity equals the quantity bought by the retailers ($\sum_{l=1}^m q_l^P = \sum_{k=1}^n q_k^D$). The inverse demand function in the downstream market is assumed to be linear, the size of consumers' potential demand being normalized to 1:

$$p\left(\sum_{k=1}^n q_k^D\right) = 1 - \sum_{k=1}^n q_k^D$$

In order to compute the wholesale price w on the intermediate market, we have to solve the downstream Cournot game. Retailer k maximises his profit. Solving the necessary and sufficient first order conditions gives the optimal quantity sold by a retailer, as a function of the wholesale price w :

$$q^D = \frac{1 - \gamma - w}{n + 1}$$

Moreover, the equilibrium between supply and demand on the intermediate market allows us to write the inverse demand function as follows, with $Q = \sum_{k=1}^n q_k^D$:

$$w(Q) = 1 - \gamma - \frac{n + 1}{n} Q$$

Producer l has then to solve the following program:

$$\text{Max}_{q_l^P} q_l^P (w - c) = q_l^P \left(1 - \gamma - c - \frac{n + 1}{n} \sum_{i=1}^m q_i^P \right)$$

This yields the following results, with Π^P the profit of each producer and Π^D the profit of each retailer:

$$q^P = \frac{n}{(n + 1)(m + 1)} (1 - \gamma - c)$$

$$\begin{aligned}
q^D &= \frac{m}{(n+1)(m+1)}(1-\gamma-c) \\
Q &= \frac{nm}{(n+1)(m+1)}(1-\gamma-c) \\
w &= \frac{(1-\gamma)+mc}{m+1} \\
p &= 1 - \frac{nm}{(n+1)(m+1)}(1-\gamma-c) \\
\Pi^P &= \frac{n}{n+1} \frac{(1-\gamma-c)^2}{(m+1)^2} \\
\Pi^D &= \frac{m^2}{(m+1)^2} \frac{(1-\gamma-c)^2}{(n+1)^2}
\end{aligned}$$

We hereafter study the incentives to merge respectively for the producers and the retailers. For this purpose, let us consider an initial configuration where the number of firms is the same in both markets ($m = n$). Both markets are then equally concentrated.

2.1 Upstream Merger

Suppose that s producers merge. The n retailers which still compete à la Cournot, now face $n - s + 1$ producers, whose strategic variables remain the produced quantities.

The profits are then given by:

$$\begin{aligned}
\Pi_{Um}^P &= \frac{n}{n+1} \frac{(1-\gamma-c)^2}{(n-s+2)^2} \\
\Pi_{Um}^D &= \frac{(n-s+1)^2}{(n-s+2)^2} \frac{(1-\gamma-c)^2}{(n+1)^2}
\end{aligned}$$

2.2 Downstream Merger

In the case where s retailers merge, and no merger occurs among the producers, the results are as follows:

$$\begin{aligned}
\Pi_{Dm}^P &= \frac{(n-s+1)}{(n-s+2)} \frac{(1-\gamma-c)^2}{(n+1)^2} \\
\Pi_{Dm}^D &= \frac{n^2}{(n+1)^2} \frac{(1-\gamma-c)^2}{(n-s+2)^2}
\end{aligned}$$

2.3 Interpretation

The aim of this section is to compare from a static point of view the incentives to merge up- and downstream. For this purpose, assume that s producers (resp. s retailers) merge. This merger is supposed to be exogenous. The following table summarizes the profit levels upstream and downstream in the different cases.

	Profit of a producer	Profit of a retailer
Initial Situation	$\Pi^P = \frac{n}{n+1} \frac{(1-\gamma-c)^2}{(n+1)^2}$	$\Pi^D = \frac{n^2}{(n+1)^2} \frac{(1-\gamma-c)^2}{(n+1)^2}$
Upstream Merger	$\frac{(n+1)^2}{(n-s+2)^2} * \Pi^P$	$\frac{(n-s+1)^2 (n+1)^2}{(n-s+2)^2 n^2} * \Pi^D$
Downstream Merger	$\frac{(n-s+1)}{(n-s+2)} \frac{n+1}{n} * \Pi^P$	$\frac{(n+1)^2}{(n-s+2)^2} * \Pi^D$

The merger of s firms brings the same gain for the involved firms whether they are producers or retailers. The gain brought by a merger then is the same for the retailers and the producers. Indeed, the profits are both multiplied by the same factor $\frac{(n+1)^2}{(n-s+2)^2} > 1$ after a merger.

Despite this similarity, there exist some important differences. On the one hand, when s retailers merge, the profit of a producer is multiplied by $\frac{(n-s+1)(n+1)}{(n-s+2)n} < 1$.⁴ On the other hand, when s producers merge the profit of a retailer is multiplied by $\frac{(n-s+1)^2 (n+1)^2}{(n-s+2)^2 n^2} < \frac{(n-s+1)(n+1)}{(n-s+2)n}$. As a consequence, a merger between producers is more harmful for the retailers than is a merger between retailers for the producers, whatever the number s of merging firms.

This simple model gives the following intuition: the retailers may fear more from a merger among producers than the producers do from a merger between retailers. For this reason, they may have a greater incentive to protect themselves from this merger risk. This suggests that the incentives may be naturally more important at the downstream level than they are at the upstream level. But of course, the comparative static analysis should take into account both the expected profits post merger and the anticipated losses due to a merger at the other level.

Yet that first model has obvious limits. In particular, in a Cournot framework with constant marginal costs we come up against the paradoxical result illustrated for the first time by Salant et al. (1983): any coalition of firms, behaving as a merged

⁴ Because $s \geq 2$ and $f : x \mapsto \frac{x+1}{x}$ is decreasing on $]1, +\infty[$.

entity, consisting of fewer than 80 per cent of the present firms in the industry, is disadvantageous for the firms involved. Our investigation shows that this result still holds in our two-level competition model between producers on the one hand and retailers on the other.⁵ This kind of result is to a large extent due to the fact that these modellings only take into account the reduction of the number of active firms in the industry. It is also due to the fact that the firms compete à la Cournot.⁶ Yet possible synergies may exist when a merger takes place, and our model does not allow us to take them into account. Moreover, it is plausible to suppose that a "size effect" also plays a role in favor of the merging firms. Perry and Porter (1985), using different market structures and cost structures (as compared to Salant et al. [1983]), integrate this "size effect". They show that there are then greater incentives to merge for the firms. These results seem to be more plausible.

In the following section, we present the model on which relies our analysis of the incentives to merge at both upstream and downstream levels, taking this "size effect" into account, by introducing synergies effects for the merging firms.

3 The model

Consider a market for a homogeneous good produced by m producers and sold by n retailers. We assume that $n \geq 3$ and $m \geq 3$. The consumers' inverse demand function is linear: $P(q) = 1 - q$, with P the resale price of the good at the retailers' shops and q the total quantity of the good sold by all the retailers. Retailers compete in quantities in a Cournot game to sell the good to consumers. They buy the good on an intermediate market where producers also compete in quantities in a Cournot game. We denote w the wholesale price on the intermediate market. Producers are not able to sell directly to the consumers.

We use a quadratic cost structure, following McAfee and Williams (1992), and

⁵For example, a merger involving s retailers among n is profitable if and only if the following inequality holds $s(n - s + 2)^2 \leq (n + 1)^2$. This is exactly the condition found by Salant et al. (1983).

⁶In the case of price competition in a differentiated products market, Deneckere and Davidson (1985) show that a horizontal merger is always profitable. Nonetheless, the outsider firms gain more from this merger than do the firms involved in the merger.

Perry and Porter (1985). We assume that there is a factor whose total supply is fixed to the industry and which is a necessary input in the production process. For example, this factor might be the firm's capital investment. All retailers have quadratic costs: retailer i ($i \in \{1, \dots, n\}$) sells quantity q_i^D with a retailing cost⁷ of $c_i^D(q_i^D) = \frac{a(q_i^D)^2}{2k_i^D}$, where $0 \leq a \leq 1$. Producers have quadratic costs as well: producer j ($j \in \{1, \dots, m\}$) produces quantity q_j^P with a cost of $c_j^P(q_j^P) = \frac{(q_j^P)^2}{2k_j^P}$. Parameters k_i^P and k_j^D represent the amount of capital owned respectively by a producer and a retailer.⁸ Furthermore, the assumption that, given equal capital stocks, retailing costs can be lower than producers' costs seems rather realistic, so that we assume that $a \leq 1$. This allows us to introduce an asymmetry between retailers' and producers' costs.

We assume that the total quantity of capital of an industry (upstream or downstream) is fixed and we normalize it to be unity. We thus assume that no entry can occur, neither upstream nor downstream. Analytically this is equivalent to:

$$\sum_{i=1}^n k_i^D = \sum_{j=1}^m k_j^P = 1$$

To model the increase of productivity due to a merger, we assume that when s firms merge, the total amount of capital owned by the merged firm is the sum of the amounts of capital of the firms who have merged. For instance, if retailers $i \in \{1, \dots, s\}$ merge,

$$k_{merged}^D = \sum_{i \in \{1, \dots, s\}} k_i^D$$

The merged firm's marginal production or retailing costs are then lower than the marginal costs of a single firm before the merger. This cost structure allows us to integrate the synergies created by a merger, and more particularly the "size effect" described in the previous section. The merger of s firms not only reduces the number of firms in the industry, as in Salant et al. (1983), but it also changes the cost function of the new firm, introducing size asymmetries on the market. Consequently, incentives to merge can exist even for a small number of firms.

⁷Of course, this retailing cost adds to the wholesale price the retailer has to pay to buy the good on the intermediate market.

⁸Another possible interpretation would be to consider that these parameters are proportional to the firms' capacities.

3.1 Cournot equilibrium

In this section we determine the equilibrium profits of the firms in the Cournot game with m producers and n retailers. First, we make the simplifying assumption that before the merger occurs, capital is symmetrically distributed among firms:

$$\forall i \in \{1, \dots, n\}, k_i^D = k^D = \frac{1}{n}$$

$$\forall j \in \{1, \dots, m\}, k_j^P = k^P = \frac{1}{m}$$

The inverse demand function on the intermediate market is given by the retailers' profit maximization. Retailer i ($i \in \{1, \dots, n\}$) chooses the quantity q_i^D to maximize its profit:

$$\text{Max}_{q_i^D} [1 - \sum_{k=1}^n q_k^D - a \frac{q_i^D}{2k_i^D} - w] q_i^D$$

We introduce the following simplifying notation:

$$\beta_i^D = \frac{k_i^D}{k_i^D + a}$$

$$B^D = \sum_{i=1}^n \beta_i^D$$

Solving the first-order conditions gives the following equation:

$$w = 1 - \frac{1 + B^D}{B^D} Q^D$$

Furthermore, in equilibrium, retailer i purchases a quantity q_i^D , such that $q_i^D = \frac{\beta_i^D}{B^D} Q^D$. Finally, we find that:

$$q_i^D = \frac{\beta_i^D}{1 + B^D} (1 - w)$$

Anticipating this inverse demand function, producer j ($j \in \{1, \dots, m\}$) maximizes his profit by setting q_j^P to solve:

$$\text{Max}_{q_j^P} [1 - \frac{1 + B^D}{B^D} \sum_{k=1}^m q_k^P - \frac{q_j^P}{2k_j^P}] q_j^P$$

Similarly as above, we define the following notation:

$$b = \frac{1 + B^P}{B^D}$$

$$\beta_j^P = \frac{bk_j^P}{1 + bk_j^P}$$

$$B^P = \sum_{j=1}^m \beta_j^P$$

Producer j then offers the following quantity:

$$q_j^P = \frac{1}{1 + B^P} \frac{\beta_j^P}{b}$$

We thus deduce that:

$$Q^P = Q^D = \frac{1}{b} \frac{B^P}{1 + B^P}$$

In the next section we characterize the conditions under which there exists an incentive to merge for s producers and for s retailers. The merger of s firms is profitable if and only if the profit of the merged firm exceeds the sum of the profits of its constituent firms. Before-merger profits are given by the following expressions:

$$\Pi^D = \left(\frac{B^P}{(1 + B^P)(1 + B^D)} \right)^2 \frac{k^D(2k^D + a)}{2(k^D + a)^2}$$

$$\Pi^P = \frac{1}{(1 + B^P)^2} \frac{k^P(2bk^P + 1)}{2(bk^P + 1)^2}$$

3.2 Retailers' incentives to merge

In the Cournot equilibrium we defined previously, the sum of the profits of s independent retailers is equal to:

$$s\Pi^D = s \left(\frac{B^P}{(1 + B^P)(1 + B^D)} \right)^2 \frac{k^D(2k^D + a)}{2(k^D + a)^2}$$

A merged firm consisting of s retailers ($s \leq n$) faces a lower marginal cost than its competitors. We have to take this decrease of the marginal cost into account to determine the firms' post-merger profits. The merged firm sells the quantity q with a marginal retailing cost of $\frac{aq^2}{2sk^D}$.

We define the following notation;

$$\begin{aligned}
 B_S^D &= \frac{(n-s)k^D}{k^D+a} + \frac{sk^D}{sk^D+a} \\
 b_S &= \frac{1+B_S^D}{B_S^D} \\
 B_S^P &= \frac{m(b_S \times k^D)}{1+b_S \times k^D}
 \end{aligned}$$

The profit of the merged firm competing with the $(n-s)$ remaining retailers then is:

$$\Pi_S^D = \left(\frac{B_S^D}{(1+B_S^D)(1+B_S^D)} \right)^2 \frac{sk^D(2sk^D+a)}{2(sk^D+a)^2}$$

The merger of s retailers thus is profitable if the following condition, which we will refer to as "Condition D", is verified:

$$\left(\frac{B_S^D}{(1+B_S^D)(1+B_S^D)} \right)^2 \frac{(2sk^D+a)}{(sk^D+a)^2} \geq \left(\frac{B^P}{(1+B^P)(1+B^D)} \right)^2 \frac{(2k^D+a)}{(k^D+a)^2}$$

3.3 Producers' incentives to merge

In the Cournot equilibrium defined previously, the sum of the profits of s independent producers is equal to:

$$s\Pi^P = s \frac{1}{(1+B^P)^2} \frac{k^P(2b.k^P+1)}{2(b.k^P+1)^2}$$

We define the following notation:

$$\begin{aligned}
 \beta_S^P &= \frac{sb.k^P}{1+sb.k^P} \\
 B_S^{P'} &= \beta_S^P + (m-s)\beta^P
 \end{aligned}$$

The merged firm composed of s producers earns a profit of:

$$\Pi_S^P = \left(\frac{1}{1+B_S^{P'}} \right)^2 \frac{sk^P(1+2sb.k^P)}{2(1+sb.k^P)^2}$$

The merger of s producers is profitable if and only if the following condition, which we will refer to as "Condition P", is verified:

$$\left(\frac{1}{1+B_S^{I''}}\right)^2 \frac{(1+2sb.k^{I'})}{(1+sb.k^{I'})^2} \geq \frac{(2b.k^{I'}+1)}{(1+B^P)^2(b.k^{I'}+1)^2}$$

3.4 Comparing the incentives

Condition D (resp. Condition P) determines the threshold $s_D^*(n, m, a)$ (resp. $s_P^*(n, m, a)$) above which the merger of s retailers (resp. s producers) is profitable: $\forall s \in \{s_D^*, \dots, n\}$, the merger of s retailers is profitable, and similarly $\forall s \in \{s_P^*, \dots, m\}$, the merger of s producers is profitable.

Lemma 1 *There exists a unique threshold $s_D^*(n, m, a)$ in $]1, n[$ above which the merger of s retailers is profitable, and a unique threshold $s_P^*(n, m, a)$ in $]1, m[$ above which the merger of s producers is profitable.*

Proof: see appendix 1. ■

Now we focus on the case where $a = 1$. Initially, producers' and retailers' situations are symmetrical. Following the conclusions of the first model, we could expect that incentives to merge are the same upstream and downstream. However, we can prove that there are significant differences. Proposition 2 gives bounds on the thresholds $s_D^*(n, m, a)$ and $s_P^*(n, m, a)$ for situations where there are as many producers as retailers, i.e. for $m = n$.

Proposition 2 *When $a = 1$ and $m = n$, we have: $s_D^*(n, n, 1) \leq 3$ and $\frac{n}{4} < s_P^*(n, n, 1) < \frac{7n}{10}$. Furthermore, for $n \geq 9$, $s_P^*(n, n, 1) < \frac{n}{2}$.*

Proof: See appendix 2. ■

Whatever the number n of firms before the merger, retailers have an incentive to merge for very small values of s (3, and even 2 if the number n of firms at both levels of the market is less than 15). Mergers involving a small number of retailers are

profitable, whereas mergers among producers have to involve more than a quarter of the firms to be profitable. Our results differ quite strongly from Salant et al. (1983), who showed that if any given fraction α ($\alpha < 1$) of an industry is assumed to merge, there is an industry size n large enough for this merger to be unprofitable.

In this setting, it therefore appears clearly that incentives to merge are stronger for retailers than for producers. The efficiency gains due to the reduction of costs have a stronger impact on retailers' decisions than on producers'. An intuitive explanation for that could be that, for the same number of upstream and downstream firms, the retailers' intermediate position in the vertical structure implies that they are in a position of inferiority, as they are price-takers on the intermediate market. They are the last players in a Stackelberg game: their margins, even if strictly positive, are less than producers' margins, and their profits are smaller than those of producers.

To analyse this result, we compare the variation of the firms' profits after a merger. We observe that mergers involving a "reasonable" number of firms (that is, less than 80% of the firms) bring about greater profit gains for retailers than for producers.

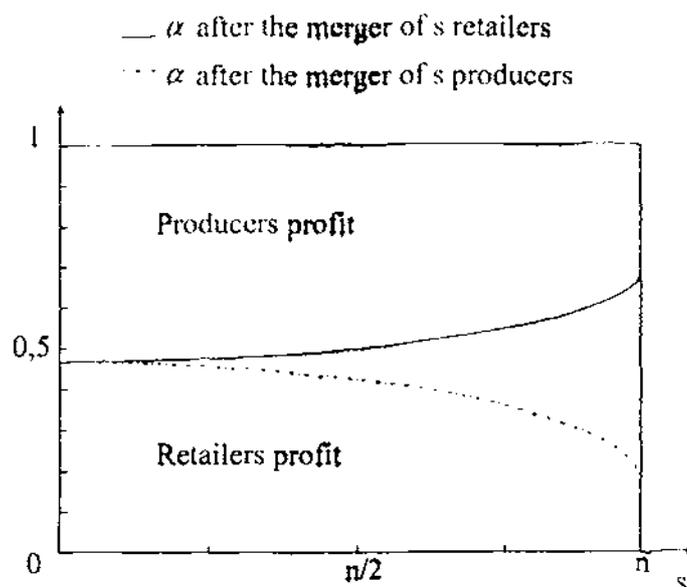
Proposition 3 For all $n \geq 3$, for $m = n$ and $a = 1$, $\forall s \leq \frac{4n}{5}$,

$$\frac{\pi_S^P}{s\pi^P} \geq \frac{\pi_S^R}{s\pi^R}$$

Proof: See appendix 3. ■

A merger brings about a multiplication of the firms' profit by a factor $\frac{11^s}{s11^n}$. We can see that, for a merged entity composed of less than 80% of the retailers, this multiplier is greater than the multiplier of the producers, when the merger occurs among less than 80% of the manufacturers. This result is different from the results we obtained in the first simple model, where we showed that the multiplier was the same for producers and retailers.

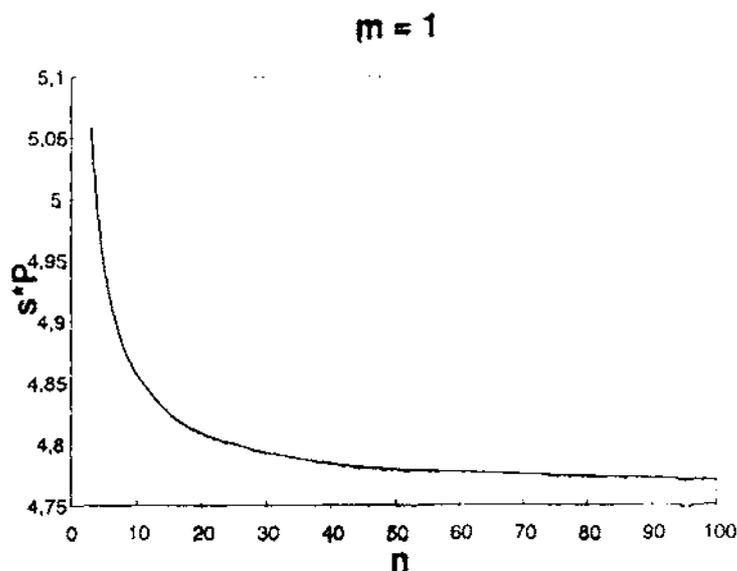
It is then interesting to study the division of the profits among the firms. The following figure shows the evolution of that division when s retailers (resp. s producers) merge. We denote α the share of retailers' profits in the sum of producers' and retailers' profits. To study the evolution of α graphically, we have fixed $m = n = 20$.



We can see that, before a merger, retailers have a smaller part of the whole profit than producers. A merger among retailers increases downstream profits and decreases upstream profits, so that α increases. On the contrary, a merger among producers decreases downstream profits and increases upstream profits, so that α decreases. However, notice that when less than half of the retailers have merged, producers still have the greatest part of the whole profit. This confirms the fact that retailers are in a position of inferiority.

We now try to analyse a more general framework. The general resolution of *Condition D* and *Condition P* is not easy. First we keep the simplifying assumption that $a = 1$, and we study the variations of $s_p^*(m, n, a)$ and $s_D^*(m, n, a)$ with $m \neq n$. We can then simulate the values of the thresholds. Table 1 in appendix 4 gives an overview of the simulations' results for m and n varying between 3 and 100.

For instance, fixing the number of producers on the market equal to 10, we can see the variations of the threshold s_p^* , when the number of retailers varies on the following figure:



We observe that $s_D^*(m, n, a)$ is decreasing with m and $s_p^*(m, n, a)$ is decreasing with n . This result may seem counterintuitive as we might first expect that the concentration at the retailers' level increases the incentives to merge for producers. On the contrary, we can see that, fixing the number m of producers, the more concentrated the retailers, the less profitable a merger of s producers. Similarly, when a merger among producers occurs, it reduces the incentives to merge for the retailers.

Finally, we can test the importance of the assumption that $a = 1$. Keeping the earlier assumption that $m = n$, we study the evolution of the thresholds when a varies. Table 2 in appendix 4 gives an overview of the results we obtained by simulations, for n varying between 3 and 100 and a between 0 and 1.

The threshold $s_D^*(n, n, a)$ is increasing with n and decreasing with a . The fact that the threshold is increasing with n is a classical result of Cournot competition models: the more firms there are on a market, the less profitable the merger of s firms. Moreover, when the marginal retailing cost is very low, retailers have few incentives to merge. But when a is near 1, retailers face costs that are comparable to producers costs and they have incentives to merge in order to increase their buying power.

Moreover, $s_p^*(n, n, a)$ is increasing with n and with a . When retailers' costs are low, their bargaining power is more important and they threaten producers who face

greater incentives to merge. On the contrary, when retailers' costs are higher they have less bargaining power, and producers' individual profit is greater: a merger between a given number s of producers is thus less profitable, and that is why the threshold $s'_p(n, n, a)$ is increasing with a .

3.5 Welfare analysis

Before any merger, when all the firms are symmetric at both levels of the market, welfare is defined by:

$$W(n, m) = \sum_{j=1}^m \Pi_j^P + \sum_{i=1}^n \Pi_i^D + \frac{(1-p)^2}{2}$$

We can easily show that welfare is increasing in the number of producers and in the number of retailers. This is a classical result, as competition increases welfare. Moreover, keeping the assumptions that $a = 1$ and $m = n$, we can show that mergers occurring at both levels of the market decrease welfare (see appendix 5).

Furthermore, if the number of firms is the same at both levels of the market before any merger occurs, we can compare the loss in welfare induced by a merger involving s retailers and a merger involving s producers. We show that welfare loss due to a merger among s producers is greater than welfare loss due to a merger among s retailers. This is not surprising, as we have already shown that profit gains involved by a merger occurring at the same level of the market is greater among retailers than among producers, and that profit losses involved by a merger occurring at the other level of the market are lower for the producers than for the retailers.

Finally, we show that, even if mergers among retailers are more likely to happen than mergers among producers, they are less harmful in terms of welfare. It is then difficult to know whether competition authority should maintain stricter control over mergers among retailers or mergers among producers.

These results motivate us to proceed with a study of an acquisition game, which allows us to relax the hypothesis that mergers are exogenous. We try to determine whether a concentration process is more likely to take place among retailers than producers, and what its impact would be on the balance of power in the vertical structure.

and on total welfare.

4 A Simple Merger Game

Our purpose in this section is to examine what happens if the merger process is endogenous. If the firms have the ability to acquire other firms, we ask whether it is possible for a retailer (resp. a producer) to acquire t retailers (resp. t producers). We do not consider vertical integration, and we assume that retailers can only acquire retailers and producers can only acquire producers. We keep the assumption that the number of firms is the same at both levels of the market before the beginning of the game ($m = n$). And as the basic philosophy behind merger control is the prevention of monopolization of an industry, we ask whether a monopoly may emerge at one of the two levels ($t = n - 1$).

4.1 The game

Assuming that the assumptions of the previous model still hold, we consider a merger game, as used by Kamien and Zang (1990) or Gaudet and Salant (1992). In fact, as we focus on mergers at one level of the market, we consider two similar merger games: the first one where the players are the retailers, and the second one where the players are the producers. We keep the assumption that $a = 1$. Initially each firm is owned by a single owner. Each owner can purchase other firms or sell his own firm. Each merger game consists in two stages.

In the first stage, each owner announces publicly a bid for each of the other firms and an asking price for his own firm. A retailer can only bid for retailers, and the same assumption is also made at the upstream level, so that the number of producers is fixed in the first game (among retailers) and the number of retailers is fixed in the second game (among producers). If an asking price is less than a bid announced by a rival then the firm is sold to the bidder who offered the highest price. If two (or more) bidders announce the same bid for a firm, and if this bid is higher than the asking price, then the targeted firm is allocated randomly among the two bidders. The allocation rule is common knowledge. By acquiring other firms, a firm will benefit from

synergies in costs as previously described.

In the second stage, the firms with the new amounts of capital compete à la Cournot.

4.2 Conditions for the existence of merged configurations in equilibrium

The question we ask then is whether certain merged configurations (including a monopoly) may appear as a result of competitive acquisition of retailers (resp. producers) by their rivals. The analysis done hereafter will then give us necessary conditions for a retailer (resp. a producer) to own t retailers (resp. t producers) at the end of the first stage, given that the $n - (t + 1)$ other firms do not merge.

Let us denote π_{t+1}^n the profit of the retailer who acquired t retailers at the first stage, $\pi_{refusal}$ the deviation profit of an acquired firm, *i.e.* the profit an acquired retailer could earn by raising his asking price above the purchaser's bid such that he would not be bought, and finally π_1 the profit of a single retailer before any acquisition.

A purchaser is unwilling to pay the t other retailers more than (on the whole) $\pi_{t+1}^n - \pi_1$ as the last price represents his opportunity cost. So a necessary condition for the equilibrium where a retailer acquires t other retailers is given by the rationality constraint of the acquirer:

$$\pi_{t+1}^n - \pi_1 \geq t\pi_{refusal}$$

As a consequence, this configuration cannot exist in equilibrium if this condition does not hold. Let us now characterize situations where the condition is verified.

For a downstream firm, the necessary condition to the acquisition of t other firms is given by:

$$\left[\frac{B_{t+1}^P}{(1 + B_{t+1}^P)(1 + B_{t+1}^D)} \right]^2 \frac{(t + 1)k^D(2k^D(t + 1) + a)}{2[(1 + t)k^D + a]^2} \geq t \left[\frac{B_t^P}{(1 + B_t^P)(1 + B_t^D)} \right]^2 \frac{k^D(2k^D + a)}{2[k^D + a]^2} + \left[\frac{B^P}{(1 + B^P)(1 + B^D)} \right]^2 \frac{k^D(2k^D + a)}{2[k^D + a]^2} \quad (1)$$

The upstream condition can as well be defined by:

$$\frac{1}{(1 + B_{t+1}^{P'})^2} \frac{(t+1)k^{t'}(2(t+1)k^{t'}b+1)}{2[(1+t)k^{t'}b+1]^2} \geq t \frac{1}{(1 + B_t^{P'})^2} \frac{k^{t'}(2bk^{t'}+1)}{2[bk^{t'}+1]^2} + \frac{1}{(1 + B^{P'})^2} \frac{k^{P'}(2bk^{P'}+1)}{2[bk^{P'}+1]^2} \quad (2)$$

Inequations (1) and (2) cannot be solved easily. However, we can show that condition (2) is never verified.

Lemma 4 $\forall t \leq n - 1$, the upstream condition is never verified in the game among producers.

Proof : see appendix 6 ■

Consequently, we can show that the issues of the game where one producer buys t other producers and the $(n - t - 1)$ other producers remain unmerged are not Nash equilibria of the game. This results from the fact that, in such cases, the outsiders' profit is greater than the insiders' profit. However, to find the equilibria of the game among producers, we have to study all possible configurations, including multiple coalitions issues. Then we try to determine whether s firms may be interested in merging if the remaining $n - s$ firms are partially or totally merged. First, we study the possibility of duopolization of the sector. Lemma 5 shows that duopoly cannot emerge in equilibrium in the game among producers.

Lemma 5 The necessary condition for s producers to merge given that the $n - s$ other producers have merged is never verified in the game among producers, so that duopolization of the producing industry cannot emerge in equilibrium.

Proof : See appendix 7. ■

We thus can see that if a merged equilibrium exists in the game for producers, it involves multiple coalitions, and the resulting configuration can be neither a monopoly,

nor a duopoly. To study the ability of a producer to buy $s - 1$ other producers, whatever the $n - s$ remaining firms do, we have to build an index of the other firms' concentration, which we call Δ :

$$\Delta = \sum_{j=s+1}^n \frac{bk_j^P}{1 + bk_j^P}$$

It is easy to show that Δ is decreasing in the concentration of the $n - s$ firms: when two firms or two coalitions of firms among the $n - s$ merge, Δ decreases. Moreover, Δ varies from $\frac{b(n-s)}{n+b(n-s)}$ when the $n - s$ other firms are merged, to $\frac{b(n-s)}{n+b}$ when the $n - s$ other firms are separate. We can show that, whatever the value of this index, the merger of s firms cannot happen in equilibrium, because the rationality constraint of the acquirer is never verified.

Proposition 6 *The only equilibrium of the game among producers is the statu quo equilibrium.*

Proof: see appendix 8. ■

This result is quite different from the usual literature on merger games. The existence of the downstream level totally suppresses the incentives to merge for the producers. However, this result depends strongly on our assumption that $m = n$. When the number of retailers is sufficiently high compared to the number of producers, it is easy to show that producers may merge in equilibrium. But our purpose here is to compare the incentives to merge for producers and retailers in a symmetric setting. On the contrary, we can show that, in the game among retailers, the downstream level can be much more concentrated in equilibrium: condition (1) is more often verified than condition (2). The general resolution of the game is complicated, and our purpose is not to solve it completely, but to compare the level of concentration that can be reached in equilibrium in the game among producers and in the game among retailers. In particular, one can ask whether it is possible for a retailer to monopolize the downstream level of the chain, i.e. to buy all his competitors.

Proposition 7 *An equilibrium with monopolization exists in the game for the retailers if $n \leq 5$. Moreover, if we relax the assumption that $m = n$, such an equilibrium exists for the retailers if $n \leq 4$, whatever m , and if $n = 5$, whatever $m \geq 4$.*

Proof: see appendix 9. ■

Mergers thus may lead to a monopolization in equilibrium in the game among retailers, if the number of firms is not too high at the beginning of the game, which is impossible in the game among producers. Formally, this result is more classical⁹, but it shows the differences in incentives between producers and retailers and confirms the intuitions we had in the previous section.

These results have implications in terms of competition policy. We have shown that no merger may occur in equilibrium in the game among producers, whereas a monopoly can emerge from a merging game among retailers if they are at the most 5 at the beginning of the game, which is a realistic condition, as we have seen earlier¹⁰. However, there may exist other equilibria without monopolization in the game among retailers. It is difficult to solve the game for all its equilibria, in the general case. But we can solve it in a simple case, fixing the number of firms at both levels of the market.

4.3 Application : the four firms case

Now we focus on a particular case. We assume that, at the beginning of the game, there are four firms at each level of the market: $m = n = 4$. We first examine the merger game among retailers.

The profits of the retailers in the different configurations of the market that may exist after the merger game are presented in appendix 6. We can show that there exist many equilibria, and that all configurations of the market can emerge in equilibrium.

Proposition 8 *If $m = n = 4$, all market configurations may emerge in equilibrium in the retailers' game. The total surplus of the retailing sector is maximum in the equilibrium with monopolization.*

⁹In a single-stage model, Kamien and Zang (1990) show that monopolization cannot happen if $n \geq 5$.

¹⁰In most of the European countries, the retailing scene is dominated by an oligopoly of retailers. In Finland, for instance, 5 retailers shared 96% of the market in 1997. This explains why the European Commission has not allowed the merger proposal between Kesko and Tuko, arguing that the retailing sector was enough concentrated in Finland.

Proof : see appendix 10. ■

We thus can see that there exist a lot of Nash equilibria in that game, and that we cannot exclude any market configuration. However, the equilibrium with monopolization Pareto-dominates the others in the sense that the total profit of the retailing sector is maximum in that configuration.

Finally we analyse the game among producers. As we have shown previously in proposition 6, there are no merged configuration in equilibrium: the only equilibrium in that game is the *statu quo* equilibrium.

In terms of welfare, we can show that the loss can be greater in the game for retailers than in the game for producers. If the retailers choose the *statu quo* equilibrium, there is no loss of welfare, but all of the other equilibria bring about welfare losses, whereas the only equilibrium of the game among producers is the *statu quo* equilibrium which involves no welfare loss. In terms of anti-trust policy, our model supports the idea that competition authority should maintain strict control over mergers among retailers.

5 Concluding remarks and further research

In this paper, we revisit the question of the observed high concentration of retailers relatively to producers. We give a new interpretation of this concentration by comparing the incentives to merge for producers and retailers in a two-level model of Cournot competition. First, a comparative static analysis allows us to show that retailers have stronger incentives to merge than producers, insofar as an exogenous merger of s retailers among n is profitable for smaller values of s than for producers. This is due to the fact that for an equal number of producers and retailers, the latter have an intermediate position in the vertical structure which makes them earning smaller profits than the producers. For that reason, a merger is more likely to enhance the profits of retailers than those of producers. And as the first model showed, the retailers may fear more from a merger among producers than the producers do from a merger between retailers. Moreover, mergers involving a reasonable number of firms bring about greater profit gains for retailers than for producers. We also show that producers' incentives to merge increase with the number of retailers, and *vice versa*.

Finally, the merger game studied in the last section strengthens these results, by analysing whether certain merged configurations are possible when acquisition of retailers (resp. producers) by their rivals is allowed. We show that mergers never occur in equilibrium in the game among producers, whereas many merged configurations emerge in equilibrium in the game among retailers. In particular, we show that, when there are the same number of firms at both levels of the market at the beginning of the game, a monopoly can emerge under certain conditions at the retail level, but never at the upstream level. In terms of welfare, we show that a merger involving the same number of firms among the same number is worse when it occurs among producers than among retailers. But as mergers cannot appear in equilibrium in the game among producers, whereas they can appear in equilibrium in the game among retailers, we show that, when there are few firms at both levels of the market, the equilibria of the merger game among retailers are worse in terms of welfare than the equilibria in the merger game among producers.

Our results can be extended and strengthened in different directions. The first one is to consider - as the retailers may fear more a merger between producers than the opposite - a merger race between retailers and producers and to see which mergers are more likely to happen. One can ask the crucial question of whether a merger at some level will block mergers at the other level. This is possible in our framework since we have shown that the threshold above which a merger between retailers (resp. producers) is profitable decreases with the number of producers m (resp. with the number of retailers n). An interesting direction for further research thus would be to model the dynamics of the merger process, in order to study whether producers or retailers would merge faster, and what would be the outcomes of a dynamic merger game.

Finally, another important aspect is left for future research. It is well known that mergers do not have the same impact on welfare whether they occur among symmetrical firms or not.¹¹ It would then be interesting to relax the hypothesis that firms are symmetric. We could therefore determine, in an asymmetric setting, which firms (small ones or big ones) have the strongest incentives to merge.

¹¹ See for example McAfee and Williams (1992).

6 References

- Comanor, W. S. and P. Rey (1995) "Vertical restraints and the market power of large distributors", *CREST Working Paper N. 9537*.
- Deneckere, R. and C. Davidson (1985) "Incentives to Form Coalitions with Bertrand Competition", *Rand Journal of Economics*, 16, 473-486.
- Dobson, P. and M. Waterson (1999) "Retailer power ; recent developments and policy implications", *Economic Policy* No. 28, 135-164.
- Farrell, J. and C. Shapiro (1990) "Horizontal mergers: an equilibrium analysis", *American Economic Review*, Vol. 80 No. 1, 107-126.
- Gaudet, G. and S. W. Salant (1991) "Increasing the Profits of a Subset of Firms in Oligopoly Models with Strategic Substitutes", *American Economic Review*, Vol. 81 No. 3, 658-665.
- Gaudet, G. and S. W. Salant (1992) "Mergers of producers of perfect complements competing in prices", *Economics Letters* No. 39, 359-364.
- Kamien, M. I. and I. Zang (1990) "The limits of monopolization through acquisition", *Quarterly Journal of Economics*, No. 105, 465-499.
- Kamien, M. I. and I. Zang (1993) "Monopolization by sequential acquisition", *Journal of Law, Economics, & Organization* Vol. 9 No. 2, 205-229.
- McAfee, P. and M. A. Williams (1992) "Horizontal mergers and antitrust policy", *Journal of Industrial Economics*, Vol. XL No. 2, 181-187.
- Mills, D. (1995) "Why retailers sell private labels", *Journal of Economics and Management Strategy*, Vol. 4, 3, 509-528
- Perry, M. and R. H. Porter (1985) "Oligopoly and the incentive for horizontal merger". *American Economic Review*, Vol. 75 No. 1, 219-227.

Salant, S. W., S. Switzer and R. J. Reynolds (1983) "Losses from horizontal merger : the effects of an exogenous change in industry structure on Cournot-Nash equilibrium", *Quarterly Journal of Economics*, No. 390, 185-199.

Steiner, R. L. (1985) "The nature of vertical restraints", *The Antitrust Bulletin*, Spring 1985, 143-197.

7 Appendix

7.1 Appendix 1: Existence and uniqueness of the thresholds

We can easily prove that, when s retailers merge, the difference between their *post merger* profit and the sum of their profits before the merger, denoted "*Condition D*", is given by the following expression:

$$\text{Condition } D = (an + s)^2 \left[-\frac{2+an}{(1+an)^2(an+2s)} + \frac{(1+m+n+an+2mn+amn)^2}{(a^2(1+m)n^2+(2+n+m(3+2n-2s)-s)s+an(1+n+s+m(1+2n+s)))^2} \right]$$

Similarly, when s producers merge, the difference of their *post-merger* profit and the sum of their profits before the merger, denoted "*Condition P*", is given by the following expression:

$$\text{Condition } P = (mn + (1 + n + an)s)^2 \times \left[-\frac{2+(2+2a+m)n}{(1+(1+a+m)n)^2(mn+2(1+n+an)s)} + \frac{(1+m+n+an+2mn+amn)^2}{(m^2n(1+(2+a)n) - (1+n+an)^2(s-2)s + m(1+n+an)(n+s+2ns+ans))^2} \right]$$

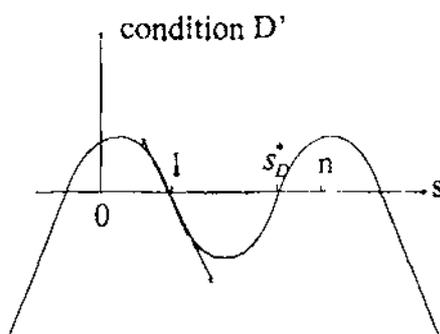
Let us fix m , n and a . Simplifying the previous conditions allows us to show that their sign is given by a polynomial of degree four, depending on s . The sign of *Condition D* is thus given by the sign of:

$$\frac{(1+m+n+an+2mn+amn)^2(1+an)^2(an+2s) - (2+an)\{a^2(1+m)n^2+(2+n+m(3+2n-2s)-s)s+an(1+n+s+m(1+2n+s))\}^2}{(1+an)^2(an+2s)(a^2(1+m)n^2+(2+n+m(3+2n-2s)-s)s+an(1+n+s+m(1+2n+s)))^2}$$

whose denominator is positive whatever positive value of β . The numerator of this fraction, which we will denote by $\phi_D(n, m, s, a)$, is a polynomial of degree four depending on s . We can easily check that $\phi_D(n, m, 1, a) = 0$ whatever the values of

n , m and a . Finally, $\phi_D(n, m, s, a) = (s - 1)\psi_D(n, m, s, a)$ where ψ_D is a polynomial function of the third degree depending on s .

We then prove that $\frac{\partial \phi_D}{\partial s} |_{s=1} < 0$ whatever $n > 2$. Similarly, $\phi_D(n, m, n, a) > 0$ whatever m, n and a . ϕ_D has then at least one root between 1 and n . Yet we can show that $\lim_{s \rightarrow -\infty} \phi_D(n, m, s, a) = \lim_{s \rightarrow +\infty} \phi_D(n, m, s, a) = -\infty$. According to the intermediate values theorem, ϕ_D has thus one root in $] -\infty, 1[$ and another root between n and $+\infty$. We know that ϕ_D is of degree four depending on s , so it cannot have more than four real roots. We thus deduce the shape of the curve "Condition D" when s varies, and the existence of a unique root $s_D^*(n, m, a)$ in the interval $]1, n[$.



The same proof applies for "Condition P".

7.2 Appendix 2: proof of proposition 2

We suppose that $m = n$ and $a = 1$. In that case, for $s = 3$, Condition D has the same sign as the following polynomial function:

$$-3 + 4n + 75n^2 + 138n^3 + 54n^4$$

It is straightforward to show that this polynomial is positive whatever $n \geq 3$, so that Condition D is always positive for $s = 3$. A merger between three retailers is then always profitable.

Furthermore, if $s = n/4$, the sign of condition P is given by the sign of the following function:

$$128 + 1152n + 3872n^2 + 6446n^3 + 5108n^4 + 1055n^5 - 684n^6 - 4n^7$$

It is easy to check that this polynomial is negative for all $n \geq 4$, so that a merger between producers involving less than a quarter of the upstream firms is never profitable.

Finally, if $s = 7n/10$, the sign of *Condition P* is given by the sign of the following function:

$$14000 + 82800n + 163040n^2 + 100958n^3 - 64972n^4 - 98545n^5 - 12600n^6 + 16856n^7$$

It is easy to check that this polynomial is positive for all $n \geq 3$, so that a merger involving more than 70% of the producers is always profitable. Similarly, we show that for $s = \frac{n}{2}$, *condition P* is positive for all $n \geq 9$.

7.3 Appendix 3: proof of proposition 3

We assume that $m = n$ and $a = 1$. Let $\xi(s) = (\frac{\pi_s^D}{s\pi^D} - \frac{\pi_s^P}{s\pi^P})$. We show by the same method as in Appendix 2 that the sign of $\xi(s)$ is given by the sign of a polynomial of degree 5, which we denote by $\zeta(s)$. Whatever n , $\zeta(1) = 0$ and $\frac{\partial \zeta}{\partial s}(s=1) > 0$. Furthermore, $\lim_{s \rightarrow -\infty} \zeta(s) = -\infty$, $\lim_{s \rightarrow +\infty} \zeta(s) = +\infty$. We show that $\zeta(s)$ has always two negative roots, and two roots greater than 1. Finally we show that whatever $n \leq 9$, the two roots that are greater than 1 are greater than n , and if $n \geq 9$, $\zeta(s)$ has a unique root lying between 1 and n .

So, whatever $n \leq 9$, a merger of s retailers among the n multiplies their profit by a greater coefficient than the multiplier of producers' profit.

Furthermore, whatever $n \geq 9$, there exists a threshold $s^{**}(n) \in [1, n]$ above which $\xi(s)$ becomes negative. We can show that this threshold is always greater than $\frac{8n}{10}$. This means that any merger involving less than 80% of the firms is more profitable at the retailers' level than at the producers' level.

7.4 Appendix 4: simulations

Table 1: variations of the thresholds with $m \neq n$.

		m=3	m=5	m=7	m=10	m=30	m=50	m=100
n=3	s_P^*	2.07	3.00	3.86	5.06	12.18	18.87	35.21
	s_D^*	1.10	1.06	1.04	1.03	1.01	1.00	1.00
n=5	s_P^*	2.05	2.96	3.79	4.95	11.75	18.11	33.6
	s_D^*	1.40	1.35	1.33	1.31	1.29	1.28	1.28
n=7	s_P^*	2.04	2.94	3.76	4.89	11.56	17.76	32.87
	s_D^*	1.61	1.56	1.53	1.51	1.48	1.47	1.47
n=10	s_P^*	2.03	2.92	3.73	4.86	11.41	17.50	32.3
	s_D^*	1.85	1.78	1.75	1.73	1.69	1.68	1.67
n=30	s_P^*	2.02	2.9	3.69	4.79	11.17	17.07	31.38
	s_D^*	2.58	2.45	2.40	2.35	2.29	2.27	2.26
n=50	s_P^*	2.02	2.90	3.68	4.78	11.12	16.98	31.20
	s_D^*	2.87	2.72	2.65	2.60	2.51	2.50	2.48
n=100	s_P^*	2.02	2.89	3.68	4.77	11.08	16.91	31.05
	s_D^*	3.18	2.99	2.91	2.84	2.74	2.72	2.7

Table 2: variations of the thresholds with $m = n$ and $a \neq 1$.

		a=0.1	a=0.3	a=0.5	a=0.7	a=0.8	a=0.9	a=1
n=3	s_P^*	1.91	1.96	1.99	2.03	2.04	2.05	2.07
	s_D^*	1.94	1.66	1.44	1.28	1.21	1.15	1.1
n=7	s_P^*	3.06	3.26	3.43	3.57	3.64	3.70	3.76
	s_D^*	4.46	3.34	2.58	2.04	1.84	1.67	1.53
n=10	s_P^*	3.7	4.02	4.31	4.55	4.66	4.76	4.86
	s_D^*	6.35	4.51	3.29	2.46	2.16	1.92	1.73
n=30	s_P^*	6.62	7.88	8.98	9.94	10.38	10.79	11.17
	s_D^*	18.69	11.51	7.02	4.21	3.34	2.72	2.29
n=50	s_P^*	8.76	11.01	13.02	14.77	15.56	16.30	16.98
	s_D^*	30.76	18.13	10.18	5.32	3.95	3.07	2.5
n=100	s_P^*	13.04	18	22.44	26.28	27.99	29.57	31.05
	s_D^*	60.6	34.31	17.42	7.18	4.78	3.46	2.71

7.5 Appendix 5 : welfare analysis

Let us first assume that the two levels of the market are symmetrical before any merger occurs: $m = n$. We keep the assumption that $a = 1$. After a merger of s

retailers, welfare becomes :

$$W_{SD} = m \frac{1}{(1+B_S^P)^2} \frac{k^P(2b_s.k^P+1)}{2(b_s.k^P+1)^2} + (n-s) \left(\frac{B_S^P}{(1+B_S^P)(1+B_S^D)} \right)^2 \frac{k^D(2k^D+a)}{2(k^D+a)^2} \\ + \left(\frac{B_S^P}{(1+B_S^P)(1+B_S^D)} \right)^2 \frac{sk^D(2sk^D+a)}{2(sk^D+a)^2} + \frac{1}{2} \left(\frac{B_S^P}{b_s(1+B_S^D)} \right)^2$$

We can also study analytically W_{SD} . Fixing n , we can show that $\frac{\partial W_{SD}}{\partial s} = \frac{nN(s)}{D(s)}$, where $D(s)$ is a polynomial of degree three depending on s , always positive for any $s \in [1, n]$. Furthermore, analytical study of $N(s)$ shows that it is negative for any $s \in [1, n]$. Consequently, W_{SD} is decreasing in s for $s \in [1, n]$.

Similarly, after the merger of s producers, the welfare becomes

$$W_{SP} = (m-s) \frac{1}{(1+B_S^{P'})^2} \frac{k^P(2b.k^P+1)}{2(bk^P+1)^2} + \frac{1}{(1+B_S^{P'})^2} \frac{sk^P(2sb.k^P+1)}{2(skb^P+1)^2} \\ + n \left(\frac{B_S^{P'}}{(1+B_S^{P'})(1+B^D)} \right)^2 \frac{k^D(2k^D+a)}{2(k^D+a)^2} + \frac{1}{2} \left(\frac{B_S^{P'}}{b(1+B_S^{P'})} \right)^2$$

We can study W_{SP} by the same way than W_{SD} , and we show that it decreases with s for $s \in [1, n]$.

We cannot study analytically the welfare in the general case, with $m \neq n$. But simulations show that, even then, given m and n , any merger (occurring among retailers or among producers) reduces welfare.

Finally, we can compare the variations in welfare after a merger occurring among retailers and after a merger occurring among producers. We study the difference $W_{SD} - W_{SP}$. The sign of this difference is given by a polynomial of degree four in s , and, by the same way than in appendix 1, we show that this polynomial is negative for any $s \in [1, n]$.

7.6 Appendix 6: proof of lemma 4

Let us denote $\varphi(n, t)$ the difference $\pi_{t+1}^n - \pi_t - t\pi_{refusal}$. The necessary condition for the equilibrium where a retailer acquires t other retailers to be possible is given by

$$\varphi(n, t) \geq 0$$

We can easily show that, $\forall n$, $\varphi(n, t = 0) = 0$ and $\varphi(n, t = n - 1) < 0$. Furthermore, we can show that $\varphi(n, t)$ is decreasing with respect to t for $t \in [0, n - 1]$:

$$\frac{\partial \varphi}{\partial t} = \frac{N(n, t)}{D^3(n, t)}$$

where $N(n, t)$ is a polynomial of degree 10 in t which is always negative $\forall t \in [0, n - 1]$, and $D(n, t)$ is a polynomial of degree 4 in t which is always positive $\forall t \in [0, n - 1]$. We thus deduce that, $\forall n$, $\varphi(n, t)$ is decreasing with respect to t for $t \in [0, n - 1]$, so that $\forall n$, $\forall t \in [0, n - 1]$, $\varphi(n, t) \leq 0$. The necessary condition is never verified.

7.7 Appendix 7: non-existence of equilibria with duopolization in the game among producers.

A necessary condition for the existence of an equilibrium with duopolization is that there exists $s \in \{1, \dots, n\}$ such that the merger of s producers is profitable given that the $(n - s)$ other producers have merged, and that the merger of $(n - s)$ producers is profitable given that the s others have merged.

Let us assume that $(n - s)$ producers merge. We can write a necessary condition for the s remaining producers to merge in equilibrium. The rationality constraint of the acquirer is given by

$$\frac{1}{(1 + B_s^{P'})^2} \frac{sk^P(2sk^P b + 1)}{2[sk^P b + 1]^2} - (s - 1) \frac{1}{(1 + B_s^{P''})^2} \frac{k^P(2bk^P + 1)}{2[bk^P + 1]^2} + \frac{1}{(1 + B_{n-s}^P)^2} \frac{k^P(2bk^P + 1)}{2[bk^P + 1]^2} \geq 0$$

$$\text{with } B_s^{P''} = \frac{bk_P}{1 + bk_P} + \frac{(n-s)bk_P}{1 + (n-s)bk_P} + \frac{(s-1)bk_P}{1 + (s-1)bk_P}$$

$$\text{and } B_{n-s}^P = \frac{(n-s)bk_P}{1 + (n-s)bk_P} + \frac{sbk_P}{1 + sbk_P}$$

The sign of the previous difference is given by the sign of a polynomial function depending on n and s , which study is tedious but simple and shows that it is always negative. Thus the necessary condition for the duopolization of the upstream level is never verified.

7.8 Appendix 8: proof of proposition 6

A necessary condition for a producer to be able to buy $s - 1$ other producers, given that the $n - s$ remaining producers' concentration is indexed by Δ , is

$$\begin{aligned} \pi_s^n - \pi_1 &\geq (s - 1)\pi_{refusal} \\ &\iff \varphi(n, s, \Delta) \geq 0 \end{aligned}$$

The sign of φ is given by a polynomial function $P(n, s, \Delta)$. The study of the successive derivatives¹² of P with regards to Δ allows us to show that $\forall n, \forall s \geq 7, \forall \Delta, \frac{\partial \varphi}{\partial \Delta} \geq 0$. Yet we have shown in lemma 4 that, when $\Delta = \frac{b(n-s)}{n+b}$, i.e. when the $n - s$ remaining firms are separate, $\forall n, \forall s, \varphi(n, s, \Delta = \frac{b(n-s)}{n+b}) \leq 0$. As $\Delta \in [\frac{b(n-s)}{n+b(n-s)}, \frac{b(n-s)}{n+b}]$, we thus deduce that $\forall n, \forall s \geq 7, \forall \Delta \in [\frac{b(n-s)}{n+b(n-s)}, \frac{b(n-s)}{n+b}]$, $\varphi(n, s, \Delta) \leq 0$.

Furthermore, $\forall s \in \{1, \dots, 6\}$, it is easy to show¹³ that $\forall n, \forall \Delta, \varphi(n, s, \Delta) \leq 0$.

We can thus conclude that the necessary condition is never verified.

7.9 Appendix 9: sketch of the proof of proposition 7

Simple calculations imply that the inequality (1) which gives the necessary condition for the acquisition of $n - 1$ retailers to be profitable is given by:

$$\frac{n^2(-117 - 456n + 80n^2 + 2205n^3 + 1800n^4 - 2847n^5 - 3725n^6 + 760n^7 + 2124n^8 + 320n^9 - 144n^{10})}{[2(3+4n)^2(1+3n+3n^2)^2(3+2n-8n^2-4n^3)^2]} \geq 0$$

A simple study shows that this inequality is verified for $n \leq 5$ and is not for $n > 6$. If $m = n$, the necessary condition for a monopoly to emerge in equilibrium in retailers' game, given by equation 1, is thus verified whatever $n \leq 5$.

Furthermore, even if $m \neq n$, we can show that this necessary condition holds (unless when $n = 5$ and $m = 3$). Indeed, one can easily show that this inequality is equivalent to:

¹² P is of degree 6 in Δ . We first study $\frac{\partial^6 P}{\partial \Delta^6}$, and show that this polynomial function depending on s and n is positive $\forall n, \forall s \in [1, n]$. Thus we know that $\frac{\partial^5 P}{\partial \Delta^5}$ is increasing with respect to Δ , and we show that $\frac{\partial^5 P}{\partial \Delta^5}(\Delta = 0)$ is positive $\forall n, \forall s \in [1, n]$, etc...

¹³ All the calculations are available from the authors upon request.

($n = 3$)

$$\frac{m^2}{2} \left[\frac{3}{(3+4m)^2} - \frac{5}{(7+10m)^2} - \frac{250}{(33+46m)^2} \right] > 0$$

It is easy to show that this inequality holds whatever $m > 0$.

($n = 4$)

$$3m^2 \left[\frac{1}{2(3+4m)^2} - \frac{1}{(9+13m)^2} - \frac{147}{(57+79m)^2} \right] \geq 0$$

This inequality is verified if $m \geq 1$.

($n = 5$)

$$\frac{4m^2(-2766 - 10917m - 13190m^2 - 3104m^3 + 2176m^4)}{(3+4m)^2(11+16m)^2(29+40m)^2} \geq 0$$

This inequality holds whatever $m \geq 4$.

We have now to show that there exist some strategies that sustain equilibrium with monopolization in the game among retailers. Suppose that a retailer bids $\frac{\pi^m - \pi_1}{n-1}$ for the other retailers and that his asking price is π^m (where π^m is the monopoly profit he can get when facing m producers). Suppose the other retailers' asking prices are $\frac{\pi^m - \pi_1}{n-1}$ and their bids are π_1 . Let us show that these strategies form a subgame perfect Nash equilibrium of the merger game.

First of all, since $n \leq 5$ we have $\frac{\pi^m - \pi_1}{n-1} \geq \pi_{refusal} \geq \pi_1$. From the buyer's point of view, lowering his asking price down to π_1 has no effect since he still earns the same profit. Moreover, he has no incentive to lower it to a lesser price than π_1 since he will get less than π_1 . He has no incentive to increase his bids for the other firms, and he is not incited to decrease one or more of these bids.

If he deviated from his strategy and diminished his bid for s retailers to buy only $n - s - 1$ firms, he would get $\pi^D(n-s, 1, \dots, 1)$. Simple calculations then show that $\pi^D(n-s, 1, \dots, 1) - (n-s-1) \frac{\pi^m - \pi_1}{n-1} \leq \pi_1$ for every $s \in \{1, 2, \dots, n-1\}$ and for $n \leq 5$. The buyer thus has no incentive to deviate from this strategy.

Let us now show that no acquired firm is incited to change its bids. It is straightforward to see that lowering or increasing its asking price is not interesting for an

acquired firm, given the other firms' strategies. However, such a firm could prefer to deviate and buy s other firms. If so, it would have to increase its bids for s firms from π_1 to at least $\frac{\pi^m - \pi_1}{n - 1}$. Its profit would then be at most:

$$\pi^D(n - s - 1, s + 1) - s \frac{\pi^m - \pi_1}{n - 1}$$

which is less than $\frac{\pi^m - \pi_1}{n - 1}$ for every $n \leq 5$.

For example, for $n = 5$, it can be easily shown that the difference $\pi^D(n - s - 1, s + 1) - s \frac{\pi^m - \pi_1}{n - 1} - \frac{\pi^m - \pi_1}{n - 1}$ equals:

(for $s = 1$):

$$\frac{m^2(21411 + 236106m + 745608m^2 + 934344m^3 + 412304m^4)}{(3 + 4m)^2(11 + 16m)^2(93 + 130m)^2} \leq 0$$

($s = 2$)

$$\frac{3m^2(30852 + 230385m + 607532m^2 + 685028m^3 + 281928m^4)}{(3 + 4m)^2(11 + 16m)^2(93 + 130m)^2} \leq 0$$

($s = 3$)

$$\frac{2m^2(6447 + 44958m + 113188m^2 + 123328m^3 + 49408m^4)}{(3 + 4m)^2(11 + 16m)^2(29 + 40m)^2} \leq 0$$

Finally, no acquired firm has any incentive to deviate from the strategy we previously defined, and these strategies define an equilibrium with monopolization for the retailers.

7.10 Appendix 10 : sketch of the proof of proposition 8

Let us first analyse retailers' game.

Retailers' pay-offs in the first game with $m = n = 4$ are as follows :

The profit of any retailer at the beginning of the game is $\Pi_1 = \frac{48}{3721}$.

The profit of the firm resulting of the merger of all the retailers is $\Pi_{1234}^{1234} = \frac{24}{361}$.

The profit of the merged entity composed of three firms is $\Pi_{123}^{123} = \frac{6000}{139129}$.

The profit of the firm competing with the merged entity composed of three firms is $\Pi_4^{123} = \frac{2352}{139129}$.

The profit of the merged entity composed of two firms, when the two other firms did not merge, is $\Pi_{12}^{12} = \frac{200}{7569}$

The profit of a single firm facing one single firm and another firm resulting from the merger of two firms is $\Pi_3^{12} = \frac{12}{841}$.

The profit of a firm resulting from the merger of two retailers, facing a firm resulting from the merger of the two remaining retailers is $\Pi_{12}^{12,34} = \frac{32}{1089}$.

Proposition 6 shows that there exists an equilibrium with monopolization in the game for retailers. The following strategies define such an equilibrium :

- player 1 bids $\frac{\Pi_{1234}^{1234} - \Pi_1}{2}$ for player 2, 3 and 4, and asks Π_{1234}^{1234} for himself,
- the other players ask Π_{1234}^{1234} for themselves and bid 0 for each other firms.

An equilibrium also exists where three firms merge. Simple calculations show that the following strategies define a Nash equilibrium:

- player 1 bids $\frac{\Pi_{123}^{123} - \Pi_1}{2}$ for player 2 and for player 3, 0 for player 4, and asks Π_{1234}^{1234} for himself,
- players 2 and 3 ask $\frac{\Pi_{123}^{123} - \Pi_1}{2}$ and bid 0 for each other firms,
- player 4 asks Π_{1234}^{1234} and bids 0 for each other firms.

Similarly, simple calculations show that an equilibrium exists where only two retailers merge. The following strategies define such an equilibrium:

- player 1 bids $\Pi_{12}^{12} - \Pi_1$ for player 2, 0 for players 3 and 4, and asks Π_{1234}^{1234} for himself,
- player 2 bids 0 for each other firm and asks $\Pi_{12}^{12} - \Pi_1$ for himself,
- players 3 and 4 bid 0 for each other firm and ask Π_{1234}^{1234} for themselves.

There also exists an equilibrium where two merged entity emerge. The following strategies define such an equilibrium:

- player 1 bids $\Pi_{12}^{12,34} - \Pi_1$ for player 2, 0 for players 3 and 4, and asks Π_{1234}^{1234} for himself,
- player 2 bids 0 for each other firm and asks $\Pi_{12}^{12,34} - \Pi_1$ for himself,
- player 3 $\Pi_{12}^{12,34} - \Pi_1$ for player 4, 0 for players 1 and 2, and asks Π_{1234}^{1234} for himself,
- player 4 bids 0 for each other firm and asks $\Pi_{12}^{12,34} - \Pi_1$ for himself.

Finally, the *statu quo* equilibrium may also emerge, for instance with the following strategies:

- each player bids 0 for the others and asks Π_{1234}^{1234} for himself.

ANNEXE 4

Intégration verticale avec source alternative d'approvisionnement

Intégration verticale en présence d'une source alternative d'approvisionnement

Stéphane Caprice *†
CREST-LEI, EUREQUA Univ. de Paris I.

15/12/98

Résumé :

Cet article se propose de réexaminer les questions de l'intégration verticale et de l'exclusion stratégique lorsque les distributeurs de par l'existence de marques propres disposent de fait d'un pouvoir de négociation avec le secteur amont.

Formellement, l'existence de marques propres se traduit par l'introduction de contraintes de participation endogènes dans le problème de la firme amont. Le profit de réservation d'un distributeur, défini par le profit qu'il réalise sur sa marque propre, est affecté par les produits que son concurrent distribue. Une conséquence en est que la firme amont peut favoriser la concurrence sur le marché aval afin de réduire les profits de réservation des distributeurs. Les effets pro-concurrentiels liés à l'existence de marques propres peuvent alors favoriser la création d'ententes.

L'intégration verticale de la firme amont par l'un des distributeurs permet à ce dernier de détenir des droits sur les profits de la firme amont et d'internaliser par là-même le coût de vente à l'autre firme aval. La firme aval non-intégrée est également servie par la firme intégrée. L'absence d'exclusion stratégique va dans le sens d'une appréciation favorable de l'intégration verticale. A l'opposé, l'intégration verticale réduit la concurrence sur le marché aval. L'appréciation de l'intégration verticale consiste finalement à affronter ces deux effets.

* Adresse postale : CREST-LEI / ENPC / 28 rue des Saint-Pères, 75007 Paris. Tél : 01.44.58.27.56 / 01.41.17.58.96 . E-mail : caprice@ensae.fr / stephane.caprice@insee.fr.

† Je remercie particulièrement Anne Perrot pour ses conseils dans le cadre de ce travail.

1 Introduction

L'évaluation des effets de l'intégration verticale par les autorités concurrentielles consiste à confronter les gains d'efficacités et les risques anticoncurrentiels attendus. D'un côté, l'intégration verticale peut se justifier par la spécificité de certains actifs et le caractère souvent incomplet des formules contractuelles de long terme. Dans les secteurs où les activités d'ingénierie sont importantes, comme dans le secteur de l'aluminium¹, les gains d'efficacité sont souvent mis en avant pour apprécier favorablement l'intégration verticale. D'un autre côté, la situation concurrentielle des marchés peut être évaluée par un certain nombre de facteurs : intensité des barrières à l'entrée tant en amont qu'en aval, conditions du marché excluant ou non la possibilité de pratiques discriminatoires. En particulier, un des soucis majeurs de la politique de la concurrence est de s'assurer que la firme intégrée ne pratique la forclusion à l'égard de firmes en aval non-intégrées². Dans ce contexte, l'existence d'une source alternative d'approvisionnement permet à la firme non-intégrée d'échapper à une dépendance trop étroite à l'égard de la firme intégrée et va dans le sens d'une appréciation favorable de l'intégration verticale³.

Cet article se propose de réexaminer les questions de la forclusion et de l'intégration verticale dans un contexte où les firmes en aval à savoir les distributeurs bénéficient d'un "pouvoir de négociation" dans leurs relations avec le secteur amont. Ce pouvoir leur est conféré du fait de l'existence de marques propres⁴⁵.

Formellement, l'existence de marques propres se traduit par l'introduction de contraintes de participation endogènes dans le problème du producteur :

¹Voir à ce sujet, les décisions de la Commission dans les secteurs d'activité de l'aluminium (Gencor/Shell, M470, 29/08/94) ou de l'acier (Mannesmann/Hoesch, M222, 12/11/92).

²L'analyse juridique rejoint ici l'analyse économique : voir par exemple Ordovery, J. A., Saloner, G. et S. C Salop (1990) et Rey, P. et J. Tirole (1996) pour des développements récents.

³En matière d'exemples européens, voir les décisions : ABB/Brel, M221, (26/05/92), Viag/Continental Can, M81, (06/06/91).

⁴L'émergence de la grande distribution s'est effectivement accompagnée d'un développement croissant des "marques de distribution", voir à ce sujet Vigny, J. (1994).

⁵Mills (1995) a d'ailleurs utilisé cette idée pour expliquer en quoi l'introduction d'une marque de distributeur, à un niveau de qualité relativement proche de la marque de production, peut permettre de réduire les problèmes de double marginalisation issus de la relation producteur de marque-distributeur.

en effet, le profit de réservation d'un distributeur est affecté par les produits qu'il peut distribuer⁶, mais également par les produits que ses concurrents peuvent distribuer.

Cette approche permet d'affiner la littérature théorique s'intéressant à l'introduction d'un pouvoir de négociation en aval. En effet, l'optique jusqu'ici dominante de la littérature sur les relations entre producteurs et distributeurs consiste à privilégier une approche de type principal-agent et à ignorer le pouvoir de négociation des distributeurs⁷. La mise en place de restrictions verticales telles que l'intégration verticale relève alors de motivations propres aux producteurs. Le pouvoir accru de la distribution est généralement omis ou simplement illustré par un renversement des rapports de forces⁸.

Nous considérons dans cet article une structure verticale composée d'un producteur et de deux distributeurs. Le producteur fabrique un produit de marque de qualité haute et les distributeurs ont accès à un marché concurrentiel de production sur lequel ils peuvent s'approvisionner en marques propres. Le niveau de qualité des marques propres est exogène et est supposé inférieur à celui de la marque de production. Le producteur fait des offres de type "à prendre ou à laisser" aux distributeurs, l'acceptation d'au moins un des deux distributeurs lui permet d'avoir accès au marché final. De par l'existence des marques propres, les contraintes de participation des distributeurs sont désormais endogènes dans le programme de la firme amont. Cette dernière peut par exemple choisir d'intensifier la concurrence sur le marché aval afin de réduire les profits de réservation des distributeurs. Dans ce contexte, certaines formes d'ententes peuvent être mises en place par les acteurs afin de préserver les profits de l'industrie. L'intégration verticale entre la firme amont et l'un des distributeurs en est un exemple. Elle permet d'internaliser les actions de la firme amont et de relâcher la concurrence sur le marché aval.

Reexaminer le débat autour de l'intégration verticale en présence d' "un pouvoir de négociation" des firmes en aval a des implications significatives en matière de politique de la concurrence. Il apparaît notamment qu'interdire l'intégration verticale domine dans de nombreux cas en termes de surplus social, en effet la séparation verticale est le vecteur d'une intensification de la concurrence.

La section 2 présente le cadre du modèle. La section 3, dans laquelle nous effectuons la résolution du jeu, met l'accent sur l'arbitrage auquel la firme

⁶Voir à ce sujet Shaffer, G. (1991a), O'Brien, D. P. et G. Shaffer (1997).

⁷Pour une revue de littérature sur les relations verticales : Katz, M.L. (1989), Rey, P. (1994), Caprice, S. et J. Philippe (1997).

⁸Voir notamment Shaffer, G. (1991b), et Comanor, W.S. et P. Rey (1994).

amont est confrontée dans la détermination de sa politique de distribution lorsque les firmes en aval disposent de marques propres. La section 4 étudie les implications de l'existence de marques propres en matière d'intégration verticale. La section 5 présente les implications de l'intégration en matière de bien-être. La section 6 conclut.

2 Le modèle

On considère une structure verticale composée d'un producteur et de deux distributeurs notés 1 et 2. Le producteur fabrique un produit appelé par la suite, "*marque de production*", moyennant un coût unitaire de production $c_H = c > 0$; la qualité est normalisée à 1 ($q_H = 1$). Chaque distributeur, outre ce produit, a également accès à un produit substitut de qualité plus faible q_L avec $q_L \in]0, 1[$. La qualité de ce produit est supposée commune aux deux distributeurs et exogène. On considère que ce bien est fabriqué indifféremment par un marché concurrentiel de petits producteurs produisant un bien homogène, auquel chaque distributeur a accès, ou par une unité de production intégrée verticalement à chaque distributeur. Ce bien est obtenu à prix coûtant et son coût unitaire de production c_L est négligé et fixé à 0⁹. Dans la mesure où le bien de qualité faible constitue une source d'approvisionnement alternative à la marque de production, nous choisissons de l'interpréter comme une marque propre à chaque enseigne, ou "*marque de distribution*". La différence de qualités entre la marque de production et les marques internes aux enseignes peut résulter de phénomènes de réputation ou de capacités technologiques de production différentes.

On conserve la structure principal-agent habituelle. Celle-ci donne l'avantage au producteur de proposer aux distributeurs des contrats d'approvisionnement de type "*à prendre ou à laisser*". Cependant, l'accès à un produit substitut de qualité plus faible permet à chaque distributeur de disposer d'un profit non nul lorsqu'il refuse de s'approvisionner en marque de production. Son profit est alors égal au profit qu'il réalise sur sa marque propre. L'alternative d'approvisionnement confère à un distributeur le "*pouvoir de dire non*", s'il juge les conditions d'achat de la marque de production peu intéressantes. Le distributeur dispose ainsi d'un pouvoir de négociation accru face au producteur. Dans ces conditions, l'accès au marché final pour le producteur de marque de production n'est pas automatique mais suppose l'obtention du référencement de sa marque par au moins un des deux distributeurs. Celui-ci est conditionné par l'acceptation du contrat par au moins un des deux distributeurs.

Nous supposons que le producteur ne peut pas discriminer entre les dis-

⁹ c s'interprète alors comme le différentiel de coûts de production entre la qualité haute et la qualité basse. Le fait de fixer certains paramètres n'enlève rien à la généralité des propositions qui vont suivre. Des résultats similaires peuvent être mis en évidence en exprimant les propositions en fonction du différentiel de coût ($c_H - c_L$) et du différentiel de qualité ($q_H - q_L$).

tributeurs, et que le contrat proposé est un tarif binôme de la forme $T(y_{H_i}) = wy_{H_i} + F$ où w désigne le prix de gros, F la franchise et y_{H_i} la quantité mise sur le marché par le distributeur i ($i = 1, 2$). Les décisions des distributeurs en matière de référencement de la marque de production sont notées d_i , et dépendent des conditions d'achat proposées par la firme amont. On a ainsi

$$d_i = \begin{cases} A_i & \text{si acceptation} \\ R_i & \text{sinon} \end{cases}$$

Les consommateurs sont supposés hétérogènes dans leur goût pour la qualité et l'utilité d'un consommateur de type θ s'écrit :

$$U(\theta, p_j, q_j) = \begin{cases} (\theta q_j - p_j)_{j=H,L} & \text{s'il consomme une unité de bien } j \\ 0 & \text{sinon} \end{cases}$$

Le paramètre θ , qui représente la disponibilité marginale à payer la qualité est distribué uniformément sur l'intervalle $[0, 1]$ et la taille du marché est normalisée à 1.

Un distributeur, lorsqu'il ne référence pas la marque de production, s'engage à vendre une quantité nulle de marque de production. Il est défini comme une enseigne monoproduit au sens où il dispose d'une unique source d'approvisionnement. Un distributeur, lorsqu'il référence la marque de production, est désormais défini comme une enseigne multiproduits au sens où il a accès à deux sources d'approvisionnement, l'une en bien de qualité haute (la marque de production), l'autre en bien de qualité basse (la marque de distribution). Une enseigne multiproduits ne commercialise pas nécessairement une quantité strictement positive des deux biens, la terminologie utilisée signifie seulement qu'elle a accès aux deux biens en matière d'approvisionnement. La quantité de marque propre distribuée par cette dernière peut par exemple être nulle à l'équilibre. Les fonctions inverses de demande sont conditionnelles aux décisions de référencement de la marque de production, trois cas doivent être distingués¹⁰.

a) Si les décisions prises en matière de référencement sont $\{A_1, A_2\}$, où A_1 et A_2 désignent l'acceptation des distributeurs 1 et 2, le marché aval est un duopole caractérisé par deux enseignes multiproduits.

b) Si les décisions sont de la forme $\{A_1, R_2\}$ ou $\{R_1, A_2\}$, seul un distributeur a référencé la marque de production. Nous sommes en présence d'une enseigne multiproduit et d'une enseigne monoproduit.

¹⁰L'expression des fonctions de demande inverse est donnée en annexe 1, pour des décisions de référencement données $\{d_1, d_2\}$.

c) Si les décisions sont $\{R_1, R_2\}$, aucun distributeur n'a référencé la marque de production. Le marché aval est un duopole représenté par deux enseignes monoproduits distribuant exclusivement leurs marques propres.

A décisions de référencement données $\{d_1, d_2\}$, la concurrence entre les distributeurs est de type Cournot. Les quantités mises sur le marché par les distributeurs sont notées $\{y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2}\}$.

Le jeu se déroule de la façon suivante.

1) A la première étape du jeu, le producteur fait une offre "à prendre ou à laisser" à chaque distributeur i ($i = 1, 2$), $T(y_{H_i}) = wy_{H_i} + F$. Les contrats proposés sont rendus publics à la fin de la première étape du jeu et ne sont pas renégociables.

2) Lors de la deuxième étape, les distributeurs, au vu des contrats proposés acceptent ou refusent simultanément de référencer la marque de production. Les décisions d'acceptation ou de refus $d_i \in \{A_i, R_i\}$ sont ensuite rendues publiques.

3) Les distributeurs se font concurrence en quantités.

Nous déterminons l'équilibre parfait en sous-jeux. La section suivante présente la résolution du jeu par induction amont.

3 Résolution du jeu

Les sous-sections suivantes présentent successivement la résolution des différentes étapes du jeu.

3.1 Détermination de l'équilibre sur le marché aval

Les quantités proposées sur le marché par les distributeurs dépendent des décisions prises par les distributeurs en matière de référencement mais également des prix d'approvisionnement contenus dans les contrats de référencement. Il faut distinguer quatre sous-jeux selon les décisions prises en matière de référencement de la marque de production

$$\{A_1A_2, A_1R_2, R_1A_2, R_1R_2\}$$

On présente ensuite pour des décisions de référencement données, les programmes des distributeurs et l'équilibre sur le marché aval.

3.1.1 Les deux distributeurs ont référencé la marque de production

On considère le sous-jeu consécutif à une décision de référencement de la marque de production par les deux distributeurs. Le profit d'un distributeur i est composé du profit réalisé sur sa marque propre et du profit réalisé sur la marque de production. Son programme s'écrit alors :

$$\begin{aligned} \max_{y_{L_i}, y_{H_i}} \pi_{D_i} &= [p_L(y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2})] y_{L_i} + [p_H(y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2}) - w] y_{H_i} - F \\ \text{avec } i &= 1, 2 \end{aligned}$$

La résolution de ce programme pour chacune des deux entreprises nous donne les fonctions de meilleures réponses en quantités des distributeurs et nous permet de déterminer l'équilibre de Cournot-Nash, à contrat donné du producteur.

$$\begin{cases} y_{H_1} = y_{H_2} = \text{Max} \left\{ 0, \frac{1-q_L-w}{3(1-q_L)} \right\} \\ y_{L_1} = y_{L_2} = \text{Min} \left\{ \frac{1}{3}, \frac{w}{3(1-q_L)} \right\} \end{cases}$$

Le profit d'un distributeur i ($i = 1, 2$) est de la forme

$$\begin{cases} \pi_{D_i}^{A_1 A_2}(w, F) = \frac{(1-w)^2 - q_L(1-2w)}{9(1-q_L)} - F & \text{si } w < 1 - q_L \\ \pi_{D_i}^{A_1 A_2}(w, F) = \frac{q_L}{9} - F & \text{si } w \geq 1 - q_L \end{cases}$$

où $\pi_{D_i}^{d_i d_j}$ représente le profit du distributeur i pour une décision de référencement de la marque de production d_i , à décision d_j donnée de l'autre distributeur j .

Le prix de gros inclus dans le contrat de référencement w s'interprète comme un différentiel de coûts marginaux d'approvisionnement entre les deux qualités. Ainsi lorsque les deux distributeurs ont référencé la marque de production, $\{A_1 A_2\}$, si ce différentiel est relativement faible ($w < 1 - q_L$), la marque de production est écoulee par les deux distributeurs, chacun distribuant également sa marque propre. Si ce n'est pas le cas, le différentiel est trop grand et les distributeurs mettent sur le marché une quantité nulle de la marque de production, seules les marques propres sont écoulees.

3.1.2 Un seul distributeur a référencé la marque de production

Seul le distributeur i a référencé la marque de production, si bien que ce dernier peut proposer à la vente deux produits : la marque de production et sa marque propre, tandis que le distributeur j ne peut proposer que sa

marque propre. Le programme de chacun des distributeurs s'écrit alors :

$$\begin{cases} \max_{y_{L_i}, y_{H_i}} \pi_{D_i} = [p_L(y_{L_i}, y_{H_i}, y_{L_j})] y_{L_i} + [p_H(y_{L_i}, y_{H_i}, y_{L_j}) - w] y_{H_i} - F \\ \max_{y_{L_j}} \pi_{D_j} = [p_L(y_{L_i}, y_{H_i}, y_{L_j})] y_{L_j}, \text{ avec } i \neq j = 1, 2 \end{cases}$$

Si le distributeur i est le seul à avoir référencé la marque de production, $\{A_i R_j\}$, l'équilibre sur le marché aval est de la forme :

$$\begin{cases} y_{H_i} = \text{Max} \left\{ 0, \text{Min} \left\{ \frac{1-q_L-w}{2(1-q_L)}, \frac{2-q_L-2w}{4-q_L} \right\} \right\} \\ y_{L_i} = \text{Min} \left\{ \frac{1}{3}, \text{Max} \left\{ 0, \frac{-1+q_L+3w}{6(1-q_L)} \right\} \right\} \\ y_{L_j} = \text{Min} \left\{ \frac{1}{3}, \frac{1+w}{4-q_L} \right\} \end{cases}$$

et les profits des distributeurs s'écrivent :

$$\left(\pi_{D_i}^{A_i R_j}(w, F), \pi_{D_j}^{A_i R_j}(w) \right) = \begin{cases} \left(\left(\frac{2-q_L-2w}{4-q_L} \right)^2 - F, \left(\frac{1+w}{4-q_L} \right)^2 q_L \right) & \text{si } w \leq \frac{1-q_L}{3} \\ \left(\frac{5q_L^2 - q_L(14-18w) + 9(1-w)^2}{36(1-q_L)} - F, \frac{q_L}{9} \right) & \text{si } w \in \left] \frac{1-q_L}{3}, 1 - q_L \right[\\ \left(\frac{q_L}{9} - F, \frac{q_L}{9} \right) & \text{si } w \geq 1 - q_L \end{cases}$$

Le distributeur i a accès à deux sources d'approvisionnement, l'une en qualité haute (marque de production), l'autre en qualité basse (marque de distribution). Il peut alors choisir de vendre une quantité positive des deux biens ou d'un seul bien. Ce choix dépend du différentiel de coûts marginaux d'approvisionnement entre les biens.

Lorsque ce différentiel est faible, ($w \leq \frac{1-q_L}{3}$), la quantité de marque propre écoulee par le distributeur i est nulle. Le distributeur i dispose de deux biens différenciés verticalement dont les coûts d'approvisionnement sont proches, il choisit de n'écouler que le bien de qualité haute. Dans cette configuration, la qualité haute (marque de production) est écoulee par le distributeur i tandis que la qualité basse (marque propre) est écoulee par le distributeur j . La structure de marché aval est ici équivalente à deux firmes monoproduits dont l'une distribue une qualité haute tandis que l'autre distribue une qualité basse. Pour cette structure de marché, une baisse du prix de gros de la marque de production référencée par le distributeur i induit une réduction du profit du distributeur j . En effet, lorsque la quantité de bien de qualité basse vendue par le distributeur i est nulle, le distributeur j réagit à une hausse de la quantité offerte par le distributeur i en bien de qualité haute par une baisse de sa quantité offerte en bien de qualité basse¹¹.

¹¹ Les biens sont des substituts stratégiques.

En termes d'effet partiel, le producteur peut donc manipuler par le choix du prix de gros la quantité offerte par le distributeur j .

Pour des valeurs intermédiaires du différentiel de coûts d'approvisionnement entre la marque propre et la marque de production, ($w \in]\frac{1-q_L}{3}, 1 - q_L[$) le distributeur i écoule une quantité positive des deux biens. Dans ce cas, le marché aval est équivalent à un duopole asymétrique dans la mesure où le distributeur i est une firme multiproduit qui offre une quantité strictement positive des deux qualités, le distributeur j étant avec sa marque propre une firme monoproduit. Pour cette structure de marché, le profit du distributeur j est indépendant du prix de gros de la marque de production référencée par le distributeur i . Lorsque la quantité de bien de qualité basse vendue par le distributeur i est positive, le producteur ne peut manipuler en termes d'effet partiel par le choix du prix de gros la quantité offerte par le distributeur j , seule la répartition des ventes au sein du distributeur i entre qualité haute et qualité basse est affectée.

Enfin, lorsque le différentiel en coût d'approvisionnement est trop élevé, ($w \geq 1 - q_L$), la marque de production n'est plus écoulée, la structure de marché est équivalente à un duopole de distributeurs monoproduits, chacun distribuant sa marque propre. L'équilibre sur le marché aval est indépendant du prix de gros de la marque de production.

3.1.3 Aucun des distributeurs n'a référencé la marque de production

Le profit d'un distributeur i est uniquement composé du profit réalisé sur sa marque propre. Son programme est alors :

$$\max_{y_{L_i}} \pi_{D_i} = (p_L(y_{L_1}, y_{L_2})) y_{L_i} \text{ avec } i = 1, 2$$

l'équilibre sur le marché aval est de la forme :

$$\left\{ y_{L_1} = y_{L_2} = \frac{1}{3} \right.$$

et le profit d'un distributeur i est :

$$\pi_{D_i}^{R_1, R_2} = \frac{q_L}{9}$$

La marque de production n'ayant pas été référencée, l'équilibre sur le marché aval est indépendant du prix de gros de la marque de production.

3.2 Détermination de l'équilibre en référencement

Trois types d'équilibres peuvent apparaître en fonction du niveau de la franchise : $\{A_1, A_2\}$, $\{A_i R_j\}$, $\{R_1, R_2\}$. Si la franchise est relativement faible, chaque distributeur est assuré de réaliser un profit plus grand en référençant la marque de production, et ce, quelle que soit la stratégie de l'autre en matière de référencement. Par conséquent, il accepte. Pour des valeurs intermédiaires de la franchise, un distributeur n'accepte le référencement que si l'autre refuse. En cas d'acceptation de l'autre distributeur, le profit supplémentaire réalisé en distribuant la marque de production ne lui permet pas de recouvrer le paiement de la franchise. Dans ce cas, coexistent deux équilibres de Nash symétriques dans lesquels un distributeur accepte le référencement et l'autre le refuse. Finalement, si la franchise est élevée, la marque de production n'est pas référencée. Les équilibres en référencement sont répertoriés dans le tableau suivant :

valeur de la franchise	équilibre
$F \leq \pi_{D_i}^{A_i A_j}(w)^{Brut} - \pi_{D_i}^{R_i A_j}(w)$	$\{A_1, A_2\}$
$F \in \left[\pi_{D_i}^{A_i A_j}(w)^{Brut} - \pi_{D_i}^{R_i A_j}(w), \pi_{D_i}^{A_i R_j}(w)^{Brut} - \pi_{D_i}^{R_i R_j}(w) \right]$	$\left\{ \begin{array}{l} 2 \text{ équilibres symétriques de} \\ \text{Nash } \{A_1 R_2, R_1 A_2\} \end{array} \right\}$
$F > \pi_{D_i}^{A_i R_j}(w)^{Brut} - \pi_{D_i}^{R_i R_j}(w)$	$\{R_1, R_2\}$

où $\pi_{D_i}^{d_i d_j}(w)^{Brut}$ représente le profit du distributeur i avant paiement de la franchise.

Pour des valeurs intermédiaires de la franchise, nous avons deux équilibres symétriques $\{A_1 R_2\}$ et $\{R_1 A_2\}$ en permutant les indices des firmes. Dans ce qui suit, on appelle 1 le distributeur qui accepte le contrat et on retient donc l'équilibre $\{A_1 R_2\}$.

3.3 Politiques de distribution choisies par le producteur

Nous avons montré précédemment que le contrat offert par le producteur détermine le nombre de distributeurs dépositaires de la marque de production. L'équilibre en référencement est conditionnel au contrat proposé par la firme en amont. On étudie donc successivement la politique tarifaire de la firme amont selon qu'elle propose un contrat aboutissant à un référencement auprès du seul distributeur 1 ou auprès des deux distributeurs. La politique de distribution optimale se déduit alors d'un choix entre ces deux politiques tarifaires.

3.3.1 Seul le distributeur 1 propose la marque de production

Nous considérons le cas où seul le distributeur 1 propose la marque de production.

La fonction objectif de la firme en amont est :

$$\begin{aligned} & \underset{w, F}{Max} (w - c) y_{H_1} + F \\ & \text{s.c.} \quad \left\{ \begin{array}{l} F \in \left[\pi_{D_i}^{A, A_j}(w)^{Brut} - \pi_{D_i}^{R_i, A_j}(w), \pi_{D_i}^{A, R_j}(w)^{Brut} - \pi_{D_i}^{R_i, R_j} \right] \\ \pi_{D_1}^{A_1, R_2}(w, F) \geq \pi_{D_1}^{R_1, R_2} \end{array} \right\} \end{aligned}$$

Le niveau de la franchise se déduit directement de la contrainte de rationalité individuelle du distributeur 1. La saturation de cette contrainte permet de réécrire le programme sous la forme

$$\begin{aligned} & \underset{w}{Max} A(w) - B \\ & \text{où} \quad \left\{ \begin{array}{l} A(w) = (w - c) y_{H_1} + \pi_{D_1}^{A_1, R_2}(w)^{Brut} \\ B = \pi_{D_1}^{R_1, R_2} \end{array} \right. \end{aligned}$$

Le terme B représente le profit de réservation du distributeur 1. Ce terme est égal au profit réalisé par le distributeur 1 sur sa marque propre lorsque le distributeur 2 ne référence pas la marque de production. Le terme B est donc indépendant du prix de gros proposé. Le terme $A(w)$ représente le profit de la structure verticale (producteur-distributeur 1). L'étude de ce terme permet de définir la politique tarifaire de la firme amont dans le cadre d'un référencement par un unique distributeur, cette dernière est donnée dans la proposition 1.

Proposition 1 ¹² Lorsque la firme amont vend à un seul distributeur, il existe un seuil du différentiel de coût de production $c_a(q_L)$ tel que

- (i) si $c < c_a(q_L)$, la firme amont fixe un prix de gros $w = w_{1D}^*$ inférieur au coût marginal de production (et $w < \frac{1-q_L}{3}$), avec $\begin{cases} w_{1D}^* > 0 & \text{si } c \geq c_1(q_L) \\ w_{1D}^* = 0 & \text{sinon} \end{cases}$
- (ii) si $c \geq c_a(q_L)$, la firme amont fixe un prix de gros égal au coût marginal de production.

L'étude du terme $A(w)$ dépend des configurations de concurrence sur le marché aval.

¹² Annexe 2.

$$\begin{aligned} c_a(q_L) &= \begin{cases} c_2(q_L) & \text{si } c \geq c_1(q_L) \\ c_3(q_L) & \text{sinon} \end{cases} / \begin{cases} A(w_{1D}^*(c), c) = A(w_{1D}^{**}(c), c)|_{c=c_2(q_L)} \text{ avec } w_{1D}^*(c) > c \\ A(w_{1D}^*(c) > 0, c) = A(w_{1D}^{**}(c), c)|_{c=c_2(q_L)} \text{ avec } w_{1D}^*(c) = 0 \end{cases} \\ c_1(q_L) &= \left. \frac{\partial A(w, c_1(q_L))}{\partial w} \right|_{w=0} - 0 \end{aligned}$$

Lorsque le différentiel de coûts de production est faible (situation (i)), la firme amont fixe un prix de gros inférieur au coût marginal de production lorsqu'elle choisit de vendre à un seul distributeur. Le distributeur 1 a accès à deux produits différenciés verticalement dont les coûts marginaux d'approvisionnement sont proches ($w \leq \frac{1-q_L}{3}$), il choisit alors d'écouler une quantité positive d'un seul bien, celui de qualité haute. Dans cette configuration, une baisse du prix de gros induit une baisse de la quantité offerte par le distributeur 2 sur sa marque propre. Une politique tarifaire, dans laquelle le prix de gros est inférieur au coût marginal, permet d'accroître le profit de la structure verticale (producteur-distributeur 1) et de récupérer une franchise plus élevée. Une telle politique est qualifiée d'agressive à l'égard du distributeur 2.

Lorsque le différentiel de coûts de production est important (situation (ii)), la firme amont fixe un prix de gros égal au coût marginal de production lorsqu'elle choisit de vendre à un seul distributeur. Le distributeur 1 a accès à deux produits différenciés verticalement et choisit d'écouler une quantité positive des deux ($w \in]\frac{1-q_L}{3}, 1 - q_L[$). Ici, la quantité offerte de marque propre par le distributeur 2 est indépendante du prix de gros de la marque de production. La firme amont réalise un transfert d'objectif au distributeur 1, le profit de la structure verticale (producteur-distributeur 1) est récupéré par la firme amont via la franchise (déduction faite du profit de réservation du distributeur 1).

3.3.2 Les deux distributeurs proposent la marque de production

Nous considérons le cas où les deux distributeurs proposent la marque de production.

La fonction objectif de la firme amont est de la forme :

$$\begin{aligned} & \underset{w, F}{\text{Max}} (w - c)(y_{H_1} + y_{H_2}) + 2F \\ & \text{s.c.} \quad \left\{ \begin{array}{l} F \leq \pi_{D_1}^{A_1 A_2} (w)^{\text{Brut}} - \pi_{D_1}^{R_1 A_2} (w) \\ \pi_{D_1}^{A_1 A_2} (w, F) \geq \pi_{D_1}^{R_1 A_2} (w) \\ \pi_{D_2}^{A_1 A_2} (w, F) \geq \pi_{D_2}^{A_1 R_2} (w) \end{array} \right. \end{aligned}$$

Comme précédemment, le niveau de la franchise se déduit directement des contraintes de rationalité individuelle des distributeurs. Par symétrie, et après saturation de ces contraintes, le programme de la firme amont s'écrit

$$\begin{aligned} & \underset{w}{\text{Max}} \quad A(w) - B(w) \\ & \text{où} \quad \left\{ \begin{array}{l} A(w) = 2(w - c)y_{H_1} + 2\pi_{D_1}^{A_1 A_2} (w)^{\text{Brut}} \\ B(w) = 2\pi_{D_1}^{R_1 A_2} (w) \end{array} \right. \end{aligned}$$

En l'absence d'une source alternative d'approvisionnement, l'objectif de la firme amont se réduit au terme $A(w)$. Dans ce cas, la firme amont fixe un prix de gros supérieur au coût marginal de production afin de relâcher la concurrence que les distributeurs se livrent sur son produit¹³. Mais l'étude de la politique tarifaire de la firme amont en présence d'une source alternative d'approvisionnement suppose également d'explicitier le lien entre le terme $B(w)$ et le prix de gros.

Le terme $B(w)$ représente les profits de réservation qu'il convient de laisser aux distributeurs afin de s'assurer de leur collaboration. L'étude de ce terme permet d'explicitier la politique tarifaire de la firme amont avec deux distributeurs. Cette dernière est donnée en proposition 2.

Proposition 2 ¹⁴ *Lorsque la firme amont vend aux deux distributeurs, il existe différentes valeurs du différentiel de coût de production telles que*

(i) *si $c < c_b(q_L)$, la valeur prise par le prix de gros w_{2D}^* permet de réduire les profits de réservation des distributeurs, on a*

si $c < c_4(q_L)$, le prix de gros w_{2D}^ est inférieur au coût marginal de production, avec*

$$\begin{cases} w_{2D}^* = 0 & \text{si } c \leq c_5(q_L) \\ w_{2D}^* > 0 & \text{sinon} \end{cases}$$

si $c \geq c_4(q_L)$, le prix de gros w_{2D}^ est supérieur ou égal au coût marginal de production,*

(ii) *si $c \geq c_b(q_L)$, les profits de réservation des distributeurs sont indépendants du prix de gros, ce dernier w_{2D}^{**} ($> w_{2D}^*$) est supérieur au coût marginal de production.*

Les profits de réservation des distributeurs dépendent des régimes de concurrence aval dans lesquels un seul distributeur a référencé la marque de production.

En effet, pour un équilibre du sous-jeu de référencement dans lequel les deux distributeurs référencent la marque de production, le profit de réservation laissé à un distributeur i est égal au profit qu'il réalise sur sa marque propre tandis que le distributeur j référence la marque de production. Les régimes avals de concurrence sont alors décrits :

¹³ On retrouve cet élément lorsqu'une firme amont s'adresse à un secteur aval oligopolistique.

¹⁴ Annexe 3.

$$c_b(q_L) / A(w_{2D}^*(c), c) - B(w_{2D}^*(c), c) = A(w_{2D}^{**}(c), c) - B(c) \Big|_{c=c_b(q_L)}$$

$$c_4(q_L) / w_{2D}^*(c) = c$$

$$c_5(q_L) / \left. \frac{\partial [A(w, c_5(q_L)) - B(w, c_5(q_L))]}{\partial w} \right|_{w=0} = 0$$

soit par un duopole de firmes monoproduits dans lequel la firme i vend un bien de qualité basse, la firme j un bien de qualité haute ($w \leq \frac{1-q_L}{3}$)¹⁵, soit par un duopole asymétrique dans lequel la firme i vend un bien de qualité basse, la firme j une quantité positive des deux biens ($w \in]\frac{1-q_L}{3}, 1 - q_L[$)¹⁶.

Des structures de marché aval, on déduit que pour un prix de gros faible, ($w \leq \frac{1-q_L}{3}$), le profit de réservation de chaque distributeur est une fonction croissante du prix de gros. Tandis que pour des valeurs intermédiaires du prix de gros, ($w \in]\frac{1-q_L}{3}, 1 - q_L[$), le profit de réservation de chaque distributeur est indépendant du prix de gros.

Finalement, lorsque le différentiel de coûts de production est faible (situation (i)), la politique tarifaire de la firme amont consiste à réduire le prix de gros, cela lui permet de réduire les profits de réservation des distributeurs. Une telle politique est qualifiée d'agressive à l'égard du secteur aval. Lorsque le différentiel est plus grand (situation (ii)), les profits de réservation des distributeurs sont indépendants du prix de gros proposé par la firme amont, le prix de gros est donc choisi au dessus du coût marginal de production indépendamment des profits de réservation.

3.3.3 Choix de la politique de distribution du producteur

La détermination de la politique de distribution relève maintenant d'un choix entre deux politiques tarifaires, l'une aboutissant au référencement par un seul distributeur, l'autre par les deux distributeurs. La firme amont dispose d'un avantage en termes de différenciation verticale lié à la maîtrise d'un bien de qualité haute. Mais cet avantage ne présente un intérêt pour la firme amont que si le différentiel de coûts de production avec le bien de qualité basse est grand.

Pour des valeurs élevées du différentiel de coûts de production, $c > \text{Max}\{c_a(q_L), c_b(q_L)\}$ ¹⁷, une politique de distribution avec un ou deux distributeurs n'a pas d'impact sur le profit de la firme amont¹⁸. Avec un seul

¹⁵Lorsque la quantité vendue par le distributeur j sur sa marque propre est nulle, une hausse de la quantité offerte en bien de qualité haute par ce dernier implique une baisse de la quantité offerte par le distributeur i sur le bien de qualité faible. A une baisse du prix de gros est associée une réduction du profit du distributeur i . Le profit de réservation du distributeur i défini par cette configuration décroît lorsque le prix de gros diminue.

¹⁶Lorsque le distributeur j vend une quantité positive des deux biens, le profit du distributeur i sur sa marque propre est indépendant du prix de gros. En conséquence, le profit de réservation du distributeur i est indépendant du prix de gros.

¹⁷Comme $c_a(q_L) > c_b(q_L)$, on a $c > \text{Max}\{c_a(q_L), c_b(q_L)\} \Leftrightarrow c > c_a(q_L)$.

¹⁸Lorsque les distributeurs se livrent une concurrence à la Cournot en biens homogènes,

distributeur, le prix de gros est égal au coût marginal¹⁹ ; avec deux distributeurs, le prix de gros est supérieur au coût marginal et est fixé indépendamment des profits de réservation des distributeurs.

Lorsque le différentiel de coûts de production entre la qualité haute et la qualité basse est faible, la firme amont peut bénéficier de la maîtrise d'un bien de qualité haute. Elle fait désormais l'objet de deux incitations privées opposées²⁰. D'une part, en fixant une franchise suffisante, elle peut exclure l'une des deux firmes aval. Un prix de gros inférieur au coût marginal de production joue alors le rôle d'un engagement à distribuer une quantité élevée via le distributeur 1, et permet de réduire la part de marché du distributeur 2 sur sa marque propre. Le choix d'une politique de distribution avec un distributeur permet d'accroître le profit de la structure verticale (producteur-distributeur dépositaire) mais impose en contrepartie de laisser un profit de réservation élevé au distributeur dépositaire. D'autre part, la firme amont peut choisir d'être présente chez les deux distributeurs. La politique tarifaire adoptée joue alors le rôle d'un engagement à réduire les profits de réservation des distributeurs. Le choix d'une politique de distribution avec deux distributeurs permet de réduire les profits de réservation des distributeurs mais impose en contrepartie de réduire les profits des structures verticales (producteur-distributeur dépositaires). Ainsi, pour un différentiel de coûts de production faible, le choix de la politique de distribution est issu de cet arbitrage. La politique de distribution optimale de la firme amont est décrite en proposition 3.

Proposition 3 ²¹ *Lorsqu'une firme amont s'adresse à un secteur aval oligopolistique, disposant d'une source alternative d'approvisionnement de qualité moindre, la politique de distribution optimale de cette entreprise dépend du différentiel de coûts de production entre les biens.*

en l'absence de source alternative d'approvisionnement, sous l'hypothèse d'offres publiques non renégociables, le profit de la firme amont est le même avec un ou deux distributeurs. Le prix de gros est une fonction croissante et la franchise une fonction décroissante du nombre des distributeurs. Dans le cas considéré, l'introduction d'une source alternative pour les distributeurs modifie seulement la répartition des profits de l'industrie entre l'amont et l'aval.

¹⁹ La valeur du prix de gros ne revêt pas un caractère d'annonce mais répond seulement à un transfert d'objectif entre la firme en amont et le distributeur 1.

²⁰ Les effets stratégiques mis en évidence découlent d'une annonce publique des contrats et sont sensibles à leur renégociation secrète. Voir à ce sujet, Caillaud B. et P. Rey (1995) ou Rey P. et J. Stiglitz (1995). De même, la valeur d'engagement disparaît dans le cas d'une délégation avec contrats secrets (Rey, P. et J. Tirole (1996)).

²¹ Annexe 4.

- (i) si $c \in]0, c_a(q_L)]$, il existe un seuil du différentiel de coûts de production $c_c(q_L)$ ²², tel que
- (ia) si $c \leq c_c(q_L)$, la firme amont sert les deux distributeurs,
 - (ib) si $c \in]c_c(q_L), c_a(q_L)]$, la firme amont sert un seul distributeur,
- (ii) si $c \in]c_a(q_L), (1 - q_L)]$, la firme amont est indifférente entre servir un ou deux distributeurs,
- (iii) si $c > (1 - q_L)$, la marque de production n'est pas distribuée.

Les distributeurs, en disposant de marques propres peuvent échapper à une dépendance trop étroite à l'égard du producteur et disposer d'un "pouvoir de négociation" dans leurs relations avec le secteur amont. Formellement, cela se traduit par l'introduction de contraintes de participation endogènes dans le problème du producteur. Le différentiel de coût de production mesure en quelque sorte la répartition des pouvoirs de négociation entre l'amont et l'aval. Lorsque le différentiel de coût de production est faible (situation (ia)), on dit que la firme amont dispose d'un pouvoir de négociation accru avec le secteur aval. La mise en concurrence des distributeurs lui permet de réduire leurs profits de réservation. Pour des valeurs intermédiaires du différentiel de coûts de production (situation (ib)), on dit que la firme amont dispose d'un pouvoir de négociation moindre. A défaut de pouvoir réduire les profits de réservation des distributeurs, la firme amont choisit d'exclure l'un des deux distributeurs. Cela lui permet de récupérer une franchise plus élevée par le biais d'une augmentation du profit de la structure verticale (producteur-distributeur dépositaire) aux dépens du distributeur non dépositaire. Dans les deux situations, la présence de marques propres implique un conflit d'intérêts entre l'amont et l'aval auquel la firme amont répond par une politique agressive. Il en résulte des conditions favorables à l'émergence d'ententes au sein de l'industrie, dont l'une consiste à voir la firme amont s'intégrer avec l'un des deux distributeurs.

4 Intégration verticale

En présence d'une source alternative d'approvisionnement, l'analyse se doit de prendre en compte dans l'appréciation des effets de l'intégration verticale le conflit d'intérêts entre l'amont et l'aval.

²² Avec $c_c(q_L) \in]0, c_a(q_L)]$, et $c_c(q_L) = \begin{cases} c_6(q_L) & \text{si } c \geq c_1(q_L) \\ c_7(q_L) & \text{sinon} \end{cases}$.

$c_6(q_L) / \pi_P(w_{1D}^*(c), c) = \pi_P(w_{2D}^{**}(c), c)|_{c=c_6(q_L)}$ avec $w_{1D}^*(c) > 0$
 $c_7(q_L) / \pi_P(w_{1D}^*(c), c) = \pi_P(w_{2D}^{**}(c), c)|_{c=c_7(q_L)}$ avec $w_{1D}^*(c) = 0$

Dans ce contexte, nous supposons que la firme amont et l'une des firmes aval (le distributeur 1) peuvent former une filière intégrée. L'analyse suppose d'ajouter une étape initiale pendant laquelle la firme amont et le distributeur 1 décident de former ou non une filière intégrée.

Dans le cas d'une intégration verticale, la décision de servir ou non le distributeur 2 est prise sur la base des profits joints. Les entreprises se font ensuite une concurrence en quantités.

En l'absence d'intégration, la firme amont fait une offre aux deux distributeurs, ces derniers acceptent ou refusent, et se font une concurrence en quantités sur le marché aval.

La problématique est double. Premièrement, on étudie l'impact d'une intégration verticale sur le marché aval, en particulier la filière intégrée a-t-elle intérêt à servir le distributeur 2 en bien de qualité haute ? On résout le sous-jeu consécutif à une décision d'intégration. Deuxièmement, du fait d'un conflit entre l'amont et l'aval, l'intégration est-elle plus probable ? Le jeu est résolu complètement.

4.1 Forclusion ou non-forclusion ?

On étudie le sous-jeu dans lequel la firme amont et le distributeur 1 ont formé à l'étape initiale du jeu une filière intégrée. Le vocabulaire "marque de production" et "marque propre" n'ayant pas de sens au sein de la filière intégrée, on parle désormais de bien de qualité haute et de bien de qualité basse pour désigner successivement la marque de production et la marque propre d'un distributeur.

Avant d'étudier les incitations privées de la firme intégrée à forclore ou pas le distributeur 2 du marché de la qualité haute, nous explicitons l'équilibre sur le marché aval.

4.1.1 Le distributeur 2 référence le bien de qualité haute

Pour un contrat de référencement, $T_2(y_{H_2})$ donné, on résout l'équilibre sur le marché aval. La structure aval est un duopole multiproduits où les coûts d'approvisionnement du bien de qualité haute entre la filière intégrée et le distributeur 2, sont différents. Le prix de cession interne du bien de qualité haute au sein de la filière intégrée est $T_1(y_{H_1}) = cy_{H_1}$, le prix de cession externe pour le distributeur 2 est $T_2(y_{H_2}) = w_2y_{H_2} + F_2$.

Le profit de la firme intégrée est composé de celui réalisé sur le bien de qualité basse et de celui réalisé sur le bien de qualité haute à la fois au sein

du distributeur 1 et au sein du distributeur 2. Pour des quantités données $\{y_{L_2}, y_{H_2}\}$ du distributeur 2, le programme de la firme intégrée est :

$$\max_{y_{L_1}, y_{H_1}} \pi^I = [p_L(y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2})] y_{L_1} + [p_H(y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2}) - c] y_{H_1} + (w_2 - c) y_{H_2} + F_2$$

soit encore

$$\max_{y_{L_1}, y_{H_1}} \pi^I = [p_L(y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2})] y_{L_1} + [p_H(y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2}) - c] y_{H_1}$$

Le profit du distributeur 2 est composé de celui réalisé sur le bien de qualité basse et de celui réalisé sur le bien de qualité haute, à contrat de référencement donné. Pour des quantités données $\{y_{L_1}, y_{H_1}\}$ du distributeur 1, l'écriture du programme du distributeur 2 est :

$$\max_{y_{L_2}, y_{H_2}} \pi_{D_2} = [p_L(y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2})] y_{L_2} + [p_H(y_{L_1}, y_{H_1}, y_{L_2}, y_{H_2}) - w_2] y_{H_2} - F_2$$

La résolution de ces deux programmes nous donne les fonctions de meilleure réponse en quantités de chaque firme et nous permet de déterminer l'équilibre de Cournot-Nash à contrat de référencement donné.

Les configurations d'équilibre sont définies à partir du prix de gros w_2 .

Par exemple, si le prix de gros défini par le contrat de référencement du distributeur 2 est grand, la quantité de bien de qualité haute offerte par le distributeur 2 est nulle²³.

Soit A l'ensemble des prix de gros qui permettent au distributeur 2 d'écouler une quantité de bien de qualité haute non nulle. On a :

$$A = \begin{cases} w_2 < w_2^c \text{ pour } c < \frac{(1-q_L)}{3} \\ w_2 < w_2^d \text{ pour } c > \frac{(1-q_L)}{3} \end{cases}$$

Lorsque le prix de gros appartient à cet intervalle, le distributeur 2 référence le bien de qualité haute sous réserve de disposer de son profit de réservation.

23

$$\begin{aligned} \text{si } c < \frac{(1-q_L)}{3}, \text{ pour } w_2 \geq w_2^c \text{ avec } w_2^c &= \frac{2(1-q_L)(1+c)}{4-q_L}, \text{ on a } y_{H_2} = 0 \\ \text{si } c > \frac{(1-q_L)}{3}, \text{ pour } w_2 \geq w_2^d \text{ avec } w_2^d &= \frac{c}{2} + \frac{(1-q_L)}{2}, \text{ on a } y_{H_2} = 0 \end{aligned}$$

Lorsque le prix de gros est grand, l'acceptation du contrat par le distributeur 2 impose un niveau de franchise égal à 0. Un contrat de cette forme est équivalent à une stratégie d'exclusion du marché de la qualité haute.

4.1.2 Le distributeur 2 ne référence pas le bien de qualité haute

Le distributeur 2 ne vend que le bien de qualité basse. Son profit s'écrit

$$\pi_{D_2}^{I,R_2} = \begin{cases} \left(\frac{1+c}{4-q_L}\right)^2 q_L & \text{si } c < \frac{(1-q_L)}{3} \\ \frac{q_L}{9} & \text{si } c > \frac{(1-q_L)}{3} \end{cases}$$

Il s'agit du profit de réservation du distributeur 2.

La condition sous laquelle le distributeur 2 accepte de référencer le bien de qualité haute s'écrit

$$\pi_{D_2}^{IA_2}(w_2, F_2) \geq \pi_{D_2}^{I,R_2} \text{ avec } w_2 \in A$$

où $\pi_{D_2}^{IA_2}$ désigne le profit du distributeur 2, lorsqu'il référence le bien de qualité haute, déduction faite de la franchise.

4.1.3 Politique de distribution de la firme intégrée

Si la firme intégrée décide de servir le distributeur 2, le contrat proposé est solution du programme

$$\begin{aligned} \max_{w_2 \in A, F_2} \pi^I &= [p_L(\cdot)] y_{L_1}(w_2) + [p_H(\cdot) - c] y_{H_1}(w_2) + [w_2 - c] y_{H_2}(w_2) + F_2 \\ \pi_{D_2}^{IA_2}(w_2, F_2) &\geq \pi_{D_2}^{I,R_2} \end{aligned}$$

La firme intégrée peut par le choix de tarification autoriser ou bloquer l'entrée du distributeur 2 sur le bien de qualité haute. Ce choix est donné dans la proposition 4.

Proposition 4 ²⁴ *La politique de distribution de la firme intégrée dépend du différentiel de coûts entre les biens de qualité haute et de qualité basse,*

(i) *si* $c \in \left[\frac{(1-q_L)}{3}, (1-q_L)\right]$, *la firme intégrée ne sert pas le distributeur 2. Il y a forclusion.*

(ii) *si* $c < \frac{(1-q_L)}{3}$, *la firme intégrée sert le distributeur 2. Il n'y a pas forclusion. Le prix de gros proposé au distributeur 2 dépend du différentiel de coûts,*

(iia) *si* $c \leq \frac{2(1-q_L)}{3(2+q_L)}$, $w_2^I = \frac{2-2q_L+c(2+7q_L)}{4+5q_L}$, $F_2 = \pi_{D_2}^{IA_2}(w_2^I) - \pi_{D_2}^{I,R_2}$,

(iib) *si* $c \in \left[\frac{2(1-q_L)}{3(2+q_L)}, \frac{(1-q_L)}{3}\right]$, $w_2^I = 2c$, $F_2 = \pi_{D_2}^{IA_2}(w_2^I) - \pi_{D_2}^{I,R_2}$.

²⁴Annexe 5.

La firme intégrée, dans sa décision de servir ou non le distributeur 2 est soumise à deux effets opposés. D'un côté, si elle ne sert pas son concurrent, elle est en monopole sur le bien de qualité haute. D'un autre côté, si elle sert son concurrent, elle récupère un revenu supplémentaire lié à la commercialisation du bien de qualité haute par le concurrent.

La firme intégrée a accès à deux sources d'approvisionnement, l'une de qualité haute, l'autre de qualité basse.

Lorsque le différentiel de coûts de production entre les biens est élevé, elle choisit de commercialiser une quantité positive des deux biens (situation (i)). Dans ce cas, il n'est pas dans son intérêt de servir le distributeur 2 en bien de qualité haute. En effet, servir le distributeur 2 impose une pression concurrentielle supplémentaire dommageable à la firme intégrée.

Au contraire, lorsque le différentiel de coûts de production entre les biens est faible, elle commercialise uniquement du bien de qualité haute (situation (ii)). Dans ce cas, elle est incitée à servir le distributeur 2 en bien de qualité haute²⁵. En effet, le distributeur 2 réduit sa part de marché en bien de qualité basse lorsqu'il vend également du bien de qualité haute. De ce fait, la quantité vendue en bien de qualité haute augmente. D'un côté, la firme intégrée perd en part de marché au sein du distributeur 1, mais d'un autre côté récupère les gains liés à un accroissement de la demande sur le bien de qualité haute via la tarification appliquée au distributeur 2.

Nous résolvons maintenant l'étape initiale du jeu.

4.2 L'intégration verticale est-elle plus probable ?

L'intégration verticale émerge dès lors que les profits joints de la structure intégrée sont supérieurs à la somme des profits de chacun des acteurs en l'absence d'intégration. On montre en particulier que l'existence d'un conflit d'intérêts entre la firme amont et les firmes aval rend l'intégration plus probable.

²⁵La zone du différentiel de coûts correspondant à une situation de non-forclusion s'agrandit avec le degré de différenciation. Ainsi, plus les biens sont différenciés, plus la zone du différentiel de coûts pour laquelle la firme intégrée sert son concurrent est grande.

Proposition 5 ²⁶ *La décision d'intégration verticale dépend du différentiel de coûts de production,*

(i) si $c \in \left] 0, \frac{(1-q_L)}{3} \right]$, l'intégration verticale accroît les profits du producteur et du distributeur 1,

(ii) si $c \in \left] \frac{(1-q_L)}{3}, c_a(q_L) \right]$, la séparation verticale accroît les profits du producteur et du distributeur 1,

(iii) si $c \in]c_a(q_L), (1 - q_L)]$, la firme amont et le distributeur 1 sont indifférents entre l'intégration et la séparation verticale.

Si l'intégration verticale n'est pas autorisée, les relations entre l'amont et l'aval peuvent faire l'objet d'un réel conflit. C'est le cas notamment, lorsque la firme amont applique une politique de distribution agressive à l'égard du secteur aval. La politique de distribution consiste alors soit à l'exclusion de l'un des distributeurs afin de réduire la part de marché de ce dernier, soit à un référencement par les deux distributeurs afin de réduire leurs profits de réservation. L'intégration verticale implique au contraire une coordination des actions entre le secteur amont et le secteur aval.

L'intégration verticale peut permettre à la firme intégrée, par la tarification appliquée au distributeur 2, de récupérer les gains liés à un accroissement de la demande sur le bien de qualité haute. Dans ce cas (situation (i)), l'intégration verticale accroît les profits du producteur et du distributeur 1.

Dans les autres cas, l'intégration verticale ne permet pas d'accroître la demande sur le bien de qualité haute, sans créer une pression concurrentielle dommageable à la firme intégrée. Le distributeur 2 est exclu du marché de la qualité haute. Le prix de transfert interne du bien de qualité haute est égal au coût marginal de production.

Pour des valeurs intermédiaires du différentiel de coûts de production (situation (ii)), lorsque la firme amont et le distributeur 1 sont séparés, la firme amont exclut le distributeur 2 du marché de qualité haute et fixe un prix de gros inférieur au coût marginal de production. La séparation verticale permet, par un effet d'annonce de réduire la part de marché du distributeur

²⁶ Si l'on introduit un coût à l'intégration verticale A , la proposition se réécrit : il existe une valeur du différentiel de coûts $c(A, q_L) \in \left] 0, \frac{(1-q_L)}{3} \right]$ pour laquelle

(i) si $c \in]0, c(A, q_L)]$, l'intégration verticale accroît les profits du producteur et du distributeur 1,

(ii) si $c \in]c(A, q_L), (1 - q_L)]$, l'intégration verticale diminue les profits du producteur et du distributeur 1

2. Dans ce cas, la séparation verticale accroît les profits du producteur et du distributeur 1.

Pour un différentiel élevé de coûts de production (situation (iii)), les relations entre l'amont et l'aval ne font pas l'objet d'un réel conflit. La firme amont est dans une situation où elle ne peut tirer profit de la maîtrise d'un bien de qualité haute, la firme amont et le distributeur 1 sont indifférents entre l'intégration et la séparation verticale.

Finalement, lorsque la solution d'intégration émerge, celle-ci s'accompagne d'une non-exclusion de la firme non-intégrée.

Le jeu avec possibilité d'intégration verticale résolu, nous pouvons mener une comparaison entre la situation où l'intégration verticale est permise et celle où elle est interdite.

5 analyse normative

L'intégration verticale émerge pour un différentiel de coûts de production faible. La comparaison est donc menée pour ces valeurs $\left(c \in \left]0, \frac{(1-q_L)}{3}\right]\right)$.

Dans un premier temps, nous étudions les conséquences pour la firme non-intégrée d'une intégration verticale entre la firme amont et le distributeur 1. Dans un deuxième temps, nous concluons sur les effets d'une intégration verticale.

5.1 Implications pour la firme non-intégrée

Nous étudions le profit du distributeur 2 avec ou sans intégration verticale.

Proposition 6 *L'intégration verticale n'est pas nécessairement dommageable à la firme non-intégrée.*

- (i) Lorsque, $c \in]0, c_c(q_L)]$, on a
 - (ia) si $c < c_4(q_L)$, l'intégration verticale est dommageable au distributeur 2,
 - (ib) si $c > c_4(q_L)$, l'intégration verticale est bénéfique au distributeur 2.
- (ii) Lorsque, $c \in \left]c_c(q_L), \frac{(1-q_L)}{3}\right]$, l'intégration verticale est bénéfique au distributeur 2.

Cette proposition se déduit de la comparaison des prix de transfert entre la firme amont et le distributeur suivant qu'on leur impose ou non de rester

séparées.

Lorsque la firme amont et le distributeur 1 forment une filière intégrée, le prix de transfert interne du bien de qualité haute est égal au coût marginal. Le profit du distributeur 2 est égal à son profit de réservation. Ce dernier est défini pour une structure de marché dans laquelle il vend un bien de qualité basse, le concurrent un bien de qualité haute dont le coût d'approvisionnement est égal au coût marginal de production.

Lorsque l'intégration n'est pas permise, la politique de distribution de la firme amont peut prendre deux formes différentes. Deux cas doivent alors être considérés.

Premièrement, la politique de distribution de la firme amont consiste à inclure les deux distributeurs. La valeur prise par le prix de gros dépend du degré de substituabilité entre les biens. Si les biens sont des substituts proches, le prix de gros est inférieur au coût marginal de production. Le profit du distributeur 2, égal à son profit de réservation, est défini pour une structure de marché dans laquelle il vend un bien de qualité basse, le concurrent un bien de qualité haute dont le coût d'approvisionnement est inférieur au coût marginal de production. Dans ce cas, l'intégration verticale est bénéfique au distributeur 2 (inversement si les biens ne sont pas des substituts proches).

Deuxièmement, la politique de distribution de la firme amont consiste à exclure le distributeur 2. Le prix de gros est inférieur au coût marginal de production. Le profit du distributeur 2 est défini pour une structure de marché dans laquelle il vend un bien de qualité basse, le concurrent un bien de qualité haute dont le coût d'approvisionnement est inférieur au coût marginal de production. Dans ce cas, l'intégration verticale est bénéfique au distributeur 2.

Laisser à la firme intégrée la possibilité d'abuser de sa position dominante sur le marché de qualité haute est dans certains cas favorable à la firme non-intégrée au sens l'intégration peut être une situation Pareto dominante pour les secteurs amont et aval²⁷.

5.2 Effets sur le surplus social

Si la firme amont et le distributeur 1 forment une filière intégrée, il n'y a plus de conflits entre l'amont et l'aval. Si au contraire, les deux firmes sont

²⁷C'est le cas dans les situations (ib) et (ii).

séparées, la firme amont mène une politique agressive à l'égard du secteur aval. L'intégration verticale, en relâchant la concurrence permet de fait d'accroître les profits de l'industrie.

En ce qui concerne les effets sur le surplus des consommateurs, nous détaillons successivement les effets liés à une intégration verticale et à une séparation verticale.

Si la firme amont et le distributeur 1 forment une filière intégrée, la firme non-intégrée est servie en bien de qualité haute. Par rapport à une situation d'exclusion, la quantité vendue de bien de qualité haute est élevée. L'intégration favorise la commercialisation du bien de qualité haute et est donc le vecteur d'un effet positif sur le surplus des consommateurs.

Si la firme amont et le distributeur 1 sont séparés, la politique de distribution de la firme amont peut prendre deux formes différentes.

Si les deux distributeurs sont servis, la firme amont réduit le prix de gros afin de réduire les profits de réservation des distributeurs. De par la réduction du prix de gros, la quantité vendue de bien de qualité haute augmente. Dans ce cas, la séparation verticale est le vecteur d'un effet positif sur le surplus des consommateurs. L'effet lié à la séparation verticale domine l'effet lié à l'intégration verticale, le surplus des consommateurs est plus élevé dans le cas de séparation verticale.

Si un des distributeurs est exclu de la politique de distribution de la firme amont, la variation du surplus des consommateurs est soumis à deux effets opposés. D'une part, l'exclusion de l'un des distributeurs du marché de la qualité haute induit une réduction de la quantité commercialisée en bien de qualité haute, effet défavorable pour les consommateurs. D'autre part, la firme amont lorsqu'elle exclut l'un des distributeurs, fixe un prix de gros inférieur au coût marginal ; la part de marché du distributeur non-dépositaire sur le bien de qualité basse diminue au profit du bien de qualité haute, effet favorable pour les consommateurs. Cet effet positif est d'autant plus sensible que le différentiel de qualité est faible entre les biens²⁸. Finalement, lorsque les biens sont des substituts proches, l'effet positif l'emporte, l'effet lié à la séparation verticale domine l'effet lié à l'intégration verticale, le surplus des consommateurs est plus élevé dans le cas de séparation verticale. Lorsque les biens ne sont pas des substituts proches, l'effet négatif l'emporte, l'intégration verticale est préférable du point de vue des consommateurs.

²⁸ Les gains pour la firme amont à annoncer un prix de gros inférieur au coût marginal sont croissants lorsque le différentiel de qualité diminue.

L'effet global résulte ensuite des effets partiels évoqués, la proposition suivante fait le récapitulatif des différents effets.

Proposition 7 ²⁹ Les effets sur le surplus social sont liés au différentiel de coût de production,

Δ de surplus liée à l'intégration verticale différentiel de coûts $c \in$	ΔS_{P+D}	ΔS_C	ΔS
$]0, c_c(q_L)]$ (i)	(+)	(-)	(-)
$]c_c(q_L), c_d(q_L)]$ (ii)	(+)	(+)	(+)
$]c_d(q_L), c_e(q_L)]$ (iii)	(+)	(-)	(+)
$]c_e(q_L), \frac{(1-q_L)}{3}]$ (iv)	(+)	(-)	(-)

En présence d'une source alternative d'approvisionnement, les relations entre l'amont et l'aval font l'objet d'un conflit d'intérêts. Plus le degré du conflit est grand, ce qui est le cas lorsque le différentiel de qualité ou le différentiel de coût sont faibles, plus il est probable que l'intégration verticale apparaisse défavorable en termes de surplus social.

Si l'on impose aux firmes de rester séparées, la firme amont répond au pouvoir de négociation des firmes aval par une politique agressive. La conséquence en est une intensification de la concurrence sur le marché aval, élément positif pour les consommateurs et en termes de surplus social. L'intégration verticale est alors dommageable en termes de surplus social.

Néanmoins, il convient de veiller à ce que la politique de distribution de la firme amont (lorsqu'on impose aux firmes de rester séparées) ne prenne la forme d'une exclusion de l'un des distributeurs³⁰. Dans ce cas, l'appréciation des effets de l'intégration verticale dépend du différentiel de qualité entre les biens. Si la source alternative d'approvisionnement est un substitut proche du bien de la firme amont, l'intégration est dommageable en termes de surplus social. La politique de distribution de la firme amont (lorsqu'on impose aux firmes de rester séparées), se traduit certes par une forclusion du marché de la qualité haute mais également par une intensification de la concurrence entre enseigne dépositaire et enseigne non-dépositaire. Si le différentiel de qualité est grand, la concurrence entre enseigne dépositaire et enseigne non-dépositaire s'affaiblit, et l'effet lié à l'exclusion du marché de qualité haute

²⁹ Voir annexe 6.

ΔS_{P+D} (secteurs amont et aval)

ΔS_C (consommateurs)

ΔS (total)

³⁰ Puisque, inversement l'intégration ne conduit pas à une exclusion de la firme non-intégrée.

l'emporte. Dans ce cas, l'intégration apparaît au contraire souhaitable en termes de surplus social.

6 Conclusion

Pour les autorités en charge de la politique de la concurrence, le contrôle des opérations d'intégration verticale s'explique par la crainte de voir la firme intégrée pratiquer la forclusion vis à vis des firmes non-intégrées³¹. Dans cette optique, la présence d'une source alternative d'approvisionnement permet alors à la firme non-intégrée d'échapper à une dépendance trop étroite à l'égard de la firme intégrée et va dans le sens d'une appréciation favorable de l'intégration verticale. Mais l'analyse de l'intégration verticale en présence d'une source alternative d'approvisionnement pour les firmes en aval n'est pas si simple et nous a montré qu'elle méritait un examen plus approfondi.

En particulier, l'objet de l'article a été de réexaminer les questions de la forclusion et de l'intégration verticale dans un contexte où les distributeurs disposent d'une source alternative d'approvisionnement sous la forme de marques propres. L'existence de marques propres permet aux distributeurs d'échapper à une dépendance trop étroite à l'égard de la firme amont et leur confère un "pouvoir de négociation" dans leurs relations avec l'amont.

Formellement, l'existence de marques propres se traduit par l'introduction de contraintes de participation endogènes dans le programme du producteur. En effet, le profit de réservation d'un distributeur est affecté par les produits que le distributeur concurrent peut distribuer. Dans ce contexte, l'étude de la politique de distribution de la firme amont a permis de mettre en évidence deux effets auxquels elle est soumise. D'une part, la mise en concurrence des distributeurs permet de réduire leurs profits de réservations mais impose en contrepartie de réduire les profits des structures verticales (producteur-distributeurs dépositaires). D'autre part, l'exclusion de l'un des distributeur permet d'accroître le profit de la structure verticale (producteur-distributeur dépositaire) mais impose en contrepartie de laisser un profit de

³¹Dans cet esprit, les autorités françaises de la concurrence sont intervenues pour empêcher la prise de contrôle, sous la forme d'une intégration verticale, d'une société occupant une position clé au sein de la filière de la pêche du thon tropical. L'acquisition de cette société (la Cobrecaf), par une entreprise de la filière aurait donné à celle-ci le contrôle des approvisionnements de la majeure partie des conserveries de thon surgelé. Il était à craindre notamment qu'en cas de pénurie, l'entreprise intégrée propriétaire de la Cobrecaf, privilégie ses propres conserveries et contraigne ses concurrents à acheter sur le marché mondial à des conditions moins favorables que celles fournies initialement par la Cobrecaf. (Avis n°95-A-01)

réserve élevée au distributeur dépositaire. La politique de distribution de la firme amont est issue de cet arbitrage³². L'existence d'une source alternative d'approvisionnement sous contrôle des distributeurs implique un conflit d'objectif entre l'amont et l'aval et une intensification de la concurrence sur le marché aval. Il résulte de ce phénomène un terrain favorable à l'émergence d'ententes dont l'objet est de préserver les profits de l'industrie, l'intégration verticale en est un exemple. Dans ce contexte, l'intégration verticale favorise la coordination des actions entre l'amont et l'aval et permet de relâcher la concurrence sur le marché aval.

Les conclusions que nous avançons sont doubles. Premièrement, la crainte selon laquelle l'intégration verticale est le vecteur d'une stratégie d'exclusion n'est pas fondée dans le cas étudié. En effet, lorsque l'intégration émerge, la firme non-intégrée ne fait pas l'objet d'une stratégie d'exclusion. Le fait de servir la firme non-intégrée permet à la firme intégrée d'accroître la demande en bien de qualité haute et de récupérer un revenu supplémentaire. Deuxièmement, le fait pour les entreprises en aval de disposer d'une source alternative d'approvisionnement soulève un conflit d'objectifs entre l'amont et l'aval, vecteur d'une intensification de la concurrence. Dans ces conditions, l'intégration verticale se doit d'être analysée par les autorités en charge de la politique de la concurrence comme une forme d'entente susceptible de préserver les profits de l'industrie.

³²Bien évidemment, les effets stratégiques de la politique de distribution adoptée par la firme amont sont liés à une annonce publique des contrats et sont sensibles à leur renégociation secrète. L'absence d'engagement sur les termes des contrats pose de nouveaux problèmes et constitue une des voies dans lesquelles il convient de poursuivre l'analyse. L'introduction d'une délégation avec contrats secrets peut alors permettre de mettre en évidence d'autres motivations privées propres à soutenir d'autres types d'entente verticale.

Annexe 1 : fonctions inverses de demande

Soit $\theta_{H,L}$ le consommateur indifférent entre la qualité haute et la qualité basse

$$\theta_{H,L} = \frac{p_H - p_L}{q_H - q_L}$$

Soit $\theta_{L,\odot}$ le consommateur indifférent entre la qualité basse et ne rien consommer

$$\theta_{L,\odot} = \frac{p_L}{q_L}$$

On a $\begin{cases} \sum y_{H_k} = 1 - \theta_{H,L} = 1 - \frac{p_H - p_L}{q_H - q_L} \\ \sum y_{L_k} = \theta_{H,L} - \theta_{L,\odot} = \frac{p_H - p_L}{q_H - q_L} - \frac{p_L}{q_L} \end{cases}$, les fonctions inverses de demande sont obtenues en résolvant le système en $\{p_L, p_H\}$.

$$\text{a) } \{A_1, A_2\}, \text{ on a } \begin{cases} p_L = \left(1 - \sum_{k=1}^2 y_{L_k} - \sum_{k=1}^2 y_{H_k}\right) q_L \\ p_H = \left(1 - \sum_{k=1}^2 y_{L_k} - \sum_{k=1}^2 y_{H_k}\right) q_L + \left(1 - \sum_{k=1}^2 y_{H_k}\right) (q_H - q_L) \end{cases}$$

$$\text{b) } \{A_i, R_j\}_{i \neq j=1,2}, \text{ on a } \begin{cases} p_L = \left(1 - \sum_{k=1}^2 y_{L_k} - y_{H_i}\right) q_L \\ p_H = \left(1 - \sum_{k=1}^2 y_{L_k} - y_{H_i}\right) q_L + (1 - y_{H_i})(q_H - q_L) \end{cases}$$

$$\text{c) } \{R_1, R_2\}, \text{ on a } \sum y_{L_k} = 1 - \theta_{L,\odot} = 1 - \frac{p_L}{q_L}, \text{ d'où } p_L = \left(1 - \sum_{k=1}^2 y_{L_k}\right) q_L$$

Annexe 2 : politique tarifaire de la firme amont avec un distributeur

Le programme de la firme amont s'écrit

$$\begin{aligned} \text{Max}_w \quad & \pi^{1D}(w) \\ \text{avec } \pi^{1D}(w) = & \begin{cases} (w - c) \left(\frac{2 - q_L - 2w}{4 - q_L}\right) + F & \text{si } w < \frac{1 - q_L}{3} \\ \text{avec } F = \left(\frac{2 - q_L - 2w}{4 - q_L}\right)^2 - \frac{q_L}{9} \\ (w - c) \left(\frac{1 - q_L - w}{2(1 - q_L)}\right) + F & \text{si } w > \frac{1 - q_L}{3} \\ \text{avec } F = \frac{5q_L^2 - q_L(14 - 18w) + 9(1 - w)^2}{36(1 - q_L)} - \frac{q_L}{9} \end{cases} \end{aligned}$$

La fonction objectif est concave par morceaux, on note $\left. \frac{\partial \pi^{1D}}{\partial w} \right|_G$ la dérivée si $w < \frac{1 - q_L}{3}$, et $\left. \frac{\partial \pi^{1D}}{\partial w} \right|_D$ si $w > \frac{1 - q_L}{3}$.

$$\left. \frac{\partial \pi^{1D}}{\partial w} \right|_G \Big|_{w = \frac{1 - q_L}{3}} \text{ s'annule pour } c_i(q_L) = \frac{2 - q_L}{6}$$

$\frac{\partial \pi^{1D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^D$ s'annule pour $c_j(q_L) = \frac{1-q_L}{3}$ avec $c_j(q_L) < c_i(q_L)$.

$c < c_j(q_L)$, $\frac{\partial \pi^{1D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^G < 0$, $\frac{\partial \pi^{1D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^D < 0$, la fonction objectif présente un maximum pour $w < \frac{1-q_L}{3}$.

$c \in [c_j(q_L), c_i(q_L)]$, $\frac{\partial \pi^{1D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^G < 0$, $\frac{\partial \pi^{1D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^D > 0$, la fonction présente deux maxima locaux.

$c > c_i(q_L)$, $\frac{\partial \pi^{1D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^G > 0$, $\frac{\partial \pi^{1D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^D > 0$, la fonction présente un maximum pour $w > \frac{1-q_L}{3}$.

On a

$$\text{si } c < c_a(q_L), \text{ avec } \begin{cases} c_a(q_L) = \text{Min} \{c_2(q_L), c_3(q_L)\} \in [c_j(q_L), c_i(q_L)] \\ c_2(q_L) = \frac{\sqrt{2-3q_L+q_L^2}}{3\sqrt{2}} \\ c_3(q_L) = \frac{(1-q_L)\sqrt{q_L}(3\sqrt{q_L}+2\sqrt{2+q_L})}{3(4-q_L)} \end{cases},$$

$$w = w_a = \text{Max} \left\{ \frac{2c(-4+q_L)-(-2+q_L)q_L}{4(-2+q_L)}, 0 \right\} \left(< \frac{1-q_L}{3} \right)$$

$$\text{avec } w_a = 0 \text{ si } c < c_1(q_L), c_1(q_L) = \frac{(2-q_L)q_L}{2(4-q_L)}$$

$$F = \pi_{D_1}^{A_1 R_2}(w_a)^{\text{Brut}} - \pi_{D_1}^{R_1 R_2}$$

si $c > c_a(q_L)$,

$$w = w_b = c \left(> \frac{1-q_L}{3} \right)$$

$$F = \pi_{D_1}^{A_1 R_2}(w_b)^{\text{Brut}} - \pi_{D_1}^{R_1 R_2}$$

Annexe 3 : politique tarifaire de la firme amont avec deux distributeurs

Le programme de la firme amont s'écrit

$$\text{Max}_w \pi^{2D}(w)$$

$$\text{avec } \pi^{2D}(w) = \begin{cases} 2(w-c) \left(\frac{1-q_L-w}{3(1-q_L)} \right) + 2F \\ \text{et } F = \begin{cases} \frac{(1-w)^2 - q_L(1-2w)}{9(1-q_L)} - \left(\frac{1+w}{4-q_L} \right)^2 q_L & \text{si } w < \frac{1-q_L}{3} \\ \frac{(1-w)^2 - q_L(1-2w)}{9(1-q_L)} - \frac{q_L}{9} & \text{si } w > \frac{1-q_L}{3} \end{cases} \end{cases}$$

$$\frac{\partial \pi^{2D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^G \text{ s'annule pour } c_k(q_L) = \frac{4+13q_L-17q_L^2}{9(4-q_L)}$$

$$\frac{\partial \pi^{2D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^D \text{ s'annule pour } c_l(q_L) = \frac{1-q_L}{9} \text{ avec } c_l(q_L) < c_k(q_L).$$

$c < c_l(q_L)$, $\frac{\partial \pi^{2D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^G < 0$, $\frac{\partial \pi^{2D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^D < 0$, la fonction objectif présente un maximum pour $w < \frac{1-q_L}{3}$.

$c \in [c_l(q_L), c_k(q_L)]$, $\frac{\partial \pi^{2D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^G < 0$, $\frac{\partial \pi^{2D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^D > 0$, la fonction présente deux maxima locaux.

$c > c_k(q_L)$, $\frac{\partial \pi^{2D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^G > 0$, $\frac{\partial \pi^{2D}}{\partial w} \Big|_{w=\frac{1-q_L}{3}}^D > 0$, la fonction présente un maximum pour $w > \frac{1-q_L}{3}$.

On a

si $c < c_b(q_L)$, avec $c_b(q_L) = \frac{1}{9} \left(-15 + 3q_L + 2\sqrt{2}\sqrt{32 - 7q_L - 7q_L^2} \right) \in [c_l(q_L), c_k(q_L)]$,

$$w = w_c = \text{Max} \left\{ \frac{-16 - 3c(-4 + q_L)^2 + 42q_L - 27q_L^2 + q_L^3}{2(-32 + 7q_L + 7q_L^2)}, 0 \right\}$$

$$F = \pi_{D_1}^{A_1 A_2}(w_c)^{\text{Brut}} - \pi_{D_1}^{R_1 A_2}(w_c)$$

si $c > c_b(q_L)$,

$$w = w_d = \frac{1}{4} (1 + 3c - q_L) \left(> \frac{1-q_L}{3} \right)$$

$$F = \pi_{D_1}^{A_1 A_2}(w_d)^{\text{Brut}} - \pi_{D_1}^{R_1 A_2}(w_d)$$

Annexe 4 : politique de distribution de la firme amont

On a $c_b(q_L) < c_a(q_L)$.

si $c \leq c_b(q_L)$ $\pi^{2D}(w_c) > \pi^{2D}(w_d)$ et $\pi^{1D}(w_a) > \pi^{1D}(w_b)$, la solution est définie par $\text{Max} \{ \pi^{2D}(w_c), \pi^{1D}(w_a) \}$. Il existe $c_c(q_L) < c_b(q_L)$,³³ tel que pour $c \leq c_c(q_L)$, $\text{Max} \{ \pi^{2D}(w_c), \pi^{1D}(w_a) \} = \pi^{2D}(w_c)$, sinon $c \in]c_c(q_L), c_b(q_L)]$, $\text{Max} \{ \pi^{2D}(w_c), \pi^{1D}(w_a) \} = \pi^{1D}(w_a)$.

si $c \in]c_b(q_L), c_a(q_L)[$ $\pi^{2D}(w_d) > \pi^{2D}(w_c)$ et $\pi^{1D}(w_a) > \pi^{1D}(w_b)$, $\pi^{2D}(w_d) = \pi^{1D}(w_b)$ d'où $\pi^{1D}(w_a)$.

si $c \in [c_a(q_L), (1 - q_L)]$ $\pi^{2D}(w_d) > \pi^{2D}(w_c)$ et $\pi^{1D}(w_b) > \pi^{1D}(w_a)$, $\pi^{2D}(w_d) = \pi^{1D}(w_b)$ d'où indifférence entre $\pi^{1D}(w_b)$ et $\pi^{2D}(w_d)$.

Annexe 5 : choix de la firme intégrée de servir ou non la firme non intégrée

$$^{33}\text{avec } c_c(q_L) = \begin{cases} c_6(q_L) & \text{si } c > c_1(q_L) \\ c_7(q_L) & \text{sinon} \end{cases}$$

$$c_6(q_L) = \left(\frac{1}{-42 - 60q_L + 48q_L^2} \right) \left(\frac{-90 + 153q_L - 72q_L^2 + 9q_L^3 + \sqrt{3}}{\sqrt{2560 - 9520q_L + 13112q_L^2 - 7290q_L^3 + 277q_L^4 + 1162q_L^5 - 301q_L^6}} \right)$$

$$c_7(q_L) = \frac{(1-q_L)\sqrt{q_L}}{3(-4+q_L)^4} \left(\sqrt{q_L} (-138 + 51q_L + 6q_L^2) + \sqrt{3} \sqrt{-640 + 3980q_L - 2140q_L^2 - 557q_L^3 + 322q_L^4 + 7q_L^5} \right)$$

Configurations sur le marché aval

* Si $c < \frac{(1-q_L)}{3}$

$$w_2 < w_2^a \text{ avec } w_2^a = \frac{c}{2},$$

$$\begin{cases} y_{L1} = \frac{c}{2(1-q_L)} & y_{L2} = 0 \\ y_{H1} = \frac{2-4c+2w_2+q_L(c-2(1+w_2))}{6(1-q_L)} & y_{H2} = \frac{1}{3}(1+c-2w_2) \end{cases}$$

$$w_2 \in [w_2^a, w_2^b] \text{ avec } w_2^b = 2c,$$

$$\begin{cases} y_{L1} = \frac{2c-w_2}{3(1-q_L)} & y_{L2} = \frac{2w_2-c}{3(1-q_L)} \\ y_{H1} = \frac{1-q_L-2c+w_2}{3(1-q_L)} & y_{H2} = \frac{1-q_L-2w_2+c}{3(1-q_L)} \end{cases}$$

$$w_2 \in [w_2^b, w_2^c] \text{ avec } w_2^c = \frac{2(1-q_L)(1+c)}{4-q_L}$$

$$\begin{cases} y_{L1} = 0 & y_{L2} = \frac{w_2}{2(1-q_L)} \\ y_{H1} = \frac{1}{3}(1+w_2-2c) & y_{H2} = \frac{2-4w_2+2c+q_L(w_2-2(1+c))}{6(1-q_L)} \end{cases}$$

$$w_2 > w_2^c,$$

$$\begin{cases} y_{L1} = 0 & y_{L2} = \frac{1+c}{4-q_L} \\ y_{H1} = \frac{2-q_L-2c}{4-q_L} & y_{H2} = 0 \end{cases}$$

Le programme de la firme intégrée s'écrit

$$\text{Max}_{w_2} \pi^I(w_2)$$

$$\text{avec } \pi^I(w_2) = (p_L(\cdot) - c_L) y_{L1}(w_2) + (p_H(\cdot) - c_H) y_{H1}(w_2) + (w_2 - c_H) y_{H2}(w_2) + \pi_{D_2}^{IA_2}(w_2)^{\text{Brut}} - \pi_{D_2}^{IR_2}$$

$$\frac{\partial \pi^I(w_2)}{\partial w_2} \Big|_{w_2=w_2^a}^G > 0, \frac{\partial \pi^I(w_2)}{\partial w_2} \Big|_{w_2=w_2^b}^D > 0, \frac{\partial \pi^I(w_2)}{\partial w_2} \Big|_{w_2=w_2^b}^G > 0, \frac{\partial \pi^I(w_2)}{\partial w_2} \Big|_{w_2=w_2^c}^G < 0$$

et $\frac{\partial \pi^I(w_2)}{\partial w_2} \Big|_{w_2=w_2^b}^D$ s'annule pour c_i avec $c_i = \frac{2(1-q_L)}{3(2+q_L)}$.

si $c < c_i$, $\frac{\partial \pi^I(w_2)}{\partial w_2} \Big|_{w_2=w_2^b}^D > 0$ la fonction objectif admet une solution intérieure à l'intervalle $[w_2^b, w_2^c]$

$$w_2 = w_2^d = \frac{2-2q_L+c(2+7q_L)}{4+5q_L}$$

$$F_2 = \pi_{D_2}^{IA_2}(w_2^d)^{\text{Brut}} - \pi_{D_2}^{IR_2}$$

si $c \in [c_i, \frac{(1-q_L)}{3}]$ $\frac{\partial \pi^I(w_2)}{\partial w_2} \Big|_{w_2=w_2^b}^D < 0$, la fonction objectif admet une solution en coin $w_2 = w_2^b$

$$w_2 = w_2^b = 2c$$

$$F_2 = \pi_{D_2}^{IA_2}(w_2^b)^{\text{Brut}} - \pi_{D_2}^{IR_2}$$

* Si $c > \frac{(1-q_L)}{3}$, la fonction objectif est concave et croissante par morceaux

$w_2 = w_2^c - \varepsilon$ avec $w_2^c = \frac{c}{2} + \frac{(1-q_L)}{2}$ et $\varepsilon (> 0)$ petit,

$$\begin{cases} y_{L1} = \frac{2c-w_2}{3(1-q_L)} & y_{L2} = \frac{2w_2-c}{3(1-q_L)} \\ y_{H1} = \frac{1-q_L-2c+w_2}{3(1-q_L)} & y_{H2} = \frac{1-q_L-2w_2+c}{3(1-q_L)} \end{cases}$$

$\frac{\partial \pi^I(w_2)}{\partial w_2} \Big|_{w_2=w_2^c-\varepsilon} > 0$, il y a forclusion.

Annexe 6 : étude normative

Δ de surplus liée à l'intégration verticale

$\Delta S_C > 0$ si $c \in]c_a(q_L), c_d(q_L)]$ ³⁴

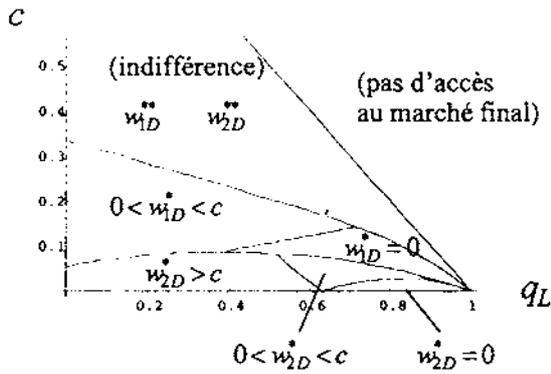
$\Delta S > 0$ si $c \in]c_a(q_L), c_e(q_L)]$ ³⁵

$${}^{34}c_d(q_L) = \frac{(2-q_L)(16-31q_L+15q_L^2-\sqrt{64-64q_L+240q_L^2-316q_L^3+q_L^4+75q_L^5})}{(48-34q_L+10q_L^2-6q_L^3)}$$

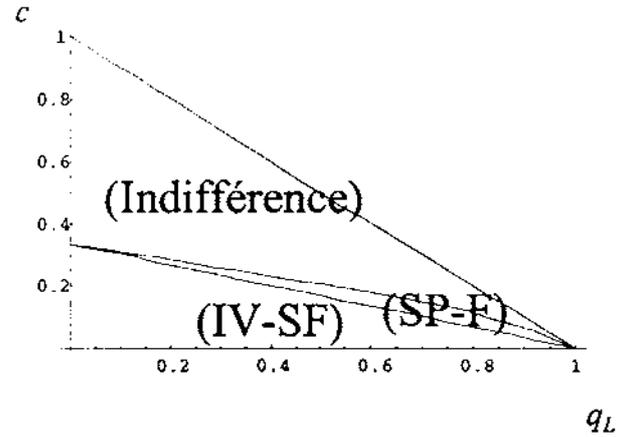
$${}^{35}c_e(q_L) = \frac{(2-q_L)(16-17q_L+q_L^2-\sqrt{64-64q_L+208q_L^2-276q_L^3+53q_L^4+15q_L^5})}{2(24+9q_L-5q_L^2-q_L^3)}$$

Annexes graphiques

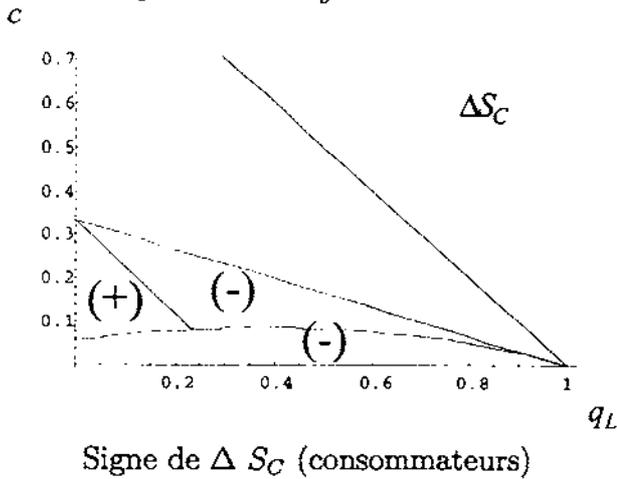
Politique de distribution de la firme amont



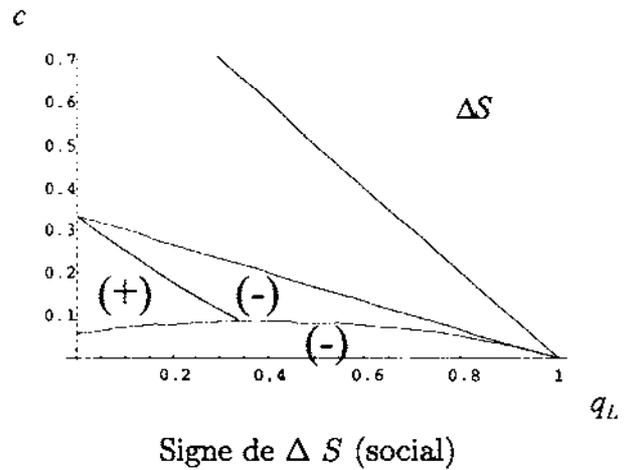
Choix d'intégration



ΔS_C liée à l'intégration verticale



ΔS liée à l'intégration verticale



Bibliographie

Caillaud B. et P. Rey (1995) : "Strategic Aspects of Vertical Delegation", *European Economic Review*, 39, 421-431.

Caprice, S. et Philippe J. (1997) : "Les relations verticales", dans "Réglementation et concurrence", Ed. A. Perrot, Collection ESA, Economica.

Comanor, W. S. et P. Rey (1995) : "Vertical Restraints and the Market Power of Large Distributors", Document de travail CREST/INSEE, 9537.

Katz, M.L. (1989) : "Vertical Contractual Relations", dans R. Schmalensee et R.D.Willig, "*Handbook of industrial Organization*", Vol. I, 655-721.

Mills D. E. (1995) : "Why Retailers Sell Private Labels", *Journal of Economics and Management Strategy*, 4(3), 509-528.

O'Brien, D. P. et G. Shaffer (1997) : "Nonlinear Supply Contracts, Exclusive Dealing, and Equilibrium Market Foreclosure", *Journal of Economics and Management Strategy*, 755-785.

Ordover, J. A., Saloner, G. et S. C Salop (1990) : "Equilibrium Vertical Foreclosure", *American Economic Review*, 80(1), 127-142.

Rey, P. (1994) : "Impact des accords verticaux entre producteurs et distributeurs", mimeo CREST-LEI.

Rey, P. et J. Tirole (1996) : "A Primer on Foreclosure", mimeo I.D.E.I., Toulouse.

Shaffer, G. (1991a) : "Capturing Strategic Rent : Full-Line Forcing, Brand Discounts, Aggregate Rebates, and Maximum Resale Price Maintenance", *Journal of Industrial Economics*, 39(5), 557-575.

Shaffer, G. (1991b) : "Slotting Allowances and Resale Price Maintenance : a Comparison of Facilitating Practices", *Rand Journal of Economics*, 22(1), 120-135.

Vigny, J. (1994) : "La distribution : Structures et pratiques", Ed. Dalloz Gestion Marketing.

ANNEXE 5

Mutualisation des risques et concurrence sur le marché de l'assurance

Mutualisation des risques et concurrence sur le marché de l'assurance

Marie-Cécile Fagart, CREST-LEI, Université de Rouen
Nathalie Fombaron, THEMA, CREST-LEI, Université Paris X
Meglana Jeleva, CREST-LEI, EUREQua, Université de Nantes

Février 2000
Version préliminaire

1 Présentation du modèle et contrats efficaces

1.1 Le cadre d'analyse

Sur le marché de l'assurance existent des agents identiques et adversaires du risque. Les agents peuvent se rassembler pour constituer des mutuelles, groupes d'individus pratiquant la co-assurance entre eux. Ils prévoient avant la réalisation de l'état de la nature un contrat qui organise les paiements et remboursements en fonction de la réalisation des accidents de chacun des membres. Ces paiements doivent être réalisables, la somme des remboursements doit être négative ou nulle quels que soient les sinistres subis.

Sur ce marché existent aussi des compagnies d'assurance. Une compagnie est constituée par un groupe d'individus (les actionnaires), neutres vis à vis du risque, et possédant un capital. La compagnie propose à ses clients des contrats d'assurance qui, comme pour la mutuelle, peuvent dépendre de la réalisation des accidents du groupe d'assurés. Ces paiements doivent être réalisables, c'est à dire que les fonds propres de l'entreprise doivent suffire à couvrir des engagements (les remboursements prévus par le contrat).

Les contrats proposés doivent être non discriminants et robustes à la renégociation. Ceci implique qu'ils sont optimaux au sens de Pareto. Pour la mutuelle, le contrat proposé doit être réalisable et il rend maximale l'espérance d'utilité de chacun de ses membres. Il en est de même pour les compagnies d'assurance, sachant que l'espérance de profit des actionnaires est supérieure à une certaine borne.

Remarquons qu'une compagnie d'assurance est modélisée comme une mutuelle dont certains membres sont neutres vis-à-vis du risque et prêts à engager une partie limitée de leur richesse. Une mutuelle est une compagnie qui ne possède ni fonds propres, ni actionnaires.

Par définition, les contrats peuvent dépendre du nombre d'individus qui s'assurent auprès de la société qui les proposent. Il est aussi possible que les actionnaires perdent leur fonds propres. On appelle cette situation faillite ou insolvabilité. Toutefois cette faillite ne signifie pas que les assurés ne seront pas remboursés, mais tout simplement que le montant des avoirs fixes de la compagnie d'assurance sont insuffisants pour couvrir les dommages.

Le jeu concurrentiel étudié est un jeu de concurrence en contrats en deux étapes dans lequel, dans une première étape, les sociétés présentes sur le marché proposent des contrats d'assurance. Dans la seconde étape, les agents choisissent simultanément le contrat qu'ils préfèrent dans l'ensemble proposé.

Toutes les compagnies d'assurance sont identiques pour les agents, et quand il n'existe pas de mutuelle, le processus concurrentiel est de type Bertrand.

On étudiera successivement le cas du monopole, puis le cas du monopole concurrencé par une mutuelle. On s'intéressa ensuite à l'oligopole, sans mutuelle puis avec mutuelle.

1.2 Les contrats efficaces

Soit $X_i \in [0, \infty[$ la variable aléatoire qui représente la perte monétaire d'un agent i . Sa densité est notée $f(\cdot)$ et soient $E(X_i) = m$ et $V(X_i) = \sigma^2$. Les accidents individuels sont des variables indépendantes et de même loi. Les agents sont adversaires du risque et leur fonction d'utilité de Von Neumann Morgenstern est $U(\cdot)$, avec $U' > 0$ et $U'' < 0$.

Supposons qu'une société (compagnie ou mutuelle) assure N agents. Un état de la nature est la description des accidents de chaque agent, c'est la réalisation d'un N -uplet (X_1, X_2, \dots, X_N) . Si l'on note $\beta_i(x_1, x_2, \dots, x_N) \in \mathcal{R}$ le paiement fait à l'agent i si l'état de la nature est (x_1, x_2, \dots, x_N) , un contrat est décrit par $C(N) = \{\beta_i(x_1, x_2, \dots, x_N), i = 1, \dots, N, (x_1, x_2, \dots, x_N) \in [0, \infty[^N\}$. Un contrat est réalisable dès lors que quels que soient les accidents, les paiements qu'il prévoit sont inférieurs aux fonds propres K , c'est à dire que $\forall (x_1, x_2, \dots, x_N)$, on doit avoir $K \geq \sum_{i=1}^N \beta_i(x_1, x_2, \dots, x_N)$. Pour une mutuelle, celle-ci ne possédant pas de fonds propres, cette inégalité doit être vérifiée pour $K = 0$.

Un contrat $C(N)$ donne à un agent i une espérance d'utilité

$$\Psi_i(\beta_i) = \int U(w - x_i + \beta_i(x_1, x_2, \dots, x_N)) f(x_1) f(x_2) \dots f(x_N) dx_1 dx_2 \dots dx_N$$

et engendre une espérance de profit

$$\Phi(\beta_1, \beta_2, \dots, \beta_N) = \int (K + \sum_{i=1}^N \beta_i(x_1, x_2, \dots, x_N)) f(x_1) f(x_2) \dots f(x_N) dx_1 \dots dx_N$$

Un contrat est efficace s'il maximise $\sum_{i=1}^N \gamma_i \Psi_i(\beta_i) + \delta \Phi(\beta_1, \beta_2, \dots, \beta_N)$ dans l'ensemble des contrats réalisables. Pour une compagnie d'assurance le paramètre δ sera choisi strictement positif. Pour une mutuelle, le stock de capital K et le paramètre δ sont nuls. Plus particulièrement, on s'intéresse à des contrats efficaces non discriminants, c'est à dire vérifiant $\gamma_i = \gamma$ pour $i = 1, \dots, N$.

Proposition 1 *Les contrats efficaces et non discriminants sont, pour la mutuelle,*

$$\beta_i(x_1, x_2, \dots, x_N) = x_i - \frac{\sum_{j=1}^N x_j}{N},$$

et pour la compagnie d'assurance

$$\beta_i(x_1, x_2, \dots, x_N) = x_i - \text{Max}\left(\alpha, \frac{\sum_{j=1}^N x_j - K}{N}\right), i = 1, \dots, N.$$

Conformément au principe de mutualité de Borch (1962), la richesse de l'agent dépend, non pas de son propre sinistre mais du sinistre agrégé. Pour la mutuelle, cette richesse est égale à $w - \frac{\sum_{j=1}^N x_j}{N}$, pour la compagnie, $w - \text{Max}\left(\alpha, \frac{\sum_{j=1}^N x_j - K}{N}\right)$. Notons toutefois

que pour des montants faibles de sinistralité on a $\sum_{j=1}^N x_j \leq K + N\alpha$, et la richesse de l'agent assuré par une compagnie est constante.

Tout se passe comme si la mutuelle s'engageait à rembourser totalement les pertes individuelles, mais calcule la prime d'assurance ex post et la choisit pour équilibrer ses

comptes égale au sinistre moyen $\frac{\sum_{j=1}^N x_j}{N}$. Si l'on note $f_N(\cdot)$ la fonction de densité des pertes agrégées des N agents, l'espérance d'utilité d'un membre de la mutuelle est alors fonction de la taille du groupe, on la note

$$V_M(N) = \int U(w - \frac{x}{N}) f_N(x) dx$$

En ce qui concerne la compagnie, le contrat fait apparaître une prime d'assurance α fixe payée par l'agent si le sinistre moyen est suffisamment faible, c'est à dire si les fonds propres et le paiement de cette prime suffisent à rembourser la totalité des dommages subis, si $K + N\alpha \geq \sum_{i=1}^N x_i$. Dans le cas contraire, la prime fixe étant insuffisante, le contrat prévoit un rappel de prime ex post pour équilibrer les comptes de l'entreprise.

L'espérance d'utilité d'un agent assuré par cette compagnie peut alors être écrite:

$$V(\alpha, N, K) = \int U(w - \text{Max}(\alpha, \frac{x}{N} - \frac{K}{N})) f_N(x) dx$$

L'espérance de profit de la compagnie d'assurance est:

$$\Pi(\alpha, N, K) = \int \text{Max}(K + N\alpha - x, 0) f_N(x) dx$$

Le graphique 1 représente la richesse de l'agent comme fonction du sinistre moyen, en trait pointillé s'il est assuré auprès d'une mutuelle, en trait plein auprès d'une compagnie. Pour des niveaux faibles de sinistralité, l'agent qui s'assure auprès de la mutuelle paye une prime plus faible que celle de la compagnie, mais variable tandis que celle de la compagnie est stable. Pour des niveaux de sinistralité élevée (faillite), les fonds propres sont consacrés au remboursement des agents, et ceux-ci payent dans ce cas une prime plus faible. Les fonds propres de la compagnie permettent un échange entre les assurés et les actionnaires assureurs, qui obtiennent une rémunération positive si le sinistre est faible mais renoncent à leur capital investi quand l'accident agrégé est trop grand.

Insérer le graphique 1

Ces contrats peuvent être décrits par les clauses suivantes. La société prévoit un remboursement des dommages individuels, moyennant le paiement d'une prime fixe, α . Ce paiement est ajusté ex post par un rappel de primes, qui dans le cas de la mutuelle équilibre les comptes (le trop perçu est reversé) et, dans le cas de la compagnie, ce rappel ex post n'arrive qu'en cas de faillite (c'est à dire quand fonds propres et primes sont insuffisants pour faire face aux engagements) et doit alors équilibrer les comptes.

Apparaît alors une asymétrie entre mutuelle et compagnie. Parce que la mutuelle tarifie ex post, en fonction du nombre d'assurés qu'elle obtient (en seconde étape du jeu),

elle n'a aucun choix stratégique en première étape. La compagnie, quant à elle, propose une prime ex ante, indépendante du nombre d'assurés qu'elle obtiendra en seconde étape, puis ajustera ex post cette prime en cas de faillite. Choisir un contrat pour la compagnie revient à sélectionner une prime.

Le jeu de concurrence peut alors être modélisé ainsi : Dans la première étape, les n compagnies d'assurance présentes proposent des primes fixes $\alpha_i, i = 1, \dots, n$. Dans une seconde étape, les agents observent ces offres puis décident simultanément de s'adresser à l'une des compagnies présentes ou de s'assurer auprès de la mutuelle (si cette possibilité existe) ou encore de ne pas s'assurer. L'espérance d'utilité d'un individu assuré auprès de la compagnie j est $V(\alpha_j, N_j, K)$, avec N_j le nombre d'agents de la société. L'espérance de profit est alors $\Pi(\alpha_j, N_j, K)$. De même, l'espérance d'utilité d'un membre de la mutuelle est $V_M(N)$, avec N le nombre des co-assurés. Remarquons qu'un agent non assuré obtient une espérance d'utilité $V_M(1)$.

1.3 L'effet taille : comment varie l'espérance d'utilité de l'agent quand le nombre de ses co-assurés augmente ?

Proposition 2 Soit $Y_N = \frac{\sum_{i=1}^N X_i}{N}$ le montant moyen de sinistres pour N individus. La variable aléatoire Y_{N+1} DS2 Y_N .

La proposition 2 est la clef de notre modèle. Elle nous garantit que le sinistre moyen est une variable aléatoire d'autant moins risquée (au sens de la dominance stochastique d'ordre 2, voir Rothschild et Stiglitz (1970 et 1971)) que le nombre d'individus pour lequel est considérée cette moyenne est grand. Cette proposition engendre les résultats ci dessous.

Lemma 1 $V_M(N)$ est une fonction strictement croissante de N .

L'utilité d'un membre de la mutuelle est une fonction strictement croissante du nombre de mutualisés. L'utilité est en effet une fonction strictement concave du montant moyen de sinistres, et le sinistre moyen est une variable aléatoire d'autant moins risquée que le nombre de mutualisés est important. Ce résultat est évident dans un modèle moyenne-variance, puisque l'espérance du sinistre moyen ne dépend pas du nombre d'assurés, et que la variance diminue avec ce nombre (elle est égale à $\frac{\sigma}{(N)^{1/2}}$). Si la loi des grands nombres s'applique, l'espérance d'utilité tend vers sa borne supérieure $U(w - m)$.

Lemma 2 $\frac{\Pi(\alpha, N, Nk)}{N}$ est une fonction décroissante de N et $V(\alpha, N, Nk)$ est une fonction croissante de N .

La preuve de ce lemme repose sur le fait que $\frac{\Pi(\alpha, N, Nk)}{N}$ est l'espérance d'une fonction convexe du sinistre moyen et $V(\alpha, N, Nk)$ celle d'une fonction concave.

Lemma 3 Si $\alpha \geq m$, $\Pi(\alpha, N, K)$ est une fonction (strictement si $\alpha > m$) croissante de N , si $\alpha < m$, et N grand $\Pi(\alpha, N, K) < K$.

Remarquons que le profit est toujours plus grand que $K + N(\alpha - m)$. Si de plus N est grand, la loi des grands nombres s'applique et l'espérance de profit de l'entreprise tend vers $K + N(\alpha - m)$. Autrement dit si $\alpha = m$, l'espérance de profit de l'entreprise est très proche de K .

Si la prime est plus grande que la prime actuarielle, la compagnie préfère avoir beaucoup de clients. Si les primes sont trop faibles, la compagnie fera des pertes et préfère n'assurer personne (elle conserve alors ses fonds propres K).

Le problème est plus complexe en ce qui concerne l'utilité d'un individu assuré par une compagnie. En effet, d'une part augmenter le nombre d'agents fait courir moins de risques à l'assuré. Mais augmenter le nombre d'agents diminue le capital par tête et donc augmente la prime payée dans le cas de faillite. Si le capital par tête reste constant, seul joue l'effet mutualisation et l'espérance d'utilité de l'agent est une fonction strictement croissante de la taille de la compagnie. Mais l'effet capital par tête vient s'opposer à l'effet mutualisation, et la monotonie n'est plus garantie.

Le lemme suivant nous permet de préciser cette intuition.

Lemma 4 Si $\alpha \geq m$, \bar{N} grand, $V(\alpha, \bar{N}, K) \geq V(\alpha, N, K)$ quel que soit $N \leq \bar{N}$. Il existe deux seuils $N^{**} \leq N^*$ tel que $V(\alpha, N, K)$ est une fonction croissante de N , $\forall N \geq N^*$ et si K est suffisamment grand $V(\alpha, N, K)$ est une fonction décroissante de N , $\forall N \leq N^{**}$.

L'espérance d'utilité de l'agent est une fonction non monotone du nombre d'assurés, décroissante quand ce nombre est petit, croissante s'il est grand. Elle est maximale si tous les agents sont groupés au sein de la même compagnie. Ce résultat est du à la loi des grands nombres, l'utilité de l'agent quand le nombre de clients devient suffisamment grand tend vers sa borne supérieure $U(w - \alpha)$ si $\alpha \geq m$.

Hypothèses et notations supplémentaires

H_1 – Il existe \bar{N} agents sur le marché et ce nombre est grand, de même $\frac{\bar{N}}{n}$ est grand.

H_2 – L'espérance d'utilité $V(\alpha, N, K)$ est une fonction décroissante de N en dessous d'un certain seuil puis croissante au-dessus de ce seuil.

H_3 – $V(m, 1, K) \leq V_M(\bar{N})$

Enfin on note

$U(w - m - \rho) = V_M(1)$

r est le taux de profit de réservation des actionnaires.

Les preuves omises sont reportées en annexe.

1.4 Optimalité des fusions, qu'elles résultent des structures concurrentielles ou des structures mutualistes

La prise en compte de la mutualisation des risques conduit à des résultats très tranchés en terme d'optimalité des fusions. Les fusions dominent au sens de Pareto. Ce résultat

est évident pour les mutuelles puisque l'espérance d'utilité d'un agent mutualisé est une fonction strictement croissante de la taille de cette mutuelle. Autrement dit regrouper deux mutuelles augmente l'espérance d'utilité des membres de chaque organisation. Cette intuition reste valide dans le cas des compagnies d'assurance.

Proposition 3 *La fusion de deux compagnies identiques est efficace au sens de Pareto.*

Proposition 4 *Si l'Etat peut réguler le marché de l'assurance, il groupe tous les agents au sein de la même société, soit une mutuelle, soit une compagnie. Ce choix dépend du taux de profit de réservation des actionnaires. Il existe un taux d'intérêt $r^* > 0$, tel que si $r \leq r^*$, une compagnie est préférable, et si $r \geq r^*$, la mutuelle sera choisie.*

Ces deux propositions nous livrent une conclusion très claire. En assurance, comme l'affirment les professionnels, "il faut être gros". Cette remarque est confirmée par le modèle théorique dans toutes ses interprétations possibles:

- si deux compagnies assurent un nombre identique d'individus et réalisent la même espérance de profit, il existe un contrat pour l'entité fusionnée qui augmente l'utilité des assurés (strictement) sans réduire l'espérance de profit des entreprises. La fusion est donc profitable;

- grouper les agents au sein d'une même société est efficace, celle-ci peut être, soit une mutuelle, soit une compagnie. Ce choix dépend du niveau du taux de profit r . Ainsi par exemple si $r = 0$, l'utilité d'un individu assuré par une compagnie privée qui propose une prime $\alpha = m$ est plus grande (quel que soit le nombre d'assurés) que son espérance d'utilité au sein d'une mutuelle (le contrat proposé par la mutuelle appartient à l'ensemble des possibles de la compagnie). Dans ce cas une compagnie régulée est plus efficace qu'une mutuelle;

- si l'Etat peut choisir le stock de capital optimal, celui-ci est une fonction décroissante de r .

2 Equilibre sur le marché de l'assurance

2.1 Présentation du concept d'équilibre

On cherche à déterminer l'ensemble des équilibres en stratégies pures de Nash parfaits en sous-jeu du jeu.

Remarquons qu'en seconde période, les décisions des agents sont simultanées, chaque assuré choisit une société, prévoyant les stratégies (pures) des autres clients du marché. On note $N_i, i = 1, \dots, n$ le nombre de clients de la compagnie $i = 1, \dots, n$ à l'équilibre de seconde période du jeu. On appelle configuration du marché un n -uplet (N_1, N_2, \dots, N_n) qui vérifie $N_i \geq 0$ et $\sum_{i=1}^n N_i \leq \bar{N}$.

(N_1, N_2, \dots, N_n) est une configuration stable du marché pour les offres $(\alpha_1, \alpha_2, \dots, \alpha_n)$ si les choix des agents que décrit ce partage du marché forment un équilibre de Nash de la seconde étape du jeu. Ces conditions sont données par $\forall i, \forall j \neq i, j, j = 1, \dots, n$

- a) Si $N_i > 0, V(\alpha_i, N_i, K) \geq V(\alpha_j, N_j + 1, K)$,
- b) Si $N_i > 0, V(\alpha_i, N_i, K) \geq V_M(1)$ si la mutuelle n'existe pas, et $V(\alpha_i, N_i, K) \geq V_M(\bar{N} + 1 - \sum_{j=1}^{j=n} N_j)$ si le jeu admet une mutuelle,
- c) Si $\bar{N} - \sum_{j=1}^{j=n} N_j > 0$ s'il existe une mutuelle $V_M(\bar{N} - \sum_{j=1}^{j=n} N_j) \geq V(\alpha_i, N_i + 1, K)$
- d) Si $\bar{N} - \sum_{j=1}^{j=n} N_j > 0$, et s'il n'existe pas de mutuelle $V_M(1) \geq V(\alpha_i, N_i + 1, K)$.

Si une compagnie i assure effectivement des agents à l'équilibre du jeu, alors l'espérance d'utilité obtenue par chacun de ses clients, $V(\alpha_i, N_i, K)$ doit être supérieure à l'espérance d'utilité qu'obtiendrait ce client en s'adressant à une autre compagnie (condition a), en s'assurant lui même s'il n'existe pas de mutuelle ou en s'assurant par le biais de la mutuelle (condition b).

Si la mutuelle existe et est active, ses membres préfèrent s'assurer auprès de la mutuelle qu'auprès des compagnies présentes (c). S'il n'existe pas de mutuelles et que certains agents restent sans assurance, ces individus préfèrent la non assurance plutôt que n'importe laquelle des compagnies (d).

Ce concept d'équilibre parfait en sous jeu laisse apparaître à la fois une multiplicité d'équilibres et une sous-optimalité de certains équilibres en seconde étape. Cette multiplicité est due au fait que l'espérance d'utilité d'un agent dépend du nombre d'individus qui s'assurent dans la même société. La sous-optimalité illustre la difficulté qu'ont les agents à se coordonner sur l'un des équilibres.

Dans un premier temps, on étudie les équilibres du jeu qui vérifient que dans la seconde étape quelles que soient les offres de première période, les agents se coordonnent sur la configuration qui maximisent leur espérance d'utilité. On les appelle comportement de type "boule de neige". Puis on étudie tous les autres équilibres.

Plus précisément, les offres $(\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)$ des compagnies et la configuration

$(N_1^*, N_2^*, \dots, N_n^*)$ forment un équilibre du jeu si :

e) $(N_1^*, N_2^*, \dots, N_n^*)$ est une configuration stable du marché pour les offres $(\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)$;

f) si la compagnie i dévie et choisit d'offrir une prime α_i plutôt que α_i^* , il existe une configuration de marché (N_1, N_2, \dots, N_n) stable pour les offres

$$(\alpha_1^*, \alpha_2^*, \dots, \alpha_{i-1}^*, \alpha_i, \alpha_{i+1}^*, \dots, \alpha_n^*)$$

qui rende la déviation non profitable c'est à dire telle que $\Pi(\alpha_i^*, N_i^*, K) \geq \Pi(\alpha_i, N_i, K)$.

2.2 Le monopole

Proposition 5 *Si les agents adoptent un comportement de type boule de neige, le jeu admet un seul équilibre dans lequel le monopole assure tous les agents. Si le monopole n'est pas concurrencé par la mutuelle, il propose la prime α^m qui vérifie $V(\alpha^m, \bar{N}, K) = V_M(1)$. S'il existe une mutuelle, le monopole propose la prime α_M^m telle que $V(\alpha_M^m, K, \bar{N}) = V_M(\bar{N})$.*

Preuve.

Rappelons qu'une proposition α^* et un nombre de clients M^* est un équilibre du jeu de monopole si M^* est une configuration stable pour l'offre α^* et si pour toute offre $\alpha \neq \alpha^*$, il existe une configuration stable $M(\alpha)$ telle que $\Pi(\alpha^*, M^*, K) \geq \Pi(\alpha, M(\alpha), K)$.

Il existe une seule configuration stable vérifiant l'hypothèse "boule de neige", elle est donnée par:

- si la mutuelle n'existe pas, $M(\alpha) = \bar{N}$ si $V(\alpha, \bar{N}, K) \geq V_M(1)$, et $M(\alpha) = 0$ si $V(\alpha, \bar{N}, K) < V_M(1)$;

- si la mutuelle existe, $M(\alpha) = \bar{N}$ si $V(\alpha, \bar{N}, K) \geq V_M(\bar{N})$, et $M(\alpha) = 0$ si $V(\alpha, \bar{N}, K) < V_M(\bar{N})$.

Le profit maximum que peut obtenir le monopole suite à une déviation est alors

- si la mutuelle n'existe pas, $\Pi(\alpha^m, \bar{N}, K)$ avec $V(\alpha^m, \bar{N}, K) = V_M(1)$;

- si la mutuelle existe, $\Pi(\alpha_M^m, \bar{N}, K)$ avec $V(\alpha_M^m, \bar{N}, K) = V_M(\bar{N})$.

On en déduit que si la mutuelle n'existe pas, $M^* = \bar{N}$ et $\alpha^* = \alpha^m$, si la mutuelle existe, $M^* = \bar{N}$ et $\alpha^* = \alpha_M^m$.

Notons que $\alpha_M^m > m$, puisque $V(m, \bar{N}, K) > V_M(\bar{N})$. La mutuelle ne sera pas active à l'équilibre. \square

Sous l'hypothèse "boule de neige" l'équilibre du jeu est unique. Si le monopole n'est pas concurrencé il facture la prime la plus élevée possible. Le pouvoir concurrentiel de la mutuelle est très grand. En effet, le rôle de concurrent potentiel qu'elle joue suffit à diminuer considérablement les primes, et si la taille du marché est suffisamment importante, la prime est proche de la prime actuarielle.

Si l'on suppose que les agents ne se coordonnent pas sur l'équilibre qui leur est le plus favorable, on obtient les résultats suivants.

Proposition 6 *Le jeu de monopole admet un grand nombre d'équilibres qui vérifient les propriétés suivantes:*

i) *il existe $\alpha^m > \alpha_0 > m$ telle que toutes les offres $\alpha^* \in [\alpha_0, \alpha^m]$ et les configurations $M^* = \bar{N}$ sont des équilibres du jeu;*

ii) *il est possible qu'il existe une prime $\alpha^* \in]m, m + \rho]$ et une configuration stable vérifiant $0 < M^* < \bar{N}$ et $V(\alpha^*, M^*, K) \geq V_M(1) > V(\alpha^*, M^* + 1, K)$, qui soit un équilibre du jeu.*

Preuve.

Comme on l'a remarqué précédemment, une proposition α^* et un nombre de clients M^* est un équilibre si M^* est une configuration stable pour l'offre α^* et si pour toute offre $\alpha \neq \alpha^*$, il existe une configuration stable $M(\alpha)$ telle que $\Pi(\alpha^*, M^*, K) \geq \Pi(\alpha, M(\alpha), K)$. Pour déterminer tous les équilibres du jeu, il suffit de trouver la configuration stable la pire pour le monopole s'il dévie. On note $\text{Min}_M V(\hat{\alpha}, M, K) = V_M(1) = V(\hat{\alpha}, 1, K)$.

Posons:

si $\alpha \leq \hat{\alpha}$, $M(\alpha) = \bar{N}$;

si $\hat{\alpha} \geq \alpha \geq \hat{\alpha}$, $V(\alpha, M(\alpha), K) \geq V_M(1) > V(\alpha, M(\alpha) + 1, K)$;

si $\alpha \geq \tilde{\alpha}$, $M(\alpha) = 0$.

On peut vérifier que $M(\alpha)$ est une configuration stable, qui minimise le nombre de clients de la compagnie. Toute configuration stable qui donne plus au monopole que son profit de réservation $\bar{\pi} = \text{Max}_{\alpha} \Pi(\alpha, M(\alpha), K)$ est alors un équilibre du jeu. Définissons $\Pi(\alpha_0, \bar{N}, K) = \bar{\pi}$, on a bien sûr $\alpha_0 > m$ et pour toute prime acceptable $\alpha^* \in [\alpha_0, \alpha^m]$ et $M^* = \bar{N}$, on a $\Pi(\alpha^*, M^*, K) \geq \bar{\pi}$.

Si le monopole n'assure pas tout le marché, alors la seule possibilité est $\tilde{\alpha} \geq \alpha^* \geq \hat{\alpha}$ et $M^* = M(\alpha^*)$. Dans ce cas l'espérance de profit du monopole est $\bar{\pi}$ et $\alpha^* = \text{ArgMax}_{\alpha} \Pi(\alpha, M(\alpha), K)$. Notons toutefois que $\Pi(\alpha, K, M(\alpha))$ peut atteindre son maximum en un point où $M(\alpha) = \bar{N}$. Dans ce cas, l'équilibre de type 2 n'existe pas. L'existence d'un équilibre de type 2 dépend des valeurs des paramètres du modèle. \square

Proposition 7 *Le jeu admet un grand nombre d'équilibres quand la mutuelle existe. Soit M^* le nombre d'assurés de la compagnie. La mutuelle assure tous les autres agents $\bar{N} - M^*$. Il existe trois types d'équilibres:*

- i) seul le monopole est actif à l'équilibre $M^* = \bar{N}$ et il propose une prime $\alpha^* \in [m, \alpha^m]$;*
- ii) seule la mutuelle est active à l'équilibre, $M^* = 0$ et $\alpha^* \geq m$,*
- iii) les deux sociétés sont actives à l'équilibre $M^* \in]0, \bar{N}[$, et la prime vérifie*

$$V(\alpha^*, M^*, K) \geq V_M(N - M^* + 1)$$

et

$$V_M(N - M^*) \geq V(\alpha^*, M^* + 1, K), \Pi(\alpha^*, M^*, K) \geq K$$

Le graphique (2) représente ces trois types d'équilibres.

Preuve.

Si la mutuelle existe, et si $V(m, 1, K) \leq V_M(\bar{N})$, dès lors que le monopole propose une prime $\alpha \geq m$, la configuration dans laquelle tous les agents s'adressent à la mutuelle est stable, même si elle n'est pas la meilleure pour les agents. Toute déviation pour une prime supérieure à la prime actuarielle engendre donc un profit égal à K pour la compagnie. Si par contre elle propose une prime $\alpha < m$, la configuration dans laquelle tous les agents s'adressent à elle est stable, mais induit une espérance de profit strictement inférieure à K . Elle n'est donc pas souhaitable. Par conséquent on a trouvé des configurations qui rendent les déviations non souhaitables dès lors que la compagnie obtient à l'équilibre une espérance de profit plus grande que K , qui est son profit de réservation (profit qu'il obtient sans client). Par conséquent toutes les primes et configurations stables associées qui donnent à la compagnie une espérance de profit plus grande que K sont des équilibres de ce jeu.

Ces équilibres peuvent être classés en trois familles. \square

2.3 L'oligopole

L'analyse quand il existe plusieurs entreprises repose sur les mêmes mécanismes.

Proposition 8 *Si les agents se coordonnent spontanément sur l'équilibre qui leur est le plus favorable, le jeu avec n compagnies admet un seul équilibre. Au moins deux compagnies proposent la même prime $\alpha^* = m$, et tous les agents s'assurent auprès d'une seule d'entre elles. Les compagnies concurrentes offrent une prime plus grande. L'introduction d'une mutuelle n'affecte pas l'équilibre.*

Preuve.

Supposons dans un premier temps qu'il n'existe pas de mutuelle.

Sous l'hypothèse boule de neige les agents s'adressent spontanément et tous ensemble à l'entreprise qui a offert la prime la plus faible, à condition toutefois que la prime d'assurance leur assure une utilité supérieure à leur utilité sans assurance $V_M(1)$. Par conséquent une configuration stable $M_i(\alpha_1, \dots, \alpha_n)$ vérifie

si $\alpha_i > \alpha^m$, $M_i(\alpha_1, \dots, \alpha_n) = 0$

si $\alpha_i \leq \alpha^m$, $\alpha_i < \text{Min}(\alpha_j, j \neq i)$ alors $M_i(\alpha_1, \dots, \alpha_n) = \bar{N}$

si $\alpha_i \leq \alpha^m$, $\alpha_i > \text{Min}(\alpha_j, j \neq i)$ alors $M_i(\alpha_1, \dots, \alpha_n) = 0$

si $\alpha_i \leq \alpha^m$, $\alpha_i = \alpha_k \leq \text{Min}(\alpha_j, j \neq i)$. Soit S_i l'ensemble des indices des compagnies ayant offert la même prime que i , $M_{\text{Inf } S_i}(\alpha_1, \dots, \alpha_n) = \bar{N}$

Supposons qu'à l'équilibre la compagnie j assure tous les clients $M_j^* = \bar{N}$. α_j^* est alors la plus petite prime, et par l'application de la loi des grands nombres, $\alpha_j^* \geq m$. Remarquons qu'à l'équilibre toutes les autres compagnies concurrentes obtiennent une espérance de profit égale à K .

Supposons que $\alpha_j^* > m$. Si une compagnie concurrente dévie et offre une prime $m < \alpha < \alpha_j^*$, elle attire alors tous les clients et obtient un profit $\Pi(\alpha, \bar{N}, K) > K$. Par conséquent une telle déviation est souhaitable et à l'équilibre on a $\alpha_j^* = m$. Si toutes les offres des compagnies concurrentes sont strictement plus grandes, alors la compagnie j a intérêt à augmenter sa prime d'une toute petite quantité. Par conséquent deux compagnies au moins à l'équilibre proposent la prime actuarielle et seule l'une d'entre elles assure tout le marché.

S'il existe une mutuelle, le raisonnement est inchangé, il suffit de remplacer α^m par α_M^m dans la définition de la configuration stable. On en déduit que la mutuelle ne joue aucun rôle dans ce processus concurrentiel sauvage. Sa présence n'affecte pas l'équilibre.

□

L'utilité obtenue à l'équilibre est supérieure au maximum d'utilité que peut engendrer un système d'assurance par le biais de la mutuelle. L'équilibre est efficace au sens de Pareto. Mais il détruit toute perspective de profit et toutes les entreprises sauf une jouent le rôle de fantômes.

Dans le cas où les agents n'adoptent spontanément un comportement qui conduit à se coordonner en seconde période sur l'équilibre de type boule de neige, il est plus difficile de déterminer l'ensemble des équilibres du jeu. Toutefois, la proposition suivante caractérise certains de ces équilibres.

Proposition 9 *S'il n'existe pas de mutuelle concurrençant l'oligopole, soit $V(\alpha_2, \bar{N}, K) = V(m, 1, K)$. Toute prime $\alpha^* \in [m, \alpha_2]$ et toute configuration stable associée est un équilibre de Nash du jeu étudié.*

Preuve.

Imaginons que le jeu admette un équilibre symétrique et que les compagnies proposent la prime α^* . Pour toute prime α telle que $V(\alpha, 1, K) \leq V(\alpha^*, \bar{N}, K)$, il existe une configuration stable qui donne au déviant un nombre de clients nul. Autrement dit augmenter la prime n'attire aucun client et diminuer la prime n'est rentable que si la différence est suffisamment importante. Dans un certain sens, les clients sont inertes pour répondre à une déviation qui pourrait être pour eux avantageuse.

Il existe au moins un équilibre, qui est tel que $V(\alpha^*, \bar{N}, K) \geq V(m, 1, K)$, suivie d'une configuration stable. Dans ce cas, une compagnie qui augmente la prime n'attire aucun client. Si elle souhaite diminuer la prime, il est nécessaire qu'elle offre une prime qui vérifie $V(\alpha, 1, K) > V(\alpha^*, \bar{N}, K) \geq V(m, 1, K)$, et donc $\alpha < m$. Dans ce cas, il existe une configuration stable qui donne tous les clients à l'entreprise déviante et qui rend de ce fait sa déviation non profitable. Notons qu'il existe sûrement d'autres équilibres. \square

Proposition 10 *S'il existe une mutuelle concurrençant l'oligopole, une proposition $(\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)$ et une configuration stable associée $(M_1^*, M_2^*, \dots, M_n^*)$ est un équilibre du jeu si et seulement si $\Pi(\alpha_i^*, M_i^*, K) \geq K$, $i = 1, \dots, n$, c'est à dire que chaque l'entreprise reçoit un profit supérieur ou égal à K .*

Preuve.

Si $V(m, 1, K) \leq V_M(\bar{N})$ et s'il existe une mutuelle, quelle que soient les offres des compagnies, tant que les primes offertes sont supérieures à la prime actuarielle, il existe une configuration stable telle que tous les agents s'assurent auprès de la mutuelle, qui engendre donc pour les compagnies d'assurance une espérance de profit égale à K . Si l'une des compagnies propose une prime inférieure à m , il existe une configuration stable telle que tous les agents s'adressent à la compagnie qui a offert la prime la plus faible. Celle-ci obtient alors une espérance de profit proche de $K + \bar{N}(\alpha - m) < K$.

Remarquons tout d'abord que $\Pi(\alpha_i^*, M_i^*, K) \geq K$. En effet, dans le cas inverse, la compagnie préfère n'assurer personne et elle proposera une prime par exemple inacceptable.

Si une compagnie d'assurance dévie et propose une prime $\alpha \geq m$, on sait qu'il existe une configuration stable qui ne lui donne aucun client et qui donc engendre un profit K . Si elle offre une prime $\alpha < m$, deux cas sont possibles. Soit cette prime est plus faible que les primes offertes par les compagnies concurrentes et alors il existe une configuration stable qui lui donne tous les clients mais qui engendre une espérance de profit strictement inférieure à K , soit cette prime n'est pas la plus faible et dans ce cas il existe une configuration stable qui ne lui donne aucun client.

Par conséquent, à l'équilibre, le profit de chaque entreprise doit être supérieur ou égal à K , et cette condition est suffisante pour assurer qu'aucune déviation n'est souhaitable. \square

Autrement dit sont des équilibres du jeu toutes les propositions des compagnies qui leur assurent une espérance de profit plus grande que K et qui forment des configurations stables. En particulier sont des équilibres les situations suivantes

- la mutuelle seule active est un équilibre;
- tous les équilibres du jeu de monopole sont des équilibres du jeu d'oligopole;
- les équilibres de l'oligopole sans mutuelle sont possibles;
- une situation dans laquelle une grosse compagnie facture une prime élevée et des petites compagnies facturent des primes plus faibles peut aussi être un équilibre;
- enfin des structures mixtes dans lesquelles mutuelle et compagnies co-existent, dans ce cas les compagnies restent petites (le nombre de clients qu'elles assurent doit vérifier $V(\alpha, N, K)$ est une fonction décroissante de N).

Notons que la mutuelle non seulement perd son pouvoir concurrentiel, mais, à cause de l'effet boule de neige annule la concurrence que se livrent les compagnies entre elles. La grande diversité des structures d'équilibre s'accompagne d'une conclusion surprenante: certains de ces équilibres peuvent être efficaces au sens de Pareto, et l'utilité des agents peut être très grande (par exemple est un équilibre une situation dans laquelle une seule compagnie assure tout le marché en offrant la prime actuarielle), mais il est aussi possible que le "pire" pour les agents apparaisse.

3 Equilibre en présence d'une réglementation prudentielle

Les professionnels de l'assurance (assureurs et autorité de contrôle) s'accordent sur l'idée selon laquelle les agents sous-estiment la probabilité de faillite des sociétés d'assurance. Dans cette section nous adoptons cette hypothèse, et supposons que les agents n'envisagent pas la possibilité d'insolvabilité de leur société. Si la compagnie propose une prime α , l'utilité des agents perçue est $U(w - \alpha)$ et l'espérance de profit de la compagnie d'assurance est $\Pi(\alpha, N, K)$. Dans ce cas les assurés sont insensibles à l'effet mutualisation des risques, cette hypothèse est incompatible avec l'existence des mutuelles, celles-ci faisant faillite dès qu'un agent a un accident.

L'équilibre sur ce marché avec plusieurs compagnies conduit au résultat de Bertrand dès lors qu'au moins deux compagnies sont présentes sur ce marché, à l'équilibre les compagnies présentes proposent une prime $\alpha = m$.

Enfin notons que si le stock de capital est une variable stratégique, parce que $\frac{\Pi(\alpha, K, N)}{K}$ est une fonction décroissante de K (voir annexe), le stock de capital d'équilibre est le plus petit possible. Par conséquent si les agents sont myopes, la concurrence augmente très fortement la probabilité de faillite (prime la plus faible possible, capital le plus petit possible).

Pour protéger les consommateurs d'une telle sous-capitalisation, les autorités européennes ont mis en place une réglementation prudentielle, qui exige des compagnies le respect de certaines règles financières. La première de ces règles impose que le stock de chaque compagnie soit suffisamment grand, $K \geq \bar{K}$. La seconde de ces règles introduit un lien entre le capital et les engagements des compagnies, et impose que $K \geq N \text{Max}(s_1 \alpha, s_2 m)$, avec $s_1 = 0,16$ et $s_2 = 0,23$.

La prise en compte de la réglementation prudentielle impose une modification du déroulement du jeu étudié. En effet, il est possible que les exigences réglementaires conduisent à un rationnement des agents. Un agent qui s'adresse à une compagnie peut donc se voir refuser l'accès à l'assurance, perspective que nous n'avions pas envisagé précédemment. Pour compléter le jeu, nous supposons qu'un agent qu'une société refuse d'assurer peut s'adresser à une autre entreprise, et ceci autant de fois qu'il est nécessaire. Enfin nous n'étudions dans ce jeu que les équilibres symétriques en stratégies pures, et posons deux hypothèses simplificatrices: un agent indifférent entre s'assurer ou pas s'assure, et si les agents ont indifférents entre plusieurs compagnies, ils se répartissent de façon équitable.

Supposons dans un premier temps que le stock de capital des compagnies K est donné. La réglementation prudentielle crée un lien entre taille de la compagnie et prime. Notons

$$G(\alpha) = \frac{K}{\text{Max}(s_1\alpha, s_2m)} \text{ et définissons } \hat{\alpha} = \text{Arg max}_{\alpha \leq m+\rho} \Pi(\alpha, G(\alpha), K).$$

Proposition 11 *Le jeu avec stock de capital donné admet deux équilibres symétriques. Le premier de ces équilibres existe si $\bar{N} \leq (n-1)G(m)$, les compagnies offrent la prime actuarielle $\alpha^* = m$ et assurent $M^* = \frac{\bar{N}}{n}$ agents. Le second de ces équilibres existe si $nG(\hat{\alpha}) \leq \bar{N}$, les compagnies offrent la prime $\alpha^* = \hat{\alpha}$ et assurent $M^* = M(\hat{\alpha})$ agents.*

Preuve.

Une prime $\alpha^* \leq m + \rho$ et un nombre d'assurés M^* est un équilibre du jeu étudié si $M^* \leq G(\alpha^*)$ et $nM^* \leq \bar{N}$, l'une des inégalités étant stricte.

$\Pi(\alpha^*, M^*, K) \geq \Pi(\alpha, M(\alpha), K)$ pour toute prime $\alpha \leq m + \rho$ où $M(\alpha)$ est donné par: si $\alpha < \alpha^*$, tous les assurés préfèrent la compagnie déviante, qui peut assurer $M(\alpha) = \text{Min}(G(\alpha), \bar{N})$;

si $m + \rho \geq \alpha > \alpha^*$, aucun des assurés ne préfère la compagnie déviante. Tous les assurés s'adressent aux compagnies non déviantes et il reste $\text{Max}(\bar{N} - (n-1)G(\alpha^*), 0)$ agents non assurés, donc $M(\alpha) = \text{Min}(G(\alpha), \text{Max}(\bar{N} - (n-1)G(\alpha^*), 0))$.

Etape 1 - si $nG(\hat{\alpha}) \leq \bar{N}$, le jeu admet un équilibre symétrique $\alpha^* = \hat{\alpha}$ et $M^* = M(\hat{\alpha})$.

En effet, le profit d'une compagnie qui propose une prime $\alpha \geq m$ est majorée par $\Pi(\alpha, G(\alpha), K)$. Par conséquent aucune déviation n'est souhaitable si le profit obtenu est le maximum de $\Pi(\alpha, G(\alpha), K)$.

Etape 2 - Si $\alpha^* > m$, alors $\alpha^* = \hat{\alpha}$ et $M^* = M(\hat{\alpha})$.

Si $\alpha^* > m$, une compagnie d'assurance peut offrir une prime plus faible $\alpha < \alpha^*$ et obtenir un profit $\Pi(\alpha, G(\alpha), K)$, et les conditions d'équilibre imposent alors que $\Pi(\alpha^*, M^*, K) \geq \Pi(\alpha, G(\alpha), K)$ pour toute prime $\alpha < \alpha^*$. Cette inégalité impose que $M^* = G(\alpha^*)$.

Mais alors, si la compagnie dévie et propose une prime $m + \rho > \alpha > \alpha^*$, le nombre d'assurés qui s'adresse à elle est $\bar{N} - (n-1)G(\alpha^*) \geq G(\alpha^*) > G(\alpha)$ car la fonction G est une fonction décroissante de α . Par conséquent on doit avoir $\Pi(\alpha^*, G(\alpha^*), K) \geq \Pi(\alpha, G(\alpha), K) \forall \alpha$ et donc $\alpha = \hat{\alpha}$.

Etape 3 - Supposons que $\alpha^* = m$. Une compagnie n'a pas intérêt à dévier et à offrir une prime plus faible. Si elle propose une prime plus grande le nombre des assurés qui s'adresse à elle est $\bar{N} - (n-1)G(m)$.

Si la loi des grands nombres s'applique à $\frac{\bar{N}}{n}$, le profit d'équilibre est égal à K et le profit du à une déviation est strictement plus grand que K dès lors que $\bar{N} - (n-1)G(m) > 0$. On en déduit que cet équilibre existe si et seulement si $\bar{N} \leq (n-1)G(m)$. \square

Supposons dans un second temps que les compagnies choisissent en même temps que leur prime le niveau de stock de capital. Dans ce cas, leur critère de choix est le taux de rentabilité $\pi(\alpha, N, K) = \frac{\Pi(\alpha, N, K)}{K}$.

Proposition 12 *Si les compagnies peuvent choisir leur stock de capital, à l'équilibre elles choisissent $K^* = \bar{K}$. Les équilibres sont alors identiques à ceux de la proposition précédente.*

Preuve.

La fonction $\pi(\alpha, N, K)$ est une fonction décroissante du stock de capital et croissante de N si $\alpha \geq m$.

$$\pi(\alpha, N, K) \leq \pi(\alpha, N, N \text{Max}(s_1\alpha, s_2m)) = \frac{\Pi(\alpha, N, N \text{Max}(s_1\alpha, s_2m))}{N \text{Max}(s_1\alpha, s_2m)} = v(N, \alpha).$$

Or, $v(N, \alpha)$ est une fonction décroissante de N et croissante de α . Par conséquent la compagnie, tant qu'elle ne bute pas sur la contrainte $K \geq \bar{K}$, préfère toujours augmenter la prime et diminuer le nombre de ses clients (si celui-ci reste positif).

Un équilibre symétrique du jeu est décrit par les trois paramètres α^*, M^*, K^* qui vérifie

$$\alpha^* \leq m + \rho, M^* = \text{Min}\left(\frac{\bar{N}}{n}, \frac{K^*}{\text{Max}(s_1\alpha^*, s_2m)}\right), K^* \geq \bar{K}$$

$$\pi(\alpha^*, N^*, K^*) \geq \pi(\alpha, M(\alpha), K), \forall \alpha \leq m + \rho, \forall K \geq \bar{K}$$

avec

$$\begin{aligned} &\text{si } \alpha = \alpha^*, \text{ si } nM^* = \bar{N}, M(\alpha) = \text{Min}\left(\frac{\bar{N}}{n}, \frac{K}{\text{Max}(s_1\alpha^*, s_2m)}\right) \text{ sinon } M(\alpha) = \min\left\{\bar{N} - \right. \\ &(n-1)\frac{K^*}{\text{Max}(s_1\alpha^*, s_2m)}, \left. \frac{K}{\text{Max}(s_1\alpha^*, s_2m)}\right\} \\ &\text{si } \alpha < \alpha^*, K = M(\alpha)\text{Max}(s_1\alpha, s_2m) \\ &\text{si } m+\rho \geq \alpha > \alpha^*, M(\alpha) = \min\left\{\text{Max}\left(0, \bar{N} - (n-1)\frac{K^*}{\text{Max}(s_1\alpha^*, s_2m)}\right), \frac{K}{\text{Max}(s_1\alpha, s_2m)}\right\}. \end{aligned}$$

Quelle que soit son offre, la compagnie préfère toujours choisir $K = \bar{K}$, et par conséquent les conditions d'équilibres des deux jeux coïncident \square

La réglementation prudentielle modifie considérablement l'équilibre du jeu, quand les agents sont myopes. Si le capital n'est pas une variable stratégique, la réglementation prudentielle atteint son but, diminuer la probabilité de faillite, mais par l'augmentation des primes. Cette augmentation des primes est très néfaste pour l'assuré. Si le capital est une variable stratégique au même titre que les autres (la prime), alors l'effet est pire encore, la capitalisation des entreprises d'assurance sera la plus faible possible, seule la prime, elle, reste élevée.

Notons toutefois que si la réglementation prudentielle a des effets très anti-concurrentiels, elle encourage l'entrée de petites compagnies d'assurance. Cette entrée n'implique pas une

diminution de prime. De plus, même si le capital par tête augmente sur le marché, parcel-
liser les sociétés d'assurance n'est pas souhaitable, comme nous le montrent les résultats
de la section précédente. Avoir beaucoup de petites compagnies, c'est renoncer aux effets
positifs de la mutualisation.

ANNEXE 6

Allocating and Funding Universal Service Obligations in a Competitive Network Market

n° 9955

**Allocating and Funding Universal
Service Obligations in a Competitive
Network Market**

**P. CHONÉ¹
L. FLOCHEL²
A. PERROT³**

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

¹ INSEE and CREST, Laboratoire d'Economie Industrielle.

² EPEE-Université d'Evry-Val d'Essonne and EUREQua-Université Paris 1.

³ CREST-Laboratoire d'Economie Industrielle and EUREQua-Université Paris 1.

Abstract

We examine, in a network market open to competition, various mechanisms of allocation and funding for "universal service obligations". The obligations we consider are geographic "ubiquity" and non discrimination. We analyze, both from an efficiency and an equity point of views the respective advantages of a restricted entry system, where the entrant is not allowed to serve high costs consumers and the "pay or play" system at work for instance in Australia. We show that the pay or play regulation always dominates the restricted entry regulation when the ubiquity constraint is imposed alone. This result no longer holds when the regulator imposes also the non discrimination constraint.

Résumé

Nous examinons, dans le cadre d'un secteur en réseau ouvert à la concurrence, divers modes d'allocation et de financement des obligations de service universel. Les obligations de service universel que nous considérons sont d'une part l'ubiquité géographique et d'autre part la non discrimination. On s'intéresse aux avantages respectifs de deux systèmes en particulier, le système de l'"entrée restreinte" et la régulation de type "pay or play" en vigueur par exemple en Australie.

Nous montrons que cette dernière domine toujours l'entrée restreinte lorsque seule est prise en compte la contrainte d'ubiquité. En revanche, lorsque la contrainte de discrimination est imposée également, ce résultat ne tient plus.

1 Introduction

The transition towards a more competitive regime on the markets of public utilities rises a number of new questions. In network utilities like telecommunications, electricity, gaz, the regulator often values an "equal access" of all consumers to the service at an "affordable tariff"¹. Whereas networks were previously operated by monopolies who were in charge with these universal service obligations (USOs), and used cross-subsidies between "profitable" and "unprofitable" users, the arrival of new entrants on markets that are now open to competition induces cream skimming on profitable segments of the market (see Laffont and Tirole (1998)), and makes previous monopolies unable to finance these obligations through cross-subsidies². Moreover, competition leads to outcomes that are not necessarily desirable from the regulator's point of view. Without regulatory constraints, some users would then be excluded of the market, and users with different consumption or costs characteristics would face different tariffs. If the regulator values equality with respect to tariffs and/or access of all the users to the market, he must then impose "universal service obligations" (USOs). In practice, those obligations include two types of components. The "social" component includes the fact that light users should be offered tariffs similar to heavy users, or low tariffs for targeted consumers (see Gasmi, Laffont, Sharkey (1999)). The "geographical" component can be divided into two obligations: ubiquity and non discrimination. The "ubiquity constraint" states that all consumers should be connected to a network, whatever their location. The "non discrimination" constraint states that the same tariff should be proposed to all those consumers, whatever their location or their connection cost. These constraints may be imposed together or independently³. Those constraints are particularly relevant in developed countries for promoting access of a large number of consumers to new technologies, like Internet. In developing countries, and particularly in case of privatization, those constraints are crucial to ensure a large development of networks.

The universal service obligation raise two series of questions. First, which USO's should be imposed to whom (the allocation problem) and second, who should pay for the USO's (the funding problem). The combination of various solutions to the first and to the second question define different regulatory mechanisms that have different implications in terms of global and/or partial surplus. Compared to an unconstrained competition between network providers, universal service obligations induce distortions on the competitive entry process and on the equilibrium market structure. Those distortion generate both social benefits and social costs. The question of how to share these costs and benefits becomes one of the

¹See e.g., Federal Communications Commission, "In the matter of joint Federal-Sate Joint Board on Universal Service", CC Docket, n 96-45, Nov. 8, 1996.

²For a complete theory of competition in the sectors of public utilities, see Laffont and Tirole (1993).

³For instance in France, the provision of gas is not submitted to the ubiquity constraint: the operator Gaz De France may choose not to serve a given area. But when it serves it, it is submitted to a non discrimination constraint, that is, it must offer the same tariff (or the same menu) to any consumer of the areas that are served.

main questions for regulators and has received various answers in different countries.

The main tasks for regulators consist in determining optimal rules for allocating and funding those USOs. On the allocation side, two alternative approaches exist, depending on who between the incumbent and its competitors incur USOs. In most countries, following openness to competition, only the incumbent is in charge of USOs. In certain cases, productive efficiency would require that competitors incur some of the USOs. For instance, USOs can be auctionned (see Milgrom (1996), Weller (1998))⁴. On the funding side, three methods exist. Most systems share the property that they have to be self funded. The cost of the USOs is then financed through taxes, levied on all units of goods. An alternative method consists in financing USOs through lump sum transfers. Finally, probably the less distorting scheme consists in funding USOs through cross-subsidies.

In this paper, we focus mainly on two allocation mechanisms and funding systems. The "restricted-entry" rule gives to the incumbent the obligation of serving non profitable users at the same tariff as profitable ones, whereas the pay or play regulation⁵ allows the entrant to serve the market. On the funding side, both rules allow for cross-subsidies and taxation. We compare them from a global and partial surplus point of view. In the last section of the paper, we briefly discuss other possibilities, like auctions on the allocations side and lump sum transfers on the funding side.

We examine various procedures of tax determination. Among others, we analyze the case where the tax is designed to be "competitively neutral", that is, the incumbent must obtain through taxes, if it is possible, the same amount of profits than under a competitive regime. This view corresponds to that developed in the US in the Telecommunication Act of 1996⁶. However, it differs greatly from a determination of taxes through a first best principle, and thus induces social losses. An important remark is thus that the "cost of universal service" is conditionnal to the way it is funded. As pointed out by Panzar (1999), *'any USO costing exercise must begin with a careful specification of an unsubsidized market scenario that would prevail in the absence of the USO'*. Many proposed measures of USO costs (in particular the accountable approach of *'net avoidable costs'*⁷) ignore this basic requirement. In this paper, we define precisely the benchmark situation, where no USO is imposed, and then we carefully list and evaluate the economic distortions due to the presence of the tax. In our framework, the market structure as well as the cost of the USO's are endogeneous, since they depend on the regulation rule and the tax determination procedure.

We build a model where two operators compete for a market of a network good. Consumers are heterogenous with respect to their connection costs. The regulator imposes different combinations of USOs, and funds them through taxes. The incumbent firm is

⁴For instance, USOs are auctionned for the postal service in Germany or for telecommunications in the US (and more recently in Germany and Austria).

⁵The pay or play regulation is applied in Australia and has been discussed in UK in the telecommunication sector.

⁶See Pub. L. No. 104-104, 110 Stat. 56 (1996) (1996 Act), 47 U.S.C. § 254(b).

⁷See for instance WIK (1997).

in charge with the USOs. We then compare two regulatory mechanisms. In the "pay or play" system, the entrant may choose to serve the non profitable users, whereas in the "restricted-entry" system, he is not allowed to. We examine the equilibria of the corresponding games, and compare them in terms of social welfare and in terms of transfers between agents. In a last section, we compare the funding of USOs through taxation with the case lump sum transfer scheme. Finally, we compare these two regulatory mechanisms with a second price auction on the market of non profitable users.

2 Model and notations

2.1 The framework

We consider the following situation. Two firms, denoted I and E , compete for the market of a network good. This network has a character of public utility, that is, the public authority may pursue social objectives such as "equal access" of all the consumers to the service provided by the network. Typically, this situation reflects liberalization of a network market, such as electricity, gaz, telecommunications or airlines, where an incumbent firm (I) faces a entrant (E).

There are two types of users denoted by $\mu = \underline{\mu}, \bar{\mu}$. The type of a user corresponds to its location across the geographical space: the value $\underline{\mu} = \underline{\mu}$ (resp. $\bar{\mu} = \bar{\mu}$) denotes a location where the cost of connection to the network is low (resp. high). For instance $\underline{\mu}$ is a measure of the cost of connection in high density areas (urban areas), whereas $\bar{\mu}$ corresponds to low density areas (rural areas). The proportion of consumers of type $\bar{\mu}$ is $\bar{\alpha}$ and that of consumers $\underline{\mu}$ is $\underline{\alpha}$.

A consumer who buys q units of the good and who is charged a tariff $T(q)$ ⁸ receives a net surplus equal to: $u = w(q) - T(q)$, where $w(q)$ is supposed to be increasing and concave. Here, the only element of heterogeneity concerns the connection costs. The demand addressed to firm K by a consumer facing the tariff $T(q)$ is given by $w'(q) = T'(q)$. One should note that the demand of a particular consumer is independant of its connection cost.

Firms are endowed with the following technologies. Firm $K = I, E$ incurs a cost $C_K(q, \mu)$, when providing q units to consumer $\mu = \underline{\mu}, \bar{\mu}$. We assume that

$$C_K(q, \underline{\mu}) < C_K(q, \bar{\mu})$$

for all q and for both firms $K = I, E$. If the consumers $\underline{\mu}$ and $\bar{\mu}$ are respectively urban and rural users, this assumption expresses the fact that connecting rural users to the network is more costly to any firm than connecting urban users, whatever the level of their consumption⁹. The profit of firm K is then $\pi_K(q, \mu) = T_K(q, \mu) - C_K(q, \mu)$.

⁸Under the assumption of homogeneous preferences, a two-part tariff allows to achieve the same allocation than more general non linear tariffs. In a more general model, consumers could be heterogenous in their tastes for the good.

⁹In order to focus on questions related to universal service obligations, we do not consider any problem of

Remark 1 It is equivalent to work with tariff variables $T(\mu, q)$ or with variables u and q , where u denotes the utility level offered to a consumer. We deal thereafter with (u, q) . We denote by \underline{u}_K (respectively \bar{u}_K) the utility offered by firm K to consumers of type μ (respectively $\bar{\mu}$).

Remark 2 The surplus derived from the relationship between a consumer of type μ and firm K is $S_K(\mu, q) = w(q) - T_K(q, \mu) + \pi_K(\mu, q) = w(q) - C_K(q, \mu)$.

The first best quantity is defined by $q_K^{FB} = \operatorname{argmax}_{q \geq 0} S_K(\mu, q)$. The first best surplus is $S_K(\mu) = w(q^{FB}) - C_K(q^{FB}, \mu)$ if $q_K^{FB} > 0$. This value $S_K(\mu)$ is the value to be shared between the firm and the consumer, the share of the consumer being u_K .

We shall maintain the following assumption throughout¹⁰:

$$\max(\bar{S}_I, \bar{S}_E) < 0 < \min(\underline{S}_I, \underline{S}_E), \quad (1)$$

which means that high cost consumers are not profitable and would not be served by any firm without universal service obligations.

2.2 Benchmark: The competitive case

As a benchmark, we study the case where no firm is submitted to any USO constraint, that is, each firm K is able to offer a perfectly discriminatory tariff to any consumer. Because of our assumption (1), firms then compete on user $\underline{\mu}$ only. The profit of firm E is given by:

$$\pi_E(\underline{u}_E, q_E) = \begin{cases} \underline{\alpha}(\underline{S}_E(q) - \underline{u}_E) & \text{if } \underline{u}_E > \underline{u}_I \\ 0 & \text{if } \underline{u}_E < \underline{u}_I. \end{cases}$$

The profit of firm I has a symmetric expression. It is clear that we have $q_K = q_K^{FB}$ when the consumer is served by firm K . Therefore in what follows, we will call "strategies" the choices of the variable \underline{u}_K by firm $K = I, E$.

Each firm reacts to the strategy of its rival by choosing the share of the surplus it leaves to the consumer. We can then compute the best reply function of firm E in response to the strategy \underline{u}_I of firm I .

Facing \underline{u}_I , firm E can offer $\underline{u}_I + \varepsilon$ if $\pi_E = \underline{S}_E - \underline{u}_I - \varepsilon$ remains positive, that is, as long as $\underline{u}_I < \underline{S}_E$. If $\underline{u}_I > \underline{S}_E$, firm E has no incentive to be active. Thus the best reply function of E to \underline{u}_I is given by:

$$\underline{u}_E(\underline{u}_I) = \begin{cases} \underline{u}_I & \text{if } \underline{u}_I < \underline{S}_E \\ \text{any value in } [0, \underline{u}_I] & \text{if } \underline{u}_I > \underline{S}_E. \end{cases}$$

interconnection between both networks, which amounts to assume that the interconnection charge between networks is zero. For more details on the interconnection problems, see Armstrong, Doyle and Vickers (1996).

¹⁰except in section A.6, where we relax this assumption, in order to study the impact of imposing a non discrimination constraint alone.

In what follows, we will use as a benchmark the sequential game in which firm I is the leader. This representation of the competition process suits better to the case where a (dominant) incumbent faces a new entrant. Finally, as we will see in the next section, the allocation of universal service obligations gives to the incumbent a first mover advantage.

The sequential game (where firm I acts as a leader) may lead to a continuum of subgame perfect equilibria (SPE). When $0 < \underline{S}_I < \underline{S}_E$, any value $u_I \in [0, \underline{S}_E]$ leads to the same profit for the firm I , namely $\Pi_I = 0$ since E serves the consumer).

Lemma 1 *The subgame perfect equilibria of the game where firm I acts as a leader are defined as follows:*

- if $0 < \underline{S}_E < \underline{S}_I$, there is a unique SPE: firm I serves the consumer, offers him $\underline{u}_I = \underline{S}_E$ and $\underline{q}_I = q_I^{\text{FB}}$, and makes profit $\pi_I = \underline{S}_I - \underline{S}_E$;
- if $0 < \underline{S}_I < \underline{S}_E$, there is a continuum of SPE, parametrized by $\underline{u}_E \in [0, \underline{S}_E]$: firm E serves the consumer, who obtains any value $\underline{u}_E \in [0, \underline{S}_E]$. Firm E gets $\Pi_E = \underline{S}_E - \underline{u}_E$ and firm I stays out of the market.

2.3 Taxes and distortions

In what follows, we will often consider the case where firm E pays a tax t on each unit sold to consumer $\underline{\mu}$. We now state a few useful results concerning the surplus of the relationship between firm E and the consumer $\underline{\mu}$, in presence of the tax t . In the sequel, the presence of a tilde indicates the dependance of the variable with regard to the tax rate t .

Let $T(q)$ be the tariff posted by the firm and q be the quantity of good purchased by the consumer $\underline{\mu}$. The profit of the firm writes:

$$\tilde{\pi}_E = T(q) - C_E(q, \underline{\mu}) - tq,$$

where t is the tax level. We denote \tilde{S}_E the surplus from the relationship between E and $\underline{\mu}$

$$\tilde{S}_E(t, q) = \tilde{\pi}_E + w(q) - T(q) = w(q) - C_E(q, \underline{\mu}) - tq.$$

Let $\tilde{S}_E(t)$ be the maximum value over $q \geq 0$ of the surplus:

$$\tilde{S}_E(t) = \max_{q \geq 0} \{w(q) - C_E(q, \underline{\mu}) - tq\}.$$

Let \tilde{q}_E be the value of q for which the maximum is attained. The following lemma is proved in the appendix (section A.1).

Lemma 2 *The following properties are true:*

1. *The total surplus function \tilde{S}_E is convex;*
2. *The functions \tilde{S}_E and $\tilde{S}_E + t\tilde{q}_E$ are nonincreasing with respect to t ;*

3. There exists a tax level t^e such that for $t \geq t^e$, firm E can extract no surplus from consumers $\underline{\mu}$: $\tilde{S}_E(t^e) = 0$.

The sum $\tilde{S}_E(t) + t\tilde{q}_E(t)$ represents the total amount that can be extracted from the relation between E and the consumer $\underline{\mu}$, that is, the whole amount of what the firm E and the consumer $\underline{\mu}$ receive on the one hand, $\tilde{S}_E(t)$, and the amount of tax collected on the other hand $t\tilde{q}_E$. The higher the tax, the lower this total amount. This partial effect should lead the regulator to choose not too high taxes.

Lemma 2 shows that the entrant is only active on the market when the tax level is not too high ($t < t^e$). Very high level of taxes eject firm E from the competition with firm I .

2.4 Definitions of and interactions between USOs

The Universal Service Obligations consist in two different constraints imposed by the regulator to the incumbent I :

- Ubiquity constraint (subscript U will identify thereafter the variables under this constraint): Each consumer should have access to the good, that is, each consumer must face a tariff such that he obtains a nonnegative level of utility;
- Non discrimination constraint (ND): If the incumbent chooses to serve both types of consumers, it must offer the same tariff to both. The incumbent may prefer to serve only one type;
- Ubiquity and non discrimination constraints (UND): the incumbent must offer to all consumers the same tariff and this tariff must be such that all the consumers have a nonnegative utility level.

Recall that we model the competition between firms I and E by a sequential game, where firm I is the leader: whatever the regulation framework, firm I first announces a pair $(\underline{u}_I, \bar{u}_I)$. The USO's may thus be represented by constraints on the space of strategies of the incumbent:

$$\begin{aligned} \text{USO=U} : & \quad \underline{u}_I \geq 0, \bar{u}_I \geq 0 \\ \text{USO=ND} : & \quad \text{if } \underline{u}_I \geq 0, \bar{u}_I \geq 0 \text{ then } \underline{u}_I = \bar{u}_I \\ \text{USO=UND} : & \quad \underline{u}_I = \bar{u}_I \geq 0. \end{aligned}$$

Note, however, that these restrictions on I 's strategies imply (at equilibrium) restrictions on E 's strategies. Suppose for example that under USO=UND, firm E serves all the consumers. Then, it is clear that the consumers $\underline{\mu}$ and $\bar{\mu}$ get the same level of utility at equilibrium ($\underline{u}_E = \bar{u}_E$).

Imposing ND alone may lead to some paradoxical effects. It may happen that with no constraint, the incumbent would have served both types of consumers, but submitted to the ND constraint, it stops serving type $\bar{\mu}$ consumers. Therefore, a constraint U may be required together with ND in order to make ND effective (see section A.6 for more details). In this paper, we focus on constraints U and UND.

2.5 Regulatory rules

The purpose of the paper is to examine the consequences on consumer's surplus, on profits and on global surplus, of different ways of allocating and funding the USOs (U or UND). We consider essentially two regulatory frameworks, denoted by "restricted-entry regulation" and "pay or play regulation".

Whatever the regulatory mechanism, the timing of the game is the following. The regulator first announces the level of the tax t paid on each unit of the good. It also announces that at the end of the game the amount of tax collected will be allocated to the incumbent, if it serves the high cost consumer¹¹. Then firms compete, that is, they choose the share of the surplus they will leave to consumers, the tax rate being taken as exogenous.

2.5.1 The various regimes

As we will see further, the regulatory rules may lead at equilibrium to one of the four following situations:

- Firm I serves both types of consumers, and cross-subsidizes between them. We call this regime "I cross-subsidizes" (ICS). No tax is raised nor perceived;
- Firm E serves both types of consumers, and cross-subsidizes between them; we call this regime "E cross-subsidizes" (ECS). No tax is raised nor perceived;
- Firm I serves the high costs consumers, E serves the low cost consumers and pays a tax; we call this situation the "taxation regime" (TR);
- High costs consumers are excluded from the consumption of the good, low cost consumers being served by either of the two firms; we call this regime "exclusion regime" (ER).

2.5.2 'Restricted-entry' regulation versus 'pay or play' regulation

Under the 'restricted-entry' regulation, firm E is not allowed to serve high cost consumers $\bar{\mu}$. Under USO=U or UND, these consumers are served by the incumbent. Two regimes may occur: ICS and TR.

Under the 'pay or play' regulation (POP), firm E is allowed to serve consumer $\bar{\mu}$. When E chooses to serve $\bar{\mu}$, it does not pay nor receive any tax. Therefore, firms' payoffs are not symmetric¹². Under the 'play or pay' regulation, the identity of the firm who serves $\bar{\mu}$ is endogeneously determined. Three regimes may occur: ICS, TR, and ECS.

¹¹If the incumbent serves both types of consumers, it pays the tax but then it also receives the amount of the tax; it is thus neutral for its profit.

¹²By contrast, auction mechanisms generally lead to symmetric payoffs (see for example Anton and alii [1]).

Under the restricted-entry regulation, firm E 's strategy reduces to the choice of \underline{u}_E . Under the POP regulation, firm E may in addition choose to serve $\bar{\mu}$ consumers and thus also chooses \bar{u}_E .

2.5.3 The choice of the tax level

The choice of the tax level depends on the objectives pursued by the regulator. We focus here on three particular procedures: the first best tax, the balanced-budget tax and the competitively neutral tax.

The *first best tax* is the level of tax that maximizes the social welfare. However, this first best level may lead to a loss in profit for the incumbent.

Therefore, we introduce a second-best tax, namely the *balanced-budget tax*, defined by the maximization of the welfare, under the constraint:

$$\tilde{\pi}_I^{USO} = 0.$$

We also analyse the "*competitively neutral*" tax, where firm I is compensated and receives, when it is possible, the profit of the benchmark case. Let $\tilde{\pi}_K^{USO}$ be the profit of firm K at the equilibrium of this competitive process with tax, when the set of obligations is $USO=U$ or UND . Under competitive neutrality, the tax rate t^{USO} is determined by the equality

$$\tilde{\pi}_I^{USO} = \pi_I \quad (\text{"competitive neutrality condition"})$$

The idea of competitive neutrality is that the funding and the presence of USOs are designed not to affect the profit of firm I with respect to the unconstrained situation (benchmark)¹³.

Budget balance and competitive neutrality may or may not coincide, according on what happens in the benchmark case.

2.6 Welfare criteria and productive efficiency

Generally, USOs are imposed in a purpose of enhancing the situation of unfavored consumers (by the constraint U) or to achieve equity objectives (by the constraint ND). In our model, "unfavored" consumers correspond to high connection cost consumers (the $\bar{\mu}$ type). We assume that the public authority values by itself the access to the network of all consumers and denote by k the valuation of the access by the regulator. The welfares in the different regimes are

$$\begin{array}{ll} W_{II} = \underline{\alpha} \underline{S}_I + \bar{\alpha} \bar{S}_I + k & \text{in the ICS regime} \\ W_{EE} = \underline{\alpha} \underline{S}_E + \bar{\alpha} \bar{S}_E + k & \text{in the ECS regime} \\ \tilde{W}_{EI} = \underline{\alpha} (\underline{S}_E + t \tilde{q}_E) + \bar{\alpha} \bar{S}_I + k & \text{in the taxation regime} \\ W_{EI} = \underline{\alpha} \underline{S}_E + \bar{\alpha} \bar{S}_I + k & \text{in the taxation regime with } t = 0 \\ W_{K0} = \underline{\alpha} \underline{S}_K + \underline{\alpha} k & \text{in the exclusion regime (firm K serves } \underline{\mu} \text{).} \end{array}$$

¹³Firstly, remark that the competitive neutrality tax depends on the endogeneous market structure when USOs are imposed. Secondly, this criteria compensates perfectly the firm who incurs USOs but not its competitors.

In the benchmark case analysed above, and under our assumption $\bar{S}_K < 0$, the welfare is thus W_{I0} or W_{E0} according to which firm is the most efficient on the market of low costs consumers. The following remark follows readily from Lemma 2.

Remark 3 *The welfare in the taxation regime \widetilde{W}_{E1} is nonincreasing with respect to the tax level t .*

Let W be the value of the global welfare at the competitive equilibrium, and $\widetilde{W}^{USO}(t)$ the value of the global welfare at equilibrium with USO and tax. Then the difference $\Delta W^{USO} = \widetilde{W}^{USO}(t) - W$, that may be positive or negative, measures the variation of welfare due to the *USO* constraint. The variations of the global welfare may be decomposed into individual contributions, which allows to evaluate the transfers associated with a particular regime. Both constraints also induce welfare losses. According to the way USOs are attributed and funded, these losses may be incurred by firms and/or by $\underline{\mu}$ type consumers. As a whole, the global welfare may increase or decrease, according to which partial effect dominates. Moreover, the two systems that we examine may lead to opposite conclusions. An interesting question is to compare the welfares under the restricted-entry and the pay or play regulation.

3 The restricted-entry regulation: Firm I serves high cost consumers

Recall that under the restricted-entry rule, the USOs are imposed to the incumbent (firm I) and that the entrant (firm E) cannot choose to serve the consumers relevant for the USO instead of paying the tax. We consider successively the cases where the USO is restricted to U only, and defined by *UND*. We derive first the equilibria of the game, identify thereafter the possible inefficiencies generated by the situation and proceed finally with the welfare and redistribution analysis.

3.1 Equilibria of the competition process under USOs

The constraints U and *UND* that the regulator can impose on firm I have different consequences on the equilibrium configurations and on their welfare implications. We examine successively these various constraints. Recall that we restrict ourselves to the case where $\bar{S}_I < 0$ and $\bar{S}_E < 0$.

3.1.1 Equilibria under ubiquity constraint alone

As explained in the introduction, the competition process (for a given tax level t) is modelled by a sequential game, where the incumbent acts as a leader: firm I announces two values \underline{u}_I and \bar{u}_I such that: $\underline{u}_I \geq 0$ and $\bar{u}_I \geq 0$ (ubiquity constraint). Since firm E is not

allowed to serve consumer $\bar{\mu}$, we have clearly: $\bar{u}_I = 0$. Firm E 's profit function is

$$\bar{\pi}_E = \begin{cases} 0 & \text{in the ICS regime} \\ \underline{\alpha}(\bar{S}_E - \underline{u}_I) & \text{in the taxation regime.} \end{cases}$$

Firm E chooses to serve $\bar{\mu}$ if and only if $\underline{u}_I < \bar{S}_E$. Therefore the announcement of \underline{u}_I determines the regime that will prevail: $\underline{u}_I < \bar{S}_E$ leads to the taxation regime, $\underline{u}_I > \bar{S}_E$ leads to the ICS regime. Firm I 's profit function is then given by

$$\bar{\pi}_I = \begin{cases} \underline{\alpha}(\underline{S}_I - \underline{u}_I) + \bar{\alpha}\bar{S}_I & \text{if } \underline{u}_I > \bar{S}_E \text{ ICS regime} \\ \underline{\alpha}t\bar{q}_E + \bar{\alpha}\bar{S}_I & \text{if } \underline{u}_I < \bar{S}_E \text{ taxation regime} \end{cases}$$

Then the strategy of the incumbent reduces to the choice of the regime: In the taxation regime, firm I 's profit does not depend on $0 \leq \underline{u}_I < \bar{S}_E$ (which leads to a multiplicity of equilibria); in the ICS regime, firm I 's optimal announcement is of course $\underline{u}_I = \bar{S}_E$. Then firm I has to compare the corresponding values of its profit. We have:

$$\bar{\pi}_I = \max\left(\underline{\alpha}(\underline{S}_I - \bar{S}_E) + \bar{\alpha}\bar{S}_I, \underline{\alpha}t\bar{q}_E + \bar{\alpha}\bar{S}_I\right).$$

In all the sequel, we will express the profits of firm I in the different regimes in terms of the corresponding welfares. We can rewrite $\bar{\pi}_I$ as

$$\bar{\pi}_I = \max\left(W_{II} - k - \underline{\alpha}\bar{S}_E, \widetilde{W}_{EI} - k - \underline{\alpha}\bar{S}_E\right).$$

Up to a constant ($\underline{\alpha}\bar{S}_E + k$), the profits of I in the regimes ICS and TR coincide with the corresponding welfares W_{II} and \widetilde{W}_{EI} . We introduce the following convention to express the fact that firm I 's compares the quantities W_{II} and \widetilde{W}_{EI} when choosing the regime:

$$\widetilde{\Pi}_I = \max(W_{II}, \widetilde{W}_{EI}).$$

The notation $\widetilde{\Pi}_I$ denotes the profit of I (in presence of the tax t) in the regimes ICS and TR up to a constant. This leads to the following proposition.

Proposition 1 (Restricted-entry regulation, USO=U) *For any level of the tax, the perfect subgame equilibria lead to the regime associated to the highest value of the welfare: Firm I chooses the taxation regime when $\widetilde{W}_{EI} > W_{II}$ and it chooses to cross-subsidize when $W_{II} > \widetilde{W}_{EI}$.*

It is interesting to note that since the regulator can transfer its objective to the incumbent, the restricted-entry regulation avoids incentives problems that could appear due to informational asymmetries between parties: even if the regulator observes neither the costs of the firms nor the consumers' characteristics, delegation to the incumbent of the ubiquity constraint solves the adverse selection problems associated with unobservability of the firms' and the consumers' characteristics¹⁴.

¹⁴However, second price auctions would also solve such informational problems.

By Lemma 2, the fonction \widetilde{W}_{EI} decreases with t . Then if $W_{EI} = \widetilde{W}_{EI}(0) < W_{II}$, we have: $\widetilde{W}_{EI}(t) < W_{II}$ for all $t \geq 0$ and the ICS regime occurs for all tax level. If $W_{EI} \geq W_{II}$, there exists a tax level τ^U such that

$$\widetilde{W}_{EI}(\tau^U) = W_{II}.$$

The taxation regime occurs for $t \leq \tau^U$ and the ICS regime occurs for $t > \tau^U$. It is intuitive that when the tax rate is high, the entrant prefers to stay out of the market. A given level of tax generates a given level of the sum $\tilde{S}_E + t\tilde{q}_E$, which measures the "efficiency" of the relationship between E and the $\underline{\mu}$ consumer, conditionnally to t . Recall that this quantity is a decreasing function of t .

The following result gives the consumers' surpluses in the different regimes.

Corollary 1 (SPE in the restricted-entry game, USO=U)

In both regimes, the consumer $\underline{\mu}$ gets $\underline{u} = 0$. The utility of consumer $\underline{\mu}$ depends on the regime:

$$\underline{u} = \begin{cases} \underline{u}_I = \tilde{S}_E & \text{in the ICS regime} \\ \underline{u}_E \in [0, \tilde{S}_E] & \text{in the taxation regime.} \end{cases}$$

The cross-subsidy regime corresponds to a situation where I is relatively more efficient on the market of low cost consumers than E when the latter is submitted to the tax. In this case, I announces for both types of consumers values that E cannot match because the tax is too high (and thus the quantity $\tilde{S}_E + t\tilde{q}_E$ is too small). Therefore, I serves both types of consumers and cross-subsidizes between them (that is, I finances the deficit created by the constraint U on high costs consumers by a transfer levied on low cost consumers).

In the taxation regime, by contrast, the sum of the surplus generated by the relationship between E and $\underline{\mu}$ and of the amount of taxes collected is sufficiently high to allow both firms to be active, each of them serving a different segment of the market. In this regime, there are multiple equilibria¹⁵ which can only be selected by an additionnal criterion: a criterion of dominance for firms would imply $\underline{u} = 0$, whereas if there would exist a second entrant, competition on the $\underline{\mu}$ would lead to $\underline{u} = \tilde{S}_E$. An important consequence of this result is the following.

Under ubiquity, a criterion of 'pure' productive efficiency would consist in the comparison of W_{II} and W_{EI} . By Proposition 1, we know that firm I compares W_{II} and \widetilde{W}_{EI} . Since $\widetilde{W}_{EI} < W_{EI}$, it follows that the ICS regime is "too often" implemented with regard to 'pure' productive efficiency.

However, a reasonable criterion of productive efficiency should take into account the fact that ubiquity is funded through taxes. If we define productive efficiency as the comparison between W_{II} and \widetilde{W}_{EI} , we see that the delegation of the regulator's objective function to firm I allows to choose the socially preferred regime.

¹⁵If we had used a simultaneous game instead of a sequential game, the equilibrium would have been unique. Note, however, that a Nash equilibrium may fail to exist when the regulator imposes also the non discrimination constraint.

In addition, a tax level equal to 0 allows to achieve pure productive efficiency. A zero tax, however, may fulfill neither of the tax constraints (budget balance of the incumbent or competitive neutrality).

3.1.2 Equilibria of the restricted-entry game with non discrimination constraint

The regulator imposes that each consumer receives the same nonnegative surplus. This constraint is implemented by restricting the space of strategies of the incumbent ($\underline{u}_I = \bar{u}_I \geq 0$). The proof of the following can be found in appendix (section A.2).

Proposition 2 (Restricted-entry regulation, USO=UND) *Firm I's profit is given, up to a constant, by*

$$\tilde{\Pi}_I \equiv \max(W_{II} - \bar{\alpha}\tilde{S}_E, \tilde{W}_{EI}).$$

By contrast with the case USO=U (see Proposition 1), the profit of the incumbent does not coincide with the welfares in the regimes ICS and TR. Under UND, the choice of the incumbent differs from that of the regulator (for any given tax level t).

By Lemma 2, the function $\tilde{W}_{EI} + \bar{\alpha}\tilde{S}_E$ is decreasing with respect to t . Let τ^{UND} be the value of the tax for which

$$\tilde{W}_{EI}(\tau^{\text{UND}}) = W_{II} - \bar{\alpha}\tilde{S}_E(\tau^{\text{UND}}).$$

It is worth noting that $\tau^{\text{UND}} \geq \tau^{\text{U}}$. We have the following

Corollary 2 (SPE in the restricted-entry game, USO=UND) *The ICS regime occurs if and only if*

$$\tilde{W}_{EI} \leq W_{II} - \bar{\alpha}\tilde{S}_E \quad \text{or, equivalently} \quad t \geq \tau^{\text{UND}}.$$

The taxation regime can occur only if $\underline{S}_E > \underline{\alpha}S_I$. In that case, it appears when $0 \leq t < \tau^{\text{UND}}$. Consumers' surplus depends on the regime:

- *cross-subsidy regime (ICS): $\underline{u}_I = \bar{u}_I = u_I = \tilde{S}_E$;*
- *Taxation regime (TR): $\underline{u} = \bar{u} = 0$.*

Recall that we have identified a first distortion due to the tax: Under USO=U, firm I compares W_{II} and \tilde{W}_{EI} , instead of W_{II} and W_{EI} . The tax distortion $W_{EI} - \tilde{W}_{EI}$ implies that the ICS regime is too often implemented.

Proposition 2 brings out an additional distortion under USO=UND, coming from the fact that under UND firm I has to leave $\bar{\alpha}\tilde{S}_E$ to the $\bar{\mu}$ consumers. The distortion $\bar{\alpha}\tilde{S}_E$ has the opposite impact: compared to productive efficiency, firm I chooses for too large a set of parameters to remain in the taxation regime, instead of choosing the ICS regime

($\tau^{\text{UND}} \geq \tau^{\text{U}}$). The ND constraint softens the competition: since the incumbent has to leave the same utility to both consumers, it is less aggressive on the low cost consumers' market¹⁶. By contrast, the taxation regime allows the incumbent to leave a zero utility to high costs consumers.

The distortion due to ND appears particularly clearly in the following case. Suppose that firm I is more efficient on both markets of consumers $\bar{\mu}$ and $\underline{\mu}$ ($0 < \underline{S}_E < \underline{S}_I$ and $\bar{S}_E < \bar{S}_I < 0$, so that $W_{II} > W_{EI}$). When ubiquity (U) alone is at work, firm I serves $\bar{\mu}$ and $\underline{\mu}$ (cross-subsidy regime). Under UND, however, firm E serves $\underline{\mu}$ if and only if the condition

$$W_{EI} < W_{II} < \widetilde{W}_{EI} + \bar{\alpha}\bar{S}_E$$

is satisfied. This condition expresses the fact that the ND distortion ($\bar{\alpha}\bar{S}_E$) dominates the tax distortion ($W_{EI} - \widetilde{W}_{EI}$). In that case, firm E is active on the market whereas I is more efficient.

We now can analyse the welfare and redistributive implications of imposing USOs depending on their funding is realized through cross-subsidies or through taxation.

3.2 Welfare and redistribution analysis

Consider first the case where U is imposed alone. The following proposition, which follows directly from Proposition 1, states the implications of U in terms of welfare.

Lemma 3 (Restricted-entry regulation, USO=U, welfare analysis)

The welfare under the restricted-entry regulation and with the ubiquity constraint is given by

$$W_f^U(t) = \max(\widetilde{W}_{EI}(t), W_{II}).$$

The function W_f^U is nonincreasing and continuous with respect to the tax level t .

More precisely, it is strictly decreasing in the taxation regime and then constant in the cross-subsidy regime.

Consider now the case where both constraints are imposed together (UND). We have the following lemma.

Lemma 4 (Restricted-entry regulation, USO=UND, welfare analysis) *The welfare under the restricted-entry regulation with USOs U and ND is given by*

$$W_f^{\text{UND}}(t) = \begin{cases} \widetilde{W}_{EI} & \text{if } \widetilde{W}_{EI} \geq W_{II} - \bar{\alpha}\bar{S}_E \\ W_{II} & \text{if } \widetilde{W}_{EI} \leq W_{II} - \bar{\alpha}\bar{S}_E. \end{cases}$$

The function W_f^{UND} is non increasing with respect to t and discontinuous at $t = \tau^{\text{UND}}$.

¹⁶This result is also present in Anton and alii [1]

A consequence of this proposition is that the tax level which maximizes the welfare (that is, the first best tax), in the case where the taxation regime exists, i.e. when $W_{EI} > W_{II}$, is $t^{FB} = 0$. In the cross-subsidy regime, the welfare is constant with respect to the tax rate. However, for this level t^{FB} of the tax rate, firm I incurs a deficit. Therefore it is interesting to examine what happens for the tax level t^{BB} that guarantees the budget balance of firm I : $\tilde{\pi}_I(t^{BB}) = 0$. This is done in the following proposition.

Proposition 3 *Assume $\bar{S}_E < 0, \bar{S}_I < 0$ and $\underline{S}_I < \underline{S}_E$ or $W_{II} < W_{EI}$. The profit of the incumbent in the benchmark case is $\pi_I = 0$. Then, whatever the USOs at work:*

- *Competitive neutrality and budget balance are equivalent.*
- *The taxation regime exists for small values of t and the first best tax is $t^{FB} = 0$. This first best tax implies a deficit for the incumbent: $\tilde{\pi}_I(0) = \bar{\alpha}\bar{S}_I < 0$.*
- *If $\underline{\alpha}\underline{S}_I + \bar{\alpha}\bar{S}_I = \tilde{\pi}_I(t^e) \geq 0$, that is, when the monopoly of I is viable, then there exists at least a competitively neutral tax.*

When $t = t^{BB}$ the welfare is $\widetilde{W}_{EI} = \underline{\alpha}(\tilde{S}_E + t^{BB}\tilde{q}_E) + \bar{\alpha}\bar{S}_I + k$ in the taxation regime, whereas in the case where I cross-subsidizes, it is W_{II} . In the benchmark case, it is equal to $W_{E0} = \underline{\alpha}\underline{S}_E + \underline{\alpha}k$. Clearly the comparison with the benchmark depends on the value of k .

The last point of Proposition 3 follows from the continuity of the profit function $\tilde{\pi}_I$ with respect to t . Since $\tilde{\pi}_I(0) < 0$ and $\tilde{\pi}_I(t^e) > 0$, there exists at least a solution to the equation $\tilde{\pi}_I(t) = 0$.

Remark 4 *Suppose now $\underline{S}_I > \underline{S}_E$ (and still $\bar{S}_E < 0, \bar{S}_I < 0$).*

- *The profit of firm I in the benchmark case is: $\pi_I = \underline{\alpha}(\underline{S}_I - \underline{S}_E) > 0$. Therefore, competitive neutrality implies budget balance for the firm I .*
- *If $USO = U$, the cross-subsidy regime prevails for all t . The welfare does not depend on the tax level.*
- *If $USO = UND$, the taxation regime occurs when $t < \tau^{UND}$ if $\underline{S}_E > \underline{\alpha}\underline{S}_I$.*
- *In both cases, if $\underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_I \geq 0$, then there exists a competitively neutral tax.*

At this stage, we can discuss the social benefits and costs of the USO's.

Assume for instance that $\underline{S}_E < \underline{S}_I$ (and still $\bar{S}_I, \bar{S}_E < 0$). Then, in the benchmark case, firm I serves the low cost consumers and high costs consumers are not served; the welfare is then W_{I0} . If UND is imposed, then the welfare is given by the above proposition. In the taxation regime (see point 3 of the above remark), the variation of welfare with regard to the benchmark case can be written as:

$$\Delta W^{UND} = \widetilde{W}_{EI} - W_{I0} = \bar{\alpha}k + \bar{\alpha}\bar{S}_I + \underline{\alpha}(\tilde{S}_E + t\tilde{q}_E - \underline{S}_E) + \underline{\alpha}(\underline{S}_E - \underline{S}_I).$$

The first term, which is positive, represents the gain, in terms of social welfare, due to the connection of the high costs consumers. The other terms are negative. The term $\bar{\alpha}\bar{S}_I$ represents the cost associated with the connection of these consumers. The third term represents the loss in *allocative efficiency* (distortion of the surplus due to the taxation). The last term represents the *productive inefficiency*, due to the fact that the most efficient firm for serving these consumers, that is, firm I is replaced by firm E (less efficient).

The choice of a tax level has also important consequences in terms of redistribution. We now compare the benefits obtained by the various classes of consumers across the various possible regimes.

Suppose first that UND is at work. If the tax rate t leads to the regime where I cross-subsidizes, then both types of consumers obtain a positive surplus $\underline{u}_I = \bar{u}_I = \bar{S}_E > 0$, which depends on t . As Gasmi et alii (1999) pointed out, cross-subsidization corresponds to implicit taxation and has important redistributive effects.

In the taxation regime, where the tax is perceived, all consumers have $\underline{u}_E = \bar{u}_I = 0$. Therefore, the choice between cross-subsidies or taxation regime raises an equity issue, that has to be solved by political choices and not only through *ex-ante* criteria, such as competitive neutrality.

Suppose now $USO=U$. In the taxation regime, the sharing of the surplus between μ consumers and E cannot be determined. However, they have $\underline{u}_I = \bar{S}_E$ in the cross-subsidies regime and $\underline{u}_E \leq \bar{S}_E$ in the taxation regime. High costs consumers are indifferent, as whatever the regime, they have $\bar{u}_I = 0$. The ubiquity constraint generates no redistributive effect between consumers. However, both consumers are better in the ICS regime.

4 Pay or Play: the entrant may serve high cost consumers

In this section, we investigate the alternative allocation rule where the entrant may choose to serve the non profitable users instead of paying the tax. As in the previous section we consider separately the effects of U and of UND .

4.1 The ubiquity constraint under the pay or play rule

We first compute the equilibria of the pay or play game and thereafter examine its welfare properties.

4.1.1 Equilibria of the competition process

The first important result shows that the profit of firm I coincides again with that of the regulator and that the social objective can thus be decentralized. But compared to the restricted-entry regulation, the pay or play regulation allows to achieve another regime

where firm E serves all the consumers (the ECS regime) and finances its deficit on high costs consumers through cross-subsidies. Let τ_{pop} be the value of the tax for which

$$\underline{\alpha} (\underline{S}_E - \bar{S}_E) + \bar{\alpha} \bar{S}_E = 0$$

that is, $\underline{\alpha} \bar{S}_E = W_{EE} - k$. If t is lower than this threshold, then E has no incentive to serve the non profitable users whereas if it is higher, E can profitably serve both classes by cross-subsidization.

The proposition and the corollary below are proved in appendix (section A.3).

Proposition 4 (POP regulation, USO=U) *Under the pay or play regulation, when only the ubiquity constraint is at work, the ECS regime may appear only for $t > \tau_{pop}$. The profit of the incumbent is given (up to a constant) by*

$$\tilde{\Pi}_I = \begin{cases} \max(W_{II}, \widetilde{W}_{EI}) & \text{if } t \leq \tau_{pop} \\ \max(W_{II}, \widetilde{W}_{EI}, W_{EE}) & \text{if } t > \tau_{pop}. \end{cases}$$

Therefore the profit is identical, up to a constant, to the welfare in the corresponding regime.

As in the restricted-entry regulation, the choice of the regime by firm I is thus that of the regulator (for a given level of the tax) who can thus decentralize the social optimum.

However, the pay or play regulation now introduces, for some values of the tax rate, the possibility that firm E serves both classes of consumers, a situation which was not achievable under the restricted-entry regulation. The following lemma now gives the equilibria of the game.

Corollary 3 (POP, USO=U, Perfect equilibria) *At equilibrium of the pay or play regulation with ubiquity constraint, two cases may appear, according to the value of the tax.*

(i) *When $t \leq \tau_{pop}$, the situation is identical to the restricted-entry regulation. The regime ICS and TR may occur. The welfare is given by*

$$W_{pop}^U = W_f^U = \max(\widetilde{W}_{EI}, W_{II}).$$

The equilibria are the same as in the restricted-entry game:

- *if $\widetilde{W}_{EI} < W_{II}$: I cross-subsidizes (ICS), $\underline{u}_I = \bar{S}_E$, $\bar{u}_I = 0$*
- *if $\widetilde{W}_{EI} > W_{II}$: Taxation regime (TR), $\underline{u}_I \in [0, \underline{S}_I - t\bar{q}_E]$, $\bar{u}_I = 0$.*

(ii) *When $t > \tau_{pop}$, three regimes (ICS, TR, ECS) may occur. Then the welfare is given by*

$$W_{pop}^t = \max(\widetilde{W}_{EI}, W_{II}, W_{EE}).$$

The equilibria are given by

- if $W_{pop}^U = W_{II}$, I cross-subsidizes (ICS) and $(\underline{u}_I, \bar{u}_I) \in D_{II}$.
- if $W_{pop}^U = \widetilde{W}_{EI}$, the taxation regime occurs (TR), and $(\underline{u}_I, \bar{u}_I) \in D_{EI}$.
- if $W_{pop}^U = W_{EE}$, E cross-subsidizes (ECS) and $(\underline{u}_I, \bar{u}_I) \in D_{EE}$ where

$$D_{EE} = \left\{ (\underline{u}_I, \bar{u}_I) / \bar{\alpha}\bar{u}_I \leq \underline{\alpha}(\underline{S}_E - \widetilde{S}_E) + \bar{\alpha}\widetilde{S}_E \text{ and } \underline{\alpha}\underline{u}_I + \bar{\alpha}\bar{u}_I \leq W_{EE} - k \right\}$$

$$D_{II} = \left\{ (\underline{u}_I, \bar{u}_I) / \underline{u}_I \geq \widetilde{S}_E \text{ and } \underline{\alpha}\underline{u}_I + \bar{\alpha}\bar{u}_I = W_{EE} - k \right\}$$

$$D_{EI} = \left\{ (\underline{u}_I, \bar{u}_I) / \bar{\alpha}\bar{u}_I = \underline{\alpha}(\underline{S}_E - \widetilde{S}_E) + \bar{\alpha}\widetilde{S}_E \text{ and } \underline{u}_I \leq \widetilde{S}_E \right\}.$$

The pay or play regulation has a number of welfare implications that are now examined more in detail.

4.1.2 Welfare consequences of the ubiquity constraint: A comparison between the restricted-entry and the pay or play regulations

The virtue of the pay or play regulation, compared to the restricted-entry regulation, is that it allows, in situations where E is more efficient, to allocate the market of high costs consumers to firm E . It results that the pay or play regulation offers an additional possibility for the allocation of the users compared to the restricted-entry one. It thus enhances productive efficiency. Moreover, it is easy to check the following points:

- In the taxation regime, high costs consumers obtain $\bar{u} = 0$ in the restricted-entry regulation, and $\bar{u} = \frac{1}{\bar{\alpha}} \left[\underline{\alpha}(\underline{S}_E - \widetilde{S}_E) + \bar{\alpha}\widetilde{S}_E \right] > 0$ in the pay or play regulation. Therefore the high cost consumers prefer the pay or play regulation.
- In the regime where I cross-subsidizes, consumers $\bar{\mu}$ have $\bar{u} = 0$ in the restricted-entry regulation and a value $\bar{u} \in \left[0, W_{EE} - k - \underline{\alpha}\widetilde{S}_E \right]$ in the pay or play regulation; consumers $\underline{\mu}$ have $\underline{u} = \widetilde{S}_E$ under the restricted-entry regulation, whereas they obtain a value $\underline{u} \in \left[\widetilde{S}_E, W_{EE} - k \right]$ in the pay or play regulation. Thus both classes of consumers benefit from the pay or play regulation instead of the restricted-entry one.
- In the regime where E cross-subsidizes, there is a multiplicity of equilibria that share differently the surplus between consumers and firm E .

An important consequence of the first remark is that the pay or play regulation involves redistribution between consumers, although the ND constraint is not imposed (this redistribution does not appear in the restricted-entry regulation, under $USO=U$). These remarks lead to the following proposition.

Proposition 5 (USO=U, Welfare and utility comparison) *When the ubiquity constraint alone is at work, the welfare and the consumers'surplus are both always at least higher under the pay or play regulation than under the restricted-entry regulation.*

We have seen that the welfare configuration depends on the comparison between three terms: W_{II} , W_{EI} , W_{EE} . More precisely, at equilibrium, the welfare exhibits the following properties.

- If W_{II} is the highest, then the welfare is flat (constant with respect to t), the equilibrium configuration is that where I cross-subsidizes. Whatever the value of t , the profit of firm I is $\pi_I = W_{II} - W_{EE} > 0$. Every t is then a first best tax.
- If W_{EE} is the highest, then the first best is achieved for all $t > \tau_{\text{pop}}$. For such values of t , we have $\pi_I = 0$, and these tax rates verify the budget constraint and are competitively neutral.
- If W_{EI} is the highest, the first best is achieved for $t = 0$ and for that value, the profit of firm I is $\tilde{\pi}_I(0) = \bar{\alpha}\bar{S}_I < 0$. The balanced-budget tax t^{BB} is thus a second best. Budget balance and competitive neutrality coincide, but this (common) level of the tax may lead to any of the possible regimes.

4.2 Ubiquity and non discrimination under pay or play regulation

We now examine the case where the incumbent is tied down to a non discrimination constraint. Firm E is allowed to serve the high costs consumers $\bar{\mu}$. Since we assume $\tilde{S}_E < 0$, it is clear that firm E never serves consumer $\bar{\mu}$ only. Consider the case (ECS) where firm E serve both types of consumers. Recall that I must announce $\underline{u}_I = \bar{u}_I \geq 0$ because of the ubiquity and non discrimination constraints. It follows that, in the case (ECS)

$$\underline{u}_E = \bar{u}_E \geq 0.$$

In other words, firm E has to fulfill both constraints also (in case it serves $\underline{\mu}$ and $\bar{\mu}$). Therefore, in that case, there is no reason for E to pay a tax when it serves both markets and cross-subsidizes (ECS regime).

Recall that we have defined τ_{pop} by

$$\underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E = \underline{\alpha}\tilde{S}_E.$$

We introduce another threshold $\bar{\tau}$, defined by

$$\underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E = \tilde{S}_E.$$

Since \tilde{S}_E is decreasing with respect to t , we have: $\tau_{\text{pop}} \leq \bar{\tau}$. As usual, the regimes at equilibrium follow immediately from the profit function of firm I . The proposition below (proved in appendix section A.4) gives the profit of the incumbent in the different regimes.

Proposition 6 (USO=UND, POP regulation, SPE) *Under the pay or play regulation and USO=UND, the profit function of firm I is given (up to a constant) by*

$$\tilde{\Pi}_I = \begin{cases} \max(W_{II} - \bar{\alpha}\tilde{S}_E, \tilde{W}_{EI}) & \text{if } t \leq \tau_{\text{pop}} \\ \max(W_{II} - [\tilde{S}_E - (\underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E)], \tilde{W}_{EI}, W_{EE}) & \text{if } \tau_{\text{pop}} \leq t \leq \bar{\tau} \\ \max(W_{II}, W_{EE}) & \text{if } \bar{\tau} \leq t. \end{cases}$$

Three configurations may prevail at equilibrium, according to the level of the tax rate:

- when the tax rate is very low, it is never profitable for firm E to serve both markets. Therefore, two regimes may appear, ICS or TR, according to the value of the tax;
- for intermediate values of the tax rate, letting E cross-subsidize may appear at equilibrium, depending on the efficiency of both firms;
- when the tax rate is very high, the total amount generated by the taxation regime is so low that this regime disappears, and the equilibrium is either ICS or ECS, according to the respective efficiency of both firms.

Note that with regard to "pure" productive efficiency, the first two configurations involve a distortion, due to the non discrimination constraint. However, this distortion has not the same expression in both cases. More precisely, the distortion induced by the non discrimination constraint writes $\bar{\alpha}\tilde{S}_E$ for $t \leq \tau_{\text{pop}}$ (as under the restricted-entry regulation), $\tilde{S}_E - (\underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E)$ for $\tau_{\text{pop}} \leq t \leq \bar{\tau}$ and cancels for $t \geq \bar{\tau}$. It is easy to check that the distortion is continuous and decreasing with respect to the tax rate.

In the following proposition (see section A.4 for the proof), we compare the welfares under the restricted-entry and the POP regulation (for USO=UND).

Proposition 7 (USO=UND, Welfare comparison) *If $t \leq \tau_{\text{pop}}$, the welfare is the same in the POP regulation as in the restricted-entry regulation. If $t > \tau_{\text{pop}}$, this is no longer the case. The welfare under the POP regulation may be strictly lower than under the restricted-entry regulation.*

(i) *Suppose first*

$$\begin{cases} \tau_{\text{pop}} \leq t \leq \bar{\tau} \text{ or, equivalently, } \underline{\alpha}\tilde{S}_E \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E \leq \tilde{S}_E \\ W_{EE} < W_{II} < W_{EE} + [\tilde{S}_E - (\underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E)] \\ \tilde{W}_{EI} + \bar{\alpha}\tilde{S}_E < W_{II}. \end{cases} \quad (2)$$

Then under the restricted-entry regulation, the ICS regime prevails and the welfare is W_{II} , while the POP regulation leads to the ECS regime, where the welfare is $W_{EE} < W_{II}$.

(ii) *Suppose now*

$$\begin{cases} t > \bar{\tau} \text{ or, equivalently, } \tilde{S}_E \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E \\ \tilde{W}_{EI} > \max(W_{EE}, W_{II}). \end{cases}$$

Then under the restricted-entry regulation, the taxation regime prevails and the welfare is \tilde{W}_{EI} , while the POP regime leads to the regimes ECS or ICS, and thus to a lower welfare.

Unlike what happens when the ubiquity constraint alone is at work (see Proposition 5), the welfare in the POP regulation may be strictly lower than in the restricted-entry regulation. The loss of welfare when it appears may come from the distortion $\bar{S}_E - (\underline{\alpha}S_E + \bar{\alpha}\bar{S}_E)$ due to the ND constraint (case (i) above) or to the disappearance of the taxation regime in the POP regulation for high values of the tax level (case (ii) above).

According to Proposition 6, *ex-ante* criteria (competitive neutrality,...) may rule out some socially optimal regimes. The regulator should take this into account when determining the tax level and choosing the regulation rules.

5 Discussion

In this paper, we have analysed and compared the properties of the restricted-entry regulation and the pay or play regulation for allocating and funding universal service obligations in a network market open to new competitors. Of course, other mechanisms may be implemented to fill these universal service obligations both on the allocation and on the funding sides. We now briefly mention two alternative regulatory frameworks : lump sum transfers and second-price auctions.

5.1 Lump sum transfers

In this paper, we have concentrated here on the mechanisms that are financed through taxation. Losses associated with serving the high costs consumers could also be funded through lump sum transfers levied on tax payers.

Suppose for instance $USO=U$ and a transfer T is used to fund the USO ¹⁷. We briefly describe the situation in case the entrant is not allowed to serve high costs consumers. The profit of the incumbent is

$$\pi_I = \begin{cases} \underline{\alpha}(S_I - S_E) + \bar{\alpha}\bar{S}_I + T & \text{if I serves } \underline{\mu} \text{ and } \bar{\mu} \\ \bar{\alpha}\bar{S}_I + T & \text{if E serves } \underline{\mu} \text{ and I serves } \bar{\mu}. \end{cases}$$

The welfares in the different regimes are

$$\begin{aligned} W_{II}^T &= \underline{\alpha}S_I + \bar{\alpha}\bar{S}_I + k + T - (1 + \lambda)T = W_{II} - \lambda T \\ W_{EI}^T &= \underline{\alpha}S_E + \bar{\alpha}\bar{S}_I + k + T - (1 + \lambda)T = W_{EI} - \lambda T, \end{aligned}$$

where λ denotes the cost of transfer of public funds (a transfer equal to T thus generates a dead weight loss equal to λT). The profit of the incumbent is given, up to a constant, by

$$\Pi_I = \max(W_{II}, W_{EI}).$$

Recall that with the unit tax, we had $\tilde{\Pi}_I = \max(W_{II}, \tilde{W}_{EI})$ (see Proposition 1). Funding the ubiquity constraint through lump sum transfers thus guarantees "pure" productive efficiency.

¹⁷Remark that in such a case, the distortion on demand disappears.

Now suppose that the regulator chooses a second-best criterion and defines the transfer T^{BB} so as to compensate the incumbent if it is necessary

$$T^{\text{BB}} = \begin{cases} \max(-\alpha(\underline{S}_I - \underline{S}_E) - \bar{\alpha}\bar{S}_I, 0) & \text{when } W_{II} > W_{EI} \\ -\bar{\alpha}\bar{S}_I & \text{when } W_{EI} > W_{II}. \end{cases}$$

Note that $T^{\text{BB}} \geq 0$. The comparison between the regulation modes depends on the value of t^{BB} , T^{BB} and λ and is given in the following proposition.

Proposition 8 (USO=U, restricted-entry regulation, taxation v.s. transfers)

Funding the ubiquity constraint through taxation is better than through lump sum transfers if

- $T^{\text{BB}} > 0$, when $W_{II} > W_{EI}$;
- $\max(\bar{W}_{EI}(t^{\text{BB}}), W_{II}) > W_{EI}^T(T^{\text{BB}}) = W_{EI} - \lambda T^{\text{BB}}$, when $W_{EI} > W_{II}$.

In the first case ($W_{II} > W_{EI}$), both taxation and lump sum transfers lead to the ICS regime. The welfare is W_{II} with the taxation rule and $W_{II} - \lambda T^{\text{BB}} \leq W_{II}$ with the lump sum transfer. Funding USOs through taxation is thus strictly better than through lump sum transfers if $T^{\text{BB}} > 0$, that is, if

$$\alpha(\underline{S}_I - \underline{S}_E) + \bar{\alpha}\bar{S}_I < 0.$$

This condition expresses the fact that the incumbent makes losses when it serves both markets, facing the competitive pressure of the entrant. Moreover, the taxation rule provides the regulator with a tool to share the surplus between the incumbent's profit and the consumers' surplus: the tax rate influences the potential competitive pressure \bar{S}_E (and does not modify the welfare W_{II}). By contrast, with lump sum transfers, the regulator lacks such instruments to implement the sharing of the global surplus between agents.

Suppose now that productive efficiency requires that low costs consumers are served by the entrant ($W_{EI} > W_{II}$). We know that this market structure cannot be implemented by the tax if t^{BB} is too high, that is if $\bar{W}_{EI}(t^{\text{BB}}) < W_{II}$. In that case, financing this USO through lump sum transfers is optimal if $W_{EI} - \lambda T^{\text{BB}} > W_{II}$.

5.2 Second-price auctions

Concerning the allocation of the USOs, regulators often use second price auctions (see for example [1]): in such a mechanism, the ubiquity constraint may be sold to the competitors through an auction mechanism. Each competitor bids for a subsidy for serving the $\bar{\mu}$ consumers. The firm that requires the lowest subsidy wins the auctions (i.e. serves the market). Assume that this auction mechanism is financed through transfers. Then the government transfers to the winning firm the value required by the other firm to serve the high cost consumers.

Clearly, at equilibrium of this auction mechanism, each firm K requires a transfer equal to its loss $\bar{\alpha} |\bar{S}_K|$, the most efficient firm on the market of $\bar{\mu}$ consumers (i.e. the firm K for which \bar{S}_K is maximum) wins the auction and receives $|\bar{\alpha} \bar{S}_{K'}|$ where K' is the other firm.

The comparison between auction and pay or play depends crucially on whether the auction mechanism allows cross-subsidization (no subsidy is perceived if the firm serves both consumers, see Anton et alii, 1999).

Assume that cross-subsidization is not allowed. Suppose for instance that the taxation regime prevails under the pay or play regulation. Then the welfare is \bar{W}_{EI} , whereas with a second price auction it is $W_{EI} - \lambda \bar{\alpha} |\bar{S}_E|$ where I is the winner of the auction. The comparison thus depends on the values of λ and of the taxes: the second price auction dominates the pay or play regulation if

$$\frac{W_{EI} - \bar{W}_{EI}}{\bar{\alpha} |\bar{S}_E|} > \lambda,$$

that is, if λ is not too high.

5.3 Concluding remarks

This paper is a first attempt to characterize the regulation systems from a welfare and a distributional point of view. However, we have only considered here the case of geographical differences between users. But users can also differ in their demand function, due either to various preferences for the network good or to the dispersion of their revenues. Light users (low income agents) can then be disadvantaged (or even excluded from the network) compared to heavy ones, because of the competition on the latter market. Another goal of the regulator could then be to protect the interests of light users. This requires to take into account a more complex structure of demand and another form of USO. We will examine this problem in a future paper.

References

- [1] Anton J., J. Weide and N. Vettas (1999), "Strategic Pricing and Entry under Universal Service and Cross-Market Price Constraints", mimeo, Duke University.
- [2] Armstrong J. C. Doyle and J. Vickers (1996), "The Access Pricing Problem: A Synthesis", *Journal of Industrial Economics*, June, vol XLIV, n°2, 131-150.
- [3] Baumol W.J. (1999), "Having your Cake: How to Preserve Universal Service Cross-Subsidies While Facilitating Competitive Entry", *Yale Journal on Regulation*, vol 16, n°1, Winter.
- [4] FCC (1996) "In the Matter of the Federal-State Joint Board on Universal Service", *CC Docket*, n°96-45, nov., 8, 1996.
- [5] Gasmi F., J.-J. Laffont, Sharkey W.W, (1999) "Competition, Universal Service and Telecommunications Policy in Developing Countries", mimeo IDEI, march.
- [6] Laffont J.-J., J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*, MIT Press.
- [7] Laffont J.-J., J. Tirole (1998), "Competition in Telecommunications", *Munich Lectures, Part VI, Universal Service*.
- [8] Milgrom P. (1996) "Procuring Universal Service: Putting Auction Theory to Work", Lecture at the *Royal Swedish Academy of Sciences*, december.
- [9] Panzar J. (1999), "A Methodology for Determining USO Costs and USF Payments", mimeo, North Western University", Chicago.
- [10] WIK (1997), *Costing and Financing Universal Service Obligations in a Competitive Telecommunications Environment in the European Union*, Study for the DG XIII of the European Commission, october.

A Appendix

A.1 Proof of Lemma 2

The function \tilde{S}_E is a supremum of affine functions, therefore it is a convex function. By the envelop theorem, we have

$$\frac{d\tilde{S}_E(t)}{dt} = -\tilde{q}_E.$$

The function \tilde{q}_E is therefore decreasing and we have

$$\frac{d}{dt} \left[\tilde{S}_E + t\tilde{q}_E \right] = t \frac{d\tilde{q}_E}{dt} \leq 0.$$

To see the last point, take $\bar{t} = w'(0) - c_E$, where $c_E > C'_E(q, \underline{\mu})$ for all q (the marginal cost $C'_E(q, \underline{\mu})$ is assumed to be bounded). Then the function

$$w(q) - C_E(q, \underline{\mu}) - \bar{t}q$$

is decreasing for $q \geq 0$. Therefore, it attains its maximum at $q = 0$ and $\tilde{S}_E(\bar{t}) = -\underline{f}_E < 0$. The continuity of \tilde{S}_E gives the existence of t^e such that: $\tilde{S}_E(t^e) = 0$.

A.2 Proof of Proposition 2 and Lemma 2

We consider the restricted-entry regulation, with USO=UND. Recall that the incumbent announces one value $u_I \geq 0$. Since firm E is not allowed to serve $\underline{\mu}$, its profit function is

$$\tilde{\pi}_E = \begin{cases} 0 & \text{ICS regime} \\ \tilde{S}_E - \underline{u}_I & \text{taxation regime.} \end{cases}$$

Firm E chooses to serve $\underline{\mu}$ if and only if $\underline{u}_I < \tilde{S}_E$. Therefore the announcement of \underline{u}_I determines the regime that will prevail: $\underline{u}_I < \tilde{S}_E$ leads to the taxation regime, $\underline{u}_I > \tilde{S}_E$ leads to the ICS regime. Firm I 's profit function is then given by

$$\tilde{\pi}_I = \begin{cases} \underline{\alpha}\underline{S}_I + \bar{\alpha}\bar{S}_I - u_I & \text{if } \underline{u}_I > \tilde{S}_E \text{ ICS regime} \\ \underline{\alpha}t\tilde{q}_E + \bar{\alpha}(\bar{S}_I - u_I) & \text{if } \underline{u}_I < \tilde{S}_E \text{ taxation regime} \end{cases}$$

Then the choice of the incumbent reduces to the choice of the regime: in the taxation regime, firm I 's profit is maximum for $u_I = 0$ (unique equilibrium); in the ICS regime, firm I 's optimal announcement is of course $\underline{u}_I = \tilde{S}_E$. Then firm I has to compare the corresponding values of its profit. We have:

$$\tilde{\pi}_I = \max \left(\underline{\alpha}\underline{S}_I + \bar{\alpha}\bar{S}_I - \tilde{S}_E, \underline{\alpha}t\tilde{q}_E + \bar{\alpha}\bar{S}_I \right).$$

The profit $\tilde{\pi}_I$ can be written as

$$\tilde{\pi}_I = \max \left(W_{II} - k - \underline{\alpha}\tilde{S}_E - \bar{\alpha}\tilde{S}_E, \bar{W}_{EI} - k - \underline{\alpha}\tilde{S}_E \right).$$

which gives Proposition 2 and Lemma 2.

A.3 Proof of Proposition 4 and Lemma 3

We consider the POP regulation, with $USO=U$. Recall that the incumbent announces two values $\underline{u}_I \geq 0$ and $\bar{u}_I \geq 0$. The profit of firm E is given by

$$\tilde{\pi}_E = \begin{cases} 0 & \text{ICS regime} \\ \underline{\alpha}(\bar{S}_E - \underline{u}_I) & \text{taxation regime} \\ \underline{\alpha}(\underline{S}_E - \underline{u}_I) + \bar{\alpha}(\bar{S}_E - \bar{u}_I) & \text{ECS regime.} \end{cases}$$

Firm E prefers ECS to TR if and only if

$$\bar{\alpha}\bar{u}_I \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E - \underline{\alpha}\bar{S}_E.$$

The preceding inequality is possible only if

$$\underline{\alpha}\bar{S}_E \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E$$

or, equivalently, $t \geq \tau_{\text{pop}}$. Therefore, if $t < \tau_{\text{pop}}$, the ECS regime cannot occur and the POP regulation leads to the same equilibria as the restricted-entry regulation.

For $t \geq \tau_{\text{pop}}$, the three regimes may appear, according to the value $\underline{u}_I, \bar{u}_I$:

$$\begin{aligned} \underline{\alpha}\underline{u}_I + \bar{\alpha}\bar{u}_I \geq \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E \text{ and } \underline{u}_I \geq \bar{S}_E &\implies \text{ICS} \\ \bar{\alpha}\bar{u}_I \geq \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E - \underline{\alpha}\bar{S}_E \text{ and } \underline{u}_I \leq \bar{S}_E &\implies \text{TR} \\ \bar{\alpha}\bar{u}_I \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E - \underline{\alpha}\bar{S}_E \text{ and } \underline{\alpha}\underline{u}_I + \bar{\alpha}\bar{u}_I \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E &\implies \text{ECS.} \end{aligned}$$

The profit function of the incumbent is

$$\tilde{\pi}_I = \begin{cases} \underline{\alpha}(\underline{S}_I - \underline{u}_I) + \bar{\alpha}(\bar{S}_I - \bar{u}_I) & \text{ICS regime} \\ \underline{\alpha}t\bar{q}_E + \bar{\alpha}(\bar{S}_I - \bar{u}_I) & \text{taxation regime} \\ 0 & \text{ECS regime.} \end{cases}$$

In the ICS regime, it is optimal for the incumbent to announce $\underline{u}_I, \bar{u}_I$ such that $\underline{\alpha}\underline{u}_I + \bar{\alpha}\bar{u}_I = \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E$. In the taxation regime, it is optimal for I to announce \bar{u}_I such that $\bar{\alpha}\bar{u}_I = \underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E - \underline{\alpha}\bar{S}_E$. In the ECS regime, all the possible values of $(\underline{u}_I, \bar{u}_I)$ lead to the same profit, namely $\tilde{\pi}_I = 0$. Finally firm I 's profit is

$$\tilde{\pi}_I = \max \left(\underline{\alpha}\underline{S}_I + \bar{\alpha}\bar{S}_I - \underline{\alpha}\underline{S}_E - \bar{\alpha}\bar{S}_E, \underline{\alpha}t\bar{q}_E + \bar{\alpha}\bar{S}_I - [\underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E - \underline{\alpha}\bar{S}_E], 0 \right),$$

which can be written as

$$\tilde{\pi}_I = \max \left(W_{II} - W_{EE}, \bar{W}_{EI} - W_{EE}, W_{EE} - W_{EE} \right).$$

A.4 Proof of Proposition 6

We consider the POP regulation with USO=UND. Recall that the incumbent announces one value u_I . The profit of the other firm is

$$\tilde{\pi}_E = \begin{cases} 0 & \text{ICS regime} \\ \underline{\alpha}(\tilde{S}_E - u_I) & \text{taxation regime} \\ \underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E - u_I & \text{ECS regime.} \end{cases}$$

Firm E prefers ECS to TR if and only if

$$\bar{\alpha}u_I \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E - \underline{\alpha}\tilde{S}_E,$$

which is possible only if

$$\underline{\alpha}\tilde{S}_E \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E$$

or, equivalently, $t \geq \tau_{\text{pop}}$. Therefore, if $t < \tau_{\text{pop}}$, the ECS regime cannot occur and the POP regulation leads to the same equilibria as the restricted-entry regulation.

For $\tau_{\text{pop}} \leq t \leq \bar{\tau}$, we have:

$$\begin{aligned} 0 \leq u_I \leq [\underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E - \underline{\alpha}\tilde{S}_E]/\bar{\alpha} &\implies \text{ECS} \\ (\underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E - \underline{\alpha}\tilde{S}_E)/\bar{\alpha} \leq u_I \leq \tilde{S}_E &\implies \text{TR} \\ \tilde{S}_E \leq u_I &\implies \text{ICS} \end{aligned}$$

The profit function of the incumbent is

$$\tilde{\pi}_I = \begin{cases} \underline{\alpha}\underline{S}_I + \bar{\alpha}\tilde{S}_I - u_I & \text{ICS regime} \\ \underline{\alpha}t\tilde{q}_E + \bar{\alpha}(\tilde{S}_I - u_I) & \text{taxation regime} \\ 0 & \text{ECS regime.} \end{cases}$$

The optimal announcement leading to TR is of course $u_I = [\underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E - \underline{\alpha}\tilde{S}_E]/\bar{\alpha}$ and the optimal announcement leading to ICS is $u_I = \tilde{S}_E$. This gives the profits

$$\tilde{\pi}_I = \max \left(\underline{\alpha}\underline{S}_I + \bar{\alpha}\tilde{S}_I - \tilde{S}_E, \underline{\alpha}t\tilde{q}_E + \bar{\alpha}\underline{S}_I - (\underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E - \underline{\alpha}\tilde{S}_E), 0 \right),$$

which can be written

$$\tilde{\pi}_I = \max \left(W_{II} - [\tilde{S}_E - \underline{\alpha}\underline{S}_E - \bar{\alpha}\tilde{S}_E] - W_{EE}, \widetilde{W}_{EI} - W_{EE}, W_{EE} - W_{EE} \right).$$

For $\bar{\tau} \leq t$, we have:

$$\begin{aligned} 0 \leq u_I \leq \underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E &\implies \text{ECS} \\ \underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E \leq u_I &\implies \text{ICS.} \end{aligned}$$

The optimal announcement in ICS is of course $\underline{\alpha}\underline{S}_E + \bar{\alpha}\tilde{S}_E$, which gives the profit

$$\tilde{\pi}_I = \max \left(\underline{\alpha}\underline{S}_I + \bar{\alpha}\tilde{S}_I - \underline{\alpha}\underline{S}_E - \bar{\alpha}\tilde{S}_E, 0 \right).$$

which can be written

$$\tilde{\pi}_I = \max \left(W_{II} - W_{EE}, W_{EE} - W_{EE} \right).$$

A.5 Proof of Proposition 7

(i) We first show that for each $(\underline{S}_E, \bar{S}_E)$ there exists $(\underline{S}_E, \bar{S}_I)$ such that the conditions (2) are satisfied. The third inequality of (2) is obtained by taking \underline{S}_I large enough. Now we can choose \bar{S}_I to ensure the second condition of (2).

Since $\bar{W}_{EI} < W_{II} - \bar{\alpha}\bar{S}_E$, the restricted-entry regulation leads to ICS regime. By using assumptions (2), we obtain

$$W_{EE} > W_{II} - [\bar{S}_E - (\underline{\alpha}\underline{S}_E + \bar{\alpha}\bar{S}_E)] > W_{II} - \bar{\alpha}\bar{S}_E > \bar{W}_{EI}.$$

It follows that the POP regulation leads to ECS regime (see Proposition 6).

(ii) The proof is straightforward.

A.6 Imposing ND alone under the restricted-entry regulation

In this section, we assume: $\bar{S}_E < 0 < \bar{S}_I$. The incumbent can profitably serve the high cost consumers. We assume that the regulator imposes only the non discrimination constraint and that firm E cannot serve $\bar{\mu}$ (restricted-entry regulation). Firm E cannot be constrained by ND, since it never serves consumer $\bar{\mu}$. Firm I is compensated for the constraint ND by a tax t .

Lemma 5 (Perfect Equilibria in the restricted-entry game, USO=ND) *We make the following assumptions*

$$\begin{aligned} \bar{S}_E < 0 < \bar{S}_I \\ \underline{\alpha}\underline{S}_I > \bar{\alpha}\bar{S}_I + \underline{\alpha}\underline{S}_E \\ \bar{S}_I < \bar{S}_E. \end{aligned}$$

Then there exists a unique subgame perfect equilibrium in which firm I serves only $\underline{\mu}$ (giving them $\underline{u}_I = \bar{S}_E$), $\bar{\mu}$ consumers being then excluded from the market.

Although I would serve both types of consumers in the benchmark case, imposing ND alone has the countervailing effect of excluding high costs consumers from the market: firm I prefers to disconnect (or not to connect) these consumers rather than incurring losses due to cross-subsidization between both classes of consumers. This should lead the regulator to impose both constraints simultaneously.

Proof of Lemma 5

The possible regimes are ICS, TR and the exclusion regime, where $\underline{\mu}$ is served by I and $\bar{\mu}$ is not served. Firm E 's profit function is

$$\tilde{\pi}_E = \begin{cases} \underline{\alpha}(\bar{S}_E - \underline{u}) & \text{taxation regime} \\ 0 & \text{ICS regime} \\ 0 & \text{exclusion regime (firm I),} \end{cases}$$

It follows that the profit of I is

$$\tilde{\pi}_I = \begin{cases} \bar{\alpha}(\bar{S}_I - \bar{u}_I) + \underline{\alpha}t\tilde{q}_E & \text{taxation regime} \\ \underline{\alpha}\bar{S}_I + \bar{\alpha}\bar{S}_I - \underline{u} & \text{ICS regime} \\ \underline{\alpha}(\underline{S}_I - \underline{u}) & \text{exclusion regime (firm I)}. \end{cases}$$

Firm I 's optimal announcement in the taxation regime is of course 0, while it is \bar{S}_E in the ICS and exclusion regimes. Therefore firm I maximizes

$$\tilde{\pi}_I = \max[\bar{\alpha}\bar{S}_I + \underline{\alpha}t\tilde{q}_E, \underline{\alpha}\bar{S}_I + \bar{\alpha}\bar{S}_I - \bar{S}_E, \underline{\alpha}(\underline{S}_I - \bar{S}_E)].$$

The assumptions of the lemma ensure that the maximum is $\underline{\alpha}(\underline{S}_I - \bar{S}_E)$ (exclusion regime).