Ecole des Hautes Etudes en Sciences Sociales
CIRED

THÈSE

Pour l'obtention du grade de
Docteur de l'Ecole des Hautes Etudes en Sociales Sociales
Discipline : Sciences Economiques

Concurrence Oligopolistique et Investissement :

application aux marchés électriques

Guy Meunier

Date de soutenance prévue : le 11 12 2008

Directeur de thèse : Dominique Finon

JURY :

Rabah Amir	Professeur d'Economie à l'Université d'Arizona
Claude Crampes	Professeur d'économie à l'Université de Toulouse I (rapporteur)
Anna Creti	Professeur d'économie à l'Université de Bocconi MILAN
Dominique Finon	Directeur de recherche au CNRS, CIRED
Jean-Charles Hourcade	Directeur d'étude de l'EHESS, CIRED
Jean-Pierre Ponssard	Directeur de recherche au CNRS, Ecole Polytechnique (rapporteur)
Jean-Michel Trochet	Senior économiste, direction de la stratégie EdF

Remerciements

Je remercie toutes les personnes que j'ai rencontrées grâce à cette thèse au CIRED, à EdF, au LARSEN, au CECO, à Cambridge, pour ce qu'elles m'ont apporté humainement et scientifiquement.

Je remercie Dominique Finon pour avoir dirigé cette thèse, pour son soutien et son savoir qu'il m'a transmis avec passion. Sa très grande connaissance des systèmes électriques et son constant soucis de pertinence lui ont toujours donné un regard critique et constructif sur mes travaux.

Je remercie Christophe Defeuilley, amorce du projet, pour m'avoir fait confiance, pour ses critiques et son soutien.

Je remercie Jean-Charles Hourcade pour m'avoir accueilli au CIRED, pour sa fougue et son enthousiasme.

Je remercie Jean-Pierre Ponssard pour l'accueil dont j'ai bénéficié au CECO la dernière année. Il a su me faire diplomatiquement de dures critiques qui m'ont permis de progresser.

Je remercie la « Task Force » : Damien Demailly et François Gusdorf pour leur amitié. Leur humour et leur pertinence font de chacune de nos rencontres un instant savoureux et inoubliable.

Au CIRED, je remercie plus particulièrement Philippe, Sandrine, Tarik et Henry, pour les fructueuses conversations que nous avons eu dans un local café où la liberté d'expression est reine.

Enfin, je remercie Fanny pour avoir rempli ces années de bonheur.

Table des matières

Table des figures

Introduction

Depuis la fin du XIXème siècle et la mise en place des premiers systèmes électriques, cette forme d'énergie a joué un rôle majeur dans la seconde révolution industrielle et l'a maintenu par la suite. L'électricité est aujourd'hui un élément essentiel de l'activité économique et du mode de vie des pays industrialisés. Cette importance et les caractéristiques physiques de l'électricité expliquent que l'organisation de sa production soit sensible et complexe. Engagée à la fin du XXème siècle, la libéralisation de ce secteur est devenue sujet d'attentes et de craintes. L'effet des nouveaux modes d'organisation sur les investissements en capacités de production demeure l'une des questions essentielles et débattues.

La théorie économique explique comment des marchés de l'électricité de gros et de détail prenant en compte les caractéristiques physiques de ce bien doivent permettre une coordination optimale des décisions de production, d'investissement, de transport et de consommation via un système de prix. Cependant, la concentration industrielle observée sur ces marchés en Europe limite la pertinence de ce cadre et justifie le développement d'analyses théoriques de la dimension stratégique des choix de production et d'investissement des producteurs. L'objectif de ce travail de thèse est de contribuer à cette réflexion en analysant les stratégies d'investissement de firmes en situation d'oligopole. Avant de présenter les chapitres qui forment le corps de la thèse, la présente introduction précise ces enjeux. Après d'une brève description des caractéristiques physiques d'un système électrique, les réformes sont décrites avant de définir la problématique de l'investissement en production électrique dans des marchés en concurrence imparfaite.

1 Les spécificités de l'offre et la demande

Bien que les exercices de modélisation développés dans la thèse fassent abstraction de caractéristiques physiques importantes de l'électricité, il convient de présenter celles-ci pour apprécier la difficulté de l'introduction des marchés et les raisons pour lesquelles le secteur est caractérisé par un emboitement étroit d'activités coordonnées par des marchés à côté d'activités qui doivent rester coordonnées de façon autoritaire.

Les caractéristiques fondamentales de l'électricité - équilibre en chaque instant de la production et de la consommation et impossibilité de contrôle des flux- sont présentées avant les particularités de la consommation et de la production.

1.1 L'équilibre

Un système électrique est composé de trois activité physiques la production, le transport/distribution par réseau fixe [3], la coordination des flux et la consommation. Les centrales de production transforment de l'énergie primaire en énergie électrique injectée dans les réseaux de transport et de distribution que les consommateurs soutirent de ces réseaux.

Le système est en équilibre permanent car l'ensemble de ses éléments est synchronisé : les alternateurs tournent tous à la même fréquence. Un écart entre la puissance appelée et la puissance produite, c'est-à-dire entre consommation et production [4], modifie la fréquence du système ce qui peu endommager des composants et entraîner des incidents en chaîne jusqu'à l'effondrement du système. A chaque instant, le système fonctionne seulement si la quantité d'électricité consommée est égale à la quantité d'électricité produite. En général, cet équilibre est assuré par

3. Le transport se distingue de la distribution par la tension de l'électricité transportée. Le transport par haute tension coûte plus cher mais permet de limiter les pertes par effet joule.

4. La puissance est un flux d'énergie : une quantité d'énergie par unité de temps. Ce qui s'échange via le réseau est une puissance en chaque instant. Pour une unité de temps fixée la puissance moyenne et l'énergie sont identique.

l'ajustement de la production mais lorsqu'il n'y a pas assez d'unités de production disponibles il faut rationner certains consommateurs en les déconnectant temporairement du réseau pour maintenir le fonctionnement général.

En plus du problème de l'équilibre permanent s'ajoute celui des contraintes de réseaux et l'impossibilité de contrôler les chemins des flux. Les flux se répartissent dans le réseau selon les lois de Kirchhoff (lois des mailles et lois des noeuds). A ces lois physiques s'ajoutent les contraintes de capacités des lignes. Chaque ligne a une capacité limitée qui ne doit pas être dépassée, sinon elle se déconnecte ce qui peut provoquer l'effondrement du système. C'est en jouant sur les injections et soutirages qu'il est possible d'éviter ce type d'événement.

Ces caractéristiques des systèmes électriques sont au fondement des difficultés d'organisation des échanges. Elles sont d'autant plus contraignantes qu'elles se combinent avec d'autres telles les variations de la demande, l'absence de transmission de prix instantané au consommateur, la non stockabilité et des contraintes de capacités fortes.

1.2 La demande

La consommation d'électricité est variable et incertaine. Elle varie continuellement au cours du temps, ses variations étant caractérisées par des périodicités journalières, hebdomadaires et annuelles. La figure 1 représente les variations au cours de la journée du mercredi 9 mai 2007 en France. L'amplitude des variations est importante, de l'ordre de 35 % par exemple pour la journée représentée sur la figure 1. Elle est fortement liée à la température et à l'activité économique : selon le Réseau de Transport d'Electricité, une variation d'un degré en dessous de zéro en hiver sur l'ensemble du territoire entraine une variation de la demande correspondant à la puissance d'une centrale nucléaire (1,5GW).
La quantité de puissance appelée est dérivée de la demande de services énergétiques des différents agents économiques. Ainsi, les variations de la consommation s'expliquent par celles de cette demande qui se réper-

cutent d'autant plus qu'il n'y a pas de moyens de stockage [5] et que les prix contractuels sont rigides. En effet, l'essentiel des consommateurs payent un prix de l'électricité qui est fixe sur des périodes plurimensuelles importantes par rapport aux durées de variations des conditions d'offre et de demande.

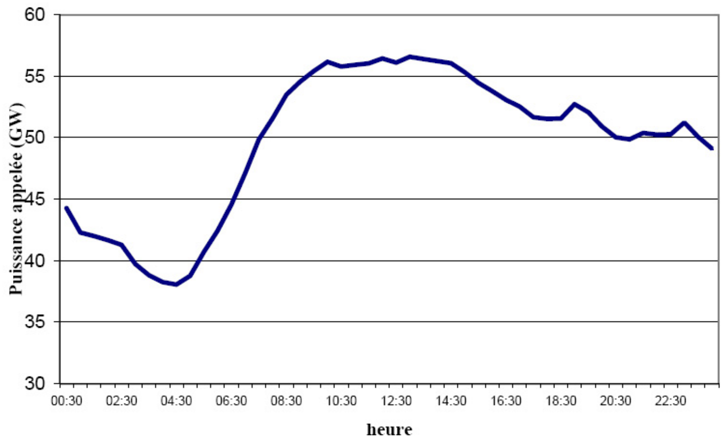

FIGURE 1 – Variation de la consommation le 9 mai 2007 [6]

Il est important de prévoir la demande d'électricité pour assurer l'équilibre du système. Les variations peuvent se décomposer en une part anticipée et une part incertaine. Au fur et à mesure que l'échéance s'approche l'incertitude est moindre, cependant même quelques heures avant l'échange physique des incertitudes persistent. Ces incertitudes sont problématiques en raison du besoin d'équilibre en temps réel et de la non stockabilité, car il est alors necessaire que des capacités soient disponibles pour faire face aux aléas de consommation.

1.3 La fonction de production d'électricité.

La production d'électricité se caractérise par la diversité des équipements utilisés. Il n'existe pas de technologie de référence mais un panel de technologies. Chaque technologie possède des caractéristiques techniques propres qui justifient son utilisation. Certaines caractéristiques

5. Le cas de l'essence illustre bien comment la présence de stocks individuels modifie la situation en lissant les variations de la demande de transport

comme la rapidité de mise en service et de montée en puissance sont pertinentes pour la gestion de l'équilibre sur le court terme et en temps réel. Pour l'analyse des stratégies d'investissement, on peut néanmoins se contenter de décrire une technologie par sa structure de coûts : ses coûts variables et ses coûts de capacités.

Toutes les technologies de production ont des contraintes de capacité fortes. Sur le moyen terme, une fois la centrale construite et dimensionnée, elle ne peut pas produire plus qu'une quantité fixée pendant un intervalle de temps donné : sa capacité de puissance. Sur le long terme, cette capacité est variable. Les délais de construction et les durées de vie des centrales sont élevés par rapport aux périodes de variation de la demande. Par exemple il faut 2 à 3 ans pour construire une centrale à cycle combiné à gaz (CCGT) et 6 à 8 pour une centrale nucléaire pour des durées de vie d'environ 25 et 60 ans respectivement.

Comme l'échelle de temps de variation de la demande est extrêmement courte par rapport à la durée de vie des centrales, il n'est pas possible d'adapter le parc de production à la demande révélée à chaque instant. Cela explique que plusieurs technologies de structure de coût différentes soient utilisées pour produire de l'électricité. Chaque technologie ne produit que pendant une fraction de l'année. Une fois l'investissement réalisé cette fraction est déterminée par les quantités de capacités installées de chaque technologie, leurs coûts variables et les variations de la demande.

Le coût d'une capacité peut se décomposer en son coût d'investissement et un coût fixe d'exploitation. Le prix des facteurs fixes nécessaires à la construction d'une centrale déterminent le coût d'investissement. Les frais d'entretien et de personnels sont des coûts fixes d'exploitation proportionnels à la quantité de capacité et non à la production réalisée. Le prix des facteurs variables, notamment le combustible, déterminent le coût variable. Le coût de capacité est dépensé même si la centrale ne produit pas alors que le coût variable n'est payé que lorsque la centrale produit [7].

Une centrale de base est caractérisée par un ratio coût variable/coût de capacité faible alors qu'une centrale de pointe est caractérisée par un ratio très élevé. Le coût total de production d'un MWh avec une technologie dépend du profil de production, c'est à dire de son facteur de charge au cours de l'ensemble de ses années d'utilisation. Une centrale de base est moins coûteuse pour produire tout au long de l'année alors qu'une centrale de pointe l'est pour une production durant quelques heures. Le tableau 1 compare les principales caractéristiques de quatre technologies classées par coût de capacité décroissant.

Caractéristiques	**Nucléaire** E.P.R.	**Charbon** pulvérisé	**CCGT**	**TAC**
Durée de vie	60 ans	35 ans	25 ans	25 ans
Délai de construction	6 ans	3 ans	2 ans	2 ans
Coût de capacité €/kW/an	200	140	65	40
Dont investissement	150	110	45	27
Dont exploitation fixe	50	30	30	13
Coût variable €/MWh	5	30	46	50
Dont combustible	4,4	11	34	30
Dont CO_2	0	15	7	10

TABLE 1 – Comparaison des technologies de production [8].

De plus, étant donné la complexité de certaines technologie, comme le nucléaire, la structure de coût de ces technologies et notamment le coût d'investissement, peut dépendre de la quantité construite. C'est-à-dire que certaines technologies présentent des économies de série expliquées

7. Cette distinction entre coûts fixes de capacité et coûts variables peut être nuancées puisque les coûts fixes d'exploitation sont des coûts « intermédiaires »qui sont variables sur le moyen terme car il est possible une fois la centrale construire de ne pas l'utiliser quelques années

8. Les calculs ont été réalisés avec un taux d'actualisation de 8% et les estimations de coûts de la DGEMP (2003) dont un prix du gaz de 4€/Mbtu, du charbon de 40€/t du CO_2 de 20€/t.

par des rendements croissants. La technologie nucléaire est la technologie de base « ultime »caractérisée par d'importants coûts fixes et des coûts opérationnels faibles. Cette technologie présente des rendements croissants puisque le coût d'une capacité d'une filière, par exemple EPR, décroit avec la capacité totale installée (DGEMP 2003). En comparaison, les centrales à cycles combinés à gaz sont des centrales de semi-base caractérisées par leur petite taille et des rendements constants.

2 La libéralisation

Avant de préciser les principes des réformes des marchés de l'électricité, un rapide survol historique permet de les mettre en perspective. Le mode d'organisation antérieur, celui de monopoles intégrés et régulés voir publics, a été mis en place progressivement alors que l'énergie électrique devenait un élément de plus en plus important de développement économique et social.

2.1 Perspective historique [9]

A la fin du XIXeme siècle, d'importantes découvertes (lampes à incandescence, électrolyse..) amorcent le développement des premiers systèmes électriques. Initialement, la production électrique est le fait de nombreuses firmes opérant des petits systèmes locaux. C'est le développement du transport d'électricité sur des distances de plus en plus élevées qui rythme l'intégration territoriale progressive des firmes dans la première moitié du XXème siècle. Le développement du réseau entraîne d'importantes baisses des coûts en permettant de profiter de l'effet de foisonnement des courbes demande [10], en reliant les grands bassins hydrauliques des zones de consommation et en augmentant la taille des unités de production. En parallèle, les techniques de production connaissent un progrès très rapide caractérisé par des effets de taille et de rendement

9. Pour une description détaillée de l'évolution des systèmes électriques dans plusieurs pays industrialisés se référer à l'ouvrage collectif *Entre monopole et concurrence* coordonné par Stoffaës (1994).

10. En connectant deux groupes de consommateurs la somme de leur demande a un profil plus plat, ce qui permet de baisser le coût total pour les servir.

amélioré.

La structure industrielle et la réglementation du secteur se développent parallèlement au réseau. Les firmes privées productrices d'électricité se concentrent par fusions et acquisitions suivant le développement du réseau pour élargir et homogénéiser leurs monopoles géographiques et développer des unités de production de plus en plus importantes. Très critiquée, en raison notamment de l'hétérogénéité des prix pratiqués, cette situation entraîne une extension du contrôle public sur la tarification et l'investissement. Après la seconde guerre mondiale, dans l'ensemble des pays industrialisés l'intégralité de la chaine est réglementée. Les formes précises de cette réglementation sont très variées (variété qui se retrouve dans les structures industrielles mises en place). La France, avec un seul monopole verticalement intégré sur l'ensemble du système, est plus l'exception que la règle. Dans la plupart des pays, plusieurs firmes régionales, souvent de propriété privée ou mixte, sont présentes. Le dénominateur commun est l'existence de monopoles verticalement intégrés en production - transport et un contrôle des tarifs et des décisions d'investissements.

Ainsi, alors qu'initialement développée sur un mode privé et parfois concurrentiel, la production d'électricité a été réglementée via le contrôle de monopoles verticalement intégrés sur de plus ou moins vastes zones géographiques. Les caractéristiques techniques de l'électricité, son rôle comme élément essentiel de développement économique et social et la nécessité de protection des consommateurs expliquent le choix de ce mode d'organisation. L'organisation intégrée permet de gérer la coordination entre production et transport sur le court comme sur le long terme. Cette coordination ainsi que la présence de rendements croissants non seulement dans les activités de transport mais aussi dans la production détermine le caractère de monopole naturel de l'activité intégrée production-transport. En effet, les centrales de production, notamment nucléaires et au charbon, présentent des tailles importantes par rapport aux marchés concernés et leurs coûts de capacités sont décroissants avec le nombre de centrales.

Au cours des années 70s, l'évolution des techniques et des concepts entraînent une réévaluation de ce mode d'organisation. Les expériences de libéralisation d'autres activités de réseaux (transports routiers, aériens et télécommunications) sont évaluées comme des réussites, tandis que l'efficacité des monopoles électriques publiques et du mode de réglementation, notamment des entreprises ('utilities') américaines, est critiquée. Cette critique s'accompagne de développements techniques favorables à la mise en place de marchés (Finon 1997) : les technologies de l'information et de la communication rendent possible la mesure précise des flux des transactions bilatérales, tandis que se sont développés les cycles combinés à gaz (CCGT) [11]. Le premier progrès permet de gérer les transactions entre acteurs bien que les transferts physiques d'électricité ne puissent être directement contrôlés. Le second progrès remet en question l'intégration de l'activité de production dans le statut de monopole naturel de l'activité intégrée production - transport. Les CCGTs sont standardisés, de petite taille et dépourvus d'économie d'échelle ce qui permet de faciliter la concurrence dans la production d'électricité par des entrées (Wilson 2002, Newbery 1998). Seule l'activité de transport et de coordination des flux est encore considérée comme un monopole naturel.

2.2 Le principe des réformes

> « *This process is usually viewed as replacing tight regulation of vertically integrated monopolies with light regulation of functionally specialized firms and supervision of competitive markets.* »(Wilson 2002)

Les activités de transport ayant conservé leur statut de monopole naturel, les réformes consistent à séparer les activités concurrentielles et de transport pour rendre l'accès au réseau ouvert et non discriminatoire et

11. La relation entre l'introduction de concurrence et le développement des CCGTs est un peu moins unilatérale puisque ce développement technique a été favorisé par l'introduction aux Etats Unis dans les années 70 de l'obligation pour les 'utilities' d'acheter l'électricité de producteurs indépendants.

de supprimer (en grande partie) l'intégration entre production et commercialisation [12]. Un marché de gros de l'électricité est mis en place, composé de plusieurs sous-marchés : une bourse ou marché organisé et un marché bilatéral. Ce marché de gros s'articule avec un marché de détail. Ces marchés peuvent eux mêmes se décomposer selon l'échéance temporelle puisque des contrats à terme y sont échangés. Les acteurs qui opèrent sur ces marchés sont soient des acteurs physiques : producteurs et gros consommateurs industriels, ou des intermédiaires : fournisseurs et négociants. Sur les marchés organisés les contrats les plus fragmentés concernent l'échange d'électricité pour une heure ou une demi-heure selon les pays la veille pour le lendemain.

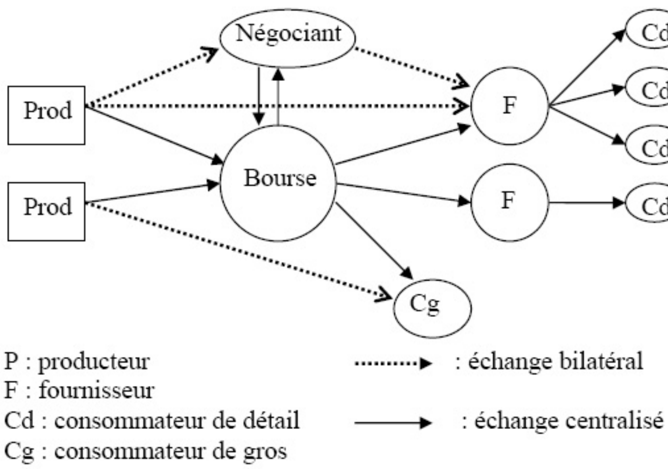

FIGURE 2 – Marchés de gros et de détail de l'électricité

La persistance, sur le très court terme, d'incertitudes sur les conditions d'offre (défaillance d'équipement), de demande (aléa météo) et de transport crée un risque d'effondrement du réseau. Les marchés ne permettant pas une coordination suffisamment rapide et complète des décisions de production et de consommation pour assurer l'équilibre en temps réel, cet équilibre est assuré par le gestionnaire du réseau de transport

12. Parfois appelée fourniture, cette activité consiste à acheter de l'électricité sur le marché de gros pour la revendre sur le marché de détail.

(GRT). Ainsi, sur le très court terme la main est passée au GRT qui assure la coordination et l'équilibre des flux physiques qui concrétisent les transactions commerciales antérieures. Des capacités de réserve doivent être mises à disposition du GRT afin qu'il corrige les éventuels écarts par rapport à ces transactions. Le caractère de bien public de ces réserves opérationnelles et du rôle du gestionnaire de réseau explique le développement de mécanismes plus ou moins marchands en plus du marché de l'énergie pour assurer la fourniture de ce service de sécurité de fourniture en temps réel. Pour assurer l'équilibre en temps réel, il existe des marchés ou mécanismes d'ajustement permettant de définir un prix en temps réel sur la base d'algorithmes et d'enchères plus ou moins complexes (car devant refléter divers caractéristiques techniques du réseau et des centrales) soumises par les producteurs et les consommateurs. Et des schémas d'incitation à l'investissement sont ajoutés au marché de l'énergie (Finon and Pignon 2008). L'étendue du contrôle du gestionnaire de réseau et l'articulation entre marchés et procédures autoritaires dépend des pays.

2.3 La mise en oeuvre des réformes en Europe

Les principes des réformes sont fixés par des directives européennes : indépendance des GRTs, accès des tiers au réseau, ouverture de la concurrence de détail, instauration d'une autorité indépendante. Les pays conservent une certaine latitude dans la mise en place des marchés et dans les choix de dé-intégration verticale et horizontale.

Les pays européens peuvent être séparés en deux groupes quant à leur attitude par rapport aux anciens monopoles. Au moment de la mise en place des marchés, certains pays forcent le développement de la concurrence en imposant des scissions (Royaume-Uni) ou des cessions d'actifs (Italie) aux anciennes entreprises intégrées. Par contre, la plupart des pays d'Europe continentale conservent la structure industrielle existante tant verticale qu'horizontale. Parmi ces pays, ces structures diffèrent. Par exemple, en Allemagne huit entreprises dominaient la production et le transport et la distribution était le fait de nombreuses régies locales alors qu'en France Electricité de France était en situation de quasi-monopole sur l'intégralité du territoire et reste intégrée verticalement le long de la chaîne.

Même si les structures industrielles étaient initialement variées en Europe, les fusions et acquisitions qui ont suivi la mise en place des marchés et les privatisations ont uniformisé les situations. Là où il y avait une dé-intégration radicale, une ré-intégration verticale s'est opérée comme au Royaume-Uni. Ainsi, chaque marché de l'électricité national présente aujourd'hui une structure très concentrée. De plus, la plupart des anciens monopoles nationaux s'étant internationalisés, cette concentration se retrouve à l'échelle européenne. Cette dynamique s'illustre bien par les cas du Royaume-Uni et de l'Allemagne. La production au Royaume-Uni est actuellement dominée par cinq firmes dont quatre étrangères (EON, EdF, RWE et Iberdrola), et la production en Allemagne par quatre firmes dont deux étrangères (Vattenfall et EdF).

3 Investissement et pouvoir de marché

Au moment de la mise en place des marchés de l'électricité, la plupart des systèmes concernés sont en situation de capacité excédentaire [13]. Le renouvellement des capacités de production n'étant pas à l'ordre du jour au moment de l'introduction des marchés, les efforts des concepteurs des réformes se focalisent sur leur fonctionnement sur le court terme. Il s'agit de mettre en place des marchés permettant l'utilisation la plus efficace des moyens de production disponibles en faisant confiance aux signaux prix des marchés de court terme pour inciter aux investissements et parvenir au mix technologique optimal.

Ainsi, dans un premier temps, la plupart des préoccupations autant pratiques que théoriques portent sur l'effet des différentes règles sur les

13. Cette situation s'explique par la baisse de la croissance de la demande dans les années 70s et l'inertie des monopoles réglementés. Cette situation est l'une des critiques des monopoles réglementés et c'est notamment pour éviter qu'elle se reproduise que les réformes ont été entreprises.

décisions de productions à capacités fixées. L'exercice du pouvoir de marché des producteurs par la manipulation de ces règles est le sujet central des analyses empiriques et théoriques. L'analyse des investissements et des risques d'exercice de pouvoir de marché de long terme se développe dans un second temps.

Avant de préciser les problématiques liées au pouvoir de marché, dans une perspective de long terme le fonctionnement d'un marché parfaitement concurrentiel est présenté ainsi que les autres explications de la difficulté des marchés à promouvoir un investissement optimal.

3.1 Concurrence et investissement optimal

Tout d'abord, il convient de souligner que les modèles théoriques développés pour l'analyse des choix d'investissements et de tarification des monopoles réglementés [14] s'adaptent directement pour décrire le fonctionnement d'un marché en concurrence pure et parfaite. Le système de prix de gros qui permet de décentraliser l'investissement dans le régime de marché correspond aux valeurs des contraintes de capacité dans la modélisation du monopole réglementé.

Le modèle théorique d'un marché concurrentiel explique comment l'introduction d'un marché de gros peut aboutir au fonctionnement optimal du secteur de l'électricité grâce à la coordination par les prix [15]. Dans le cadre de ce modèle, les prix permettent aux producteurs, firmes « price takers », de réaliser des décisions optimales sur le court et le long terme. Le choix des facteurs fixes s'appuie sur la valorisation de la production sur le court terme. Par le jeu des offres sur les marchés horaires, les différentes technologies sont utilisées sur le court terme dans l'ordre de mérite, par coût variable croissant. Une courbe d'offre horaire en escalier est construite avec les coûts variables de production et la capacité disponible de chaque technologie. Le prix de gros de l'électricité

14. Cette littérature dite du « peak load pricing »a été notamment développée par Boiteux (1960). Une revue de cette littérature est réalisée par Crew et al. (1995).

15. Une description plus détaillée est faite dans le premier chapitre et par Joskow and Tirole (2007), ou Green (2006).

est à l'intersection de cette courbe d'offre et de la courbe de demande. Ainsi, le prix de gros est le coût variable de la dernière technologie appelée lorsque la contrainte de capacité agrégée n'est pas saturée. Lorsque toutes les capacités disponibles sont utilisées, le prix est déterminé par la demande et la quantité de capacité totale. Lorsque la demande excède les capacités et que les consommateurs ne sont pas confrontés au prix de gros, c'est le régulateur qui doit fixer le prix à la valeur du dernier kwh qui n'est pas consommé : la VOLL (Value of Lost Load). Ainsi, sur le court terme les producteurs perçoivent des rentes inframarginales : différence entre les prix de gros et les coûts variables de leurs unités de production. Sur le long terme, les producteurs anticipent les prix de gros et choisissent les quantités de capacités de chaque technologie sur la base de ces anticipations. A l'équilibre, les quantités de capacités de chaque technologie sont déterminées de telle sorte que les revenus obtenus sur le court terme, les rentes inframaringales, égalisent les coûts d'investissement.

La forte variabilité de la demande d'électricité se reflète dans la forte volatilité des prix. L'absence de moyen de stockage se traduit par une forte différenciation temporel du bien ; un watt durant une heure particulière est peu substituable avec un watt durant l'heure suivante. Cette faible susbtituabilité est exacerbée par la faible élasticité des demandes horaires de gros qui est liée à l'absence de transmission des prix de gros dans les prix de détail. Les prix de gros de l'électricité sont ainsi très variables sur l'année comme sur la journée. La figure (3) représente les moyennes journalières des prix horaires aux cours de l'année 2007, et les prix horaires du 22 octobre de cette même année. Ainsi, le coût d'investissement d'une unité de pointe n'est amorti que pendant quelques heures par an où les prix sont très élevés, et les producteurs doivent anticipés ces prix ainsi que les durées d'appel.

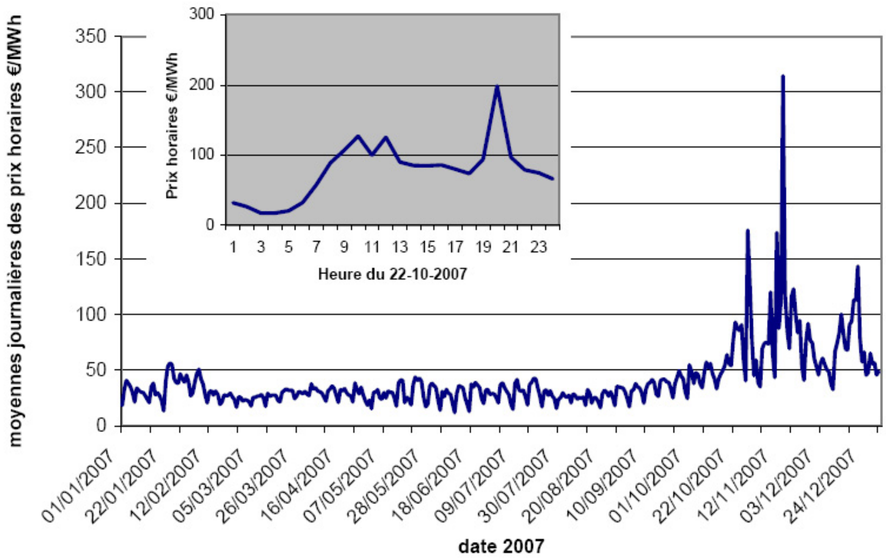

FIGURE 3 – Prix de l'électricité la veille pour le lendemain au cours de l'année 2007 [17]

3.2 Risque et « missing money »

Dans le cadre de l'organisation précédente, les décisions d'investissements sont prises au sein de l'entreprise verticalement intégrées anticipant l'évolution de la demande sur son aire de monopole de desserte. Le risque sur le niveau de celle ci, et celui sur le prix des combustibles, ne sont pas supportés par la firme car le tarif réglementé en « cost plus »assure la rentabilité de leurs investissements même si ceux-ci sont excessifs. Avec le nouveau mode d'organisation les décisions d'investissements sont décentralisées via le système de prix de gros. Ces prix, donc les revenus qu'ils assurent, sont incertains, et doivent être anticipés par les producteurs et les fournisseurs. L'incertitude sur les prix peut être intégrée dans le cadre exposé précédemment, le résultat théorique sur l'optimalité de l'équilibre des marchés reste valable à condition que les marchés soient complets [18] et que les agents aient des anticipations rationnelles. De telles conditions ne se retrouvent sur aucun marché, mais il semble que cela soit plus problématique dans le cas de l'électricité. Le problème

17. A partir d'historiques de prix de powernext : www.powernext.fr.
18. C'est-à-dire qu'il existe un prix de l'électricité pour chaque instant et contingence.

de la gestion du risque, de l'incomplétude des marchés de l'électricité et du capital sont l'une des explications d'une sous-optimalité de l'investissement en capacité et en mix technologique (Chao et al. 2005, Finon 2008). De plus, la structure de coût des différentes technologies et les relations entre le prix de l'électricité et le prix des combustibles implique des profils de risque différents selon les technologies, ce qui semble pénaliser les technologies de base efficaces mais intensives en capital (Finon and Roques 2008).

En plus des problèmes de gestions de risques, le caractère de bien public des réserves opérationnelles et l'utilisation trop fréquente de celles-ci par le GRT et d'autres services auxilliaires (qui limite la hausse des prix sur le marché de gros), sont considérées, notamment aux Etats-Unis, comme source de baisse des incitations à l'investissement en unités de pointe en érodant les revenus pendant les périodes de tensions entre l'offre et la demande (Joskow 2006, Joskow and Tirole 2006). A cela s'ajoute l'existence de prix plafonds ; mis en place pour limiter les pics de prix observés pour des raisons d'acceptabilité des réformes et de limite du pouvoir de marché de court terme, ceux-ci peuvent en limitant les revenus des producteurs entraîner un déficit d'investissement en capacités de pointe.

3.3 Pouvoir de marché

L'une des hypothèses fondamentales du modèle de référence est celle du comportement de "preneurs de prix" des firmes. Dans ce cadre, les firmes maximisent leur profit en considérant comme exogènes et fixes les prix de l'électricité. Cette hypothèse peut se justifier si les firmes sont nombreuses car l'influence de chacune d'entre elles sur le prix est alors négligeable. A l'inverse, lorsque peu de firmes sont présentes sur un marché, elles ont un « pouvoir de marché »et sont capables d'influencer le prix. L'analyse de situation de la concurrence imparfaite permet de mieux rendre compte de la réalité de certains marchés dominés par quelques firmes, ce qui est le cas de la plupart des marchés de l'électricité.

Une firme est capable d'influencer le prix soit directement en fixant celui-ci, soit indirectement en limitant les quantités de bien disponibles. La concurrence et la réaction des consommateurs limitent la capacité d'une firme à faire monter les prix car une hausse de ces prix entraine une baisse de la consommation et une hausse de la production des rivales. Sur les marchés de l'électricité, étant donné la faible élasticité de la demande et de l'offre, les firmes, mêmes petites, disposent d'un important pouvoir de marché en situation de tension entre l'offre et la demande. Plusieurs études empiriques concluent à l'exercice de pouvoir de marché sur le court terme sur les marchés d'Angleterre et du Pays de Galle (Wolfram 1999, Wolak and Patrick 2001, Green and Newbery 1992) et de Californie (Joskow and Kahn 2001, Borenstein et al. 1995).

La plupart des analyses ont initialement porté sur le design des marchés et la manipulation des règles par les producteurs. Plusieurs approches analytiques sont utilisées pour représenter le pouvoir de marché sur le court terme. Peu de temps aprés l'introduction du marché de gros d'Angleterre et du pays de Galles, Green and Newbery (1992) et von der Fehr and Harbord (1993) analysent le mécanisme d'enchères utilisé et les choix d'offres stratégiques par les producteurs [19]. Crampes and Creti (2003) analysent en détail la stratégie de rétention de capacité [20] qui selon Joskow and Kahn (2001) est l'un des ingrédients de la crise californienne.

L'exercice de pouvoir de marché peut s'étendre dans le temps jusqu'au choix des capacités de production. Le choix de long terme d'une quantité totale de capacité possède des dimensions stratégiques similaires au choix d'une quantité produite sur le court terme : une firme a intérêt à limiter sa capacité de façon à augmenter les prix pendant les périodes où

19. Green and Newbery (1992) utilisent le modèle de Klemperer and Meyer (1989) dans lequel les producteurs choisissent des courbes d'offre continues avant que l'incertitude sur la demande ne soit résolue et que le prix n'équilibre le marché alors que von der Fehr and Harbord (1993) utilisent un modèle d'enchères discrètes. Ces deux modèles entrainent des résultats qualitativement différents (Fabra et al. 2006)

20. La rétention de capacité « capacity withholding »consiste pour les producteurs à ne pas déclarer comme disponibles certaines de leurs centrales (pour des raisons de maintenance par exemple) au GRT de façon à faire monter les prix de gros

cette capacité détermine le prix de gros (lorsque la demande est élevée). Dans le cas des marchés de l'électricité, il y a non seulement l'enjeu de la capacité totale disponible, mais aussi celui du mix de technologies. Le pouvoir de marché peut non seulement expliquer un sous-investissement global, mais aussi une mauvaise répartition de la capacité totale entre les différents types de technologies.

Une revue détaillée de la littérature sur les choix de capacité avec demande variable est réalisée dans le premier chapitre, et seuls certains points sont soulignés dans cette introduction. Tout d'abord, les résultats des analyses diffèrent profondément selon le mode de concurrence sur le court terme. Notamment, les modèles avec une concurrence en prix sur le court terme (Reynolds and Wilson 2000, Fabra et al. 2008) n'ont pas d'équilibre symétrique, c'est-à-dire qu'une firme investit dans plus de capacité que l'autre à l'équilibre. Dans les modèles développés dans les chapitres, la concurrence de court terme est représentée par une concurrence en quantité. Dans ce cas, les modèles d'investissement avec demande variable et une seule technologie disponible [21] s'apparentent à un modèle de concurrence en quantité à la Cournot (1838). Un résultat fondamental de ce modèle est lié aux nombre de firmes présentes. Cournot (1838, chap8) montre qu'une augmentation du nombre de firmes améliore l'efficacité du marché. Ce résultat théorique justifie le développement de la concurrence, et il se trouve à la racine de la plupart des analyses préconisant une augmentation même artificielle (par scission d'une firme dominante) du nombre de firmes présentes. Ce résultat théorique est questionné tout au long de la thèse en identifiant, dans chaque chapitre, des situations où il n'est pas valable.

4 Contributions de la thèse

La thèse analyse les décisions d'investissement de firmes en situation d'oligopole. Elle est constituée de quatre chapitres rassemblés en deux

21. Gabsewicz et Poddar (1997) analysent le cas d'un duopole, et Zoetl (2008, Chap 1) généralise leurs résultats : il existe un unique équilibre du jeu d'investissement et les quantités de capacités sont supérieures avec incertitude que sans.

parties. Il est fait abstraction des problématiques liées au réseau électrique et à la fourniture des services auxiliaires. Les modèles développés sont relativement généraux mais présentent néanmoins des caractéristiques pertinentes pour comprendre la concurrence dans l'industrie de l'électricité en intégrant cartains caractères spécifiques de l'offre et de la demande d'électricité.

La première partie contient deux analyses de l'hétérogénéité des firmes et des choix de technologies dans des situations oligopolistiques. Dans le premier chapitre l'hétérogénéité est exogène et l'on étudie l'effet du nombre de firmes sur les décisions d'investissement, dans le second chapitre l'hétérogénéité est endogène et liée à l'existence d'une contrainte technique. La seconde partie porte sur des régulations introduites dans des industries en concurrence imparfaite : la production d'électricité par une firme publique et l'interaction entre marché de permis d'émission et investissement.

4.1 Mix technologique et pouvoir de marché

Le premier chapitre présente un modèle de système électrique avec une courbe de charge et deux technologies disponibles. Etant donné un profil de consommation exogène, il existe un mix technologique optimal qui est identifié. Bien que différentes, les deux technologies sont efficaces en ce sens qu'elles sont toutes deux utilisées à l'optimum. Elles se distinguent par leur structure de coût : l'une présente un coût de capacité important et un coût variable faible alors que l'autre a un coût d'investissement plus petit mais un coût variable élevé.

L'objectif est de comprendre comment le pouvoir de marché influence la capacité totale disponible, et surtout le mix technologique c'est-à-dire la répartition de cette capacité totale entre les deux technologies. Le marché de court terme est supposé compétitif, ou parfaitement régulé, et l'analyse se focalise sur les décisions de long terme des firmes.

L'analyse considère que les firmes productrices d'électricité ne peuvent pas toutes investir dans les deux technologies en raison de contraintes de savoir faire. Certaines sont spécialisées alors que d'autres sont généralistes. Cela permet de représenter le fait que certaines technologies nécessitent un savoir-faire particulier que seules certaines firmes possèdent alors que d'autres technologies sont plus standardisées. Cette hétérogénéité des firmes a des implications sur les choix d'investissement et le nombre de firmes de chaque type affecte le bien-être de façon non monotone.

Il est tout d'abord établi qu'une firme généraliste peut être incitée à se spécialiser à l'équilibre dans l'une des technologies : elle n'aura pas d'incitation à investir dans l'autre pour éviter de trop baisser le revenu obtenu des capacités de son choix de technologie. Par exemple, si les deux technologies sont le nucléaire et les CCGTs, l'accès au nucléaire est réservé à certaines firmes alors que l'accès aux CCGTs ne l'est pas. Il est montré que les firmes généralistes peuvent être incitées à se spécialiser dans le nucléaire si un trop grand nombre de firmes 'gazières' sont actives sur le marché. Dans ce cas, en raion d'un sur-investissement en CCGTs, les firmes généralistes n'investissent pas dans des centrales au gaz afin de préserver les profits obtenus avec leurs centrales nucléaires. Il peut alors être dommageable qu'un trop grand nombre de firmes gazières soient actives sur le marché. Une augmentation du nombre de firmes spécialisées augmente la capacité totale investie mais diminue la capacité nucléaire. La perte liée à la distorsion du mix technologique peut être supérieure au gain lié à l'augmentation de la capacité totale. Ainsi, bien que les deux technologies soient efficaces, une augmentation du nombre de firmes ayant accès à l'une de ces technologies peut diminuer le surplus collectif.

4.2 Taille minimale de production et concurrence de long terme

Le second chapitre se focalise sur une contrainte technique particulière : l'existence d'une taille minimale de production pour une technologie et l'effet de cette contrainte sur la concurrence dans un modèle de duopole. Bien que faisant abstraction de la plupart des caractéristiques de l'électricité, le modèle est né de l'observation de la situation de la technologie nucléaire par rapport à d'autres technologies plus standardisées comme les CCGTs. Le nucléaire se caractérise par sa compléxités et des économies d'échelle relativement importantes par rapport aux marchés nationaux. Il faut investir dans une certaine capacité minimale de nucléaire pour que le coût complet soit moindre que celui du gaz sachant que la technologie des CCGTs qui est plus simple et divisible bénéficie de larges effets de standardisation. Le caractère 'divisible' et standardisé des technologies de ce type a été l'un des arguments favorables à la libéralisation. Cependant l'évolution du prix du gaz et la mise en place d'une régulation des émissions de CO_2 ont remis le nucléaire au goût du jour. Il s'agit alors de comprendre comment concurrence et taille minimale cohabitent.

Dans le modèle de duopole développé, la contrainte technique porte sur la technologie efficace (la moins coûteuse). L'analyse du duopole permet de comprendre comment cette contrainte peut expliquer l'hétérogénéité des firmes de façon endogène. Bien qu'initialement identiques, les deux firmes peuvent être différentes à l'équilibre : une firme utilise à l'équilibre la technologie contrainte pour produire une quantité suffisamment importante pour que sa rivale se cantonne à la technologie coûteuse. L'effet sur le bien-être de la taille minimale n'est pas monotone car, bien que cette contrainte puisse limiter l'utilisation de la technologie efficace, elle peut aussi limiter le pouvoir de marché des firmes. Le premier effet est négatif, mais le second est positif sur le surplus collectif. De plus, si cette contrainte est trop importante, les firmes n'utilisent pas la technologie efficace en situation de duopole alors qu'un monopole l'utiliserait. Dans ce cas, le surplus collectif peut être moindre avec un duopole qu'un monopole.

Cette analyse montre que l'introduction de compétition justifiée par la disposition d'une nouvelle technologie divisible peut être remise en cause lorsque cette technologie devient plus coûteuse, à moins qu'une régulation ne soit mise en place pour corriger l'éventuel déficit d'investissement dans la technologie efficace.

4.3 Oligopole mixte et investissement

Il existe plusieurs règles pour corriger l'éventuel déficit d'investissement en capacité de production et notamment en unités de pointe. L'intervention la plus directe consiste à laisser le GRT installer des centrales de production de pointe (Finon et al. 2008). Ce mécanisme est utilisé notamment dans les pays nordiques : la Norvège, la Suède ou la Finlande. Ces centrales sont appelées lorsque les prix sont élevés. En plus de modifier la quantité de capacité totale disponible, ce type de situation modifie aussi l'équilibre en production. Bien que justifiée par le caractère de bien public de l'adéquation de capacité du système cette intervention a aussi un effet correctif global, notamment sur le pouvoir de marché.

Ce type d'intervention publique revenant à autoriser le GRT à ajouter des capacités en dernier ressort s'il anticipe un déficit d'investissement des producteurs et à les utiliser en période extrême est critiquée car en dépréciant les prix de l'électricité, elle limiterait l'investissement privé et pourrait ainsi entretenir le sous-investissement des firmes. C'est l'une des explications de la "missing money" selon Joskow (2006). Une analyse formalisée de cet argument est développée par Joskow et Tirole (2006). Ils analysent les distorsions provoquées par la production d'électricité par le GRT. Ils considèrent comme exogène la quantité de capacité contrôlée par le GRT et analysent l'effet de cette quantité sur les prix, les quantités produites et les investissements lorsque les firmes sont preneuses de prix.

Le modèle développé dans ce chapitre comporte trois étapes, deux étapes d'investissement et une étape de production. Les firmes privées investissent à la première étape puis une firme publique bienveillante (le GRT) investit et enfin toutes les firmes produisent. Les firmes privées maximisent leur profit alors que la firme publique maximise le surplus collectif. Ainsi, à l'instar de Joskow et Tirole (2006), le problème de la sécurité du système n'est pas représenté mais contrairement à eux la capacité publique est endogène. L'intervention du GRT est motivée par l'observation d'un déficit d'investissement qui s'explique par le pouvoir de marché des producteurs privés. Le GRT est suiveur dans le cadre principal mais le cas de mouvement simultanés est aussi traité.

Il est établi que : bien que la firme publique soit aussi efficace que les firmes privées et que les coûts soient linéaires, la firme publique peut être incapable d'atteindre l'optimum social en raison de l'exercice de pouvoir de marché sur le court terme. Une fois les capacités privées installées, la firme publique décide de pallier au manque d'investissement mais elle doit prendre en compte l'effet de sa décision sur les décisions de production ultérieures des firmes privées. Celles-ci ont tendance à limiter leur production suite à une augmentation de capacité publique. Ainsi, la production issue des capacités publiques est en partie compensée par la baisse de la production privée. La firme publique prenant en compte cet effet doit limiter son investissement et ne peut donc pas établir le prix au niveau du coût marginal de long terme. Sur le long terme, les firmes privées sont capables d'obtenir des profits strictement positifs en anticipant cette difficulté de la firme publique. Il est cependant montré que ce mécanisme de la "firme publique investisseuse" bien qu'imparfait, permet tout de même d'augmenter le surplus collectif en augmentant la capacité et la production totales. De plus, un tel mécanisme peut aussi encourager l'investissement privé car les firmes privées sont incitées à investir pour limiter l'intervention de la firme publique.

Ce modèle est développé pour analyser un type d'intervention publique spécifique qui concerne des capacités de pointe. Néanmoins, comme le problème du sous-investissement concerne l'ensemble des technologies, il peut être utilisé pour analyser une éventuelle action publique dans le développement d'autres types de technologies. Ceci est d'autant plus plausible que dans de nombreux pays, l'état reste actionnaire de producteurs et pourrait les mandater pour de telles opérations.

4.4 Marché de permis d'émission et concurrence imparfaite

L'objectif de réduction des émissions de CO_2 de l'Union Européenne nécessite une contribution majeure du secteur électrique, fort émetteur. L'Europe a choisi de mettre en place un mécanisme de "cap and trade" avec un marché de permis d'émissions afin de réaliser une partie de son objectif. Les entreprises de plusieurs secteurs industriels, dont l'électricité, doivent remettre à la fin de chaque période des permis pour leurs émissions de la période. Ces permis d'émissions peuvent être échangés pendant la période entre les entreprises soumises à cette contrainte. Les principaux secteurs concernés sont concentrés si bien qu'ils sont généralement considérés comme imparfaitement concurrentiels. Par contre, comme le nombre de firmes sur le marché de permis est élevé, ce marché peut être considéré comme parfaitement concurrentiel.

Le chapitre analyse l'efficacité d'un marché de permis d'émission entre des secteurs imparfaitement concurrentiels. La quantité totale d'émission est fixée, et deux marchés de biens polluants et non substituables sont considérés. Sur ces marchés, la concurrence est représentée à la Cournot. Par contre, sur le marché des permis, les firmes sont supposées preneuses de prix [22]. Il s'agit alors de comprendre comment la concurrence imparfaite sur les marchés des biens influence l'efficacité d'un marché de permis intégré et dans quelles circonstances il peut être préférable de conserver des marchés isolés ou d'introduire un mécanisme correctif.

La problématique définie par rapport à deux marchés présente l'avantage de pouvoir être spécifiée en particulier pour un marché électrique annuel dissocié en deux biens. Le marché de permis permet d'allouer la contrainte globale entre les deux marchés de biens. Comme les permis sont valables pour des émissions au sein d'une période fixée, ce marché permet aussi d'allouer les permis dans le temps. Cette allocation temporelle joue un rôle prépondérant dans le secteur électrique étant donné les

22. Il est en fait équivalent de supposer qu'à productions fixées elles négocient de façon efficace l'allocation des permis entre elles.

variations de la demande. Ainsi, le cadre développé permet d'analyser l'investissement en capacité de production et l'allocation de la contrainte d'émission entre la demande de pointe et la demande de base.

Il est établi que lorsque les marchés des biens sont imparfaitement concurrentiels, l'intégration des marchés de permis d'émission peut diminuer le surplus collectif. C'est-à-dire qu'un régulateur parfaitement informé (en l'absence d'incertitude sur les conditions des marchés de biens), peut faire strictement mieux qu'un marché de permis intégré en conservant deux marchés de permis isolés. Lorsque la concurrence est imparfaite sur les marchés des biens, la quantité d'émissions allouée à un marché doit prendre en compte le pouvoir de marché des firmes. Une allocation optimale de second rang est définie lorsque le régulateur prend en compte le pouvoir de marché des firmes sur les marchés des biens. Même si les firmes sont preneuses de prix sur le marché des permis, un marché intégré de permis ne permet d'atteindre cet optimum de second rang. Il est cependant possible de l'atteindre avec un marché de permis intégré si les émissions de l'un des secteurs sont subventionnées, par exemple via un mécanisme d'allocations gratuites.

En pratique, les permis sont alloués avant que les différents niveaux de demande des différents secteurs ne soient connus. Le régulateur n'est alors pas capable de mettre en place l'optimum de second rang parce qu'il alloue les permis avant que les niveaux des demandes de biens ne soient connus. Ainsi, afin de mieux comprendre les gains de l'intégration des marchés de permis, on introduit de l'incertitude dans le modèle. Il est alors montré que l'intégration des marchés de permis est préférable lorsque les incertitudes des secteurs considérés sont décorrélées.

Finalement, l'analyse est appliquée à la problématique de l'investissement en capacité de production dans un marché de l'électricité très stylisé avec deux états de la demande (base et pointe) et une seule technologie disponible. Dans ce cas, le marché de permis d'émission permet d'allouer la contrainte d'émissions entre la pointe et la base, et influence le choix de capacités de production. Le pouvoir de marché des firmes, plutôt que de diminuer les capacités et les productions, alloue imparfaitement la contrainte d'émission entre les deux états : les firmes sous-investissent mais sur-produisent en base. Il alors montré qu'un mécanisme d'allocations gratuites permet de corriger ces distorsions en jouant comme un mécanisme de subvention qui permet d'augmenter la capacité totale tout en diminuant la production en base.

Première partie

Pouvoir de marché et choix de technologie

Chapitre 1

Mix technologique et concurrence imparfaite

1 Introduction

One motivation behind the market reforms in electricity industries is to encourage efficient investment in the optimal mix of technologies in a timely way and to avoid situations of overcapacity observed in the former regime of regulated monopoly. But there are growing concerns about the ability of the new liberalized market regime to induce sufficient investment in building capacity, in the optimal technology mix without penalizing efficient but capital intensive technologies. Ensuring enough generation capacity to meet future electricity demand by the optimal technology mix progressively became a contentious issue in the design of the reforms, particularly since some recent crisis.

Three main reasons are highlighted in the literature to explain the sub-optimality of investment in generation mix: ”missing money”, risk management, and market power. The present chapter focuses on the third reason: the strategic choice of generating capacity. I do not consider the interrelation of long term investment choice with short term market power. It is assumed that the short term is perfectly regulated: the wholesale price is set at the marginal cost of the marginal technology or at the value of the capacity constraint when it is binding. The aim is to analyze the long term incentive for generators to underinvest in aggregate capacity and to distort the generation mix. Particularly, the influence of the number of firms that have access to various technologies is emphasized.

The current situation of electricity markets and the three explanations of underinvestment are reviewed before the theoretical literature on capacity choice in electricity markets and the contributions of the present paper.

Sub-optimality of generation investment in electricity markets

The electricity industry has the particular feature that many technologies are used to produce the same good, there is not a benchmark technology but a set of technologies. For any particular load duration curve, an optimal technology mix minimizes the cost to produce this load. A technology can be described by its ratio of variable and capacity cost [1]. A baseload technology is characterized by a high capacity cost and a low variable one, it is used to produce during a long fraction of time, whereas peaking units are characterized by a low capacity cost and high operating one and are efficient to produce during few hours per years.

A perfectly competitive wholesale electricity markets should theoretically induce efficient investment: an optimal aggregate capacity and an optimal technology mix. Scarcity rents during the peak periods are needed to ensure the profitability not only of peaking units but also of all other technologies. Moreover the short term system security of the system is a public good supplied by the system operator with operating reserves(Joskow and Tirole 2007). It implies that additive revenues are needed for the contribution of capacities to this public good.

Concerns on investment that initially focuses on peaking units are

1. It is a first approximation because others characteristics play important rules such as the ramping rate.

now extended to all technologies. Particularly, the ability of wholesale electricity markets to promote sufficient investment in capital intensive technologies, such as nuclear which is a typical baseload technology, is currently a debated issue. Three main arguments are found in the literature to explain the potential lack of investment in electricity markets.

The first explanation for underinvestment: the "missing money" refers to the deficit of revenue during peak hours. This deficit is attributed to price cap (Cramton and Stoft 2006) and to ill designed regulatory procedures related to the public good attributes of operating reserves. Despite price caps that are considered too low, the technical rules used by system operators to guarantee system reliability by calling operating reserves tend to erase revenues from the energy and reserves markets as explained by Joskow (2006). Even if the price cap can be suppressed, it remains a deficit of revenues.

The second reason is the exceptional volatility of electricity prices and markets incompleteness. The volatility of electricity prices (with a magnitude from 20€ to 5-10 000€/MW/h) is explained by the short-run inelasticity of demand and supply, given the non price transmission and the non-storable nature of electricity. In principle volatility does not deter to invest in due time and in appropriate technologies, provided that future and forward markets exist to allow investors to manage their market risks. But in electricity markets these markets are little developed, this incompleteness creates difficulty for hedging investment in generation. Risk and market incompleteness are often cited as explanations of the vertical integration observed in electricity markets (Finon 2008, Joskow 2006). Furthermore, Roques et al. (2006) argue that risk aversion can favor investment in combined cycle gas turbine (CCGT) compared to nuclear plant because of the correlation between electricity prices and gas price. By decreasing the variance of the revenue from a CCGT, this correlation favors gas plant. So risk aversion can explain a perceived lack of investment in nuclear plant.

The third reason refers to market power and is the subject of the present paper. Every generator can potentially benefit from a sub-optimal investment in generation capacity in the system, which will provoke profitable periods of price spikes. Indeed, because of the inelastic nature of demand and supply, even a slight shortfall of available capacity or unexpected high loads can provoke a dramatic increase in price in period of tight supply. Hence generators are incited to underinvest. European Commission suspects that the electricity companies delay investment in baseload and midload generation, beyond the remaining regulatory uncertainty, because in any case they would be the winners of any situation of tight supplies (D.G.COMP 2007).

Moreover, beside the incentive to underinvest in aggregate capacity firms may have also an incentive to profitably distort the technological mix. This distorsion may be amplified by the heterogeneity of firms relative to the access to technologies. Because some technologies necessitate a specific knowledge, that could have been acquired historically, all firms do not master all technologies and some are specialized whether others are "generalist". The present paper deals with this issue by analyzing the deficit of investment in several technologies related to strategic choice of capacity, and its relation with the number of firms that have access to each technology. I have in mind the current situation of nuclear technology and CCGTs, the later being a standardized technology that is perceived as the main vehicle of competition (Newbery 1998). The theoretical literature related to market power and capacity choice is reviewed next.

Capacity choice

The development of wholesale electricity markets created a renewed interest in the literature on capacity choice with demand fluctuation or uncertainty. Gabszewicz and Poddar (1997) analyze the choice of capacity by two competing producers in a linear model. They establish that firms invest more with uncertainty than without. In their model, one technology is available and short term competition is a quantity game *à la* Cournot with capacity constraints. Their results are generalized by

Zoetl (2008, chap1) who also considers the alternative assumption of a regulated, or perfectly competitive, spot market.

The issue of the technology mix has been addressed by several authors and in most papers all firms have access to all technologies. In a major contribution, von der Fehr and Harbord (1997) analyse investment by symmetric producers for different price mechanisms or regulatory regimes: they consider an efficient spot market and a non discriminatory auctions. In the later case, they state that there is no symmetric equilibrium [2] but only consider the duopoly case. With an efficient spot market, which is the case considered here, they establish that firms underinvest in aggregate capacity and profitably distort the technology mix toward peak units. A firm has an incentive to limit baseload investment in order to limit the period of marginality of this technology. An increase of the number of firms both increases the aggregate quantity of capacity and improves the technology mix. In a more recent paper Arellano and Serra (2007) establish a similar result, they consider the incentive for firms to distort the technology mix when the aggregate quantity of capacity is fixed and extend the analysis to free entry equilibrium.

Those results contrast with those obtained with short term Cournot competition. Because of the strategic effect on the short term of lower variable cost [3], there is a strategic incentive to invest in baseload capacities.Murphy and Smeers (2005) consider heterogeneous firms: a baseload and a peak producer. They emphasize the strategic effect of investment in a closed-loop equilibrium. Because of this strategic effect the baseload firm invest more in a closed loop equilibrium than in an open loop one [4]. Zoetl (2008, chap3) considers symmetric firms that have access to a continuous technology set. The continuity property of the technology set

2. This is similar to the result of Reynolds and Wilson (2000) on capacity choice under demand uncertainty and price competition. It is further analyzed by Fabra et al. (2008) who compare different auctions mechanisms.

3. The strategic effect refers to the decrease of others' productions subsequent to a decrease of one's marginal cost.

4. In the open loop equilibrium strategic effects are ignored, capacities and (conditional) production quantities are simultaneously fixed.

allows tractability of the model and explain the symmetry of firms at equilibrium. He establishes that because of a strategic incentive firms might overinvest in baseload units.

To my knowledge, only Murphy and Smeers (2005) consider asymmetric firms, and no previous analysis perform comparative statics on the number of investing firms. It might be related to the analytical difficulties of dynamic games with discrete technology sets and strategic interactions at each stages. To provide such an analysis, short term market power is ignored here and so is the strategic effect mentioned above. It is assumed that the price is set at the variable cost of the marginal technology when demand is not rationed and at the value of lost load (VoLL) when rationing occurs. Empirically, observed prices are not as high as predicted by theoretical models of imperfect competition (supply function, discrete auctions, Cournot) and more close to marginal cost than to "Cournot" prices(Wolfram 1999). It can be justified by the fact that wholesale electricity markets are highly scrutinized by regulatory authorities, or auctions are "short term" efficient [5].

I consider here heterogeneous firms: all of them have not access to all technologies, some are specialized and some are generalist. It is first shown that even if both technologies are efficient, generalist firms do not necessarily invest in both. If the number of specialized firms in a particular technology is high compared to the rest of the industry, they overinvest in this technology and it deter generalist firms from investing in this technology. In such case each generalist firm behaves as a specialist one in the other technology.

Both the aggregate quantity of capacity and the technology mix are shown to be distorted in a variety of direction that is related to the number of firms that can invest in each technology. The welfare consequences of a change in the number of firms that have access to a technology are investigated. The respective numbers of baseload and peak firms influ-

5. Another explanation is the vertical relations between electricity producers and retailers that are not considered here.

ence both aggregate capacity and the technology mix. It is established that even if both technologies are efficient an increase of the number of one kind of specialized firms can decrease welfare. If an additional firms is active despite increasing capacity it can further distort the technology mix and the cost of this distortion can offset the welfare gain from the increase in capacity. Hence, if the number of firms that have access to a particular technology is fixed, the number of firms that have access to the other technology should be limited.

For instance, the deficit of investment in nuclear can be related to the number of firms that are effectively able to invest in nuclear plants, this deficit can also explain an overinvestment in other technologies such as CCGTs.

The rest of the chapter is organized as follow, I first introduce the model and consider the first best optimum in the next section. Then the investment game is solved in the general case(section 3) before analyzing the influence of the number of firms of each type (section).

2 The model

2.1 Framework

I consider a simple electricity system without network constraints. Consumers are assumed to be insensitive to price, the demand of electricity x is uniformly distributed on the set $[0, X]$ with the density $1/X$. It is a rough representation of a load duration curve with a year duration normalized at 1. In section I discuss how a more realistic load duration curve would influence results. The surplus from each unit of electricity consumed is assumed constant and denoted v.

There are two technologies that can be used to produce electricity labeled $t = \alpha, \beta$. Each technology t is characterized by a variable cost c_t (per kwh) and a capacity cost I_t (per kW per year). Technology α is less costly to produce a unit of electricity all over the year than technology

β, but it is more costly for production over short period of time:

$$c_\alpha + I_\alpha < c_\beta + I_\beta$$
$$I_\beta < I_\alpha$$

Even if the sum of variable and capacity cost of technology β are higher than those of technology α, it is efficient for production during a short fraction of the year. The difference of capacity costs is denoted $\Delta = I_\alpha - I_\beta$ and the difference of variable costs $\delta = c_\beta - c_\alpha$. Both are positive and $\Delta < \delta$. The ratio $r = \Delta/\delta$ is the duration such that technology α (resp. β) is more efficient for production over period more (resp. less) than r. These features are illustrated on figure 1.1.

Technologies α and β are respectively called baseload and peak through the paper, but the framework can be used to consider baseload and mid-load technologies such as nuclear and CCGT.

For $t = \alpha, \beta$, the ratio $r_t = I_t/(v - c_t)$ is the minimal duration of production with technology t such that the aggregate cost is less than the consumer surplus. Both are assumed less than one i.e. $v > c_\beta + I_\beta$ and it is assumed that:

$$r_\beta < r$$

This assumption ensures that technology β is used at equilibrium. The left hand side is the minimal duration of production with technology β such that cost are below consumer surplus. To produce during a smaller period of time with this technology a unit of electricity consumed would imply a welfare loss. The right hand side is the maximal duration of production for which technology β is more efficient than technology α. So the former is smaller than the latter and it is optimal to use both technologies to produce electricity. It should be noticed that this assumption is equivalent to $r_\alpha < r$ and to $r_\beta < r_\alpha$, these two inequalities can be interpreted similarly.

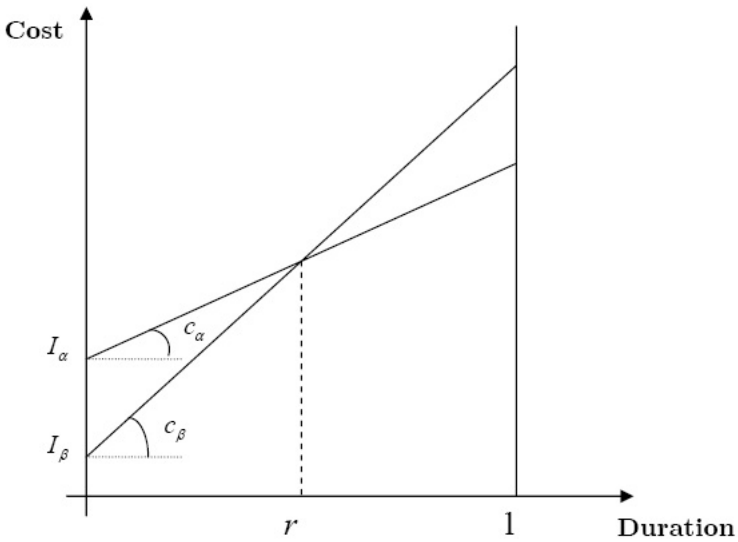

Figure 1.1: Cost and production duration

There are n firms that produce electricity indexed $i = 1..n$. Individual quantities of capacity of firm i of each technology are denoted k_α^i and k_β^i, and its aggregate quantity of capacity is denoted $k^i = k_\alpha^i + k_\beta^i$. So aggregate quantities over all firms of capacities of each technology are $k_t = \sum_i k_t^i$ for $t = \alpha, \beta$ and $k = k_\alpha + k_\beta$.

All firms have not access to both technologies, so the set of firms is divided into three subsets. There are g 'generalist' firms that have access to both technologies and $n - g$ specialized firms: s_α baseload firms have only access to technology α and s_β peak firms that have only access to technology β. So the number of firms is $n = g + s_\alpha + s_\beta$. Firms are ordered as follows: firm $i = 1, .., s_\alpha$ are baseload firms, firms $i = s_\alpha + 1, .., s_\alpha + s_\beta$ are peak firms and finally firms $i = n - g, .., n$ are generalist firms.

Each generalist firm chooses quantities of capacity of each technology, whereas a peak (resp. baseload) firms only chooses a quantity of technology β (resp. α).

Once capacities are fixed, short term is assumed 'perfectly' regulated:

there is no modeling of short term market power. The price is set at the marginal cost of the last unit called when all demand is satisfied and at v in case of rationing. Firms produce with a technology when the price is above its operating cost. Rationing occurs when the demand of electricity is higher than the aggregate capacity available. So, when demand is less than k_α the wholesale price is c_α, and the production of x is done by firms that have baseload capacities. When demand is greater than k_α and smaller than k the price is c_β, baseload capacities are fully used and the $x - k_\alpha$ remaining quantity is produced by firms with peak capacities. For higher demand the price is v and there are only k units of electricity consumed, a part $x - k$ of the demand is not satisfied [6]. Price and quantities are represented on figure 1.2.

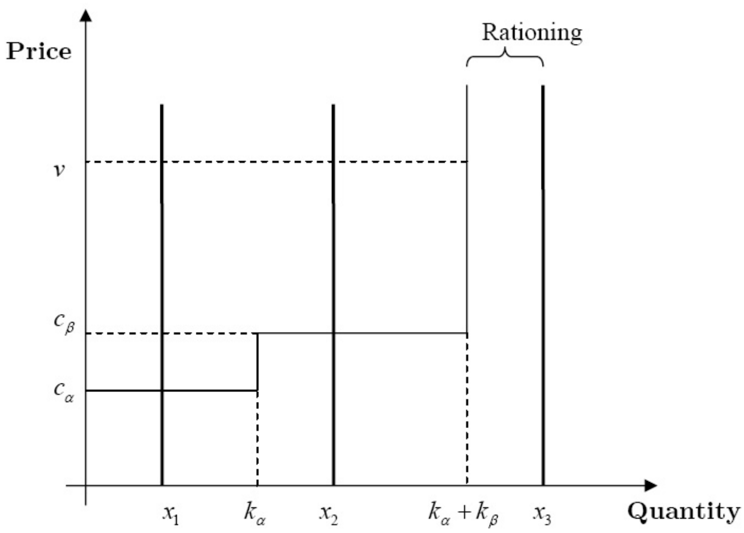

Figure 1.2: Supply curve and spot prices

Firms earn short term positive revenue from capacities of a technology type only when the aggregate quantity of capacity of this type is

6. I do not consider the cost of inefficient rationing. With a linear loss of $\gamma(x - k)$ in case of rationing the price would be $v + \gamma$ when $x > k$.

fully used. The profit of a firm $i = 1, .., n$ is:

$$\begin{aligned}\pi^i = &\frac{1}{X}\int_{k_\alpha}^{k_\alpha+k_\beta} \delta k_\alpha^i \, dx \\ &+\frac{1}{X}\int_k^X \left[(v-c_\beta)k_\beta^i + (v-c_\alpha)k_\alpha^i\right] dx \\ &-I_\alpha k_\alpha^i - I_\beta k_\beta^i\end{aligned} \tag{1.1}$$

The net revenue of a firm has two terms: the first one is the net revenue from baseload capacity when the price is set the variable cost of the peak technology and the second one is the profit obtained from both technologies when rationing occurs. Alternatively one can write the profit of a firm as a function of aggregate capacity and baseload capacity:

$$\begin{aligned}\pi^i = &\frac{1}{X}\int_{k_\alpha}^{k} \delta k_\alpha^i \, dx \\ &+\frac{1}{X}\int_k^X (v-c_\beta)k^i \, dx \\ &-\Delta k_\alpha^i - I_\beta k^i\end{aligned} \tag{1.2}$$

This writing emphasizes the relation between the technology mix and the aggregate quantity of capacity of a generalist firm. In that case, the cost of a unit of capacity of technology α is net of the cost of the unit of capacity β it replaces and similarly its short term revenue is the variable cost difference in all states where baseload capacity are fully used.

2.2 Welfare optimum

Welfare is the sum of gross consumer surplus and aggregate production cost. Gross consumer surplus is only related to the aggregate quantity of capacity, it is denoted $S(k)$; the aggregate cost of production is related to the technology mix and can be written as a function of aggregate capacity and the quantity of capacity of technology α: $C(k, k_\alpha)$. These two functions are:

$$S(k) = \frac{v}{X}\left[\int_0^k x \, dx + \int_k^X k \, dx\right] = \frac{v}{X}\left(X - \frac{k}{2}\right)k \tag{1.3}$$

$$C(k,k_\alpha) = \frac{1}{X}\left[\int_0^k c_\beta x\,dx + \int_k^X c_\beta k\,dx\right] \tag{1.4}$$
$$-\frac{1}{X}\left[\int_0^{k_\alpha} \delta x\,dx + \int_{k_\alpha}^X \delta k_\alpha\,dx\right] + \Delta k_\alpha + I_\beta k$$

And welfare is:

$$W(k_\alpha, k) = S(k) - C(k, k_\alpha) \tag{1.5}$$

The problem can alternatively be solved with respect to couple of quantities (k_α, k_β) or (k_α, k). I proceed with the second method. Welfare (1.5) is concave and first best quantities k^* and k_α^* solve following first order conditions :

$$(v - c_\beta)\,\frac{X-k}{X} = I_\beta$$

$$\delta\frac{X-k_\alpha}{X} = \Delta$$

The first relation can be rephrased with the jargon of electricity systems. It is the relation between the value of lost load v and the loss of load probability $(X-k)/X$ and the operating and capacity costs of the peak technology. The second equation is related to the optimal technology mix that minimizes the cost of production. This technology mix is such that the time of use of each unit of capacity of technology β is less than r. From these equations it is straightforward to obtain the expression of the optimal technology mix:

$$k^* = X\,(1-r_\beta) \text{ and } k_\alpha^* = X\,(1-r)\,,\ k_\beta^* = X\,(r-r_\beta) \tag{1.6}$$

And the assumption $r > r_\beta$ ensures that both technologies are used at the optimum. Furthermore, it is optimal to ration consumers during a fraction r_β of the year. This fraction is solely determined by the cost of the marginal technology and the value of lost load. The choice of first best quantities of capacities is represented on figure (1.3).

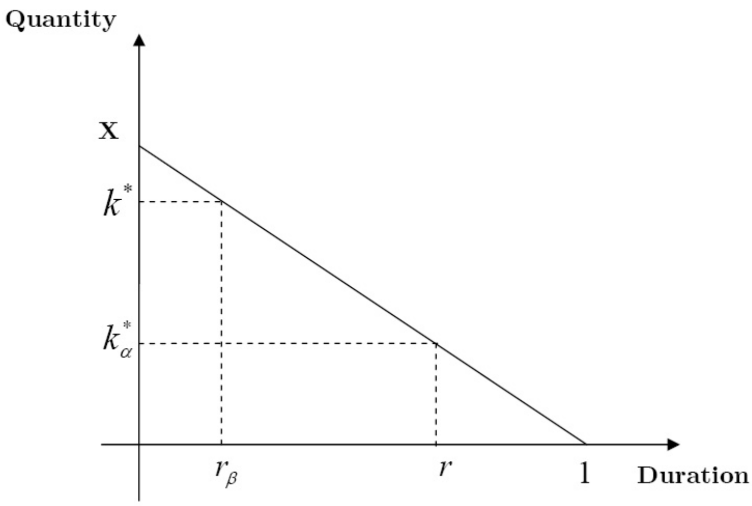

Figure 1.3: Load curve and optimal investment

I consider several industry configurations whether firms have access to both or only one technology. In the electricity industry some technologies are "standardized" and available to all firms whereas some others are not. There is an important concern on the ability of electricity markets to promote investment in peaking units, technology β in the framework. Beside the public good characteristics of operating reserves [7], few firms invest in peak capacities so they might limit their investment in order to increase the duration of periods of high prices.

Nevertheless, concern about underinvestment can be extended to all type of technologies and particularly to baseload ones. Nuclear technology is a typical baseload technology characterized and few firms are able to master this technology, the situation described above is reversed: there is a potential lack of investment in baseload technology. So, both situations are analyzed below.

First, in the next section, the general case is described, the major issue being whether generalist firms specialize. Second, in the last section welfare consequences of an increase of the number of specialized firms of

7. This public good aspect is not represented here because there is no risk of network collapse. The system operator is able to ration consumers before such event happens.

one type are described.

3 General case

Firms simultaneously choose their quantities of capacity in order to maximize their profit. Whereas specialized firms only invest in one type of capacity, generalist ones invest in both. From the expression (1.1), first order conditions of both types of specialized firms are:

$$\text{for } i = 1, .., s_\alpha : \frac{\delta}{X} k_\beta + \frac{v - c_\alpha}{X} \left(X - k - k^i_\alpha \right) - I_\alpha = 0$$
$$\text{for } i = s_\alpha + 1, .., s_\alpha + s_\beta : \frac{v - c_\beta}{X} \left(X - k - k^i_\beta \right) - I_\beta = 0$$

As in an usual quantity game, firms have an incentive to limit their investment. Here, a lower aggregate capacity increases the length of time with price at v. Baseload firms have an additive revenue when price is fixed at c_β, this revenue is proportional to the quantity of peak capacity installed.

Concerning generalist firms, each of them chooses both quantities of capacity, as they might not invest in one type of capacity positivity constraints are introduced. So the objective of a generalist firm is:

$$\max_{k^i_\alpha, k^i_\beta} \{ \pi^i (k^i_\alpha, k^i_\beta, k_\alpha, k_\beta) \}$$
$$\text{subject to } 0 \leq k^i_\alpha \ , \ 0 \leq k^i_\beta$$

The Lagrange multiplier of the positivity constraint of baseload (resp. peak) capacity is ν_i (resp. μ_i). First order conditions of a generalist firm are for $i = n - g + 1, .., n$:

$$\frac{\delta}{X} k_\beta + \frac{v - c_\alpha}{X} \left(X - k - k^i_\alpha \right) - \frac{v - c_\beta}{X} k^i_\beta - I_\alpha + \nu^i \ = \ 0$$
$$\frac{v - c_\beta}{X} \left(X - k - \left(k^i_\alpha + k^i_\beta \right) \right) - I_\beta + \mu^i \ = \ 0$$

Several relations between quantities chosen by generalist and specialized firms can be deduced from these first order conditions. If a generalist firm invests in both types of technology it chooses an aggregate quantity of capacity similar to a peak firm. The marginal capacity of a generalist firm being a peak unit its marginal revenue is similar to that of a peak firm. But part of a generalist firm's capacity is baseload.

Compared to baseload firms, generalist firms have a different incentive to invest in baseload capacity because they use baseload capacity to favorably distort the technology mix and not to set the aggregate capacity. The marginal baseload capacity of a generalist firm has the additional negative effect of decreasing its revenue from peak capacity, so generalist firms invest less in baseload capacity than baseload firms. The incentive to distort the technology mix might be best seen with the alternative writing of the first order condition :

$$\frac{\delta}{X}\left(X - k_\alpha - k_\alpha^i\right) - \Delta + \nu^i - \mu^i = 0 \tag{1.7}$$

This emphasizes that the choice of a baseload capacity is made by comparison with a peak unit, so opportunity marginal revenue and cost are δ and Δ. Equation (1.7) said that a generalist firm limits the share of baseload technology in order to increase the time of marginality of peak units.

However, whether generalist firms invest in both type of capacities depends upon industry configuration. Rewriting first order conditions of a generalist firm:

$$\begin{aligned} k_\alpha^i &= k_\alpha^* - k_\alpha + \left(\nu^i - \mu^i\right)/\delta \\ k_\beta^i &= k_\beta^* - k_\beta + \left[\frac{v - c_\alpha}{v - c_\beta}\mu^i - \nu^i\right]/\delta \end{aligned} \tag{1.8}$$

If one group of specialized firms 'overinvests', i.e. invests more than the corresponding first best quantity, generalist firms do not invest in their technology. As specialized firms have only access to one technology they have higher investment incentives than generalist firms, so such situations can occur at equilibrium. If there are too many specialized firms of one kind, these firms may over invest and deter generalist firms from investing in their technology. In such cases, the situation is similar to an entirely specialized industry.

In proposition (1) expressions of equilibrium quantities when generalist firms invest in both types of technology are established. Situations where generalist firms do not invest in one technology are precisely stated in proposition (2). Individual equilibrium quantity of capacity of a baseload (resp. peak) firm is k_α^S (resp. k_β^S), and equilibrium quantities of baseload and peak capacities of a generalist firm are k_α^G and k_β^G.

Proposition 1 *There is a unique Nash equilibrium. If at this equilibrium generalist firms invest in both types of capacity. Equilibrium quantities are:*

$$k_\alpha^S\left(s_\alpha, s_\beta, g\right) = \frac{X}{A}\left[\left(g+1\right)\left(1-r_\alpha\right) + s_\beta \frac{\delta-\Delta}{v-c_\alpha}\right]$$
$$k_\beta^S\left(s_\alpha, s_\beta, g\right) = \frac{X}{A}\left[\left(g+1\right)\left(1-r_\beta\right) + s_\alpha\left(r_\alpha - r_\beta\right)\right]$$
$$k_\alpha^G\left(s_\alpha, s_\beta, g\right) = \frac{X}{A}\left[\left(g+s_\beta+1\right)\left(1-r\right) - s_\alpha\left(r - r_\alpha\right)\right]$$
$$k_\beta^G\left(s_\alpha, s_\beta, g\right) = \frac{X}{A}\left[\left(g+s_\alpha+1\right)\left(r-r_\beta\right) - s_\beta\left(1-r\right)\right]$$

Where $A\left(s_\alpha, s_\beta, g\right) = \left(g+s_\alpha+1\right)\left(g+s_\beta+1\right) - s_\alpha s_\beta\left(v-c_\beta\right)/\left(v-c_\alpha\right)$

From expressions of the proposition 1 threshold numbers of specialized firms that deter generalist firms from investing in one of both technologies can be determined. These numbers are such that one group of specialized firms invest more than the first best corresponding quantity.

Proposition 2 *Equilibrium quantities* k_t^S *and* k_t^G *for* $t = \alpha, \beta$ *are:*

- *If* $s_\alpha \geq \left(s_\beta+g+1\right)\left(1-r\right)\left(r-r_\alpha\right)^{-1}$ *baseload firms 'over invest' and generalist firms invest only in peak capacities and:*

$$k_\alpha^S = k_\alpha^S\left(s_\alpha, s_\beta+g, 0\right) \geq k_\alpha^*/s_\alpha \text{ and } k_\alpha^G = 0$$
$$k_\beta^S = k_\beta^G = k_\beta^S\left(s_\alpha, s_\beta+g, 0\right)$$

- *If* $s_\beta \geq \left(s_\alpha+g+1\right)\left(r-r_\beta\right)\left(1-r\right)^{-1}$ *peak firms 'over invest' and generalist firms only invest in baseload capacities and:*

$$k_\alpha^S = k_\alpha^G = k_\alpha^S\left(s_\alpha+g, s_\beta, 0\right)$$
$$k_\beta^S = k_\beta^S\left(s_\alpha+g, s_\beta, 0\right) \geq k_\beta^*/s_\beta \text{ and } k_\beta^G = 0$$

– *Else quantities are those expressed in proposition 1.*

Proofs of both propositions are in appendix A-1. The term 'over-invest' is employed in a particular sense here. In situations described, firms invest more than the first best quantity of a technology but less than a second best defined with a fixed quantity of the other technology. For instance, if the quantity of peak capacity is fixed at the equilibrium quantity, the quantity of baseload capacity that maximizes welfare is always higher than $s_\alpha k_\alpha^S$.

The 'overinvestment' result is due to the limited access to a technology. It occurs if there are too many specialized firms of one kind or too few generalist firms. If the number of generalist firms increases, occurrence of overinvestment and specialization of generalist firms are less likely. Both conditions cannot be satisfied simultaneously so generalist firms always invest at least in one technology.

Conditions of overinvestment and specialization can be rewritten with the share of specialized firms in the industry. Specialization of generalist firms to baseload or peak technology occurs if respectively:

$$\frac{s_\alpha}{n+1} \geq \frac{1-r}{1-r_\alpha} \text{ or } \frac{s_\beta}{n+1} \geq \frac{r-r_\beta}{1-r_\beta}$$

These inequalities allow to analyse the effect of r on specialization. Technology β can be viewed either as a midload or a peak technology, the former case correspond to high r whereas the later a r close to r_α. So, it appears from these equation that generalist firm are more likely to specialize to technology β if this technology is midload, i.e. r is small. And conversely, specialization to baseload technology is more likely when technology β is a peak technology.

Only generalist firms

In the particular case where there are only generalist firms, they invest in both type of technologies and equilibrium quantities are:

$$k_\alpha^G(0,0,n) = \frac{1}{n+1}k_\alpha^* \text{ and } k_\beta^G(0,0,n) = \frac{1}{n+1}k_\beta^*$$

These expressions are similar to those find by von der Fehr and Harbord (1997), the oligopoly quantities are qualitatively similar to those obtain in a linear Cournot model. The aggregate capacity chosen by firms is the optimal one by a factor of $n/(n+1)$, and moreover, the technology mix is also distorted in a similar proportion.

Interesting situations are those where some specialized firms are active. I briefly discuss the two cases where there are only one kind of specialized firm before devoting the last section to the case of a entirely specialized oligopoly and the welfare consequences of an increase of the number of specialized firms. This last part is actually general because even if there are generalist firms there might specialized.

Generalist and baseload firms

Originally there was a concern on a potential lack of investment in peaking units, this concern was related to the effect of aggregate capacity on the frequency of rationing network collapse. The lack of investment in peaking units can be related to market power and the limited number of firms that invest in peakers [8]. So one should consider that all firms can invest in baseload plants whereas only a subset can invest in peakers: $s_\alpha = 0$.

In that case, when generalist firms invest in both technologies expressions are relatively simple because generalist firms 'complete' the investment of baseload firms. From equations (1.8), the following relation is satisfied by equilibrium quantities:

$$k_\alpha^G = \frac{1}{g+1}\left(k_\alpha^* - s_\alpha k_\alpha^S\right)$$
$$k_\beta^G = \frac{1}{g+1} k_\beta^*$$

Corollary 1 *If $s_\alpha (r - r_\alpha) \leq (g+1)(1-r)$ generalist firms invest in*

8. It is the assumption made by Joskow and Tirole (2007) when they analyse underinvestment in peakers and two regulations: price caps and capacity paiement.

both types of capacity, and equilibrium quantities satisfy:

$$s_\alpha k_\alpha^S + g k_\alpha^G = \frac{g}{g+1} k_\alpha^* + \frac{s_\alpha}{g+1} \frac{1-r_\alpha}{n+1}$$
$$g k_\beta^G = \frac{g}{g+1} k_\beta^*$$

Generalist and peak firms

The situation might also be reversed if there are only few firms that can invest in baseload capacity. This situation can illustrate the case of nuclear investment. There are few firms that have access to this technology but much more that can invest in gas plant, the development of competition in the electricity industry was essentially awaited from investment in CCGTs.

If there are numerous firms that can invest in peak technology, generalist firms specialized. Otherwise, expressions of aggregate and baseload capacities are simple and comparable to standard linear Cournot quantities.

Corollary 2 *If* $s_\beta (1-r) \leq (g+1)(r-r_\beta)$, *aggregate equilibrium quantities are:*

$$g k_\alpha^G = \frac{g}{g+1} k_\alpha^* \text{ and } g\left(k_\alpha^G + k_\beta^G\right) + s_\beta k_\beta^S = \frac{n}{n+1} k^*$$

In that case the distortion of investment is simply related to respective numbers of firms. Because the incentive to invest in aggregate capacity of a generalist firm is similar to the incentive of a peak firm, expressions are more simpler than in the previous case.

4 Number of firms and welfare

I consider here a specialized industry: $g = 0$ and $n = s_\beta + s_\alpha$ and consider the consequence on welfare of an increase of the number of firms of one group the other being fixed.

When firms are specialized the number of firms that have access to a particular technology influences both aggregate quantity of capacity and the technology mix. These effects explain that the number of firms that can invest in either technology has not a monotonic effect on welfare.

Let consider first that the number of baseload firms, s_α, is fixed. An increase of the number of peak firms increases quantities of aggregate capacity and peak capacity but decreases quantity of baseload capacity. So even if the aggregate quantity of capacity tends toward the optimal one as s_β grows there is a loss due to the distortion of the technology mix. More precisely, welfare is quasi concave with respect to the number of peak firms. There is an optimal number of peak firms and any increase of s_β beyond this number decreases welfare.

Some calculations (cf appendix A-2) give the following derivative of aggregate quantities with respect to the number of firms:

$$\frac{\partial}{\partial s_\beta}\left(s_\alpha k_\alpha^S\right) = -\frac{1}{A} s_\alpha \frac{v - c_\beta}{v - c_\alpha} k_\beta^S$$
$$\frac{\partial}{\partial s_\beta}\left(s_\beta k_\beta^S\right) = \frac{1}{A}\left(s_\alpha + 1\right) k_\beta^S$$

Abstracting from integer constraint, one can consider the derivative of welfare with respect to s_β. Injecting first order conditions gives the following expression :

$$\frac{dW}{ds_\beta} = \left(v - c_\alpha\right) k_\alpha^S \frac{\partial}{\partial s_\beta}\left(s_\alpha k_\alpha^S\right) + \left(v - c_\beta\right) k_\beta^S \frac{\partial}{\partial s_\beta}\left(s_\beta k_\beta^S\right)$$

As an increase of the number of peak firms has opposite effect on quantities of baseload and peak capacities the gain from the increase of the peak capacity is compensated by the loss from the decrease of baseload one. Whether one effect dominates the other depends on the relative quantity of firms.

Proposition 3 *Welfare is quasi-concave with respect to s_β, It is increasing if an only if*

$$\left(s_\alpha + 1\right) k_\beta^S(s_\alpha, s_\beta) \geq s_\alpha k_\alpha^S(s_\alpha, s_\beta)$$

Welfare is maximized at

$$s_\beta^*(s_\alpha) = \frac{v - c_\alpha}{s_\alpha\,(\delta - \Delta)}\left[1 - r_\alpha + (s_\alpha + 1)^2 (r_\alpha - r_\beta)\right]$$

This proposition is demonstrated in appendix A-3. The welfare loss is related to the asymmetry of firms but one should notice that both technologies are efficient and used at the first best optimum. The loss is not related to an inefficiency of new firms but a disequilibrium between both types of firms.

Concerning consumer surplus, at first sight it is unclear whether net consumer surplus increases with an increase of peak firms. An increase of the number of firms increases the aggregate quantity of capacity so it increases gross consumer surplus but it also decreases the quantity of baseload capacity so the price of electricity increases for some level of demand and consumers loose on these states.

Net consumer surplus is:

$$CS(k_\alpha, k_\beta) = \frac{1}{X}\left[\int_0^{k_\alpha} (v - c_\alpha)\, x dx + \int_{k_\alpha}^{k_\alpha + k_\beta} (v - c_\beta)\, x dx\right]$$

And derivatives with respect to each technology capacity are:

$$\frac{\partial CS}{\partial k_\alpha} = \frac{1}{X}\left[\delta k_\alpha + (v - c_\beta)\,(k_\alpha + k_\beta)\right]$$
$$\frac{\partial CS}{\partial k_\beta} = \frac{1}{X}\,(v - c_\beta)\,(k_\alpha + k_\beta)$$

So an increase of the number of peak firms modifies net consumers surplus of:

$$\begin{aligned}\frac{dCS}{ds_\beta} &= (v - c_\beta)\,\frac{s_\alpha}{AX} k_\beta^S \left[\left(s_\alpha k_\alpha^S + s_\beta k_\beta^S\right)\left(\frac{s_\alpha + 1}{s_\alpha} - \frac{v - c_\beta}{v - c_\alpha}\right) - s_\alpha k_\alpha^S \frac{\delta}{v - c_\alpha}\right] \\ &= (v - c_\beta)\,\frac{s_\alpha}{AX} k_\beta^S \left[s_\beta k_\beta^S \left(\frac{s_\alpha + 1}{s_\alpha} - \frac{v - c_\beta}{v - c_\alpha}\right) + k_\alpha^S\right]\end{aligned}$$

So consumers surplus is always increasing when an additive peak firm is active. The welfare loss is entirely supported by firms. However, consumers pay electricity at a higher price for some level of demand because of the decrease of baseload capacity but this loss is compensated by the overall increase of available capacity.

A similar analysis can be conducted on the number of baseload firms and similar results are obtained. If one more firm has access to baseload technology the quantity of baseload capacity increases whereas the quantity of peak capacity decreases. Even if the aggregate quantity of capacity increases welfare is not monotonic and there is an optimal number of firms that can invest in the baseload technology if the number of other firms is constant.

Proposition 4 *Welfare is quasi concave with respect to s_α , it is increasing if and only if*

$$(v - c_\alpha) \frac{s_\beta + 1}{s_\alpha} k_\alpha^S > (v - c_\beta) k_\beta^S$$

For any s_β there is an optimal number $s_\alpha^*(s_\beta)$ that writes

$$s_\alpha^*(s_\beta) = \frac{1}{s_\beta} \frac{1}{(r_\alpha - r_\beta)} \left[1 - r_\beta + (s_\beta + 1)^2 \frac{\delta - \Delta}{v - c_\beta} \right]$$

The proof is in appendix A-4. The effect on aggregate welfare of an increase of baseload or peak firms have similar qualitative properties on welfare. And consumers benefit from the entry of a baseload firm in all demand states because an increase of the number of baseload firms increases both the aggregate quantity of capacity and the quantity of baseload capacity so consumers gain in all demande states.

Those results can be compared to the analyses of an usual Cournot oligopoly with heterogeneous firms[9]. Within this standard framework an additive inefficient firm increases the aggregate production but decreases the production of efficient ones, the overall effect can be negative because of the reallocation of production from efficient to inefficient

9. In a recent paper Corchón (2008) provides a complete analysis of welfare loss with Cournot competition.

firms. Here, all firms are efficient, in the sense that both technologies are used at the first best optimum, nevertheless, an additive firm can decrease welfare by modifying the technology mix in the wrong direction.

5 Discussion

This section is devoted to the implication of a change of the load duration curve. The load duration curve used is a very rough simplification of a real one. It has been used to get explicit formula of quantities and further results. A more realistic representation that does not deeply modifies results is to consider that demand x is uniformly distributed on the set $[x^-, x^+]$ with $x^+ - x^- = X$. With this distribution of demand the first best optimum is depicted on figure (1.4). It can be seen that baseload capacity is translated of x^- whereas peak capacity is unchanged.

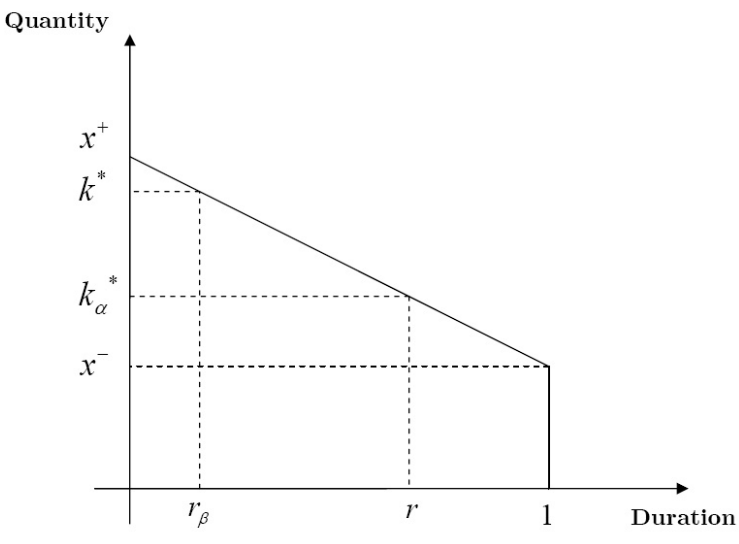

Figure 1.4: Load curve and optimal investment

Firms' choices are also modified, they all invest in greater quantities with this load duration curve. For instance, equilibrium individual capacity of a specialized baseload firm becomes:

$$k_\alpha^S = \frac{X}{A}\left[(g+1)\left(\frac{x^+}{X} - r_\alpha\right) + s_\beta\frac{\delta-\Delta}{v-c_\alpha}\right]$$

And other quantities change similarly. This change is not as benign as it might seem at first glance. The important consequence of this change is that the profit of baseload and generalist firms is not differentiable at $k_\alpha = x^-$ so they might invest in exactly the minimal quantity x^- of baseload technology and no more. In such case, on the short term, the baseload technology is never the marginal one and the price of electricity is never set at c_α.

Such situations arise if there are few baseload and generalist firms and $X = x^+ - x^-$ is sufficiently small, *i.e.* the load duration curve is sufficiently flat. Else, the analysis is not modified and results still hold.

6 Conclusion

In this chapter I analyzed investment by strategic firms in a simple electricity market perfectly regulated in the short term. This framework allowed to understand how firms have incentive both to decrease the aggregate quantity of capacity and distort the technology mix. The way the technological mix is distorted is related to the industry configuration. If one technology is not accessible to all firms there may be an overinvestment in the other technology.

Whereas both technologies are efficient, a firm that has access to both technologies does not invest in one of these technologies if there are too many specialized firms. If all firms are specialized an increase of a the number of firms of one type can decrease welfare. So, if the access to one technology is limited the number of firms active on the market via the other technology should be limited.

The development of competition via a unique technology whereas several are needed to ensure an optimal production of electricity is therefore questionable and can decrease welfare. Two major extensions are envisioned: the first would be to introduce an initial stage where firms can acquire either technologies by investing in a fixed cost, the second would be to analyse capacity markets and capacity payments that are implemented to correct investment incentives on electricity markets.

Chapter 2

Taille minimale et concurrence imparfaite

1 Introduction

This chapter deals with imperfect competition between firms when two technologies are available. One technology has a lower marginal cost than the other but is characterized by a minimum scale constraint (m.s.c.) while the other is perfectly divisible. The minimum scale is a quantity below which cost of production is prohibitive, so the low cost technology can only be used for production greater than the minimum scale. Equilibria of the duopoly game are characterized and comparative static results are derived. This game exhibits multiple equilibria among which some are asymmetric. Furthermore, the effect of the constraint on welfare is not monotonic. By constraining firms' choices the m.s.c. can improve welfare by limiting exercice of market power, but it can also prevent firms from using the low cost technology. This constraint also modifies usual comparative static results on cost shocks and on the number of firms.

This analysis originated in the observation of electricity markets and the nuclear 'renaissance'. The development and standardization of combined cycle gas turbine as small production technology was one argument pro the introduction of competition in the electricity industry. However the recent rise of gas price created a renewed interest for the nuclear technology. So the ability of electricity markets to promote investment in nuclear production plant is a sensitive issue. Nuclear production plants are characterized by a relatively high minimum scale that can prevent producers to invest in this technology. This minimum scale is not the size of one nuclear plant but the number of plants required to get a low marginal cost.

Whereas usual analysis of investment in generating capacity in an electricity market consider the interrelation between demand variability and the ratio of capacity and operating cost, the present analysis emphasizes the role of m.s.c. on competition. So the present work abstracts from main particularities of the electricity industry and focuses on the effect of this particular technological constraint on the outcome of a duopoly game. So quantities considered are production capacities and there is a constant demand function which should be interpreted as the demand for baseload production, and marginal costs are long term marginal cost (capacity cost plus variable costs).

The model used is general and related to the literature on discrete technology choice by strategic firms. Mills and Smith (1996) analyse the choice of technology in a linear duopoly game with two technologies available. Firms have to invest in a fixed cost in order to have access to a technology with a low marginal cost before they compete *à la* Cournot. Elberfeld (2003) extends the model of Mills and Smith to a n firms oligopoly. Both papers focus on welfare implications of the model. Mills and Smith's welfare analysis reveals that equilibria tend to have too little heterogeneity compared to the socially optimal industry configuration. In the general oligopoly case, Elberfeld (2003) establishes that, if there are more than three firms, firms over-invest compared to the social optimum because of a 'business stealing' effect. A firm that decides to invest do not consider the effect of its choice on the profit of others firms and this externality explains that there are two many investing firms at equilibrium. The same effect is at the root of the result of Mankiw and Whinston (1986) on excessive entry.

With the framework of Mills and Smith (1996), firms might be heterogeneous at equilibrium even though initially identical. This heterogeneity arises from the initial investment required to acquire the low cost technology [1]. This investment is a kind of scale economies. The present paper introduces them through another technological constraint. In both cases, there is a tension between competition and scale economies that explains heterogeneity. The market might not be large enough for all firms producing with the low cost technology. However, the game analyzed here is different because there is no cost to adopt the low cost technology. It is a simultaneous move game with no strategic effects related to the adoption of the technology. The m.s.c. has different effects on comparative static results than the fixed cost, because there is no cost to duplicate the adoption of efficient technology contrary there cannot be too much firms using the efficient technology.

The two modeling approaches of scale economies are used in the entry deterrence literature. Even if less popular than the fixed cost approach used by Dixit (1980) in his seminal paper, minimum scale has been introduced by Schmalensee (1981). Both explains that an incumbent can deter entry by investing in sufficient production capacity. Maskin (1999) uses this approach to establish that uncertainty decreases the attractiveness of entry deterrence. To my knowledge despite in the entry deterrence literature, this technological constraint has not been introduced in simultaneous oligopoly.

Results obtained are both positive and normative, positive ones come from the analysis of existence of asymmetric equilibria and normative ones are obtained through a comparative static study of these equilibria.

First result obtained is a positive one on equilibria of the duopoly game. Depending on the minimum scale, it is shown that there are

1. Actually even in a standard Cournot free entry game as analyzed by Novshek (1980) heterogeneity arises because firms that can initially decide whether to enter are initially identical and at equilibria firms that entered are different from those that do not.

symmetric as well as asymmetric outcomes. In a symmetric equilibrium firms use the same technology to produce the same quantity whereas in a asymmetric equilibrium one firm uses the low cost technology and the other the high cost one. Furthermore, for a range of parameter values the game exhibits three equilibria: a symmetric and two asymmetric ones.

The emergence of heterogeneity is technically explained by the particular shape of reaction functions [2]. For any strictly positive minimum scale the reaction function of a firm jumps downward because the firm switches from the low cost technology to the high cost one. The m.s.c. determines the precise position of this jump and whether there are asymmetric equilibria. The three situations that can occur are represented on figure 2.1. For small minimum scale there is a unique equilibrium which is symmetric, both firms use the low cost technology. For an intermediary range of minimum scale, two asymmetric equilibria appear and the symmetric one is lost. And finally for a sufficiently high minimum scale, a symmetric equilibrium with both firms using the high cost technology is added to the two asymmetric equilibria mentioned. At an asymmetric equilibrium, the firm that uses the high cost technology does so because it is more profitable to produce few units with the high cost divisible technology in order to keep the price high.

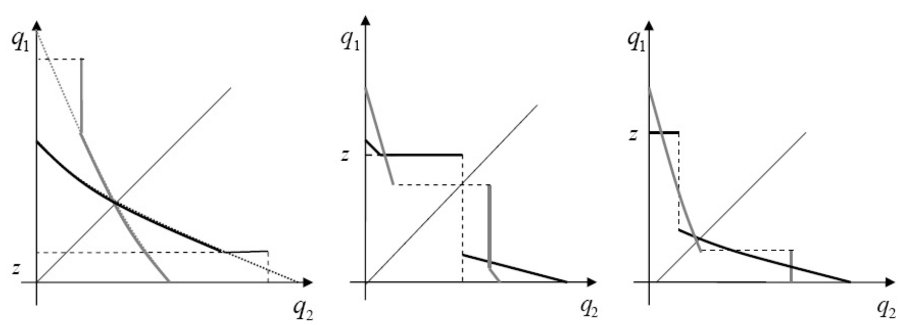

Figure 2.1: Reaction functions and equilibria

2. Amir et al. (2006) analyse a general class of game with a similar feature that explains the existence of asymmetric equilibria.

From those results, the second part of the paper analyses the welfare implications of the m.s.c. and the cost of the unconstrained technology and the effect of entry. It is established that the effect of the minimum scale constraint is not monotonic. This constraint can prevent firms from using the low cost technology but it can also prevent them from reducing quantities to increase price. This two contradictory effects explains that a decrease of the minimum scale can have positive as well as negative effect on welfare. Therefore, the standardization of the low cost technology can either increase or decrease welfare.

Furthermore, an increase in the cost of the high cost technology can increase welfare and production by increasing the incentive to use the low cost technology. Compared to previous analysis on cost shock [3], the increase in welfare is not only related to a decrease in production cost but also to an increase in quantities produced. A similar result is obtained on the number of firms. Even if firms are symmetric, the duopoly outcome can be worse than the monopoly one not only for producers but also for consumers. This result is obtained in the particular case where both firms use the high cost technology at equilibrium whereas the monopoly uses the low cost one. In other cases, even if one firm uses the high cost technology at equilibrium welfare is higher with a duopoly than with a monopoly because there is no decrease of production from the low cost technology thanks to the minimum scale constraint.

The chapter is organized as follow. The model is introduced in the following section, before solving the duopoly game (section 3) and conducting a comparative static analysis (section 4).

3. Lahiri and Ono (1988) states that the increase of the marginal cost of a firm with a low market share can increase welfare. Several authors have deepened this analysis, the most recent being Corchón (2008) that analyses welfare loss with Cournot competition.

2 The model

I analyze an homogenous good market characterized by a continuous and decreasing inverse demand function $p(q)$ where q is the aggregate quantity produced. Two technologies are available to produce the good labeled $t = \alpha, \beta$, with constant marginal cost c_t. Technology α has a lower marginal cost than technology β: $c_\alpha < c_\beta$ but characterized by a minimum production scale z. A firm cannot produce less than z units of the good with technology α.

There are two identical firms that compete *à la* Cournot by choosing individual production non cooperatively. The individual quantity chosen by a firm $i = 1, 2$ is denoted q_i, hence $q = q_1 + q_2$. Moreover, firms simultaneously choose the technology they use. The technology used by firm $i = 1, 2$ is denoted t_i. Its profit is a function of its production, its rival's one and the technology it uses so it is $\pi(q_i, q_j, t_i)$ where:

$$\pi(x, y, t) = (p(x + y) - c_t)\, x \tag{2.1}$$

A firm maximizes its profit by choosing how much to produce with which technology. The strategy of firm i is a couple (q_i, t_i) of quantity produced and technology used [4].

The following assumptions are made on the price function:

A1- It is worth producing with either technology: $p(0) > c_\beta$

A2- the price function is strictly decreasing on the set $[0, \overline{q}]$ with $\overline{q} > 0$ and null for any $q \geq \overline{q}$.

A3- It is twice differentiable on $[0, \overline{q}]$ and satisfies $p''(q) + p'(q)q < 0$ for all $q \in [0, \overline{q}[$.

The last assumption common in oligopoly literature [5] implies that the marginal revenue of a firm is decreasing with respect to the production of

4. Alternatively, one can consider that a firm chooses its technology by minimizing its cost of production, and the strategy could be limited to the choice of a quantity q_i, it would not modify the analysis.

5. Novshek (1985) introduces this assumption and establishes existence and uniqueness of Cournot equilibrium. Amir (1996; 2005) analyses in depth the implication of this assumption and its relaxation on the outcome of Cournot oligopoly.

its rival. So quantities are strategic substitute and existence and uniqueness of Cournot equilibrium are ensured if firms have convex cost. It is equivalent to assume that functions $x \rightarrow p'(x+y)x$ are decreasing for all y.

The 'unconstrained' reaction function with technology $t \in \{\alpha, \beta\}$ is denoted $r_t(y)$. It is the profit maximizing quantity of a firm with technology t when its rival produces $y \geq 0$ and without the minimum scale constraint, *i.e.* $z = 0$. For each technology let's denote $\overline{q}_t = p^{-1}(c_t)$. For all $y \geq \overline{q}_t$, a firm with technology t does not produce: $r_t(y) = 0$, and for $y \in [0, \overline{q}_t]$ the reaction function is the unique solution of the first order condition so:

$$p(r_t + y) + p'(r_t + y)r_t = c_t$$

Thanks to assumption A3 made on the price function, the reaction function is strictly decreasing with a slope above -1 on $[0, \overline{q}[$:

$$\forall y \in [0, \overline{q}_t[\,, r_t'(y) = -\frac{p' + p''r_t}{2p' + p''r_t} \in]-1, 0[$$

This property ensures existence and uniqueness of Cournot equilibria with unconstrained technology. For $t = \alpha, \beta$, equilibrium quantities when both firms have a marginal cost c_t without minimum scale constraint are denoted q_t^S. And q_t^A for $t = \alpha, \beta$ denote quantities produced at equilibrium when one firm has a marginal cost c_α and the other c_β.More specifically these quantities satisfy:

$$q_t^S = r_t(q_t^S), t = \alpha, \beta$$

$$q_\alpha^A = r_\alpha(q_\beta^A) \text{ and } q_\beta^A = r_\beta(q_\alpha^A)$$

Finally I assume that in the case of asymmetric duopoly both firms produce: $q_\beta^A > 0$ which is equivalent to $r_\alpha(0) < \overline{q}_\beta$. Reaction functions and the various quantities introduced are represented figure 2.2.

He compares this assumption to the assumption of logconcavity that is also commonly used for existence and uniqueness of Cournot equilibrium but has a different economic interpretation.

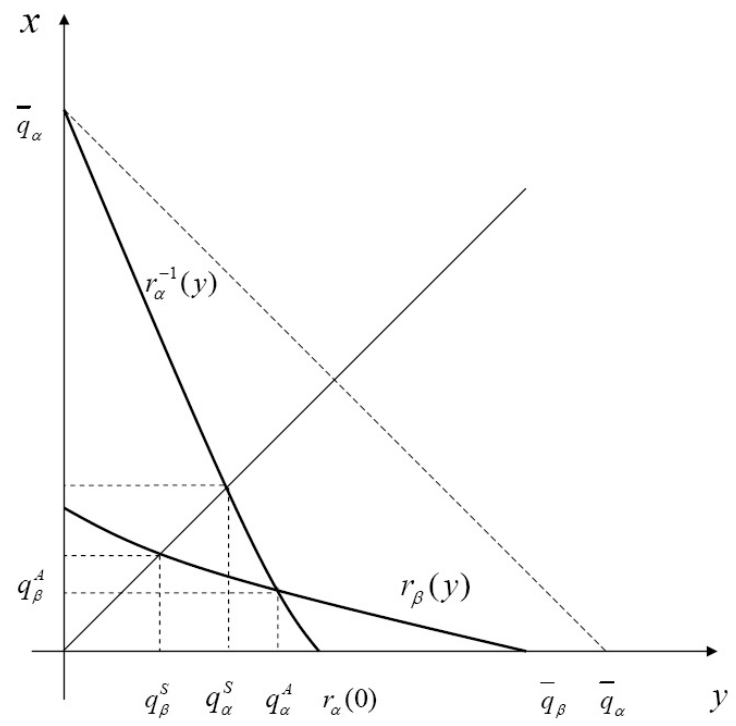

Figure 2.2: Unconstrained equilibria and notations

3 Equilibria and minimum scale

In order to solve the game I first describe the particular shape of the reaction function of a firm when minimum scale is introduced. Once reaction function is fully described the duopoly game is solved.

3.1 Reaction function

As firms are identical they have the same reaction function constituted of two components: the quantity produced and the technology used. Formally, the reaction function of a firm is $\varphi(y) = (r(y), t(y))$ where $r(y)$ is firm's production and $t(y)$ its technology when its rival produces y. It will be established that as its rival's production increases a firm first uses the low cost technology at the unconstrained level, then, for an intermediary region it produces exactly at the minimum scale and finally it switches and uses the high cost technology.

If the firm uses the high cost technology: $t(y) = \beta$, then it produces the best response: $r(y) = r_\beta(y)$ but if it uses the low cost technology several cases should be distinguished whether the m.s.c. is binding or not:

$$\text{If } t(y) = \alpha,\ \begin{array}{l} r(y) = r_\alpha(y) \text{ if } r_\alpha(y) > z \\ r(y) = z \ \text{ if } r_\alpha(y) \le z \text{ and } z \le \overline{q}_\alpha - y \\ r(y) = 0 \text{ otherwise} \end{array}$$

The m.s.c. is binding if the unconstrained best response is lower than the minimum scale: $z > r_\alpha(x) \ge 0$, because of profit concavity if $r_\alpha(y) \le z$ profit is maximized either at z or at 0. If the rival's production plus the minimum scale drive the price down to the marginal cost ($z + y \ge \overline{q}_\alpha$) the firm does not produced.

The firm chooses its technology by comparing maximum profits with each. I assume that in case of equality technology β is preferred, so if $r(y) = 0$ then necessarily $t(y) = \beta$. The choice of technology is described by the following equivalence:

$$t(y) = \alpha \text{ if and only if } \max_{x \ge z} \pi\,(z, y, \alpha) > \pi(r_\beta(y), y, \beta) \qquad (2.2)$$

If the minimum scale is prohibitively high it is never worth using technology α. It is the case if $z \ge \overline{z}$ where $\overline{z}$ is defined by:

$$\overline{z} \in [r_\alpha(0), \overline{q}_\alpha] \text{ and } (p(\overline{z}) - c_\alpha)\,\overline{z} = (p(r_\beta(0)) - c_\beta)\,r_\beta(0)$$

This quantity is well defined [6], it is the minimum quantity at which a monopoly prefers the high cost technology to the low cost one.

The choice of technology has the intuitive property that a firm switches only once from the low cost to the high cost technology when its rival production increases. If a firm chooses technology β for some rival production it chooses this technology for all greater quantities. This result is precisely set in the following Lemma:

6. Because the left hand side $(p(\overline{z}) - c_\alpha)\,\overline{z}$ is continuous and strictly decreasing on the set $[r_\alpha(0), \overline{q}_\alpha]$ and strictly greater (resp. smaller) than the right hand side at the minimum (resp. maximum) of this set.

Lemma 1 *For all $y', y'' \in [0, \overline{q}]$ with $y'' > y'$ the following implications hold:*

$$\text{If } t(y') = \beta \text{ then } t(y'') = \beta$$
$$\text{If } t(y'') = \alpha \text{ then } t(y') = \alpha$$

The proof is in appendix B-1, and relies on assumption A.3. With this assumption the profit with technology α and production z decreases more than the profit with technology β when the rival production y increases [7] so if the former is smaller than the latter at a rival production y' it would remain the case for any greater rival productions y''. Hence, a firm switches only once from one technology to the other when its rival output increases.

With this Lemma, the precise shape of the reaction function φ can be described.

Proposition 5 *There are $y^-(z), y^+(z)$ with $0 \leq y^-(z) \leq y^+(z) \leq \overline{q}_\alpha$ such that the reaction function of a firm is:*

- *If $0 \leq y \leq y^-(z) : \varphi(y) = (r_\alpha(y), \alpha)$*
- *If $y^-(z) \leq y < y^+(z) : \varphi(y) = (z, \alpha)$*
- *If $y^+(z) \leq y : \varphi(y) = (r_\beta(y), \beta)$*

This proposition is directly obtained from previous Lemma. The proposition does not precisely describe the threshold quantities y^- and y^+ at which the m.s.c. starts binding and at which the firm switches to technology β. A precise description of these quantities requires further work, and the end of this section is devoted to do so.

The low threshold can be directly derived [8] from the reaction function r_α because it is the rival production that equalizes the best response with the minimum scale :

$$y^-(z) = \begin{cases} r_\alpha^{-1}(z) & \text{if } 0 \leq z \leq r_\alpha \\ 0 & \text{if } r_\alpha(0) \leq z \end{cases}$$

7. With the alternative property of price logconcavity, assumption A3 is still satisfied for productions greater than the monopoly one (cf Amir 2005) so the result would remain valid.

8. With the slight abuse of notation r_α^{-1} is used to denote the inverse of the function r_α restricted to the set $[0, \overline{q}_\alpha]$.

The description of the high threshold is more complicated and only a partial description of it will be obtained. Three situations have to be distinguished, if z is above $\overline{z}$ the firm never uses technology α and $y^+ = 0$. And similarly, if the minimum scale is small the firm never uses technology β or more precisely it only uses it for a null production.

First, if $z \leq \overline{z}$, $y^+(z)$ is the unique solution (cf. appendix B.B-2) on the set $[y^-(z), \overline{q}_\alpha - z]$ of equation:

$$\pi\left(z, y, \alpha\right) - \pi(r_\beta(y), y, \beta) = 0 \tag{2.3}$$

As the left hand side is decreasing both in z and y, the solution y^+ is a decreasing function of z. This monotonicity is also an 'intuitive' property of the choice of technology. Furthermore, the switch to technology β occurs for a null production if z is sufficiently small. This situation arises if $\pi\left(z, \overline{q}_\beta, \alpha\right) > 0$ i.e. $z < \overline{q}_\alpha - \overline{q}_\beta$. In that case a firm switches to technology β at $y^+ = \overline{q}_\alpha - z$ to stop producing. The following lemma describes the shape of this quantity.

Lemma 2 *The quantity $y^+(z)$ satisfies:*
For $0 \leq z \leq \overline{q}_\alpha - \overline{q}_\beta$,

$$y^+(z) = \overline{q}_\alpha - z$$

For $\overline{q}_\alpha - \overline{q}_\beta \leq z \leq \overline{z}$,

$$y^+(z) \textit{ has a slope strictly smaller than } -1$$

For $z \geq \overline{z}$, $y^+(z) = 0$

The proof of the lemma is in appendix B-3, and mainly consists in differentiation of equation (2.3). In the intermediary region, an increase of one unit of the minimum scale constraint reduces the switching quantity y^+ of more than one unit. The reaction function is represented on figure 2.3 for the interesting situation where $\overline{q}_\alpha - \overline{q}_\beta \leq z \leq r_\alpha(0)$, for this intermediate range the minimum scale constraint is not always binding and the firm effectively switches to technology β for high rival production.

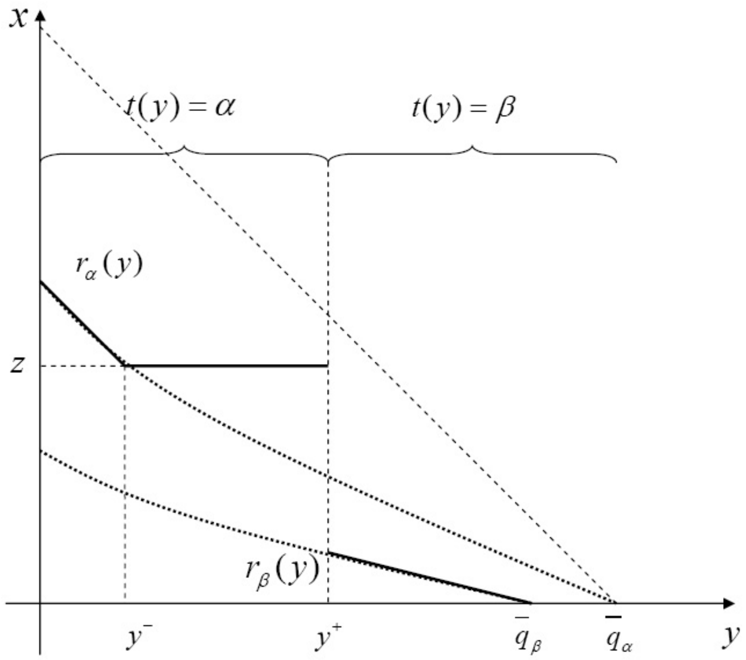

Figure 2.3: Reaction function of a firm with a minimum scale constraint

3.2 Equilibria

Starting with the analysis of reaction function made in the previous section it is now possible to graphically determine the outcome of the duopoly game by drawing both reaction functions and considering intersections. This is done on figure 2.4 for two particular values of z. The jump of the reaction function at y^+ has a great impact on the outcome of the game. As it is a downward jump the existence of at least one equilibrium is guaranteed. But the number of equilibria depends on the minimum scale z. For extreme values of z there is a unique symmetric equilibrium in pure strategies but for intermediate values there are two asymmetric equilibria.

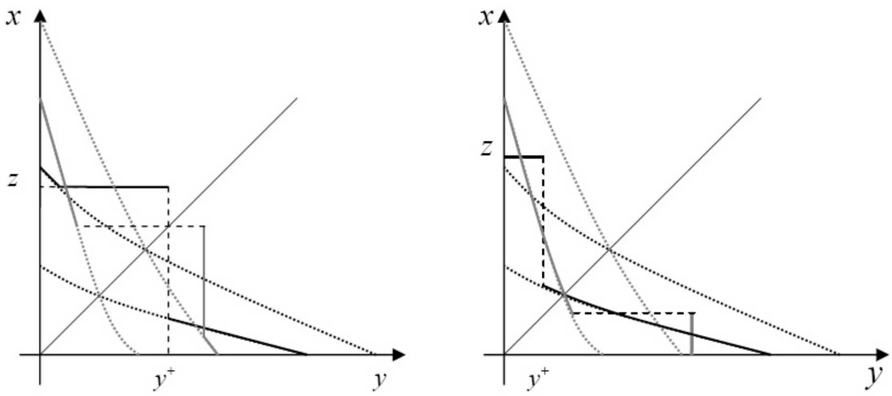

Figure 2.4: Minimum scale and equilibria

With low minimum scale the situation is similar to a duopoly with the low cost technology, the minimum scale constraint being eventually binding. Similarly for very high values of the minimum scale, technology α is not used at equilibrium and the situation is similar to a duopoly with marginal cost c_β. But, for intermediate values there are two asymmetric equilibria with one firm using the low cost technology and the other firm the high cost one. In some cases two asymmetric and one symmetric equilibria coexist as illustrated on figure 2.4. The precise occurrence of equilibria are described in the following proposition:

Proposition 6 *There are z_1, z_2, z_3, z_4 with:*

$$0 < z_1 \leq z_2 \leq z_3 < z_4 \leq \overline{z}$$

such that:

1. *There is a symmetric equilibrium with $t_1 = t_2 = \alpha$ if and only if $0 \leq z < z_2$.*
2. *There are two asymmetric equilibria with $t_i = \alpha, t_j = \beta$ for $i \neq j$ if and only if $z_1 \leq z < z_4$.*
3. *There is a symmetric equilibrium with $t_1 = t_2 = \beta$ if and only if $z_3 \leq z$.*

where z_1, z_2, z_3, z_4 *satisfy*

$$\begin{aligned} y^+(z_2) &= z_2 \\ y^+(z_1) &= \max\left\{z_2, q_\alpha^A\right\} \\ y^+(z_3) &= q_\beta^S \\ y^+(z_4) &= r_\beta(z_4) \end{aligned}$$

The proof is in appendix B-4. Notations $z_l, l = 1, .., 4$ are used through the rest of the paper, to denote the quantities defined in the proposition. Figure 2.5 graphically represents the resolution of the proposition. Both quantities y^-, y^+ are represented with reaction functions and the quantities introduced in the proposition. For instance for $z_2 \leq z \leq z_3$ the quantity $r_\beta(z)$ is less than the switching production y^+ whereas z is above, so the couple of strategies $((z, \alpha), (r_\beta(z), \beta))$ is an equilibrium. The three situations enunciated in proposition 6 are indicated and for situation 1 of symmetric equilibrium with technology α two subcases are distinguished whether the m.s.c. is binding (1a) or not (1b).

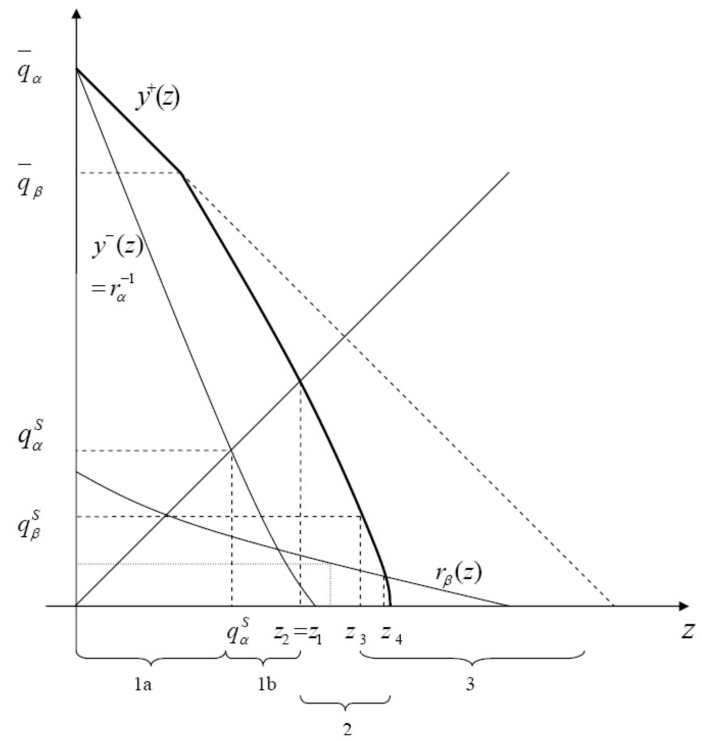

Figure 2.5: Equilibria of the duopoly game with respect to z

The coexistence of asymmetric and symmetric equilibria depends on the value of the minimum scale. Whereas there is always some coexistence of situations 2 and 3 because $z_3 < z_4$, it is unclear whether situations 1 and 2 can coexist. It is so if z_2 that solves $y^+(z) = z$, is smaller than the production q_α^A. On figure 2.5, $z_1 = z_2$ and equilibrium 1 never coexists with equilibrium 2. Even if it does not seem to be a general result, it can be stated for a convex price function.

Corollary 3 *If p is convex $z_1 = z_2$ and $\left((q_\alpha^*, \alpha), (q_\beta^*, \beta)\right)$ is an equilibrium only if $q_\alpha^* = z$ and $q_\beta^* = r_\beta(z)$.*

The proof is in appendix B-5 and shows that $\left((q_\alpha^A, \alpha), (q_\beta^B, \beta)\right)$ cannot be an equilibrium if the price function is convex. With a convex price function the effect of an increase of production on price is lower the higher production is. This feature explains that the revenue $p(2q_\alpha^A) - c_\alpha$ is higher than $p(q_\alpha^A + q_\beta^A) - c_\beta$ and this guarantees that for a minimal scale lower than the asymmetric production q_α^A a firm does not switch for a rival production of q_α^A i.e. $t(q_\alpha^A) = \alpha$. It will be assumed to hold in most results in the rest of the paper.

The analyze of equilibria can be detailed by distinguishing several subcases whether the minimum scale constraint is binding or not at equilibrium. It is done on figure 4 where 1a (resp. 1b) is used for a symmetric equilibrium with technology α and unbinding (resp. binding) m.s.c.. Similarly asymmetric equilibria can be distinguished, but in the situation depicted on figure 4 as established in corollary 1, the minimum scale constraint is always binding at asymmetric equilibria. In such case, the minimum scale constraint can eventually deter a firm from producing if $\overline{q}_\beta \leq \overline{z}$ and $\overline{q}_\beta \leq z \leq \overline{z}$.

4 Welfare implications

Several welfare implications can be derived from the analysis of the duopoly game. I consider here the effect of minimum scale, the effect of the marginal cost c_β and finally entry. Compared to previous similar comparative static analysis the game exhibits original features related to the m.s.c.. So I begin by a discussion on the effect of minimum scale on welfare. Welfare is defined as the sum of consumer net surplus and firms profit:

$$W\left(\left(q_1, t_1\right), \left(q_2, t_2\right)\right) = \left[S(q) - pq\right] + \pi_1 + \pi_2$$

Where consumer gross surplus $S(q)$ is defined as $S(q) = \int_0^q p(u)du$.

4.1 Minimum scale

The m.s.c. has two contradictory effects that explain that welfare is not monotonic with respect to z. First, by constraining the choice of firms, it limits market power and can therefore increase welfare. Though, if too constraining it can also prevent firms from using the low cost technology.

For instance, in the monopoly situation, when the minimum scale constraint is smaller than $\overline{z}$ an increase of this constraint increases welfare by increasing firm's production, it strictly increases welfare on the set $[r_\alpha(0), \overline{z}]$, but at $\overline{z}$ welfare jumps downward as the firm does not use the low cost technology for higher minimum scale.

With a duopoly, welfare evolution with respect to the minimum scale exhibits similar downward jumps and the multiplicity of equilibria raises the issue of equilibrium selection. When several equilibria coexist the one with most firms using the low cost technology is better than the other. It is not Pareto dominant because one firm has a lower profit in the latter case. It is formally stated in the following corollary for the case of coexistence of equilibria 2 and 3.

Corollary 4 *For* $z_3 \leq z < z_4$:

Welfare is higher at asymmetric equilibria than at the symmetric one but firms 2 profit is lower at equilibrium $((x, \alpha), (r_\beta(x), \beta))$ *than at equilibrium* $\left(\left(q_\beta^S, \beta\right), \left(q_\beta^S, \beta\right)\right)$ *where* $x = \max\{q_\alpha^A, z\}$.

Proof. At asymmetric equilibria aggregate production is greater and more efficiently produced than at the symmetric equilibrium and $x > q_\beta^S$ implies that the profit $\pi(r_\beta(x), x, \beta)$ is lower than the profit $\pi(q_\beta^S, q_\beta^S, \beta)$.

■

Qualitative properties of welfare evolution with respect to z are represented on figure 5 and formally described in the following corollary.

Corollary 5 *If* $z_1 = z_2$,

An increase of z *inside sets* $[0, z_2[$ *and* $[z_2, z_3[$ *increases welfare.*

There is $\varepsilon > 0$ *such that an increase of* z *from* $[z_2 - \varepsilon, z_2[$ *to* $[z_2, z_2 + \varepsilon]$ *decreases welfare.*

Any increase of z *from* $[0, z_3[$ *to* $[z_4, +\infty[$ *decreases welfare.*

Proof. Welfare can be expressed as a function of z on $[0, z_3]$:

For $z \in \left[0, q_\alpha^S\right]$ welfare is constant: $S(2q_\alpha^S) - 2c_\alpha q_\alpha^S$

For $z \in \left[q_\alpha^S, z_2\right[$ welfare is: $S(2z) - 2c_\alpha z$

And for $z \in [z_2, z_3[$ welfare is: $S(z + r_\beta(z)) - c_\alpha z - c_\beta r_\beta(z)$

And for $z \epsilon [z_4, +\infty[$ welfare is: $S(2q_\beta^S) - 2c_\beta q_\beta^S(z)$

As a consequence welfare is continuous, differentiable by part, and increasing with respect to z on each sets. As $S(2z_2) - 2c_\alpha z_2 > S(z_2 +$

$r_\beta(z_2)) - c_\alpha z_2 - c_\beta r_\beta(z_2)$ it jumps downward at z_2. And at any equilibrium with at least one firm using the low cost technology welfare is higher than at the symmetric equilibrium with both firms using technology β. ■

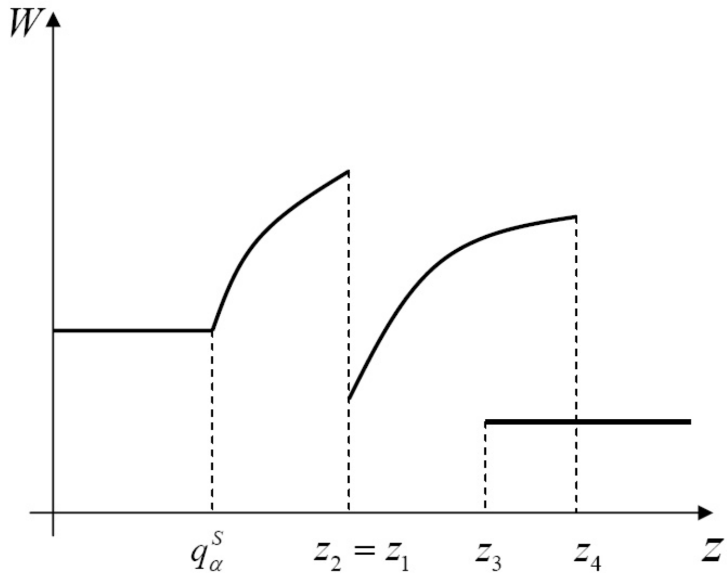

Figure 2.6: Welfare with respect to minimum scale

A decrease of the minimum scale can be interpreted as a standardization of the low cost technology. For instance, in the electricity sector, there are several projects to develop small standardized nuclear reactors that would therefore be less constraining for electricity producers. Whether this development would have a positive effect on nuclear investment depends on many phenomena not considered here, but, even if the model developed is far from being a complete representation of the electricity industry, it helps to understand the effect of such standardization on competition. The particular effect stresses here is that a standardization of a technology can increase or decrease welfare by reinforcing or softening competition between strategic firms. If the standardization leads to more firms using the low cost technology welfare is increased.

4.2 Cost increase

In this section I analyze the effect of a change of the high cost c_β on equilibria and welfare. In a framework without minimum scale, an increase of the cost of an inefficient firm can increase welfare by reallocating production from high cost to low cost firms. In such case, aggregate production decreases but the loss of consumers is more than compensated by the cost effect. The profit of the low cost firm increases at the expense of consumers and the inefficient firm (Lahiri and Ono 1988).

With the minimum scale constraint these results are modified in either directions. If the minimum scale constraint is binding a small increase of the high cost always decreases welfare because the cost reallocation effect does not occur. But, if the cost increase modifies the qualitative properties of the equilibrium of the game it can not only increases aggregate welfare but also consumer surplus. Furthermore, in the present approach firms are initially identical and asymmetry is endogenous which is not the case in usual analysis of welfare effect of cost shocks. And as results can be opposite, it is an important matter to know whether asymmetry is endogenous to conclude that 'helping minor firms reduces welfare' (Lahiri and Ono 1988).

In order to formally analyze the effect of c_β relevant quantities are written as functions of c_β.

Proposition 7 *If* $z_1(c_\beta) = z_2(c_\beta)$*, there is* z *such that for any* $\epsilon \in]0, p(r_\alpha(0)) - c_\beta[$ *an increase of the marginal cost from* c_β *to* $c_\beta + \epsilon$ *strictly increases welfare and aggregate production.*

Proof. At $z = z_2(c_\beta)$ for a marginal cost of c_β the only equilibria of the game are asymmetric $((z, \alpha), (r_\beta(z), \beta))$ and $((r_\beta(z), \beta), (z, \alpha))$ and welfare is identical at both. Whereas for a marginal cost of $c_\beta + \epsilon$ for any $\epsilon \in]0, p(r_\alpha(0)) - c_\beta[$ at $z_2(c_\beta)$ the only equilibrium is $((z, \alpha), (z, \alpha))$ because $z_2(c_\beta + \epsilon) > z_2(c_\beta)$ and $z_1(c_\beta + \epsilon) = z_2(c_\beta + \epsilon)$. So production is increased because $z > r_\beta(z)$ and welfare is increased because firms produce more at lower marginal cost. ■

Welfare and production are increased consequently of a cost increase because the low cost technology is more attractive after this shock so one firm that uses the high cost technology in the initial situation uses the low cost technology after it. It implies that, for instance, consumers can benefit of the increase of gas prices because it increases investment in nuclear plants.

4.3 Entry

For similar reasons as those mentioned about an increase of cost, an increase of the number of inefficient firms can decrease welfare in an 'unconstrained' Cournot oligopoly. The entry of an inefficient firm can decrease welfare by decreasing efficient production. Nevertheless, consumers benefit from this entry as aggregate production increases. When fixed entry cost are considered, entry can decrease welfare if welfare gains from increased production are offset by the fixed entry cost. Typically, as firms do not consider the effect of their entry on other firms profit there is excessive entry in a free entry Cournot game (Mankiw and Whinston 1986). In that case also, consumers still benefit from entry.

In the particular game analyzed here, the minimum scale constraint modifies these results in two directions. First, because it constrained the choice of firms the minimum scale constraint can prevent the production reallocation effect. Even if the firm that enters use the high cost technology, if the other keep using the low cost technology entry does not decrease welfare because it does not decrease the low cost production. But for particular parameters values, entry can not only decrease welfare but also decrease consumer surplus by decreasing production. Whether this last situation arises depends on the comparison of the aggregate duopoly production and the higher minimum scale that does not prevent a monopoly from using the low cost technology.

Corollary 6 *If $z_1 = z_2$, welfare is higher with a duopoly than with a monopoly for all $z \in [z, z_3[$*

If $2q_\beta^S < \overline{z}$, for all $z \in \left[\max\{2q_\beta^S, z_3\}, \overline{z}\right[$ welfare and aggregate production are lower at the duopoly symmetric equilibrium than with a monopoly.

To conclude, I establish that in the linear case there are parameters values such that assumptions are satisfied and $2q_\beta^S < \overline{z}$. Let consider a linear price function $p(q) = a - q$ and normalize $c_\alpha = 0$. The assumption that both firms produce at the unconstrained asymmetric equilibrium implies that $c_\beta < a/2$. The aggregate production of the Cournot duopoly with high cost technology is $2q_\beta^S = 2\left(a - c_\beta\right)/3$ and $\overline{q}_\beta < \overline{z}$ is equivalent to:

$$p(q_\beta^S)q_\beta^S > \left(p(r_\beta(0)) - c_\beta\right) r_\beta(0) = \frac{\left(a - c_\beta\right)^2}{4}$$

Some calculations lead to the equivalence: $q_\beta^S < \overline{z}$ if and only if $c_\beta \leq a/25$. So for $a/25 < c_\beta < a/2$ the situation of the corollary arise. There is a potential issue of selection of equilibria if asymmetric equilibria coexist with the symmetric one considered. Nevertheless, in the linear model $\overline{q}_\beta \leq \overline{z}$ is equivalent to $c_\beta \geq a/5$ so for $a/25 < c_\beta < a/5$ the set $\left[\max\{2q_\beta^S, z_4\}, \overline{z}\right[$ is not empty and there are values of z such that there is a unique symmetric equilibrium in the duopoly case at which both welfare and production are lower than with a monopoly.

5 Conclusion

In this chapter I analyzed how the effect of a minimum scale constraint on the choice of technology and production of strategic firms. It has been shown that heterogeneity can arise at equilibrium. Because of the minimum scale constraint firms have discontinuous reaction functions, this feature explains that asymmetric equilibria occur. Less formally, because of the minimum scale constraint on the low cost technology there might be not enough 'space' for both firms using the low cost technology at equilibrium.

Several implications on welfare have been obtained. There is an issue of equilibrium selection when several equilibria arise. Particularly when there are two asymmetric equilibria and a symmetric one, welfare is strictly lower at one of them but none is Pareto dominant. The effect of the minimum scale on welfare is not monotonic because it can either increase welfare by limiting market power, or decrease welfare by preventing at least one firm from using the low cost technology. Results obtained on a shock on the cost of the high cost technology are original because even with symmetric firms an increase of production cost can increase welfare and consumers surplus by increasing the incentive to use the low cost technology.

Part II

Pouvoir de marché et intervention publique

Chapter 3

Oligopole mixte et investissement

1 Introduction

In a number of oligopoly markets, private and public firms produce similar commodities and compete. Recently liberalized network industries are particularly relevant examples of this situation. In electricity markets there are several kinds of public firms that intervene. First, some producers are partly owned by government and second, system operators may invest in generating units in some countries (Norway, Finland and Sweden). This last kind of investment is usually justified by a particular public good: security of supply. As market power and public good effects could not be easily disentangled, system operators correct both simultaneously. I focus on the underinvestment resulting from market power exercise. In this chapter, I analyse sequential capacity choice in a mixed oligopoly, in which the benevolent public firm is a follower. The public firm is a follower for I consider that it tries to correct an observed lack of private investment but I also analyse the case of simultaneous moves.

Investment in, and production with, peak capacity by system operators (either directly or indirectly via contracts with producers) is generally criticized as decreasing incentive of private producers to invest in peaking units. It is a component of the 'missing money' argument

(Joskow 2006). It has been formally analyzed by Joskow and Tirole (2006) in a context of perfect competition. If market imperfections, the public good nature of peaking units [1] and imperfect competition, are not considered it is clear that such actions of the system operator can only impede markets' efficiency. Joskow and Tirole (2006) analyze how this intervention, and the recovery of its costs through an uplift on electricity prices, can distort the technology mix by decreasing incentive to invest in peaking units. Compared to their contribution, only one technology is considered but market power is introduced. So the intervention of the public firm is justified as it is used to correct market power on the long term by increasing capacity, and on the short term by increasing production. This work is at the crossing of the mixed oligopoly literature and the 'commitment' literature.

Initiated by the pioneering work of Merrill and Schneider (1966), mixed oligopoly models analyse situations where private profit-maximizing firms compete with public benevolent ones. They traditionally analyse the effect of concentration and of the order of move in a Cournot-Nash framework (de Fraja and Delbono 1990, Cremer et al. 1989, Pal 1998). Usually, assumptions about cost structure are made to explain that the public firm does not produce the welfare maximizing quantity alone. Disregarding this explanation, I focus on the influence of short term market power. This work is therefore more related to the literature on commitment. Indeed, dynamic models of capacity choice have been extensively analyzed in the context of duopoly games with a leader and a follower. Spence (1977) and Dixit (1979; 1980) model such 'commitment game' to analyse entry deterrence. In this line, and using a three-stage game, Ware (1984) emphasizes the commitment value of investment and the relationship with the share of sunk cost. When the follower decides how much capacity to build, he moves along the short term reaction function of the leader. Sunk costs allow the leader to shape its short term reaction function. The larger the sunk costs, the

1. The public good characteristic of peaking units is linked to the risk of network collapse. It is analyzed by Joskow and Tirole (2007) independently from the issue of 'ISO procurement'.

more 'constrained' is this shape, making it more difficult for the entrant to influence the production choice of the incumbent. I use a similar methodology with a benevolent follower and analyse how the change of the follower's objective modifies the leader strategic advantage.

The recent work of Lu and Poddar (2005) is the closest to the present one. They analyse the influence of timing in an investment game between a private and an inefficient public firm in a linear Cournot model where firms choose capacity scales before production. They analyse whether excess capacity may be installed and establish that, whatever the timing of the game is, the private firm never chooses under capacity while the public firm never chooses excess capacity. Because of its relative inefficiency, the public firm limits its investment in order to increase private production. The present approach is different as I consider strong capacity constraints and analyse the interplay between long term and short term market power. Capacity constraints play a crucial role in the present model, as in the literature on commitment. They are an important characteristic of electricity markets where short term competition is constrained by installed capacity. The share of sunk cost in capacity investment, and the time needed to develop new capacities explain that capacity could not be modified on the short term.

Indeed, since costs are assumed linear and the public firm as efficient as private ones, it may appear obvious that welfare could be maximized by the public firm intervention. One may think that the public firm only has to invest in missing capacities to reach the long term social optimum equalizing price with long term marginal cost. Actually, I establish that it may not be possible for the public firm to do so because of the short term market power of private firms. If the public firm invests in too much capacity, the private firms restrict their production, causing a public loss. At the first stage of the game, when the private firm chooses its capacity, it can gain strictly positive profit by investing in sufficient capacity to put the public firm in the situation described above. Therefore, contrary to usual mixed oligopoly models, the public firm efficiency is not sufficient to restore the social optimum and, contrary to commitment models, it is the lack of commitment ability that gives the incumbent the possibility to get strictly positive profits. If the order of moves is modified this inefficiency disappears, the public firm invests in the welfare maximizing quantity and the private firm stays inactive.

Using general assumptions, I provide a simple condition characterizing situations where the public firm is able to restore the long term optimum. This condition is an inequality involving the price elasticity of the demand function, the number of firms and the share of sunk cost. The role of these parameters is easily understood given their influence on the short term reaction function of the private firm. For example, with high share of sunk cost, the commitment ability of the private firm is counter productive, as mentioned above. This paper proposes further analysis of the role of this parameter in a linear framework, to stress that its influence is not monotonic. With simultaneous moves, I establish that the unique equilibrium is efficient: the public firm invests the long term optimal capacity and the private firm does not invest. There is therefore an incentive for the private firm to invest before the public one to limit its intervention.

In the following section the model is introduced. Then (section 3), I analyse the case of private leadership by a single private firm. In Section 4, the order of moves is considered. And the linear case is detailed to analyse the effect of the share of sunk cost (section 5) and the number of private firms (section 6).

2 The model

I consider a mixed duopoly market of a homogeneous good. A public firm denoted 0 competes with a private firm denoted 1. The inverse demand function is given by $p(q)$ where p is the market price and q the total quantity produced. This function is assumed to be twice differentiable, strictly decreasing and log concave. The log concavity of the inverse demand function means that it is less convex than exponential function. This property ensures that the private firm's profit is quasi-concave and that its reaction function is decreasing with a slope above -1 (cf Vives 1999, and appendix C-1). These characteristics of the reaction function are sufficient to existence and uniqueness of Cournot equilibrium.

The cost of production is divided into two parts: an irreversible capacity cost and an operating cost. I normalize the long run marginal cost at 1, capacity cost is denoted α and operating cost $1-\alpha$ with $\alpha \in [0,1]$.

The output of firm $i, i = 0, 1$ is denoted q_i and its capacity k_i, the firm can produce up to its capacity level ,with marginal cost $1 - \alpha$. I assume that there is two strictly positive quantities $k^*$and k^{**} such that $p(k^*) = 1$ and $p(k^{**}) = 1-\alpha$. The former is the quantity that equalizes price with long term marginal cost whereas the latter equalizes it with the variable cost.

The profits of firms are:

$$\pi_i = [p(q) - (1-\alpha)]\, q_i - \alpha k_i \text{ for } i = 1, 2 \tag{3.1}$$

The public firm is benevolent and maximizes the social surplus $W(q,k)$:

$$W(q,k) = \int_0^q p(u)du - (1-\alpha)q - \alpha k \tag{3.2}$$

This function is maximized for $q = k = k^*$ but as we shall see market power will prevent the public firm to reach this optimum. The price elasticity of the demand at the long term optimum is denoted ε, it plays a crucial role in following results. It is defined by:

$$\varepsilon = \frac{p}{p'k^*} = \frac{1}{p'(k^*)k^*} \tag{3.3}$$

3 Private leadership

In this part I analyse the situation with the private firm as a leader. The order of move is the following: the private firm chooses capacity first then the public firm invests and in the last stage they both produce. As I analyse subgame perfect equilibria (Selten 1975) I proceed by backward induction.

3.1 Production stage

At this stage capacities are fixed and both firms choose production quantities. The private firm maximizes its profit (3.1) and the public one maximizes social surplus (3.2).

The private firm maximizes its profit subject to capacity constraint. In order to describe the private firm's reaction function with capacity constraint, I introduce the unconstrained reaction function $r(q_0)$. It is a continuous differentiable function on the set $[0, k^{**}]$. It satisfies the first order condition:

$$p(r + q_0) + p'(r + q_0)r(q_0) = 1 - \alpha$$

Its slope is strictly between 0 and -1 (cf appendix C-1), and $r(k^{**}) = 0$. The reaction function of the firm with fixed capacity can be described using the unconstrained reaction function :

$$\forall q_0 \in [0, k^{**}], q_1(q_0, k_1) = \min\{r(q_0), k_1\}$$

The public firm maximizes welfare subject to its capacity constraint so its reaction function is $\max\{k^{**} - q_1, k_0\}$. There is a unique equilibrium of this game. The quantities produced at equilibrium are denoted $q_i^N(k_0, k_1), i = 0, 1$. At this equilibrium either price equals marginal cost and $q_1 = 0$, $q_0 = k^{**}$ or the public firm capacity constraint is binding $q_0 = k_0$ and $q_1 = \min\{r(k_0), k_1\}$. More precisely:

$$\begin{array}{l} \text{If } k_0 \geq k^{**},\ q_0^N = k^{**}, \text{and } q_1^N = 0 \\ \text{Else, } q_i^N(k_0, k_1) = k_0 \text{ and } q_1 = \min\{r(k_0), k_1\} \end{array}$$

3.2 Public firm choice of capacity

Once the private firm has chosen its capacity k_1, the public firm has to choose its capacity k_0. The public firm maximizes the social surplus subject to an equilibrium constraint, i.e. the production is the unique Nash equilibrium of the production subgame. I assume that for any k_1, social surplus is concave with respect to the public firm capacity. Therefore, there is a unique best response of the public firm denoted

$k_0^+(k_1)$. This quantity is defined by:

$$k_0^+(k_1) = \arg\max_{k_0} W((q^N(k_0, k_1), k_0 + k_1)$$

The choice of k_0 can be restricted to the set $[0, k^{**}]$, because any capacity above k^{**} does not influence the production stage and would be unused at the production stage. Therefore the production is $k_0 + q_1(k_0, k_1)$. Social surplus is differentiable by part and its derivative with respect to k_0 is:

$$\begin{aligned} \frac{\partial W}{\partial k_0} &= [p - (1-\alpha)] \frac{\partial q^N}{\partial k_0} - \alpha \\ &= \begin{cases} p - 1 & \text{if } 0 < k_0 < r^{-1}(k_1) \\ [p - (1-\alpha)](1 + r') - \alpha & \text{if } r^{-1}(k_1) < k_0 \leq k^{**} \end{cases} \end{aligned}$$

If the public firm chooses a small amount of capacity the private firm will bind its capacity at the production stage and it will not use its short-term market power. In this case the social surplus evolution is usual: a change in k_0 leads to a similar change in q^N and the social surplus is increased by $p - 1$. However, for k_0 greater than $r^{-1}(k_1)$, the private firm production decreases with respect to k_0 . The impact of this decrease in production on social welfare is represented by the term $[p - (1-\alpha)]\, r'$.

The long-term social optimum can be reached if the private capacity is binding for $k_0 = k^* - k_1$. In that case $q^N = k = k^*$ and $p = 1$, the price equals the long-term marginal cost. For the social optimum to be reached, the following must hold: $r(k^* - k_1) \geq k_1$. The private firm should produce at full capacity when the public firm produces the long term optimum quantity minus the private firm capacity. As the unconstrained reaction function is decreasing with a slope above -1, this inequality is satisfied for some k if and only if $r(0) \leq k^*$. This last equality is equivalent to $p(k^*) + p'k^* \leq 1 - \alpha$: the monopoly short term production is below the long term social optimum. And finally, this last inequality is equivalent to $-\varepsilon \leq 1/\alpha$. In that case, there exists a threshold quantity $l \in [0, k^*]$ solution of $r(k^* - l) = l$ such that:

$$\forall k_1 \in [0, k^*], r(k^* - k_1) \geq k_1 \Longleftrightarrow k_1 \leq l$$

The threshold can be expressed with the elasticity ε:

$$l = -\varepsilon \alpha k^* \tag{3.4}$$

Hence, if the private firm capacity is below the threshold l, the public firm is able to restore the social optimum by investing in capacity $k^* - k_1$. If the private firm capacity is above l as represented on figure 3.1, it is no longer feasible for the public firm to restore the long term optimum. In that case, the social surplus is strictly increasing as long as the private firm capacity constraint is binding i.e. for $k_0 \leq r^{-1}(k_1)$. And, similarly the social surplus is decreasing for aggregate production above k^* i.e. $k_0 > k^* - l$. Hence, the best response of the public firm is above $r^{-1}(k_1)$ and below $k^* - l$ and the aggregate production is $k_0 + r(k_0)$. If the private firm capacity is above $r(0)$, the public firm maximizes the function $W(k_0 + r(k_0), k_1 + k_0)$ which is equal to $W(k_0 + r(k_0), k_0) - \alpha k_1$. This last function is maximized at :

$$\overline{k} = \arg\max_{k} W(k + r(k), k) \tag{3.5}$$

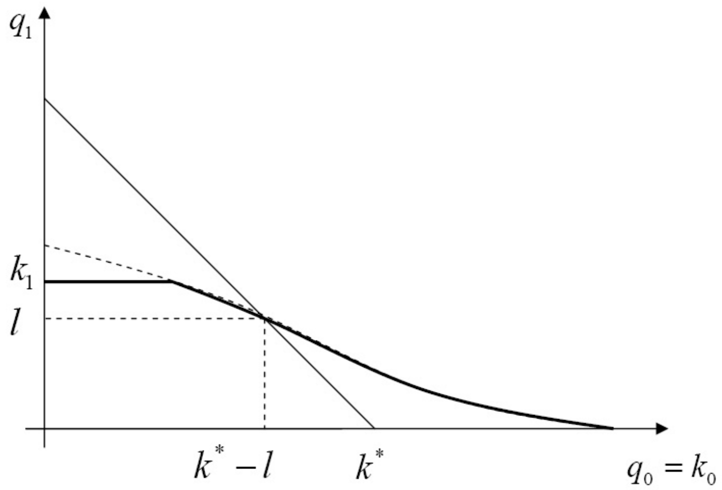

Figure 3.1: Short term market power and capacity constraint

3.3 Private firm choice of capacity

Thanks to the assumption of concavity of social surplus according to the public firm capacity, the best response of the public firm is entirely described with the quantities defined. The following lemma describes this best response.

Lemma 3 *If* $-\varepsilon \geq 1/\alpha$*, for any quantity* $k_1 \in [0, k^*]$ *the public firm best response is* $k_0^+(k_1) = k^* - k_1$

Else, three situations can be distinguished:

(i) If $k_1 \leq l$*: welfare is maximized for* $k_0^+(k_1) = k^* - k_1$*.*

(ii) If $l \leq k_1 \leq r\left(\overline{k}\right)$*: welfare is maximized for* $k_0^+(k_1) = r^{-1}(k_1)$

(iii) If $r\left(\overline{k}\right) \leq k_1$*: welfare is maximized for* $k_0^+(k_1) = \overline{k}$

In cases (ii) and (iii) the price is above the long term marginal cost.

The proof (cf Annexe C-3) uses concavity and the two monotonicity properties of the social welfare with respect to k_0: (i) it is increasing if the capacity of the private firm is binding and total capacity is less than the social optimum k_*, and (ii) it is decreasing when production is higher than k_*. As the public firm considers the loss of social surplus due to the decrease of private production, its incentive to invest is limited. Even if the public firm can force the price down to the long-term marginal cost $p = 1$, it is not socially optimal. Although the profit of the public firm is strictly positive in that case, the profit of the private firm could be negative because of unused capacities. This last situation arises in case (iii) of the lemma when the public firm best response is $\overline{k}$ and private production is $r(\overline{k})$. In intermediate situations, the private firm produces at full capacity and public capacity and production are $r^{-1}(k_1)$ because of the concavity of the welfare function.

The threshold described by formula (3.4) is linked to demand elasticity and the share of sunk cost. The less elastic the demand and the smaller the share of sunk costs, the smaller the threshold. This relation is rooted in the links between the exercise of market power and the price elasticity and variable costs. The incentive for the private firm to restrict production increases with the inelasticity of the demand and with the marginal cost of production.

The public firm best response in capacity is depicted on figure 3.2, it is composed of three parts. The first one is $k^* - k_1$, the second one $r^{-1}(k_1)$ with a slope $1/r' < -1$ and the last one is constant.

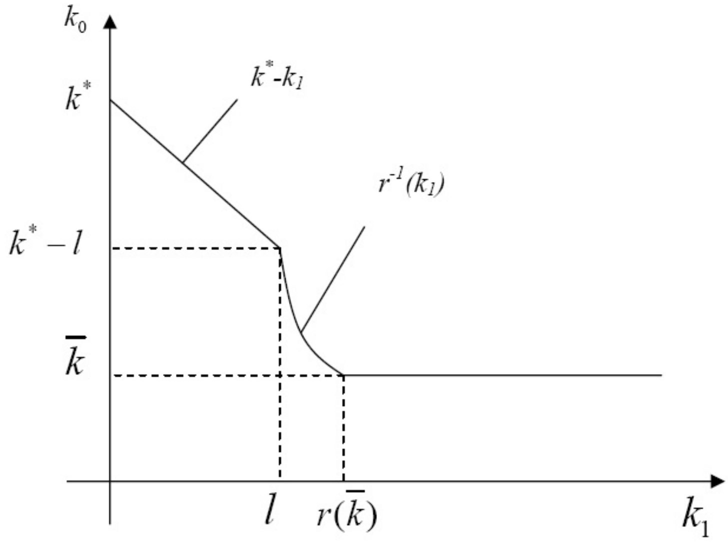

Figure 3.2: Public firm best response

3.4 Private firm choice of capacity

At the first stage of the game the private firm anticipates the reaction of the public firm when choosing its capacity k_1 . As stressed above for a quantity smaller than l, the private firm's profit is zero. At the equilibrium of the game, the firm chooses a quantity sufficiently high so that the public firm cannot restore the long-term optimum. The price is then greater than the long-term average cost and if the private firm produces at full capacity its profits are strictly positive. The following proposition describes the equilibrium paths of the game according to the value of the different parameters.

Proposition 8 *If $-\varepsilon \geq 1/\alpha$, , there exist an infinity of subgame perfect equilibria such that, along the equilibrium path, $k_1 \in [0, k^*]$, $k_0 = k^* - k_1$ and $p = 1$ and profits are null.*

Otherwise, there is a unique subgame perfect equilibrium whose equilibrium path is $k_1 = r(\overline{k})$ and $k_0 = \overline{k} < k^* - k_1$ and $q^N = k_0 + k_1$ and $p > 1$.

The proof is in the annexe C-4. The first case is trivial (cf lemma 3). In the second case the private firm is able to get strictly positive profits thanks to the public firm inability to reach the long term social optimum. All the same, the private firm incentive to limit investment is limited by the public firm intervention. If the private firm invests less than $r(\overline{k})$, the public firm's investment and the total quantity produced are increased. So the price decreases and so does the private firm's profit. The private firm is therefore induced to increase its investment. But for greater investment than $r(\overline{k})$ the private firm's choice has no influence on the public firm's one and the private firm over-invests. Part of its capacity is unused. Although the public firm's profit is strictly positive at equilibrium, the private firm cannot increase its investment to get some of this profit. This is because the firm is unable to credibly commit to a given level of production.

The total investment is superior with the public firm's intervention than without. It can be easily proved by comparing the marginal revenues in both situations. The public firm does not necessarily invest at equilibrium: $\overline{k}$ can be null for some sets of parameters as in the linear case analyzed in section 5. Even so, the private firm invests more than a monopolist would.

4 The order of moves

In the previous section private leadership was considered for the public firm intervention was seen as an ex post remedy to an observed lack of private investment. In that case, the short term market power of the private firm may prevent the public firm from restoring the long term social optimum. If the private firm invests first, it is able to exploit this weakness of the public firm intervention and get strictly positive profits. The order of moves is therefore an essential ingredient that explains the overall inefficiency.

With simultaneous moves, the private firm is no longer able to get strictly positive profits. With previous results on the public firm reaction function, the following result can be established.

Proposition 9 *With simultaneous choices of capacity, there is a unique subgame perfect equilibrium where the private firm does not invest and* *the public firm invests and produces the long term optimal quantities.*

It is clear that the situation described in this proposition is an equilibrium. The uniqueness of this equilibrium is explained by the best response in capacity of the public firm depicted in figure 3.2. The long term reaction function of the private firm is null for $k_0 \geq k^*$ and its slope is strictly between 0 and -1 for $k_0 \leq k^*$ and the long term reaction function of the private firm is strictly smaller than the short term one. From figure 3.2 it can be seen that the only intersection of reaction functions is $k_0 = k^*$ and $k_1 = 0$. Similarly, if the public firm invests first, it chooses the quantity k^* and the private firm does not invest. These two results emphasize the value of commitment for both firms. But whereas the public firm is indifferent between leadership and the simultaneous game, the private firm is strictly better of by investing first.

5 The linear case

In this section I further analyse the influence of parameters in the linear case. I consider the inverse demand function: $p = a - bq$. With this specification, the unconstrained reaction function of a private firm is:

$$r(q_0) = \max\left\{\frac{a - (1-\alpha) - bq_0}{2b}, 0\right\} \tag{3.6}$$

The quantity of capacity that equalized price with long term marginal cost is $k^* = (a-1)/b$, and the price elasticity is $\varepsilon = -1/(a-1)$. Therefore, the condition of lemma 1 and proposition 1 becomes: $\alpha \geq a - 1$ and the threshold $l = \alpha/b$. In the linear case, welfare is:

$$W(q,k) = (a - 0.5bq)q - (1-\alpha)\,q - \alpha k$$

And finally, replacing the expression (3.6) of the private firm reaction function into the welfare expression, the quantity $\overline{k}$ is:

$$\overline{k} = \max\left\{\frac{a - 1 - 3\alpha}{b}, 0\right\}$$

If the share of sunk cost is sufficiently important the loss due to the decrease of private production prevents the public firm from investing. In the sequential game, the private firm invest $r(\overline{k})$ which is $2\alpha/b$ if $a - 1 \geq$

3α and $(a-(1-\alpha))/2b$ otherwise. And finally aggregate quantities of capacity and production are:

– If $\alpha \leq (a-1)/3$, $k = r(\overline{k}) + \overline{k} = (a-1-\alpha)/b$

– If $(a-1)/3 \leq \alpha \leq a-1$, $k = r(0) = (a-(1-\alpha))/2b$

– If $a-1 \leq \alpha$, $k = k^{*}$

And the profit of the private firm is:

– If $\alpha \leq (a-1)/3$, $\pi_1 = 2\alpha^2/b$

– If $(a-1)/3 \leq \alpha \leq a-1$, $\pi_1 = \left((a-1)^2 - \alpha^2\right)/4b$

– If $a-1 \leq \alpha$, $\pi_1 = 0$

The share of sunk cost influences the set of parameters for which the public firm is able to restore the long term optimum. From previous analysis, the higher is α the greater is the set of parameters allowing the public firm to restore the long term optimum. But for suboptimal situations the influence of α is not monotonic. From previous expressions, it appears that the aggregate quantity of capacity is decreasing with respect to α for small values then it is increasing up to the long term optimum. It could be explained by two conflicting effects of this parameter, in one hand an increase of α increases the loss of replacing private production by public one, in the other hand, it decreases the variable cost and therefore increases the short term reaction function of the private firm. For small values of α the former effect dominates, and for α above $(a-1)/3$ the latter does.

Similarly, the private profit is not monotonic with respect to α, it is increasing for small share of sunk cost and decreasing for value above $(a-1)/3$. For small share of sunk cost the commitment value of investment explains the monotonicity of the private firm's profit. For higher values of α, it is the short term reaction function of the private firm that determines its investment and its profit. In those cases, it is the short term incentive of the private firm to decrease production that explains its strictly positive profit. And the short term monopoly production $r(0)$ increases with respect to the share of sunk cost. Therefore, the private firm's profit decreases when the share of sunk cost increases.

To conclude this section, I compare the private firm capacity with and without the public firm intervention because in the case of electricity market it has been argued that investment by the system operator in peaking units might drain private investment. It as already been stated that the aggregate capacity is larger with public firm intervention, but it is still unclear whether the private firm invested more. Without public firm intervention, a private monopoly would choose a quantity of capacity $(a-1)/2b = k^*/2$. For small share of sunk cost the private firm invests less with the public firm intervention but for high values it invests more.

Corollary 7 *The private firm invests more with the public firm intervention than without if and only if* $\alpha \geq (a-1)/4$.

Hence, not only aggregate capacity increases thanks to public firm intervention but so does private capacity. Public firm intervention can therefore increase private investment because of the incentive for private firms to limit the public firm capacity.

6 An oligopoly of private firms

I now move on to an oligopoly case with the linear demand introduced above. There are n private firms indexed $i = 1, .., n$, they choose their capacity k_i $i = 1..n$ simultaneously at the first stage, followed by the public firm in the second stage. As above the third stage is a production game with fixed capacities, the production of firm i is q_i. I solve the game by backward induction. I first describe the public firm's choice of capacity before analysing 'symmetric' SPE. Symmetric means that every private firm chooses the same capacity at the first stage; the public firm's choice may be different.

6.1 The production stage

At the production stage, the public firm reaction function is:

$$\max\{k^{**} - q_1, k_0\}$$

As above, if $k_0 \geq k^{**}$, the public firm produces k^{**} and the price is the short term marginal cost and no private firm produces. So the analysis can be limited to the case $k_0 \in [0, k^{**}[$ so $q_0 = k_0$ at the production stage.

Private firms play a constrained Cournot game with the translated inverse demand function $p(q + k_0)$. There is a unique equilibrium of this game, individual productions are denoted $q_i^N(k_0, k_1, .., k_n)$ and aggregate production $q^N(k_0, k_1, .., k_n)$. At this equilibrium some firms produce at full capacity and some do not. The latter ones produce the same quantity which satisfies the first order condition $p + p'q_i = 1 - \alpha$. These firms, which do not produce at full capacity, have larger capacity than the others. It means that when the quantity produced by the public firm increases, capacity constraints are relaxed successively in decreasing order. The aggregate production is increasing with respect to k_0. The production of private firms is bounded above by the unconstrained Cournot production of an oligopoly of n firm, facing the inverse demand function $p(q + k_0)$, this production is denoted $r(k_0, n)$ and:

$$r(k_0, n) = \frac{n}{n+1} \frac{a - bk_0 - (1 - \alpha)}{b} \tag{3.7}$$

Let assume that capacity are ordered $k_1 \leq k_2 \leq ... \leq k_n$. Aggregate production can be described as a function of the public capacity. First, for small capacity and production of the public firm, all firms produce at full capacity if and only if the dominant firm does and this is the case if, and only if :

$$p\left(\sum_{i=1}^{n} k_i\right) + p'\left(\sum_{i=1}^{n} k_i\right) k_n \geq 1 - \alpha$$

for any k_0 there exists j so that all firms i with $i \leq j$ produce at full capacity while the other play a Cournot game with the translated inverse demand function $p(q + \sum_{i \leq j} k_i)$ hence aggregate production is:

$$q^N = \sum_{i=0}^{j} k_i + r\left(\sum_{i=0}^{j} k_i, n - j\right)$$

With the expression (3.7) of the unconstrained Cournot production, the aggregate production is increasing with respect to the public capacity k_0 and differentiable by part with a local slope of $1/(n - j + 1)$. As k_0 increases the number j of capacity constrained firms decreases, so the slope of production decreases when k_0 increases so welfare is strictly concave[2] with respect to k_0 and there is a unique best response of the public firm $k_0^+(k_1,..,k_n)$.

6.2 Capacity choice of the public firm

I described here the best response of the public firm k_0^+, to a choice of private capacities $(k_1,..,k_n)$ with $k_1 \leq ... \leq k_n$. As I will only considered symmetric equilibria and unilateral deviation a rough description of this reaction function is sufficient.

As k_n is the largest capacity, the capacity constraint of firm n is the first relaxed as k_0 increases. For a small public capacity, all firms produce at full capacity at equilibrium if and only if $k_n \leq r(\sum_{i=2}^n k_i)$. The capacity constraint of firm n is relaxed for $k_0 \geq r^{-1}(k_n) - \sum_{i=2}^n k_i$. The social optimum can be reached if this constraint is still binding for $k_0 = k^* - \sum_{i=2}^n k_i$ if this quantity is positive.

If all capacity are large, welfare is maximized for a public capacity such that each private firm produces strictly less than its capacity. I extend previous notation to define the optimal public capacity facing an 'unconstrained' oligopoly of n firm:

$$\overline{k}(n) = \arg\max_k W(k + r(k,n), k) = \max\left\{k^* - \frac{1}{b}n(n+2)\alpha, 0\right\}$$

Lemma 4 *If $k_1 \geq r(\overline{k}(n), n)/n$ then $k_0^+ = \overline{k}(n)$ else $k_0^+ \geq \overline{k}(n)$*

And if $-\varepsilon \leq 1/\alpha$ and $\sum_{i=1}^n k_i \leq k^*$ and $k_n \leq l$ then $k_0^+ = k^* - \sum_{i=1}^n k_i$

The first part of this lemma described the reaction of the public firm if private capacities are sufficiently large. If all capacities are larger than $r(\overline{k}(n), n)/n$ agregate production is less than $k_0 + r(k_0, n)$ and welfare is maximized at $\overline{k}(n)$. If a capacity is smaller, some capacity constraint are still binding at $\overline{k}(n)$ and the public firm best response is larger (because

2. Welfare is differentiable by part with respect to k_0. Its derivative is: $\left[p - (1-\alpha)\right]/(n-j+1) - \alpha$ so it is concave because at point of non differentiability the lest derivative (with a higher j) is less than the right derivative.

of the property of the short term reaction function of the oligopoly). The second part described the situation when capacities are sufficiently so that the public firm can reach the long term optimum. One should notice that the threshold does not depend on the number of firms. This threshold should be compared with the dominant firm's capacity and not with aggregate capacity. It is due to the fact that the first constraint to be relaxed is the dominant firm's one. The optimum can be reached if this constraint is still binding for a total production of k^*. Otherwise, the short-term market power of private firms prevents the public firm from reaching the long-term optimum. It seems that increasing the number of firms does not modify the sets of parameters for which the long term optimum could be restored by the public firm. Actually, it does because of the condition that aggregate private capacity should be less than the long term social optimum. So at a symmetric equilibrium individual capacity is smaller than k^*/n and can be higher that l only if $l \leq k^*/n$.

6.3 The first stage: the choice of capacities $(k_i)_{i=1..n}$ of the private oligopoly

I construct here symmetric equilibra similar to the one analyzed in the unique private firm case. The reasoning is not very much modified by competition among private firms. If private firm short term reaction function is sufficiently small, they will be able to get strictly positive profit by investing enough so that the public firm chooses $\overline{k}(n)$. Else, there is an infinity of symmetric equilibria where the long-term social optimum is reached and all profits are null. The following proposition is a generalization of proposition 1 for the linear demand case.

Proposition 10 *If $-\varepsilon < 1/\alpha n$ i.e. $a - 1 > n\alpha$ there is a subgame perfect equilibrium such that along the equilibrium path:*

$$\text{for } i = 1,..n, k_i = \frac{r(\overline{k}(n), n)}{n}, k_0 = \overline{k}(n) \text{ and } q^N = \sum_{i=0}^{n} k_i < k^*$$

If $-\varepsilon \geq 1/\alpha n$ i.e. $a - 1 \leq n\alpha$ there is an infinity of symmetric *equilibria :*

$$\begin{gathered} for\ i=1,..n, k_i = k, k_0 = k^* - nk\ \ and\ q^N = k^* \\ where\ k \in \left[\max\left\{ \frac{a-1-\alpha}{n-1}, 0 \right\}, \frac{k^*}{n} \right] \end{gathered}$$

Proposition 2 generalizes proposition 1 in the linear. The proof (cf appendix C-6) is longer because of the complexity of the short-term reaction of private oligopoly. However, the logic is the same. If an individual firm deviates from a symmetric equilibrium by increasing its capacity, this does not modify the public firm's choice and some of the deviator's capacity would be unused. If a firm deviates by decreasing its capacity, the public firm increase capacity and consequently, both price and profits decrease. The existence of a suboptimal symmetric equilibrium depends on the number of firms, the price elasticity and the share of irreversible cost in total cost. Such equilibrium exists if $-\varepsilon < 1/\alpha n$. The less elastic the demand is the more numerous the firms should be so that the public firm can restore the long-term social optimum. A decrease of elasticity increases the incentive for firms to limit production on the short-term and the difficulty for the public firm to restore the long-term optimum. Similarly, an increase in the variable cost decreases the short-term production. The less the ratio of sunk costs over total costs, the more numerous the firms should be for the long-term optimum to be restored at equilibrium.

I focused on symmetric equilibria because of their analytical tractability. Asymmetric equilibria should exist and firms may be able to get strictly positive profits along such equilibria even in the second case of proposition 10.

This proposition explains in which situations the long term optimum can be reached. It does not give any results on the distance to this optimum when the condition is not satisfied. Actually, the evolution of the social welfare can be opposite to the evolution of the inequality. For example, when the number of firms increases, the proposition states that the set of parameters so that the optimum is reached increases, but for 'suboptimal' parameters the situation may worsen. Concerning suboptimal situations, when concentration increases there are two opposite forces that drive the evolution of the aggregate capacity and production : the effect of n on the choice of the public firm and the effect of n on the short term reaction function, the former effect is negative while the latter is positive. If the public firm production is null the second effect dominates: the aggregate capacity increases with respect to the number of firms. But, if $\overline{k}(n) > 0$, the first effect dominates and aggregate production and capacity is decreasing with respect to the number of firms. The evolution of aggregate capacity and investment with respect to the number of firms at equilibria described is explicitly written below and represented in figure 3.3.

Corollary 8 *Capacities chosen along symmetric equilibria path described in proposition 3 are:*

- *If* $n+1 \leq [(a-(1-\alpha))/\alpha]^{1/2}$, *then*

$$k_1 = .. = k_n = \frac{(n+1)\alpha}{b}, k_0 = k^* - \frac{1}{b}n(n+2)\alpha$$

  and the aggregate capacity is $k = (a-1-n\alpha)/b$
- *If* $[(a-(1-\alpha))/\alpha]^{1/2} \leq n+1 \leq [a-(1-\alpha)]/\alpha$, *then*

$$k_1 = .. = k_n = \frac{a-(1-\alpha)}{(n+1)b}, k_0 = 0$$

  and aggregate capacity is $k = n(a-(1-\alpha))/(n+1)b$
- *If* $[a-(1-\alpha)]/\alpha \leq n+1$, *aggregate investment and production are* k**.*

The effect of the number of firms is unusual in this game because of a second order effect: the decrease of the slope of the short term private reaction function. An oligopoly of n+1 firm is more sensible to an outside increase of production than an oligopoly of n firms. Therefore the adverse effect of public production increases when the number of firms increases.

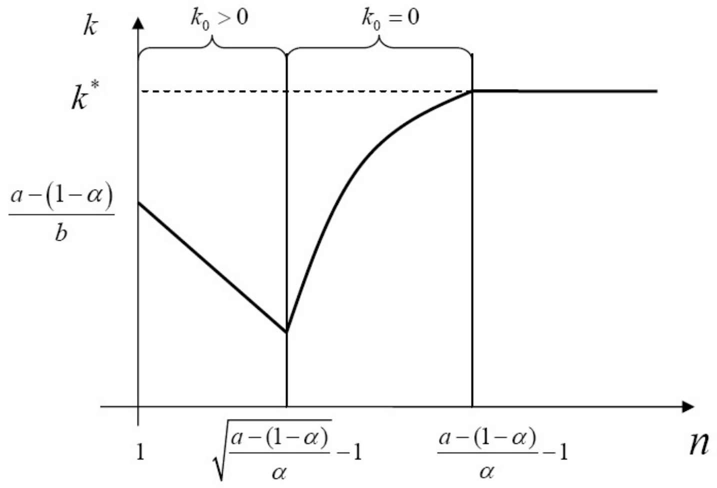

Figure 3.3: Welfare and the number of private firms

7 Conclusion

I analysed a capacity choice game with a public follower. Contrary to usual commitment games where incumbent advantages are linked to their ability to commit to a given production, incumbent inability to commit is here the key force that allows them to get strictly positive profits. I established that the short term market power of private firms impedes the public firm to establish the long term social optimum under most of the situations. As the public firm anticipates the influence of its capacity choice on aggregate production, the increase in social surplus due to the supplement of production is partly compensated by the decrease of private production induced by the exercise of market power. In the long-term, private firms remain able to get strictly positive profits and to keep the sector in a sub-optimal position. However, this situation is better than the case without a public firm intervention even if the public firm does not invest for its intervention is a threat that motivates private investment. The order of moves explains the suboptimality of the overall equilibrium, with simultaneous choice of capacity the first best optimum is the only equilibrium with the public firm being the only active firm. Hence, the private firm has an incentive to invest first in order to limit the public firm intervention. The analysis of the influence of the share of sunk cost in a linear setting shows that the influence of this parameter is not monotonic. For small values of this parameter the commitment value of investment explains that the private profit is increasing with respect to the share of sunk cost. For higher values, this influence is reversed because of the effect on the short term reaction function of the private firm. Similarly the influence of the number of firms is not monotonic because the interrelation between aggregate capacity and the short term sensitivity of private firms production. If there are few private firms, aggregate investment and production decreases with respect to the number of firms in the linear case.

Chapter 4

Marché de permis d'émission et concurrence imparfaite

1 Introduction

With the implementation of the European Union emission trading scheme (EUETS) and the Kyoto protocol emissions permits markets are used at an unprecedented scale to regulate an externality. Main sectors concerned by the EUETS are concentrated so they are often considered imperfectly competitive. In particular, several geographically isolated and concentrated electricity markets are concerned by emissions trading. The introduction of an integrated permits market creates relationship between imperfectly competitive markets initially isolated. At the same time as numerous firms are active on the emission permits market it can be assumed perfectly competitive.

This chapter deals with the efficiency of an integrated permits market between two imperfectly competitive outputs markets. The two outputs considered are non-substitutable so markets are initially isolated. The aggregate quantity of emissions is fixed and the issue is the allocation of this constraint between outputs markets, that can be done either with two isolated emissions permits markets or an integrated one. Output markets considered can be isolated in taste, space or time. Hence, the issue addressed concerns the extent of a permits market along several dimensions.

The generality of the issue addressed and the framework developed allows to analyze several aspects of emissions permits trading in the electricity industry. Electricity is a good both geographically and temporally differentiated because of networks constraints and unstorability. The present analysis is relevant to analyse the geographic extent of an emissions permits market among electricity producers on two geographically isolated markets, but also temporal aspects of emissions trading. An emission permit can be used for any emission within its period of validity, so the emissions permits market allocates emissions among subperiods of its period of validity. The demand of electricity presents important temporal and anticipated variations and each demand state can be considered as an isolated output market and the market of emissions permits allocates the emission constraint among these states. In low demand states (base period) firms do not use all their capacities to produce, whereas capacities constraints are binding in high demand state (peak period). Market power exercise and the price of emissions influence both production and capacity choices. Hence, the framework developed is used to analyze the choice of capacity in an oligopolistic electricity market.

Beside extending the literature on capacity choice with demand fluctuation and imperfect competition [1], this chapter is related to the literature on environmental policy and market power. Imperfect competition should be considered when designing an environmental policy, as Buchanan (1969) stresses for an environmental tax. He establishes that a Pigouvian tax is suboptimal when the regulated market is a monopoly. The optimal tax is less than marginal environmental damage in case of a monopoly or symmetric Cournot oligopoly (Barnett 1980), because it encompasses a subsidy that correct market power [2]. Concerning the influence of market power on the efficiency of a permit market, the litera-

1. See Requate (2005) for an extensive survey on environmental policy and imperfect competition.

2. With asymmetric firms competing *à la* Cournot, the optimal tax can be higher than environmental damage(Simpson 1995) because of an issue of production allocation among firms.

ture can be crudely divided in three whether market power is introduced on the permits market, the good market or both simultaneously. This paper belongs to the second strand.

Three previous articles address similar issues. Sartzetakis (1997; 2004) analyses the case of a duopoly exercising market power on an output market. He analyses the efficiency of a competitive emissions market to a command and control situation. In the benchmark situation emissions of each firm are exogenously given. Emissions trading modifies the allocation of emissions among firms and consequently their production choices. In Sartzetakis (2004), the author shows that welfare can decrease when emissions trading is allowed between asymmetric firms with different abatement and production technologies. The permit price that clears the market is a weighted average of the value of emissions of firms under command and control, therefore, the inefficient firm cost is reduced while the efficient one is increased when permits trade is introduced. The induced reallocation of production on the output markets can offset the efficiency gains from permits trading. Similarly the trading of permits between sectors can worsen welfare by misallocating emissions. Hung and Sartzetakis (1998) analyse the case of emissions trading between a monopolistic sector and a competitive one. Market imperfection is transmitted from the monopoly to the competitive sector via the emissions market: the monopoly consumes fewer emissions than the optimum quantity and the competitive sector more so the introduction of a permits market decreases welfare compared to an optimal command and control.

I extend this literature by considering emissions permits trading between two imperfectly competitive markets and explicitly analyzing a corrective policy and the effect of asymmetric information on the attractiveness of markets integration. Actually, if an informed regulator can achieve efficiency the introduction of market mechanisms can never outperform command and control. But, if asymmetric information is considered the regulator may rely on markets to efficiently allocate emissions among agents or at least to perform better than him. And, empirically, in the design of the EUETS, the allocation of emissions permits to producers of sectors concerned is done on the basis of anticipated demand for outputs. These demands are unknown by the regulator when setting the precise rules of the emissions markets.

Concerning environmental policy and capacity choice in the electricity, to my knowledge there is no theoretical analyze of the effect of market power. Neuhoff et al. (2005) analyze the effect of emissions allocation rules on short term production decisions and long term investment in a framework of perfect competition. Particularly, they analyse how the allocation of a fixed amount of emissions to any new capacity distort investment in a perfectly competitive electricity market. This kind of allocation rules was used during the first period of the EUETS and can be seen as an indirect subsidy of investment. Here, I do not consider the issue of the technological mix but extend the literature on strategic capacity choices by introducing an emissions market in a model of capacity choice by an oligopoly of electricity producers with a fluctuating demand.

I consider two imperfectly competitive markets and analyse how both markets imperfections influence the allocation of emissions. It is shown that the integration of permits markets can decrease welfare. I assume that the regulator cannot directly correct output markets inefficiencies so a second best allocation of emissions is introduced. The first best allocation is defined in the context of perfectly competitive output markets while the second best allocation is defined when output markets are imperfectly competitive. An increase of emissions, besides decreasing costs of production, partly corrects market power. Because of this effect, the second best optimal allocation of emissions cannot be reached with an integrated market for emissions even if firms are price takers on this market. The second best allocation can be sustained with an integrated market for emissions if a subsidy or a tax on emissions of one market is introduced. This subsidy is composed of the difference of price cost margin weighted by the sensitivity of production on emissions. Even if a direct subsidy or tax on emissions seems hardly feasible in the context of the EUETS, other indirect policies can be used such as output or capacity based allocations rules.

Next, I analyse the attractiveness of an integrated market under asymmetric information or uncertainty. The regulator does not precisely know the demand on output markets when deciding whether permits markets should be integrated. The integration of markets has two contradictory effects: the first one is the misallocation described above and the second one is the use of information about markets conditions by firms. Without any additive regulation the integration of markets is welfare enhancing if uncertainties about the two markets are sufficiently decorrelated, in that case the second (positive) effect dominates. If accompanied by an ex-ante corrective subsidy an integrated market is always welfare enhancing because the subsidy can correct the first (negative) effect.

And finally the setting is used to analyse several temporal aspects of emissions allocation and particularly the choice of capacity in a market with demand variation. It is stated that firms underinvest in capacity but overproduce in low demand states. Therefore, capacity investment should be subsidized and this subsidy would increase capacity and decrease base production. Furthermore, when the number of firms increases, contrary to standard results on Cournot competition production decreases in low demand states but capacity and welfare increase.

In the next section the model is introduced, then the influence of market power is analyzed (section 3) before considering imperfect information (section 4) and temporal aspects of emissions trading (section5). Finally, the influence of the misallocation on an optimal emission cap is analyzed (section 6).

2 The model

2.1 Set up

I consider two polluting products markets indexed $i = 1, 2$, with inverse demand function $P_i(Q_i), i = 1, 2$, where Q_i is the aggregate quantity of good i produced. Gross surplus from consumption of good i is $S_i(Q_i)$ with $S_i'(Q_i) = P_i$. Inverse demand functions satisfy the following assumptions:
For each $i = 1, 2$ there is $\overline{Q_i} > 0$ such that:

- P_i is strictly decreasing and positive on $\left[0, \overline{Q}_i\right[$
- $P_i(Q_i)$ is null for $Q_i > \overline{Q_i}$
- P_i is twice differentiable and satisfies: $P_i' + Q_i P_i'' < 0$ for $Q_i \in \left[0, \overline{Q_i}\right[$.

The last assumption, common in oligopoly literature (see Vives 1999) signifies that the marginal revenue of a firm is decreasing with respect to the production of its rival. It implies that quantities are strategic substitute and ensures existence and uniqueness of Cournot equilibrium when firms have convex cost. It is equivalent to assume that functions $x \to p'(x + y)x$ are decreasing for all y.

On market $i = 1, 2$, there are n_i firms that produce the good. The production of the good requires emissions, individual cost of production are $C_i(q_i, e_i)$ where q_i and e_i are quantities of output produced and emissions used by an individual firm. The following assumptions are made on the cost function [3]:
$\forall q_i \geq 0, e_i \geq 0$:

- Cost are increasing and convex: $\partial C_i/\partial q_i > 0, \partial^2 C_i/\partial q_i^2 > 0$
- It is worth producing: $C_i(0, e_i) = 0$ and $P_i(0) > \partial C_i/\partial q_i(0, e_i)$
- Cost and marginal cost are decreasing with respect to emissions: $\partial C_i/\partial e_i < 0$ and $\partial^2 C_i/\partial q_i \partial e_i \leq 0$
- The effect of an increase of emissions is decreasing: $\partial^2 C_i/\partial e_i^2 \geq 0$
- And $\partial^2 C_i/\partial q_i^2 . \partial^2 C_i/\partial e_i^2 - \left(\partial^2 C_i/\partial q_i \partial e_i\right)^2 \geq 0$

3. I do not introduce upper bound on emissions to ensure that at equilibria considered the emissions constraint is always binding.

Costs are increasing and convex with respect to output and decreasing and convex with respect to emissions. Marginal costs are also decreasing with respect to emissions. The last assumption ensures that the gross cost of a firm is convex (cf. appendix D-1) by limiting the effect of emissions on marginal cost.

On each market $i = 1, 2$ an emission permits market is implemented and the local price of emission is denoted σ_i, hence a firm on market i that produces q_i with e_i emissions has a profit net of initial free allocations:

$$\pi_i(q_i, Q_i, e_i, \sigma_i) = P_i(Q_i)q_i - C_i(q_i, e_i) - \sigma_i e_i \tag{4.1}$$

The aggregate quantity of emissions is $\overline{e}$, individual firms emissions is denoted e_i on market $i = 1, 2$. As I consider symmetric equilibria the aggregate quantity of emissions on a market $n_i e_i$ is symmetrically distributed among firms [4]. The aggregate quantity of emissions is therefore $\overline{e} = n_1 e_1 + n_2 e_2$. Environmental damage is assumed separable and depending only of the aggregate quantity of emissions. As this quantity is fixed I do not explicitly introduce environmental damage. In sections 3 to 5 the issue addressed is the allocation of emissions among sectors and not the choice of the aggregate constraint $\overline{e}$ that will be assumed fixed. In the last section, I consider how this quantity is affected by local market imperfections.

At all equilibria considered, thanks to firms symmetry quantities of output and emissions are equally distributed among firms on each market: $Q_i = n_i q_i$. Welfare is the sum of surpluses net of production costs on both polluting sectors:

$$W(Q_1, Q_2, e_1, e_2) = \sum_i \left[S(Q_i) - n_i C_i\left(q_i, e_i\right)\right] \tag{4.2}$$

Initial allocation to a firm on market i=1,2 is denoted $\hat{e}_i$, hence the aggregate quantity of emissions available on sector i is $n_i \hat{e}_i$ if there is no

4. I therefore avoid local cost inefficiencies that could be due to firms asymmetry under Cournot competition: for a given aggregate quantities of output and emissions Q_i and $n_i e_i$, production costs are minimized for symmetric distribution $n_i C_i(Q_i/n_i, e_i)$. Furthermore, I do not consider that the regulator can discriminate among firms in a sector by allocating different quantities of permits. Even if firms are symmetric this could increase welfare as established by Amir and Nannerup (2005).

trade of permits across markets and $\overline{e} = n_1\hat{e}_1 + n_2\hat{e}_2$.

I begin to consider the first best optimum, i.e. the quantities and emissions on each market chosen by a perfectly informed benevolent regulator able to enforce it in the next subsection. Then I introduce imperfect competition on output markets and consider a second best optimum: the allocation of emissions that maximizes welfare with market power exercice on output market.

2.2 Optimum and perfect competition

The objective of the regulator is to maximize welfare (4.2) subject to $\overline{e} \geq n_1 e_1 + n e_2$. First best quantities are denoted $Q_i^*, q_i^*, e_i^*, i = 1, 2$, they satisfy the following first order conditions:

$$S_i'(Q_i^*) = \frac{\partial C_i}{\partial q_i}\left(q_i^*, e_i^*\right), i = 1, 2 \tag{4.3}$$

$$\frac{\partial C_1}{\partial e_1}\left(q_1^*, e_1^*\right) = \frac{\partial C_2}{\partial e_2}\left(q_2^*, e_2^*\right) \tag{4.4}$$

If firms are price takers on both outputs and emissions markets, this optimum can be decentralized by an integrated emissions markets as demonstrated by Montgomery (1972). In that case the permit price is $\sigma = \sigma_1 = \sigma_2$, and on each markets price taking behavior by firms ensures that (4.3) and (4.4) are satisfied. In that case, for any initial distribution of emission permits among firms, emissions market integration always improves welfare. If the regulator does not perfectly know each market characteristics such as costs, demand and emission rates a permit market is preferred to a tax or quotas in order to minimize the cost to reach a given emission cap. The permit price at the optimum is denoted σ^*. Alternatively, if the regulator is informed he can allocate emissions $n_i\hat{e}_i = n_i e_i^*$ to each sector and not integrate emissions markets, in that case both local permit prices would be equal to σ^*.

3 Imperfect competition

3.1 General cost

Imperfect competition on output market is introduced in the model described above. Firms are assumed strategic on the output market and price takers on permits markets. Assumptions on price and cost functions ensure that there exists a unique symmetric equilibrium for any σ_i (cf appendix D-1) on each market. At this equilibrium all firms produce the same quantity of output and consume the same quantity of emissions. The permit price clears the local market for emissions. I first describe formally the equilibrium with two isolated emissions permits market then I consider a second best allocation of emissions among markets with imperfect competition. It is then shown that an integrated market of emissions permits allocates emissions differently and can therefore decrease welfare.

On output markets firms compete à la Cournot, hence, for symmetric individual quantity of emissions e_i, each firm maximizes its profit (4.1) when choosing its production. At equilibrium individual production on each market are denoted $q_i^C(e_i, n_i)$ for $i = 1, 2$. These quantities satisfy the following first order conditions:

$$P_i + P_i'.q_i^C(e_i, n_i) = \frac{\partial C_i}{\partial q}(q_i^C, e_i), i = 1, 2$$

For a local permit price σ_i at a symmetric equilibrium the individual demand of emissions is $e_i^C(n_i, \sigma_i)$ that satisfies the first order condition:

$$\sigma_i = -\frac{\partial C_i}{\partial e_i}(q_i^C(e_i^C), e_i^C), i = 1, 2 \tag{4.5}$$

As firms are price takers on the permits market, the initial distribution of permits between firms does not influence the market outcome. For symmetric equilibria, it is equivalent to consider that the regulator gives an allocation $\hat{e}_i$ to each firm or an aggregate amount of $n_i\hat{e}_i$ to all firms on market i which is allocated between firms by the emission permits market. In the latter case, the price σ_i clears the local emissions market so that $\hat{e}_i = e_i^C(n_i, \sigma_i)$.

With two isolated permits market, a regulator that can only set initial allocations $\hat{e}_i$ should consider the effect of emissions on productions. Such an allocation of emissions differs from the first best described in section 2.

Definition 1 *A second best allocation* (e_1^{**}, e_2^{**}) *of permits given market power exercise is an allocation that solves:*

$$\max_{e_1,e_2} W(Q_1^C(n_1,e_1), Q_2^C(n_2,e_2), e_1, e_2) \text{ subject to } n_1e_1 + n_2e_2 \leq \overline{e}$$

I assume that $W(Q_1^C(n_1,e_1), Q_2^C(n_2,e_2), e_1, e_2)$ is twice differentiable and concave with respect to e_1 and e_2 so that there is a unique second best allocation (e_1^{**}, e_2^{**}) with $e_2^{**} = \overline{e} - e_1^{**}$. The optimal allocation of emissions given the exercise of market power on output markets satisfies the following first order condition:

$$\frac{1}{n_1}\left(P_1 - \frac{\partial C_1}{\partial q_1}\right)\frac{\partial Q_1^C}{\partial e_1} + \frac{\partial C_1}{\partial e_1} = \frac{1}{n_2}\left(P_2 - \frac{\partial C_2}{\partial q_2}\right)\frac{\partial Q_2^C}{\partial e_2} + \frac{\partial C_2}{\partial e_2} \quad (4.6)$$

From this equation it appears that with Cournot competition on output market, the value of an incremental emission is not only the decrease of production cost but also the increase of marginal gross surplus on the output market. Individual quantities produced at the second best allocation are denoted q_1^{**} and q_2^{**} they satisfy:

$$q_i^{**} = q_i^C(n_i, e_i^{**}) \text{ for } i = 1, 2$$

Let define the difference of corrective terms:

$$s^{**} = -P_1' q_1^{**}\frac{\partial q_1^C}{\partial e_1} + P_2' q_2^{**}\frac{\partial q_2^C}{\partial e_2} \quad (4.7)$$

I then compare the result of market emissions permits markets integration with this second best policy. With an integrated market for emissions local permits prices are equalized $\sigma_1 = \sigma_2$ and the equilibrium permit price $\sigma^C(n_1, n_2)$ clears the integrated market for emissions: $\overline{e} = n_1 e_1^C(n_1, \sigma^C) + n_2 e_2^C(n_2, \sigma^C)$. At this equilibrium the following equation is satisfied:

$$-\frac{\partial C_1}{\partial e_1}(q_1^C(e_1^C), e_1^C) = -\frac{\partial C_2}{\partial e_2}(q_2^C(e_2^C), e_2^C)$$

With an integrated market for emissions, the marginal values of emissions for each firms are equalized across markets, and at equilibrium $\partial C_1/\partial e_1 = \partial C_2/\partial e_2$. Therefore, given the equilibrium productions the allocation of emissions is optimal but if one consider the effect of emissions on production it is not because of the corrective term s^{**}. Even if firms are price takers on the emissions permits market, the integration of emissions permits market does not maximize welfare in general.

Proposition 11 *If $s^{**} \neq 0$, welfare is strictly lower with an integrated market for emissions than with two isolated markets with initial allocations $\widehat{e}_i = e_i^{**}$.*

The benefit of the integration of emissions market depends of initial allocations $\hat{e}_i$, if initial allocations is the second best allocation (e_1^{**}, e_2^{**}) the integration of permits market decreases welfare. But if the initial allocation departs from this second best, the welfare effect of markets integration is ambiguous. Welfare is increased by markets integration if $\hat{e}_1$ is too low or too high. The issue of emissions permits markets integration boils down to the analysis of the choice of initial allocations. Before analyzing these choices under uncertainty, I first analyze how the regulator can correct the integrated permits market in order to reach the second best optimum.

Instead of allocating emissions to each sector and eventually keep markets isolated, the regulator can use a price instrument to correct the misallocation of emissions.

Corollary 9 *The second best optimum can be established with an integrated market for emissions permits and a subsidy s^{**} of market 1 emissions.*

The proof is in appendix D-2. The subsidy s^{**} can be either positive or negative, it is positive if market 1 is the underemitting market, similarly it is possible to tax emissions of the overemitting market. Actually there is infinity of corrective policies that could be implemented. The only relevant feature of such policy is that prices of emissions faced by firms on market 1 and on market 2 are different, and the difference should be equal to s^{**}. This subsidy reflects the value of emissions that is not considered by firms while choosing their emissions with an integrated market. Firms only consider the effect of emissions on cost of production whereas emissions have an additive value for the regulator by increasing productions. An incremental emission on market 1 beside decreasing the cost of production increases surplus of

$-P_1' q_1 \partial q_1^C / \partial e_1 = (P_1 - \partial C_1/\partial q_1)\, \partial q_1^C / \partial e_1$. This term is composed of two factors that respectively represent the effect of market power and the sensitivity of production to an increase of emissions. Hence, if both markets have similar price marginal cost margins, the market that is the more sensitive to an increase of emissions should be subsidized.

3.2 With Leontief technologies

In order to further analyze the effect of market power and derive some comparative static results I analyze the case where firms have constant marginal cost and constant emissions rates. Marginal costs are denoted c_i, $i = 1, 2$, with $P_i(0) > c_i$, and emissions rates u_i, $i = 1, 2$, the production of q_i units of good i costs $c_i q_i$ and requires $u_i q_i$ emissions. In that case the aggregate quantity of emissions is $\overline{e} = u_1 Q_1 + u_2 Q_2$. With constant emissions rates first best and second best optima are identical and characterized by the first order conditions [5]:

$$\frac{P_1(Q_1^*) - c_1}{u_1} = \frac{P_2(Q_2^*) - c_2}{u_2} \tag{4.8}$$

With this particular setting, the emissions cap is a constraint on productions and market power rather than reducing aggregate production modifies the allocation of this constraint. The misallocation of emissions can be simply analyzed with first order conditions. With an integrated market for emissions, at equilibrium both productions satisfies $P_i - c_i = \sigma u_i - P_i' q_i$ and therefore the following equality is satisfied by equilibrium quantities :

$$\frac{P_1 - c_1}{u_1} - \frac{P_2 - c_2}{u_2} = -\frac{P_1}{\eta_1} \frac{1}{u_1 n_1} - \left(-\frac{P_2}{\eta_2} \frac{1}{u_2 n_2} \right) \tag{4.9}$$

Where η_i is the price elasticity on market i: $\eta_i = P_i/(Q_i P_i')$. One of equilibrium quantities is lower than the optimal one and the other is higher.

Corollary 10 *With constant marginal costs and emissions rates, firms with market power produce more than the competitive quantity on a mar-*

5. In order to have an interior equilibrium it is assumed that $(P_i(0) - c_i)/u_i > (P_j(\overline{e}) - c_j)/u_j$ for $i, j \in 1, 2, i \neq j$.

ket and less on the other:

$$Q_1^C > Q_1^* \text{ and } Q_2^C < Q_2^* \text{ if and only if } -\frac{P_1}{\eta_1}\frac{1}{u_1 n_1} < -\frac{P_2}{\eta_2}\frac{1}{u_2 n_2}.$$

The allocation of the constraint among markets depends on the relative extent of market power on each market. For instance, the market with the lowest elasticity is more likely to be under producing and under emitting. Even if production is higher than competitive one in one market, profits of firms on both markets are strictly positive for the permit price is low.

Taking into account cross markets permits trading influences usual comparative static results and particularly those relative to the number of firms. It is no longer true that an increase of the number of firms on one market improves welfare. Actually, if an additional firm enters the over emitting market the situation is worsened.

Corollary 11 *The first best optimum is reached at:*

$$n_1^*(n_2) = n_2 P_1 \eta_2 u_2 \left(P_2 \eta_1 u_1\right)^{-1}$$

If $n_1 > n_1^*$ an increase of the number of firms on market 1 decreases welfare.

This corollary stresses that a competitive policy that aims at increasing competition in pollutant sectors should began to correct market imperfection in underemitting ones. A subsidy s^* as defined in corollary 2 on emissions of market 1 could be set in order to correct this misallocation of the CO2 constraint among markets. In the particular case of Leontief technologies, it is even possible to use a structural policy to correct the misallocation. For instance, if the number of firms on market 2 is fixed, the regulator can modify the number of firms on market 1 to reach the first best optimum.

4 Imperfect information

As stated by proposition 1, the integration of CO2 markets can decrease welfare because of market imperfection on output markets that prevents an efficient allocation of the constraint. Therefore, a perfectly informed regulatory authority is able to perform better than markets by initially allocating the constraint among markets and not creating an integrated market for emissions permits. With imperfect competition the integration of emissions markets has an ambiguous effect on welfare depending on the initial distribution of emissions among markets. If the regulator initially misallocates emissions, markets can improve efficiency even if imperfectly competitive. An initial misallocation of emissions could be explained by uncertainty or asymmetric information on markets conditions. I test how uncertainty influences the efficiency of an integrated market with an approach similar to the approach developped by Weitzman (1974) to compare price and quantity regulatory instruments.

Let assume that gross surplus on each market is random at the time of allocation of permits: $S_i(Q_i, \theta_i)$, where θ_i is a random parameter with $E\theta_i = 0$. In order to get explicit formulation marginal cost are assumed null and emission rates equal and normalized at 1: $u_1 = u_2 = 1$ and I use quadratic forms for gross surplus with an additive uncertainty :

$$S_i(Q_i, \theta_i) = (a_i + \theta_i)Q_i - \frac{b_i}{2}Q_i^2 \tag{4.10}$$

The regulator does not know the value of random parameters when deciding to allocate emissions among firms. The regulator decides whether or not to implement an integrated market for emissions permits. If markets are integrated, the initial allocation of permits does not influence the market outcome. I analyze both cases whether the regulator set a ex ante subsidy to partly compensate market power or not.

So the three regulatory options are:

1. The regulator aloocates emissions $\hat{Q}_i$ to firms of market $i = 1, 2$ and emissions permit markets are isolated so each firm produce $\hat{Q}_i/n_i$.
2. Emissions permits market are integrated and no corrective policy is implemented. Firms choose their production once θ_1, θ_2 are revealed, the permit price $\sigma(\theta_1, \theta_2)$ clears the market for emissions

so that:

$$Q_1^C(n_1,\theta_1,\sigma)+Q_2^C(n_2,\theta_2,\sigma)=\overline{e}$$

3. Emissions permits market are integrated and a subsidy s on market 1 emissions is fixed ex ante. Firms choose their production once θ_1, θ_2 are revealed, the permit price $\sigma(\theta_1,\theta_2,s)$ clears the market for emissions so that:

$$Q_1^C(n_1,\theta_1,\sigma+s)+Q_2^C(n_2,\theta_2,\sigma)=\overline{e}$$

In the benchmark situation the regulator allocates emissions $\hat{Q}_i$ to firms of market $i=1,2$. Firms may trade permits within each markets but not across markets. The regulator maximizes expected welfare: $EW(Q_1,Q_2,\theta_1,\theta_2)=E\left(\sum_i S_i(Q_i,\theta_i)\right)$ subject to the constraint $\overline{e}\geq Q_1+Q_2$. Hence, at the optimum $a_1-b_1\hat{Q}_1=a_2-b_2(\overline{e}-\hat{Q}_1)$ and the allocation to firms of market i is:

$$\hat{Q}_i=\frac{a_i-a_j+b_j\overline{e}}{b_i+b_j}, i,j=1,2, j\neq i \tag{4.11}$$

With this allocation of emissions expected welfare is noted $\hat{W}$. Emissions and productions are not influenced by random parameters because they are fixed ex ante by the regulator.

With an integrated market for emissions permits quantities produced are conditional on the random markets conditions. There are potential gains from market integration that come from the adaptation of productions to markets conditions. If an integrated market is introduced, the initial allocation of permits does not influence the outcome because firms are competitive on the permits market. In order to compare both situations with and without a corrective subsidy on the permit market, I determine a general expression for any subsidy s on market 1 emissions. First, I rewrite welfare as a function of market 1 production and emissions Q_1:

$$\begin{aligned}W(Q_1,\overline{e}-Q_1,\theta_1,\theta_2)=&\left[(a_1-a_2+\theta_1-\theta_2+b_2\overline{e})Q_1-\frac{b_1+b_2}{2}Q_1^2\right]\\&+\left[(a_2+\theta_2)\overline{e}-\frac{b_2}{2}\overline{e}^2\right]\end{aligned} \tag{4.12}$$

If a subsidy s is set on emissions on market 1, firms from market 1 face the permit price $\sigma - s$ and firms from market 2 the permit price σ. At equilibrium the quantity $Q_1^C(\theta_1, \theta_2, s)$ satisfies the equation $P_1 + P_1' \frac{Q_1^C}{n_1} = P_2 + P_2' \frac{Q_2^C}{n_2} + s$. Hence, the equilibrium production on market 1 is:

$$\begin{aligned} Q_1^C &= \frac{1}{B}\left[(a_1 - a_2) + \frac{n_2+1}{n_2} b_2 \bar{e} + s + (\theta_1 - \theta_2)\right] \\ &= \overline{Q_1^C} + \frac{\theta_1 - \theta_2}{B} \end{aligned} \tag{4.13}$$

Where $B(n_1, n_2) = (n_1 + 1)\, b_1/n_1 + (n_2 + 1)\, b_2/n_2$ can be interpreted as a measure of aggregate market power. The expression of Q_1^C consists of a fixed term and a random part. The latter term represents the adaptation of productions to revealed market conditions θ_i, i=1,2. By injecting (4.13) into the expression of welfare (4.12), expected welfare with an integrated market and a subsidy s is:

$$\begin{aligned} E\left[W\left(Q_1^C, Q_2^C, \theta_1, \theta_2\right)\right] = &E\left[W\left(\overline{Q_1^C}, \overline{Q_2^C}, \theta_1, \theta_2\right)\right] \\ &+ var(\theta_1 - \theta_2)\left[B - \tfrac{1}{2}(b_1 + b_2)\right] B^{-2} \end{aligned} \tag{4.14}$$

Hence, the effect of uncertainty and adaptation of production to markets condition is isolated in the second term. And finally, with quadratic specifications the effect of market power can also be isolated for $EW\left(\overline{Q_1^C}, \overline{Q_2^C}, \theta_1, \theta_2\right) = \hat{W} - (b_1 + b_2)\left(\overline{Q_1^C} - \hat{Q}_1\right)^2/2$ and the comparison of both situations for any level of subsidy is:

$$\Delta(s) = \frac{var(\theta_1 - \theta_2)}{B^2}\left[B - \frac{1}{2}(b_1 + b_2)\right] - \frac{b_1 + b_2}{2}\left(\overline{Q_1^C}(s) - \hat{Q}_1\right)^2 \tag{4.15}$$

With this expression, th two opposite effects of market integration are isolated: the first term represents gain due to the use of information and the second one loss due to market power. Hence if the regulator does not use any subsidy to correct market power he should compare both effects to decide whether to integrate markets. But, if an ex ante subsidy can be set on emissions the negative effect can be canceled and market integration is always welfare enhancing.

Proposition 12 *Under imperfect information* *(i) Without any subsidy an integrated market improves expected welfare if and only if:*

$$var(\theta_1 - \theta_2) > \frac{\left[b_1\left(a_1 - a_2 + b_2\bar{e}\right)/n_1 - b_2\left(a_2 - a_1 + b_1\bar{e}\right)/n_2\right]^2}{\left[\left(1+2n_1\right)b_1 + \left(1+2n_2\right)b_2\right]\left(b_1+b_2\right)}$$

(ii) With an optimal ex ante subsidy, markets integration always increases expected welfare.
The optimal subsidy is such that $\overline{Q_1^C}(s^*) = \hat{Q}_1$ *and the welfare gain is :*

$$\frac{var(\theta_1 - \theta_2)}{B^2}\left[B - \frac{1}{2}(b_1 + b_2)\right]$$

Calculations of welfare gains from markets integration under uncertainty are in appendix D-3. The use of an ex ante subsidy can correct the effect of imperfect competition while preserving the efficiency gain from the mobilization of firms' knowledge on markets conditions. This phenomenon explains that welfare gain is increasing with respect to the variance of relative efficiency. Hence, without any subsidy markets integration might be profitable if markets are sufficiently decorrelated. Most gains are obtained from the trade of permits among markets with uncorrelated trends.

Concerning the influence of concentration, some comparative static can be realized. First, it appears from the expression of production (4.13) that if market 1 underemits on average it always does. Therefore, an increase of the number of firms on market 1 increases welfare in all demand states and subsequently the appeal of market integration. But, if an ex ante subsidy is set to correct the effect of market power an increase of the number of firms in any market increases welfare.

5 Temporal aspects

5.1 Capacity choice

I extend the previous analysis to the case of capacity choice in a market with time varying demand as in the electricity sector. This sector is responsible of a large share of emissions (more than half of emissions of sectors concerned by EUETS) and characterized by high concentration and capitalistic technologies. The demand for electricity is time varying, and there are few means for intertemporal arbitrage. Therefore, each demand states can be considered as an isolated market. I consider only two demand states in order to stick on previous analysis, but it could be extended to more demand states (cf appendix D-4 for a continuum). Hence, the two markets considered are two demand states for the same good (electricity). The frequency of state i is f_i, with $f_1 + f_2 = 1$. Furthermore, the marginal surplus from consumption is higher in state 2: $S_1'(Q) < S_2'(Q)$; state 2 represents peak and state 1 base demand. I assume that the variation is additive: $P_2 = P_1 + A$ with $A > 0$ which implies that $P_1'(Q) = P_2'(Q)$.

There are n firms and the production of the good requires capacity. The cost of production is composed of two parts a variable cost and a capacity cost. Variable cost is assumed null and the marginal cost of capacity is constant and denoted c. A firm with a quantity of capacity k can produce at no cost up to k. The aggregate capacity is denoted $K(n)$. Firms invest in capacity first, and then they compete during base and peak *à la* Cournot subject to capacity constraints. The emission rate is assumed constant and normalized at one. An emission permit can be used for emissions in both states and the permits market allocates emissions among demand states. Welfare is:

$$W = f_1 S_1(Q_1) + f_2 S_2(Q_2) - cK$$

It should be maximized subject to constraints $Q_i \leq K, i = 1, 2$ and $\bar{e} \geq f_1 Q_1 + f_2 Q_2$. First best quantities are denoted $Q_i^*, i = 1, 2$ and K^*. I assume that capacity constraint is only binding during peak: $Q_1^* < K^*$and $Q_2^* = K$. Hence, optimal quantities satisfy:

$$S_1'(Q_1^*) = S_2'(K^*) - \frac{c}{f_2}$$

$$\text{and } \bar{e} = f_1 Q_1^* + f_2 K^*$$

At the non cooperative equilibrium, firms strategically choose quantities of production and capacities but they are assumed price takers on the emission permits market. I assume that demand functions and frequencies are such that at equilibrium capacities constraints are only binding during peak. Therefore, during base firms produce the Cournot production without capacity constraint and during peak they produce at full capacity. At the symmetric non cooperative equilibrium quantities satisfy:

$$P_1(Q_1^C) + P_1'(Q_1^C)\frac{Q_1^C}{n} = P_2(K^C) + P_2'(K^C)\frac{K^C}{n} - \frac{c}{f_2} \tag{4.16}$$

$$\text{and } \overline{e} = f_1 Q_1^C + f_2 K^C$$

From previous analysis, one quantity is above the optimal one whereas the other is below. The sign of the difference $P_1'(Q_1^C)Q_1^C - P_2'(K^C)K^C$ determines which quantity is above the optimal one.

Corollary 12 *At equilibrium:*

(i) Firms over produce during base and under invest in capacity: $Q_1^C(n) > Q_1^*, K^C < K^*$.

(ii) An increase of the number of firms increases welfare by reducing base production and increasing capacity.

(iii) A subsidy on capacity increases investment and decreases base production.

Proof. I only demonstrate (i) here ((ii) and (iii) are demonstrated in appendix D-4) for it is the main result and the proof reveals an interesting property of Cournot competition.
With the assumption of additive variations $P_2'(K^C) = P_1'(K^C)$ and by assumption the function $P_1'(Q)Q$ is decreasing with respect to Q. As $Q_1^C < K^C$, the difference $P_1'(Q_1^C)Q_1^C - P_2'(K^C)K^C$ is positive and $Q_1^C > Q_1^*$. ■

The surprising part of this result is that base production is above the optimal one and decreases when the number of firms increases. Once the phenomenon of misallocation of the emission constraint is understood this result is easily explained. When the number of firms increases, the aggregate quantity produced does not change but is better allocated among demand states. This misallocation can be corrected by the use of a tax on base emissions and a subsidy on capacity. As it is practically difficult to set a tax only on base consumption and not on peak consumption, it is more feasible to only subsidize investment. For instance, the use of a capacity based rule for free allocation of permits can correct market power distortion. Such a subsidy would increase permits price and decrease base production. Here, the use of an optimal subsidy on capacity is sufficient to restore the first best optimum because there are only two demand states. If more demand states were considered a subsidy on capacity would not be sufficient to correct both investment and production in all demand states but would still increases welfare.

5.2 Other temporal aspects

More generally the analysis developed in preceding sections can be used to analyze the allocation in time of a scarce resource. The analysis of capacity choice reveals that firms tend to underinvest in capacity and overproduce when demand variation is additive. When capacity constraints are introduced production costs have a particular form, but this form is not of direct influence on the result established: firms with market power tend to overemit (overconsume the scarce resource) during low demand states and underemit during high demand states. And this result is solely due to additivity of variations and the assumption on the price function that $P_i' + Q_i P_i'' > 0$.

If the two markets are two subsequent periods of a market of a perishable good and no entry occurs: $n_1 = n_2$, in that case, ignoring the discount rate, equation (4.9) becomes:

$$P_1 - P_2 = \frac{1}{n_1} \left(P_2' Q_2^C - P_1' Q_1^C \right) \tag{4.17}$$

The function $P_i'.Q_i$ is assumed decreasing, therefore, if variations are additive firms underproduce during the high demand period. So, if demand increases (resp. decreases) from period 1 to period 2 firms overconsume (resp. underconsume) the scarce resource in the first period because of market power. More precisely, the assumption that $P_i'.Q_i$ decreases implies that firms have more incentives to limit production when demand is higher. Firms produce more when demand is higher but the 'distance' to first best quantity is higher the higher the demand is. Within the issue of allocation of the emissions cap among two markets, it implies that firms allocate too much resources to the low demand state.

To consider temporal aspects, the analysis could hardly be complete if not introducing a discount rate. Even if a little far from the initial issue addressed in this chapter it is interesting to investigate how these results are modified if a discount rate is introduced, so the 'size' of the two periods is sufficiently large for that discount rate to be significant. So, if a discount rate δ common to firms and the regulator is introduced, the optimal allocation satisfies (with marginal cost normalized at 0):

$$P_1 = \delta P_2$$

Whereas with Cournot producers:

$$P_1 - \delta P_2 = \frac{1}{n}\left(\delta P_2' Q_2^C - P_1' Q_1^C\right) \tag{4.18}$$

Hence, in the case of a growing industry, the introduction of a discount rate modifies the preceding result.

Corollary 13 *If $P_2 = P_1 + A$ with $A \in \mathbb{R}$ and $i = 1, 2$ are two subsequent periods.*
If $A > 0$ there is a discount rate $\tilde{\delta}$ such that:

Firms overemit in the first period if and only if $\delta \geq \tilde{\delta}$.

If $A < 0$ firms underemit in the first period for all δ.

Proof. Q_1^C is decreasing with respect to δ (by derivation of equation (4.18), hence, $\delta P_2' Q_2^C - P_1' Q_1^C$ is decreasing with respect to δ because: i. $P_2' Q_2^C$ is negative, (ii) $P_1'(Q)Q$ is decreasing with respect to Q so $P_1'(Q_1)Q_1$ is increasing with respect to δ and (iii) $P_2' Q_2$ is decreasing with respect to δ.
If $A > 0$, it is positive for $\delta = 0$ and negative for $\delta = 1$ and by continuity it is null once at $\tilde{\delta}$.
If $A < 0$ it is always negative. ■

In the particular case of temporal allocation, a corrective policy could be either to set a rule of free allocation based on period 1 emission or a rate of permits fungibility from period 1 to period 2. Both rules would modify the intertemporal relationship between permits prices and the allocation of emissions over time. It should be stressed that, in the analysis provided, environmental damages were assumed only determined by the aggregate quantity of emissions $\overline{e}$ so that the temporal allocations of emissions between the two periods was implicitly assumed to not matter.

6 The choice of the emission cap

In previous sections the emission cap was assumed fixed and the issue addressed was how market power on output markets misallocates a fixed emission cap. A corrective policy was determined but it is not likely that such a policy will be put in place in the EUETS in order to limit possible regulatory capture or strategic interventionism by European states. Therefore, a natural question that arises is the effect of such misallocation on the optimal emission cap. Even if the political process that fixes emissions cap can hardly be seen as a benevolent regulator maximizing welfare, such a question is relevant to understand how internal imperfections influence aggregate environmental policies. The issue can be generalized: it is to understand how internal imperfections influence the choice of an aggregate constraint. And far from providing any general answer to this question, I establish here with a relatively simple example that it can be in either ways. Internal misallocations of a constraint can imply that the optimal, second best, aggregate quantity is either higher or lower than the first best one (without internal imperfections).

In the general case (i.e. with general cost function) three situations should be distinguished: with and without market power and in the former case with and without the corrective policy. I restrict here the analysis to the case of Leontief technologies so that the competitive case coincides with the case of corrected market power. Furthermore, marginal cost are normalized at 0 and emissions rate at 1: $c_i = 0$ and $u_i = 1$. As environmental damage is assumed separable and convex, it is sufficient to answer the question to analyze the derivative of welfare

(4.2) with respect to $\overline{e}$ with and without the exercise of market power. I assume that welfare is concave and differentiable with respect to $\overline{e}$ in both cases. In the case of perfect competition, the derivative of expected welfare (4.2) is $S_i'(Q_i^*)$, i=1,2 which is the value of the emission constraint and the price of emissions permits. With Cournot competition on output markets the effect of the cap on quantities produced should be considered, and the derivative of welfare is $\left(S_1'(Q_1^C) - S_2'(Q_2^C)\right) \frac{\partial Q_1^C}{\partial \overline{e}} + S_2'$. The effect of misallocation on the optimal emissions cap is determined by the sign of the difference of both derivatives:

$$\triangle = \left[P_1(Q_1^C) - P_2(Q_2^C)\right] \frac{\partial Q_1^C}{\partial \overline{e}} + P_2(Q_2^C) - P_2(Q_2^*) \tag{4.19}$$

From this equation it appears that the answer is not obvious for two contradictory effects are at stake. Let's assume that market 1 is the underemitting market, in that case the first term is positive: an increase of the emission cap has a positive effect by increasing the emissions of market 1, but the second term is negative because market 2 is overemitting. Hence, an increase of the emission cap partly corrects the misallocation but has a lower direct effect on welfare. Whether one effect dominates the other is not evident from this equation. It is unclear whether the misallocation implies a more or less stringent environmental policy.

Let's consider the linear case: $P_i = a_i - b_i Q_i$, with this specification the difference $P_2(Q_2^C) - P_2(Q_2^*)$ is simply $-b_2.\left(Q_2^C - Q_2^*\right)$ and this is equal to $-b_2\left(Q_1^* - Q_1^C\right)$. Similarly the difference $P_1 - P_2$ can be simply expressed as $(b_1 + b_2)\left(Q_1^* - Q_1^C\right)$. And finally it appears that the only relevant quantity is the sensitivity of productions to the emissions cap and this sensitivity is determined by the relative extent of market power on both markets.

Proposition 13 *The optimal emission cap is higher with market power than without if and only if:*

$$(n_1 - n_2)\left(Q_1^* - Q_1^C\right) \geq 0 \tag{4.20}$$

Calculations are straightforward to establish that $(b_1 + b_2)\frac{\partial Q_1^C}{\partial \overline{e}} - b_2$ is proportional to $n_1 - n_2$. The difference of markets concentrations is sufficient to compare both effects mentioned above. If market 1 is under emitting and $n_1 \geq n_2$ the corrective influence of the emission cap dominates and the optimal emission cap is higher with market power than without, i.e. the environmental policy is less stringent. But if the concentration on market 1 is higher than on market 2, the misallocation due to market power implies a more severe environmental policy. Hence, with two similar markets ($P_1(Q) = P_2(Q)$) the optimal emission cap is always increased by market imperfection.

7 Conclusion

In this paper I analyzed how multiple market imperfections influences the efficiency of a competitive market of emissions permits. If several output markets are imperfectly competitive an integrated emission market can decrease welfare. Imperfect competition modifies the allocation of the emissions constraint among markets and usual comparative static results could be reversed. Particularly an increase of the number of firms in an over emitting sector can decrease welfare by worsening misallocation. This misallocation can be corrected by additive policies. A subsidy (resp. a tax) on emissions of the underemitting (resp. overemitting) sector restores a second best optimum. Even if such a policy cannot be directly implemented in the EUETS, some policies currently used such as capacity based allocation have similar effects.

A perfectly informed regulator can do strictly better with a command and control regulation than an uncorrected integrated market. However, if asymmetric information is considered an integrated market can be preferred even without corrective policy if markets trends are sufficiently decorrelated. If an ex ante corrective policy is set an integrated market is always better than command and control.

The misallocation can be interpreted in a temporal way. With this interpretation the analysis can be used to study capacity investment in an market with time varying demand. In that case, the emission market allocates emissions among periods of consumption and firms underinvest in capacity but overproduce. To subsidize investment would not only increase investment but also decrease base production and re-establish the welfare optimum. Similarly, if the demand is growing on the output market firms tend to overproduce in the first period if the discount rate is high.

Many markets concerned by the regulation of emissions are concentrated such as electricity markets, and more firms are present on the emissions market than on output markets, hence, it is justified to analysis how local imperfections interacted via emissions markets. The scope of an emission market can therefore be limited in order to avoid imperfection 'contagion'. Similarly, market power influences the temporal use of emissions permits and the temporal scope of permits trading should be attentively designed. Concerning the electricity sector, further research should be done on the influence of emissions markets on capacity investment. Particularly, the choice among several technologies should be considered in order to understand how market power affects the technology mix with an emission market and how this should be corrected by additive policy such as the current European policy toward renewable energies.

Chapter 5

Conclusion

Les quatre chapitres ont analysé plusieurs aspects de l'investissement en capacités de production au sein d'un oligopole de producteurs stratégiques. Les deux premiers chapitres (première partie) ont analysé le choix de technologie de production et les deux derniers (seconde partie) des politiques publiques.

Dans le premier chapitre, les choix d'investissement de firmes hétérogènes sur un marché de gros de l'électricité avec une demande variable et inélastique ont été analysés. L'hétérogénéité introduite porte sur les technologies auxquelles les firmes ont accès. Cette hétérogénéité observée sur les marché de l'électricité s'explique par la complexité de certaines technologies et le savoir nécessaire à leur maîtrise qui a pu être acquis précédemment. Les distorsions de l'investissement en capacités de production concernent alors non seulement la capacité totale, mais aussi la répartition de celle-ci entre les deux technologies. Il a été montré qu'une augmentation du nombre de firmes peut diminuer le surplus social en aggravant la distorsion du mix technologique. Alors que les deux technologies sont efficaces, c'est-à-dire qu'elles sont utilisées à l'optimum il peut être dommageable d'augmenter le nombre de firmes ne pouvant investir que dans l'une d'entre elles. Concernant les marchés de l'électricité, cela montre que le développement de la concurrence à partir d'un type de technologie peut conduire à un sur-investissement dans cette technologie et une perte pour la collectivité.

Le second chapitre propose une explication de l'hétérogénéité des firmes en introduisant une contrainte liée à la taille sur l'une des technologies disponibles. Dans un modèle de duopole à la Cournot, les firmes peuvent utiliser deux technologies, l'une de ces technologies est moins coûteuse mais caractérisée par une taille minimale. C'est-à-dire qu'il faut produire au moins une certaine quantité avec cette technologie pour que le coût soit inférieur au coût de la deuxième technologie "standardisée". Il est montré que cette contrainte peut expliquer que les firmes soient différentes à l'équilibre: l'une d'entre elles utilise la technologie efficace (moins coûteuse) alors que l'autre firme utilise la technologie standardisée. L'hétérogénéité est donc expliquée de façon endogène par la présence d'une taille minimale. A partir de l'analyse du duopole des conclusions normatives ont été obtenues. Il a été montré que la taille minimale avait des effets non monotones sur le bien-être, car bien qu'elle puisse limiter l'emploi de la technologie efficace, elle peut aussi limiter le pouvoir de marché des firmes. Si la taille minimale est importante, un monopole peut être préférable à un duopole si, dans ce second cas, les firmes n'utilisent pas la technologie contrainte.

Ces deux premiers chapitres permettent d'éclairer la situation actuelle du nucléaire et des cycles combinés à gaz (CCGTs). Alors que cette dernière technologie est petite et standardisée le nucléaire présente des effets de taille et des économies de série importantes. Alors que les caractéristiques des CCGTs, favorables aux entrées, ont constitué un des arguments favorables à l'introduction de concurrence dans le secteur électrique, la renaissance actuelle du nucléaire soulève des questions sur l'efficacité des marchés de l'électricité et de la promotion de la concurrence par les CCGTs.

Le troisième chapitre traite des effets de la production d'électricité par une firme publique [1] sur les incitations des autres firmes à investir et produire. Le problème adressé est celui de la production d'électricité par

1. Le terme de firme publique est utilisée pour désigner en fait une firme bienveillante, c'est-à-dire une firme qui maximise le surplus collectif, il peut s'agir en fait d'une firme mandatée par la puissance publique elle-même supposée bienveillante.

le GRT et de l'investissement en capacité de pointe, l'analyse peut cependant s'étendre aux investissements dans d'autres technologies. A l'aide d'un modèle en trois étapes, deux étapes d'investissements suivi d'une étape de production, il est montré qu'une firme publique "suiveuse", même si elle est aussi efficace que les firmes privées, peut être incapable de corriger intégralement un déficit d'investissement et de rétablir l'optimum de premier rang. Les firmes privées restent capables de réaliser des profits strictement positifs malgré l'intervention de la firme publique. Néanmoins, même si l'intervention de la firme publique ne permet pas d'atteindre l'optimum, elle a cependant des effets incitatifs sur les firmes privées qui ont intérêt à limiter la capacité développée par la firme publique. De plus, cette intervention publique interagit mal avec le développement de la concurrence, puisqu'une augmentation du nombre de firme privées, en aggravant l'incapacité de la firme publique à rétablir l'optimum, peut diminuer l'efficacité de l'intervention.

Enfin, le quatrième chapitre développe un modèle général sur les échanges de permis d'émission entre deux secteurs imparfaitement concurrentiels. Si la concurrence sur les marchés de biens est imparfaite il peut être préférable de ne pas avoir un marché de permis d'émission intégré à moins d'y ajouter un mécanisme correctif. Il est établi sous quelles conditions l'intégration des marchés de permis est bénéfique dans une situation d'incertitude. L'analyse peut s'appliquer aux marchés de l'électricité de deux façons: soit pour considérer l'échange de permis d'émission entre deux marchés de l'électricité physiquement isolés, soit pour analyser l'investissement en capacité de production et la répartition de la contrainte d'émissions entre les demandes d'électricité de base et de pointe. Il est ainsi montré que la contrainte d'émission limite l'exercice de pouvoir de marché, et que les firmes stratégiques n'investissent pas suffisamment mais produisent trop en base. Cette distorsion de l'allocation de la contrainte carbone peut être corriger en subventionnant les investissements en capacités de production. Si l'on considère les échanges de permis entre deux marchés nationaux isolés, il est établi qu'une augmentation du nombre de firme sur l'un des marchés peut diminuer le bien-être en augmentant les émissions d'un secteur trop émetteur.

Ces quatre chapitres ont abordés à l'aide de modèles de concurrence imparfaite certains aspects de la concurrence et des choix d'investissement. Ils permettent de mieux comprendre les effets de long terme du nouveau mode d'organisation du secteur électrique sur les investissements en capacités de production. Plusieurs développements sont envisageables, beaucoup de questions restant encore ouvertes. Il faut souligner que les questions que posent le secteur électriques peuvent être générales, de nombreux secteurs capitalistiques partageant les mêmes “symptômes”, les effets peuvent cependant être exacerbés dans le cas de l'électricité. Les trois aspects qui devraient être le sujet des recherches ultérieures sont intimement liés: l'entrée, l'incertitude et les relations verticales. Ils renvoient à des littératures spécifiques et vastes qu'il ne convient pas de résumer ici, mais leurs relations avec les problématiques d'investissements offre des perspectives de recherche qui sont esquissées ensuite.

Tout d'abord, le mécanisme d'entrée sur ces marchés nécessite d'être mieux compris car le nombre de firmes actives sur un marché est une variable partiellement endogène. Les conclusions sur la relation entre concentration et efficacité du secteur nécessitent d'être complétées par une explication satisfaisante de cette concentration. D'autant plus qu'un résultat fondamental des modèles de concurrence imparfaite avec libre entrée est qu'il y a une entrée excessive (Mankiw and Whinston 1986), et les résultats obtenus avec un nombre de firmes fixé peuvent être inversés lorsque ce nombre est endogène (Corchón and Fradera 2002). Ainsi, ce type d'approche, à la Mankiw and Whinston (1986), mériterait d'être appliquée aux problèmes d'investissement en capacités de production afin de comprendre la robustesse des résultats. Cependant, ces modèles d'équilibre de libre entrée supposent l'existence d'un coût fixe d'entrée qui explique que celle-ci soit limitée, le contenu de ce coût fixe devrait être explicité [2]. Il pourrait notamment être lié à l'incertitude sur les con-

2. Il convient d'ajouter que ce type d'approche ne permet pas de saisir les composantes dynamiques des mécanismes d'entrée et de sortie. Un second type d'approche les prends en compte à l'aide de jeux répétés permettant d'analyser des

ditions du marché.

L'incertitude, les problèmes de gestion de risque, d'asymétrie d'information et d'anticipations sont des sujets fréquemment cités pour expliquer la structure industrielle observée sur les marchés de l'électricité, et fondamentalement liés au problème de l'investissement. Il semble que la concentration soit un moyen de gestion du risque et de coordination des anticipations et des investissements sur le long terme. Plusieurs voies de modélisation sont possibles pour expliciter cette intuition. La première, utilisée dans quelques travaux consiste à introduire de l'aversion au risque [3]. Cette approche sous entend que les firmes ont de l'aversion au risque ou, au moins, qu'elles se comportent comme ci, l'aversion au risque des firmes est alors une 'boite noire'. Une seconde approche, qui semble plus fructueuse, consisterait à se placer dans la lignée de la littérature sur la concurrence imparfaite avec asymétrie d'information, initiée par Ponssard (1979), de façon à représenter les effets de la concentration en terme de précisions des anticipations, et les avantages des firmes en place liés à leur connaissance des marchés.

Enfin, les arrangements verticaux du secteur, des contrats de long terme à l'intégration verticale de la commercialisation, nécessitent d'être mieux compris. Le marché spot de l'électricité, quoiqu'il puisse servir de référence, n'est pas le lieu principal de transactions entre producteurs et consommateurs, la plupart des échanges étant déterminés par des contrats de long terme ou au sein d'entreprises intégrées. Les relations entre contrats de long terme, investissements et concurrence sont sujets à controverse, dans le monde académique ainsi que pour les régulateurs des marchés. D'un coté, les contrats de long terme sont perçus comme un

stratégies plus riches (Ponssard 1991, Gromb et al. 1997).

3. Dans la littérature générale, cette approche est utilisées pour analyser le fonctionnement des marchés à terme et leurs effets sur les décisions de production (Hirshleifer 1988, Newbery and Stiglitz 1981), elle est utilisée notamment par Roques et al. (2006) pour analyser le choix de 'portefeuille' de technologie à l'aide de simulations de Monte Carlo, et par Ehrenmann and Smeers (2008) pour analyser l'effet de plusieurs risques (combustibles, régulation CO_2) sur les choix d'investissements.

moyen d'intensification de la concurrence [4] et de partage de risque. De l'autre, ils sont soupçonnés d'être un frein à l'entrée et au développement de la concurrence comme établi par Aghion and Bolton (1987). Ces contrats sont un élément fondamental des échanges dans les marchés de l'électricité et influencent les décisions d'investissement notamment en transmettant de l'information sur le niveau et le profil de la demande [5]. A ma connaissance, peu d'analyses formalisées étudient cette relation [6] qui devrait faire l'objet de recherches futures.

4. Ce qui s'appuie notamment sur les résultats de Allaz and Vila (1993). Cette analyse semble cependant plus convenir pour représenter des contrats 'future' de moyen ou court terme, et le résultat est trés sensible au mode de concurrence. Concernant les investissements, Adilov (2006) et Zoetl (2008, chap2) ont montré que ces contrats pouvait diminuer les incitations à investir.

5. Un tel mécanisme à été formalisé par Arrow (1975) pour expliquer l'intégration verticale. Concernant les contrats, les places de marchés organisés peuvent être le lieu d'agrégation de l'information, cependant, aucune des places boursières ne présentent des contrats à des termes supérieurs à un ou deux ans ce qui est loin des délais de construction et des durée de vie des installations.

6. Les analyses portent soit sur les investissements et les contrats de court terme, soit sur les investissements 'spécifiques' et les contrats de long terme. Dans le dernier cas, l'approche usuelle est celle des contrats incomplets, ainsi Segal and Whinston (2000) analysent l'effet des clauses d'exclusivité sur l'incitation à réaliser un investissement 'spécifique' (à la relation bilatérale). Les investissements en capacités de production n'étant pas 'spécifiques' dans le sens employé par cette littérature, ces approches ne semble pas suffisamment pertinentes pour tirer des conclusions sur le fonctionnement des marchés de l'électricité.

Appendix A

Proofs of chapter 1

The following notation is introduced to facilitate exposition of calculations:

$$\gamma = \frac{v - c_\beta}{v - c_\alpha}$$

A-1 Proof of propositions 1 and 2

Let assume that an equilibrium exists. It is clear from symmetry that all firms of a particular type invest in similar quantities at equilibrium. To establish existence and unicity of equilibrium I consider the three subcases whether generalist firms invest in both type of technologies or specialized in peak or baseload and show that these subcases cannot coexist.

Equilibrium quantities are: k_α^S, k_β^S and k_α^G, k_β^G which are respectively the individual capacity of baseload firms, the individual capacity of peak firms and the peak and baseload capacities of generalist firms.

1. Let assume that at equilibrium generalist firms invest in both technology types.

First I simplify the problem in order to get a simple linear system. As the individual aggregate quantity of a generalist firm is equal to the individual quantity of a peak firm: $k_\alpha^G + k_\beta^G = k_\beta^S$, the problem is already limited to three quantities: k_α^S, k_β^S and k_α^G.

The first order condition of peak firms writes:

$$s_\alpha k_\alpha^S + (s_\beta + g + 1)\, k_\beta^S = X(1 - r_\beta)$$

Furthermore, with the first order conditions of baseload firms and the one of baseload capacity of generalist firms it appears that $k_\alpha^S = k_\alpha^G + \gamma k_\beta^G$ so injecting the relation $k_\beta^G = k_\beta^S - k_\alpha^G$ the capacity of baseload firms is:

$$k_\alpha^S = (1-\gamma)\, k_\alpha^G + \gamma k_\beta^S.$$

This relationship can be used to set a second relation between the capacity of a baseload firm and the one of a peak firm: $(s_\alpha + g + 1)k_\alpha^S + \gamma s_\beta k_\beta = X\,(1-r_\beta)$. So the two quantities k_α^S, k_β^S satisfy the following system of equations:

$$\begin{bmatrix} s_\alpha + g + 1 & \gamma s_\beta \\ s_\alpha & s_\beta + g + 1 \end{bmatrix} \begin{bmatrix} k_\alpha^S \\ k_\beta^S \end{bmatrix} = X \begin{bmatrix} 1 - r_\alpha \\ 1 - r_\beta \end{bmatrix}$$

The determinant of the matrice is:

$$A\,(s_\alpha, s_\beta, g) = (s_\alpha + g + 1)\,(s_\beta + g + 1) - \gamma s_\alpha s_\beta$$

It is strictly positive so there is a unique solution of the system. And finally, some calculations lead to:

$$k_\alpha^S = X\frac{1}{A}\left[(g+1)\,(1-r_\alpha) + s_\beta \frac{\delta - \Delta}{v - c_\alpha}\right] \tag{A.1}$$

$$k_\beta^S = X\frac{1}{A}\left[(g+1)\,(1-r_\beta) + s_\alpha\,(r_\alpha - r_\beta)\right] \tag{A.2}$$

And for generalist firms:
Individual quantity of baseload capacity can be obtained with the relation $(1-\gamma)\,k_\alpha^G = k_\alpha^S - \gamma k_\beta^S$ by noting that $(1-r_\alpha) - \gamma\,(1-r_\beta) = (1-\gamma)\,(1-r)$.
And the peak capacity is simply $k_\beta^S - k_\alpha^G$.

$$k_\alpha^G = X\frac{g}{A}\left[(g + s_\beta + 1)\,(1-r) - s_\alpha\,(r - r_\alpha)\right] \tag{A.3}$$

$$k_\beta^G = X\frac{g}{A}\left[(g + s_\alpha + 1)\,(r - r_\beta) - s_\beta\,(1-r)\right] \tag{A.4}$$

So if there is an equilibrium with generalist firms that invest in strictly positive quantities of both type of capacities the equilibrium quantities are defined by equations (A.1), (A.2) and for generalist firms by (A.3) and (A.4).

Furthermore, if $s_\alpha (r - r_\alpha) \leq (g + s_\beta + 1)(1 - r)$ and $s_\beta (1 - r) \leq (g + s_\alpha + 1)(r - r_\beta)$ quantities defined by these equations described equilibrium strategies: each firm's profit is concave and maximimum at these quantities.

2. Let assume that generalist firms only invest in peak capacities at equilibrium.

Equilibrium quantities can be found from above calculations by replacing s_β by $g + s_\beta$ and g by 0.

So if such an equilibrium exists, it is fully described by quantities:

$$\begin{aligned}
k_\alpha^S &= X \frac{1}{A(s_\alpha, s_\beta + g, 0)} \left[1 - r_\alpha + (s_\beta + g) \frac{\delta - \Delta}{v - c_\alpha}\right] \\
k_\beta^S &= X \frac{1}{A(s_\alpha, s_\beta + g, 0)} \left[1 - r_\beta + s_\alpha (r_\alpha - r_\beta)\right] \\
k_\alpha^G &= 0 \text{ and } k_\beta^G = k_\beta^S
\end{aligned}$$

These quantities described an equilibrium only if generalist firm has an incentive not to invest in baseload capacity. And it is the case if the aggregate baseload capacity is above the first best optimal quantity: $s_\alpha k_\alpha^S > k_\alpha^* = X(1 - r)$ and this inequality is equivalent to:

$$s_\alpha \geq (g + s_\beta + 1)(1 - r) / (r - r_\alpha)$$

3. Let assume that generalist firms only invest in baseload capacities at equilibrium.

If at equilibrium generalist firms only invest in baseload capacity equilibrium strategies are:

$$\begin{aligned}
k_\alpha^S &= X \frac{1}{A(s_\alpha + g, s_\beta, 0)} \left[1 - r_\alpha + s_\beta \frac{\delta - \Delta}{v - c_\alpha}\right] \\
k_\beta^S &= X \frac{1}{A(s_\alpha + g, s_\beta, 0)} \left[1 - r_\beta + (s_\alpha + g)(r_\alpha - r_\beta)\right] \\
k_\alpha^G &= k_\alpha^S \text{ and } k_\beta^G = 0
\end{aligned}$$

These quantities described an equilibrium only if the aggregate quantity of peak capacity is above the first best quantity. It is so if and only if:

$$s_\beta \leq (g + s_\alpha + 1)(r - r_\beta) / (1 - r)$$

Propositions 1 and 2 are directly obtained from these results.

A-2 Quantities of capacity derivatives

I establish expressions of quantities derivatives with respect to the number of firms s_α and s_β. When there are only specialized firms first order conditions are:

$$\begin{aligned}(s_\alpha+1)\,k_\alpha^S+\gamma s_\beta k_\beta^S &= X(1-r_\alpha)\\ s_\alpha k_\alpha^S+(s_\beta+1)\,k_\beta^S &= X(1-r_\beta)\end{aligned}$$

– Derivatives of quantities with respect to s_β :

By derivation of first order conditions with respect to s_β :

$$\begin{aligned}(s_\alpha+1)\,\frac{\partial s_\alpha k_\alpha^S}{\partial s_\beta}+s_\alpha\gamma\frac{\partial s_\beta k_\beta^S}{\partial s_\beta} &= 0\\ s_\beta\frac{\partial s_\alpha k_\alpha^S}{\partial s_\beta}+(s_\beta+1)\,\frac{\partial s_\beta k_\beta^S}{\partial s_\beta} &= k_\beta^S\end{aligned}$$

This leads to the following expressions of derivatives:

$$\frac{\partial s_\alpha k_\alpha^S}{\partial s_\beta}=-\frac{s_\alpha}{A}\gamma k_\beta^S \text{ and } \frac{\partial s_\beta k_\beta^S}{\partial s_\beta}=\frac{s_\alpha+1}{A}k_\beta^S$$

And concerning aggregate quantity of capacity:

$$\frac{\partial k^S}{\partial s_\beta}=\frac{\partial s_\alpha k_\alpha^S}{\partial s_\beta}+\frac{\partial s_\beta k_\beta^S}{\partial s_\beta}=\frac{1}{A}k_\beta^S\left[1+s_\alpha\left(1-\gamma\right)\right]$$

– Derivatives of quantities with respect to s_α :

From first order conditions:

$$\begin{aligned}(s_\alpha+1)\,\frac{\partial s_\alpha k_\alpha^S}{\partial s_\alpha}+s_\alpha\gamma\frac{\partial s_\beta k_\beta^S}{\partial s_\alpha} &= k_\alpha^S\\ s_\beta\frac{\partial s_\alpha k_\alpha^S}{\partial s_\alpha}+(s_\beta+1)\,\frac{\partial s_\beta k_\beta^S}{\partial s_\alpha} &= 0\end{aligned}$$

This give the expression of quantities evolution with respect to the number s_α :

$$\frac{\partial s_\alpha k_\alpha^S}{\partial s_\alpha}=\frac{s_\beta+1}{A}k_\alpha^S \text{ and } \frac{\partial s_\beta k_\beta^S}{\partial s_\alpha}=-\frac{s_\beta}{A}k_\alpha^S$$

And the derivative of the aggregate quantity:

$$\frac{\partial k^S}{\partial s_\alpha}=\frac{\partial s_\alpha k_\alpha^S}{\partial s_\alpha}+\frac{\partial s_\beta k_\beta^S}{\partial s_\alpha}=\frac{1}{A}k_\alpha^S$$

A-3 Proof of proposition 3

I relax the integer constraint and consider the effect of change of s_β on welfare. Partial derivatives of welfare with respect to k_α and k_β are:

$$\frac{\partial W}{\partial k_\alpha} = \frac{v - c_\alpha}{X}(X - k) - I_\alpha + \delta k_\beta$$
$$\frac{\partial W}{\partial k_\beta} = \frac{v - c_\beta}{X}(X - k) - I_\beta$$

Injecting first order conditions give the following expression for derivative of welfare with respect to s_β:

$$\frac{dW}{ds_\beta} = \frac{v - c_\alpha}{X} k_\alpha^S \frac{\partial k_\alpha^S}{\partial s_\beta} + \frac{v - c_\beta}{X} k_\beta^S \frac{\partial k_\beta^S}{\partial s_\beta}$$

Injecting the expressions of capacities derivatives into above formula and factorizing gives:

$$\frac{dW}{ds_\beta} = (v - c_\beta)\, k_\beta^S \left[(s_\alpha + 1)\, k_\beta^S - k_\alpha^S\right] /AX$$

It is unclear whether welfare is concave or not (it is not in general) but it is quasi concave because its derivative is null only once and strictly positive (resp. negative) for smaller (resp. greater) values of s_β. I establish it directly by injecting formula of equilibrium quantities:

$$\begin{aligned} A\left[(s_\alpha + 1)\, k_\beta^S - k_\alpha^S\right] /X = {} & (s_\alpha + 1)\left[1 - r_\beta + s_\alpha\left(r_\alpha - r_\beta\right)\right] \\ & - s_\alpha\left[1 - r_\alpha + s_\beta\left(\delta - \Delta\right) / \left(v - c_\alpha\right)\right] \end{aligned}$$

As this expression is decreasing with respect to s_β welfare is quasi concave. Furthermore it is maximum at:

$$s_\beta^* = \frac{1}{s_\alpha} \frac{v - c_\alpha}{\delta - \Delta}\left[(1 - r_\alpha) + (s_\alpha + 1)^2\left(r_\alpha - r_\beta\right)\right]$$

A-4 Proof of proposition 4

The analysis is similar to the previous one. I relax the integer constraint and consider the effect of a change of s_α on welfare. The derivative of welfare with respect to s_α :

$$\frac{dW}{ds_\alpha} = \frac{v - c_\alpha}{X} k_\alpha^S \frac{\partial s_\alpha k_\alpha^S}{\partial s_\alpha} + \frac{v - c_\beta}{X} k_\beta^S \frac{\partial s_\beta k_\beta^S}{\partial s_\alpha}$$

Injecting the expressions of capacities derivatives into above formula and factorizing gives:

$$\frac{dW}{dn} = (v - c_\alpha)\, k_\alpha^S \left[(s_\beta + 1)\, k_\alpha^S - \gamma s_\beta k_\beta^S \right] / AX$$

Injecting formula of equilibrium quantities:

$$\begin{aligned}\left[(s_\beta + 1)\, k_\alpha^S - \gamma s_\beta k_\beta^S \right] / X = & \, 1 - r_\alpha + s_\beta \, (s_\beta + 2)\, (\delta - \Delta) \, / \, (v - c_\alpha) \\ & - s_\beta s_\alpha \, (r_\alpha - r_\beta)\, \gamma \end{aligned}$$

As this expression is decreasing with respect to s_α welfare is quasi concave and maximum at:

$$s_\alpha^*(s_\beta) = \frac{1}{s_\beta} \frac{1}{(r_\alpha - r_\beta)} \left[(1 - r_\beta) + (s_\beta + 1)^2 \, \frac{\delta - \Delta}{v - c_\beta} \right]$$

Appendix B

Proofs of chapter 2

B-1 Proof of Lemma 1

Let $y', y'' \in [0, \overline{q}]$ with $y'' > y'$.

– If $t(y') = \beta$:

Then $r_\alpha(y') < z$ and inequality (2.2) is reversed to:

$$(p(z + y') - c_\alpha)z - (p(r_\beta(y') + y') - c_\beta)r_\beta(y') \leq 0$$

And derivation of left hand side of this inequality with respect to y is:

$$p'(z + y)z - p'(r_\beta(y) + y)r_\beta(y)$$

This expression is negative for all $y \geq y'$ because r_β is decreasing so $r_\beta(y) \leq z$ for all $y \geq y'$ and $x \rightarrow p'(x + y)x$ is decreasing with respect to x. So at y'':

$$(p(z + y'') - c_\alpha)z - (p(r_\beta(y'') + y'') - c_\beta)r_\beta(y'') \leq 0$$

And finally as $r_\alpha(y'') \leq r_\alpha(y') \leq z$ the above inequality implies that $t(y'') = \beta$.

– If $t(y'') = \alpha$:

Either $r_\alpha(y'') > z$ and the result is straightforward or $r_\alpha(y'') = z$ and above calculations can be reproduce to establish that:

$$(p(z + y) - c_\alpha)z - (p(r_\beta(y') + y') - c_\beta)r_\beta(y') \geq 0$$

So $t(y') = \alpha$.

B-2 Existence and uniqueness of the solution of (2.3)

If $z \leq \overline{z}$, $y^+(z)$ is the unique solution on the set $[y^-(z), \overline{q}_\alpha - z]$ of the equation (2.3):

$$\pi\left(z, y, \alpha\right) - \pi(r_\beta(y), y, \beta) = 0$$

The function $\pi\left(z, y, \alpha\right) - \pi(r_\beta(y), y, \beta)$ is strictly decreasing on the set $[y^-(z), \overline{q}_\alpha - z]$ because its derivative is $p'(z+y)z - p'(r_\beta + y)r_\beta$ and $z \geq r_\alpha(y)$ and $z + y \geq \overline{q}_\alpha > \overline{q}$ on the set considered.

At $y = \overline{q}_\alpha - z$: $\pi\left(z, \overline{q}_\alpha - z, \alpha\right) = 0 \leq \pi(r_\beta(y), y, \beta)$.

And at y^- two cases should be distinguished:

- If $y^-(z) = 0$, then at $y = 0 : \pi\left(z, 0, \alpha\right) - \pi(r_\beta(0), 0, \beta) > 0$ because $r_\alpha(0) \leq z \leq \overline{z}$ and profit is concave.
- If $y^-(z) > 0$, then $z \leq r_\alpha(0)$ and $y^-(z) = r_\alpha^{-1}(z)$. And at $y = y^-$, z is the unconstrained best production: $z = r_\alpha(y^-)$ so $\pi\left(z, y^-, \alpha\right) > \pi(r_\beta(y^-), y^-, \beta)$.

Therefore, there is a unique solution of equation (2.3) on the set $[y^-(z), \overline{q}_\alpha - z]$.

B-3 Proof of Lemma 2

– For $z \in \left[0, \overline{q}_\alpha - \overline{q}_\beta\right]$:

First, $r_\alpha\left(\overline{q}_\alpha - z\right) \leq z$ because the function $x \to r_\alpha\left(\overline{q}_\alpha - x\right) - x$ is decreasing and null at 0.

Then, the quantity $\overline{q}_\alpha - z$ is greater than $\overline{q}_\beta$ so $\pi\left(z, \overline{q}_\alpha - z, \alpha\right) = 0 = \pi(r_\beta(\overline{q}_\alpha - z), \overline{q}_\alpha - z, \beta)$.

Hence, $y^+(z) = \overline{q}_\alpha - z$. Furthermore, $\overline{z} > \overline{q}_\alpha - \overline{q}_\beta$ because y^+ is decreasing.

– For $z \in \left]\overline{q}_\alpha - \overline{q}_\beta, \overline{z}\right[$:

$y^+(z)$ is the unique solution on the set $[r_\alpha^{-1}(z), \overline{q}_\alpha - z]$ of the equation:

$$\left(p(z + y^+) - c_\alpha\right) z = \left(p(r_\beta(y^+) + y^+) - c_\beta\right) r_\beta(y^+)$$

Application of the implicit function theorem and derivation of this equation with respect to z gives:

$$y^{+\prime}(z) = -\frac{p - c_\alpha + p'z}{p'z - p'r_\beta}$$

Arguments of the price function are different but omitted for ease of exposition. The denominator is negative thanks to A3. Furthermore,

$p(z+y)-c_\alpha = (p(r_\beta+y)-c_\beta)\, r_\beta/z$ and $p(r_\beta+y)-c_\beta = -p'(r_\beta+y)r_\beta$

and $r_\beta < z$ so $p-c_\alpha > -p'r_\beta(y^+)$. And finally $y^{+\prime}(z) < -1$.

– And for $z \geq \overline{z}$, $y^+(z)=0$ because $(p(z)-c_\alpha)\, z \leq (p(r_\beta(0))-c_\beta)\, r_\beta(0)$

B-4 Proof of proposition 6

I establish here that all equilibria are described in proposition 6. So there is always at least one equilibrium and depending on parameters value there is a unique equilibrium, two asymmetric ones or three equilibria.

Let z_1, z_2, z_3, z_4 be defined as in the proposition:

$$\begin{aligned} y^+(z_2) &= z_2 \\ y^+(z_1) &= \max\left\{z_2, q_\alpha^A\right\} \\ y^+(z_3) &= q_\beta^S \\ y^+(z_4) &= r_\beta(z_4) \end{aligned}$$

There are all well defined real positive numbers: It is so because $y^+(z)$ is continuous, strictly decreasing with a slope smaller than -1 on $[0,\overline{z}]$ and $y^+(0)=\overline{q}_\alpha$ and $y^+(\overline{z})=0$. Furthermore, z_4 is well defined because y^+-r_β is continuous and strictly decreasing on $[0,\overline{z}]$ and positive at $z=0$ while negative at $z=\overline{z}$.

Furthermore, it should be noticed that $y^+(z)\geq z$ if and only if $z\leq z_2$ because y^+ crosses once the 45° line. So particularly $q_\alpha^S \leq z_2$ because $q_\alpha^S = y^-(q_\alpha^S)$.

Next, I establish that $0 < z_1 \leq z_2 \leq z_3 < z_4$:

– $0 < z_1$ because $y^+(z_1) < \overline{q}_\alpha$

– $z_1 \leq z_2$ by definition of z_1: $y^+(z_1)\geq y^+(z_2)$ and y^+ decreases.

– $z_2 \leq z_3$ because $q_\beta^S \leq q_\alpha^S \leq z_2$ so $y^+(z_3) \leq y^+(z_2)$

– $z_3 < z_4$ because $y^+(z_3) = r_\beta(q_\beta^S) > r_\beta(z_3)$ the last inequality is due to $q_\beta^S \leq z_2 \leq z_3$.

Furthermore, $q_\alpha^A < z_4$ because $y^+(q_\alpha^A) > y^-(q_\alpha^A) = q_\beta^A = r_\beta(q_\alpha^A)$.

– There is a symmetric equilibrium with $t_1=t_2=\alpha$ if and only if $0\leq z\leq z_2$: Let $x=\max\{q_\alpha^S, z\}$

$\Longleftarrow$: $z \leq z_2$ implies that $x \leq z_2$ so $x \geq y^+(x)$ so $(x, \alpha), (x, \alpha)$ is an equilibrium

$\Rightarrow$: if $t_1 = t_2 = \alpha$ at equilibrium both firms produce the same quantity x and this is lower than $y^+(z)$ so $z \leq y^+(z)$ and $z \leq z_2$

- There is an asymmetric equilibrium $t_1 = \alpha$ and $t_2 = \beta$ if and only if $z_1 \leq z \leq z_4$:

  $\Longleftarrow$Two sub cases are distinguished:

  - If $q_\alpha^A \leq z_2$: $z_1 = z_2$ and $(z, \alpha), (r_\beta(z), \beta)$ is an equilibrium because $z \geq y^+(z) \geq r_\beta(z)$
  - If $q_\alpha^A \geq z_2$: $y^+(z_1) = q_\alpha^A$ and for $z \leq q_\alpha^A$, $\left(q_\alpha^A, \alpha\right), (q_\beta^A, \beta)$ is an equilibrium because $q_\alpha^A \geq y^+(z)$ and $q_\alpha^A \geq z$. And, for $z \geq q_\alpha^A$, $(z, \alpha), (r_\beta(z), \beta)$ is an equilibrium because $z \geq y^+(z) \geq r_\beta(z)$

  $\Longrightarrow$If there is an equilibrium with $t_1 = \alpha$ and $t_2 = \beta$: then $q_1 \geq y^+(z) \geq y^-(z) \geq q_2 = r_\beta(q_1)$.

  At equilibrium $q_1 = \max\{z, q_\alpha^A\}$

  - If $q_1 = z$: then $z \geq y^+(z)$ and $y^+(z) \geq r_\beta(z)$ so $z_2 \leq z \leq z_4$
  - If $q_1 = q_\alpha^A$: then $q_\alpha^A \geq y^+(z)$ and $q_\alpha^A \geq z$ so $q_\alpha^A \geq y^+(q_\alpha^A)$ and $q_\alpha^A \geq z_2$. Hence, $y^+(z_1) = q_\alpha^A$ and $q_\alpha^A \geq y^+(z)$ implies that $z \geq z_1$ and $z \leq q_\alpha^A \leq z_4$.

- There is a symmetric equilibrium with $t_1 = t_2 = \beta$ if and only if $q_\beta^S \geq y^+(z)$ if and only if $z_3 \leq z$.

B-5 Proof of corollary 3

Here, I demonstrate the following result: If p is convex, $((q_1^*, \alpha), (q_2^*, \beta))$ is an equilibrium only if $q_1^* = z$ and $q_2^* = r_\beta(z)$. I first establish that with a convex demand $y^+(q_\alpha^A) \geq q_\alpha^A$. Using the characterization of y^+ and because $q_\alpha^A \geq r_\alpha\left(q_\alpha^A\right)$ it is equivalent to:

$$(p(2q_\alpha^A) - c_\alpha)q_\alpha^A \geq \left(p(q_\alpha^A + q_\beta^A) - c_\beta\right) q_\beta^A \tag{B.1}$$

I consider the sum of marginal cost $C = c_\alpha + c_\beta$ fixed and c_β variable. So productions depend upon c_β: $q_\alpha^A(c_\beta), q_\beta^A(c_\beta)$ and $q^A = q_\alpha^A(c_\beta) + q_\beta^A(c_\beta)$. The idea is to analyze the effect of an increase of c_β in the set $[C/2, \bar{c}_\beta]$ where $\bar{c}_\beta = \min\{c_\beta \in [C/2, C] / q_\beta^A(c_\beta) = 0\}$on the difference between the two sides of the inequality above.

I use a property of interior Cournot equilibrium, the aggregate production is independent of c_β because it is the unique solution of:

$$2p(q^A) + p'(q^A)q^A = C$$

And individual productions satisfy $p(q^A)+p'(q^A)q^A_\alpha = C-c_\beta$ and $p(q^A)+p'(q^A)q^A_\beta = c_\beta$. Differentiation of these equations give:

$$\frac{\partial q^A_\alpha}{\partial c_\beta} = -\frac{1}{p'(q^A)} = -\frac{\partial q^B_\alpha}{\partial c_\beta}$$

Let consider the difference of average revenue:

$$f(c_\beta) = \big[p(2q^A_\alpha) - (C - c_\beta)\big] - \big[p(q^A_\alpha + q^A_\beta) - c_\beta\big]$$

and differentiate it with respect to c_β:

$$f'(c_\beta) = 2\left(p'(2q^A_\alpha)\frac{\partial q^A_\alpha}{\partial c_\beta} + 1\right) = 2\left(1 - \frac{p'(2q^A_\alpha)}{p'(q^A)}\right) \geq 0$$

It is positive thanks to the convexity of p and $2q^A_\alpha \geq q^A$. And $f(C/2) = 0$ so $f(c_\beta) \geq 0$ for all c_β in the set $[C/2, \bar{c}_\beta]$ so $p(2q^A_\alpha) - (C - c_\beta) \geq p(q^A_\alpha + q^A_\beta) - c_\beta$ and as $q^A_\alpha \geq q^A_\beta$, inequality (B.1) is satisfied so $y^+(q^A_\alpha) \geq q^A_\alpha$.

Next, I proceed by contradiction by assuming that $z < q^*_1$. In that case at equilibrium $q^*_1 = q^A_\alpha$ and $q^*_2 = q^A_\beta$.Therefore, $y^+(z) \leq q^A_\alpha$ and as $q^A_\alpha \leq y^+(q^A_\alpha)$ and y^+ is decreasing $z \geq q^A_\alpha$ a contradiction.

Appendix C

Proofs of chapter 3

C-1 Log concavity and best reaction function

The log concavity of the demand function is a relatively common assumption in oligopoly literature first introduced by Amir (1996). I reproduce here, a demonstration found in Vives (1999).

As *logp* is concave, the function p satisfies $p''p - (p')^2 \leq 0$. Then, if x is a solution to the first order equation: $p(x+q) + p'x = 1 - \alpha$ for $q \in [0, k^{**}]$. Hence, $p'' \leq (p')^2/(-p'x)$ so $p' + p''x \leq 0$, i.e. the cross derivative of the profit is negative, and, $2p' + p''x \leq 0$: the second order condition is satisfied. Hence, the profit function is strictly quasi-concave. It ensures the existence and uniqueness of the best response of the private firm.

The slope of the function r is given by: $(p' + p''r) / (2p' + p''r)$ which is in the set $]-1, 0[$.

C-2 Existence and value of the threshold l

As the slope of r is strictly between 0 and -1 if the solution of the equation $r(k^* - k) = k$ exists it is unique.

As $r(k^*) > 0$, the solution exits if and only if $r(0) \leq k^*$, and given quasi concavity of the profit it is equivalent to $p(k^*) + p'(k^*)k^* - (1-\alpha) \leq 0$ i.e.$p'(k^*)k^*/p \leq -\alpha$ or $-\varepsilon \leq 1/\alpha$.

The threshold can be written using the price elasticity. From the first order condition: $p(k^*) + p'(k^*)l = 1 - \alpha$, so, $l = -\alpha/p' = -\varepsilon\alpha k^*$.

C-3 Proof of Lemma 3

The result is due to the following monotonicity properties of the social surplus:

(i) $W\left(q^N(k_0,k_1),k_1+k_0\right)$ is increasing with respect to k_0 if the capacity of the private firm is binding at the production stage : $q_1^N = k_1$, and the total production is less than the long-term optimum: $k_1+k_0 \leq k^*$.

(ii) W is decreasing if the production is greater than the long-term optimum: $q^N(k_0,k_1) \geq k^*$.

For intermediary situations the monotonicity of W comes from the assumption of concavity of $W(q+r(q),q)$. This function is maximized at $\overline{k}$, so it is decreasing for $k_0 > \overline{k}$.

- If $k_1 \leq l$, the private capacity constraint is binding for $k_0 = k^* - k_1$ so $k_0^+ = k^* - k_1$.
- If $k_1 > l$, the social surplus is increasing for $k_0 \leq r^{-1}(k_1)$:
  - If $r^{-1}(k_1) \leq \overline{k}$, welfare is maximized at $k_0^+ = \overline{k}$.
  - If $r^{-1}(k_1) \geq \overline{k}$, welfare is maximized at $k_0^+ = r^{-1}(k_1)$ because of welfare concavity.

C-4 Proof of proposition 8

The first case is trivial; in that case whatever the private firm's choice smaller than k^*, the public firm is able to reach the long-term optimum.

In the second case, the threshold l is well defined. I state that $k_1 = r(\overline{k})$ is the optimal choice for the private firm.

- For $k_1 \geq r(\overline{k})$: The private firm's profit is decreasing with respect to k_1 because it over invests.
- For $l \leq k_1 \leq r(\overline{k})$: the public capacity is $r^{-1}(k_1)$ and $\overline{k} \leq r^{-1}(k_1) = k_0^+$. The production $k_0 + r(k_0)$ is increasing with respect to the public capacity so $k_0^+ + r(k_0^+) \geq \overline{k} + r(\overline{k})$ and the price is reduced: $p(k_0^+ + r(k_0^+)) \leq p(\overline{k} + r(\overline{k}))$.

  Finally by multiplying by the production and capacity, we get:

$$\begin{aligned}\left(p(k_0^+ + r(k_0^+)) - (1-\alpha)\right) q_1 - \alpha k_1 &\leq \left(p(k_0^+ + r(k_0^+)) - 1\right) k_1 \\ &\leq \left(p(\overline{k} + r(\overline{k})) - 1\right) r(\overline{k})\end{aligned}$$

So the profit of the private firm is maximized at $r(\overline{k})$.

C-5 Proof of Lemma 4

I establish here the two properties of the public firm choice of capacity. Let $(k_1, .., k_n)$ with $0 \leq k_1 \leq .. \leq k_n$ a vector of private capacities.

- If $k_1 \geq r(\overline{k}, n)/n$:
  all capacities are larger than $r(\overline{k}(n), n)/n$ and for any k_0 aggregate production is less than $k_0 + r(k_0, n)$ so welfare is smaller than
  $$W\left(k_0 + r(k_0, n), k_0\right) - \alpha \sum k_i \leq W\left(\overline{k}(n) + r(\overline{k}(n), n), k_0\right) - \alpha \sum k_i$$
  And at $k_0 = \overline{k}(n)$ all capacity constraints are relaxed so $k_0^+ = \overline{k}(n)$
- If $k_1 < r(\overline{k}, n)/n$: If $\overline{k}(n) > 0$, at $k_0 = \overline{k}(n)$ some capacity constraints are binding (at least the firm 1's one) so welfare derivative with respect to k_0 is
  $$\left[p(q^N(k_0, ..)) - (1-\alpha)\right]\frac{1}{m} - \alpha > \left[p(\overline{k}(n) + r(\overline{k}(n), n)) - (1-\alpha)\right]\frac{1}{n} - \alpha = 0$$
  Where $m \leq n$ is the number of firms whose capacity constraint is not binding at $k_0 = \overline{k}(n)$. The last equality is due to $\overline{k}(n) > 0$. If $\overline{k}(n) = 0$ the result is straightforward.

C-6 Proof of proposition 10

- If $-\varepsilon < 1/\alpha n$ i.e. $a - 1 > n\alpha$: I establish that:
  $$\text{for } i = 1, ..n, k_i = \frac{r(\overline{k}(n), n)}{n}, k_0 = \overline{k}(n) \text{ and } q^N = \sum_{i=0}^{n} k_i < k^*$$
  is an equilibrium. So I assume that $k_2 = ... = k_n = r(\overline{k}(n), n)/n$ and show that $k_1 = k_2$ maximizes firm 1's profit.
  - For $k_1 > k_2$: $k_0^+ = \overline{k}$ and aggregate production is $\overline{k} + r(\overline{k}, n)$ and $q_1^N < k_2$ so firm 1 has excess capacity, its profit is lower than at $k_1 = k_2$.
  - For $k_1 \leq k_2$: welfare is increasing as long as all capacity constraints are binding. Capacity constraints of firms $2, .., n$ are relaxed simultaneously at $k_0 = \overline{k} + k_2 - k_1$ because $r(\overline{k} + (n-1)k_2) = k_2$ because by definition $r(q, n) = r(q + (n-1)r(q, n)/n)$. Hence, $k_0^+ \geq \overline{k} + k_2 - k_1$ and aggregate production is larger than $\overline{k} + r(\overline{k}(n), n)$ so the price is lower than $p\left(\overline{k} + r(\overline{k}(n), n)\right)$ and firm 1's profit is smaller than the profit obtained for $k_1 = k_2$.

- If $-\varepsilon \geq 1/\alpha n$ i.e. $a-1 \leq n\alpha$:
  First, $[\max\left\{(a-1-\alpha)/(n-1), 0\right\}, k^*/n]$ is not empty because $(a-1-\alpha)/(n-1) \leq (a-1)/n$.
  Let $k \in [\max\left\{(a-1-\alpha)/(n-1), 0\right\}, k^*/n]$. I establish that $k_i = k$, for $i = 1, ..n$, is an equilibrium by considering deviation of firm 1.
  - For $k_1 \leq k$: $k_0^+ = k^* - (n-1)k - k_1$ because at $k_0^+ = k^* - (n-1)k - k_1$ all capacity constraints are binding because:
    $r(k^* - k) = \alpha/2b + k/2 \geq k^*/2n + k/2 \geq k$
    So all capacity constraints are binding at $k^* - (n-1)k - k_1$ so $k_0^+ = k^* - (n-1)k - k_1$ and the profit of firm 1 is zero for all $k_1 \leq k$.
  - For $k_1 > k$: I proceed by contradiction by assuming that there is a $k_1 \geq k$ such that firm 1 obtains a strictly positive profit at equilibrium. It is the case only if $q^N(k_0^+, k_1, k, .., k) \leq k^*$.
    As firm 1 has the largest capacity its capacity constraint is relaxed first at $k_0 = \max\{r^{-1}(k_1) - (n-1)k, 0\}$, and capacity constraints of other firms are relaxed simultaneously for a higher k_0. So three situations can arise: either all capacity constraints are binding, or they are all relaxed or only firm 1's capacity constraint is relaxed. I establish that in any case firm 1 cannot obtain a positive profit.
    If all capacity constraint are binding at equilibrium the aggregate production is at least $k_1 + (n-1)k$ so $k_1 \leq k^* - (n-1)k \leq \alpha/b$. The last inequality is due to the lower bound on k. And $k \leq k_1 \leq \alpha/b$ implies that $k_0^+ = k^* - (n-1)k - k_1$ and $q^N = k^*$ a contradiction.
    If none capacity constraint is binding $q^N \geq r(0, n) > k^*$.
    If firm 1 is the only firm that does not produce at full capacity, aggregate production is at least $r\left((n-1)k\right) + (n-1)k$ but $k \geq (a-1-\alpha)/(n-1)b$ implies that $r\left((n-1)k\right) + (n-1)k \geq (a-1)/b = k^*$, a contradiction.

Appendix D

Proof of chapter 4

D-1 Existence of uniqueness of Cournot equilibrium

Let $i \in \{1, 2\}$, I establish that for any permit price σ_i there exists a unique equilibrium which is symmetric on market i. To do so I consider the gross cost function $\Gamma(q_i, \sigma_i) = \min_{e_i}\{C_i(q_i, e_i) - \sigma_i e_i\}$.

This function is convex with respect to q_i by derivation:

$$\frac{\partial \Gamma}{\partial q_i} = \frac{\partial C_i}{\partial q_i}(q_i, e_i) \text{ and } \frac{\partial^2 \Gamma}{\partial q_i^2} = \frac{\partial^2 C_i}{\partial q_i^2} - \left(\frac{\partial^2 C_i}{\partial q_i \partial e_i}\right)^2 \left(\frac{\partial^2 C_i}{\partial e^2}\right)^{-1} \geq 0$$

Thanks to the assumption on the price function $P_i' + P_i'' Q_i, \forall Q_i$ there is a unique Cournot equilibrium for all σ_i (Novshek 1985).

Equilibrium quantities satisfy the two equations:

$$P_i + P_i' q_i = \frac{\partial C_i}{\partial q_i}$$
$$\sigma_i = \frac{\partial C_i}{\partial e_i}$$

Hence, for any permit price the strategy $\left(q_i^C(e_i^C), e_i^C(n_i, \sigma_i)\right)$ is the unique (symmetric) equilibrium of the game considered.

D-2 Proof of corollary 9

With a subsidy s on emissions of market 1 and an integrated market for emissions with a price σ, the price of emissions faced by firms on market 1 is $\sigma + s$ and by firm on market 2 σ, therefore, at equilibrium respective emissions per firm are $e_1^C(n_1, \sigma{+}s)$ and $e_2^C(n_2, \sigma)$and therefore:

$$\frac{\partial C_1}{\partial e_1} = \frac{\partial C_2}{\partial e_2} + s \tag{D.1}$$

For a subsidy s^*, emissions satisfy the first order condition of the second best equation (4.6), so, thanks to the assumption of welfare concavity, it is the second best allocation.

D-3 Proof of proposition 12

With quadratic expressions welfare could be written (4.12) as in the main text. When an integrated market is implemented equilibrium production on market 1 is given by expression (4.13): $Q_1^C = \overline{Q_1^C} + (\theta_1 - \theta_2)/B$ and the expected production is:

$$\frac{1}{B(n_1, n_2)}\left[(a_1 - a_2) + \frac{n_2 + 1}{n_2} b_2 \overline{e} + s\right] \tag{D.2}$$

Quadratic expressions and additive uncertainty on marginal surpluses allow to separate effects of expected production and random term as in expression (4.14). The effect of expected production is:

$$EW\left(\overline{Q_1^C}, \overline{Q_2^C}, \theta_1, \theta_2\right) = \hat{W} - \frac{b_1 + b_2}{2}\left(\overline{Q_1^C} - Q_1^*\right)^2$$

And the difference between Cournot production and optimal one is:

$$\overline{Q_1^C} - Q_1^* = \frac{1}{B}\left[(a_1 - a_2) + \frac{n_2 + 1}{n_2} b_2 \overline{e}\right] - \frac{1}{b_1 + b_2}\left[(a_1 - a_2) + b_2 \overline{e}\right]$$

$$= \frac{1}{B.(b_1 + b_2)}\left[(a_1 - a_2)(b_1 + b_2 - B) + b_2 \overline{e}\left(\frac{n_2 + 1}{n_2}(b_1 + b_2) - B\right)\right]$$

$$= \frac{1}{B.(b_1 + b_2)}\left[(a_2 - a_1)\left(\frac{b_1}{n_1} + \frac{b_2}{n_2}\right) + b_1 b_2 \overline{e}\left(\frac{1}{n_2} - \frac{1}{n_1}\right)\right]$$

Therefore, the difference between command and control and an integrated permits market without subsidy is:

$$\begin{aligned}\triangle(0) = {} & \frac{var(\theta_1 - \theta_2)}{B^2}(B - \frac{1}{2}(b_1 + b_2)) \\ & - \frac{1}{2B^2(b_1 + b_2)}\left[(a_2 - a_1)\left(\frac{b_1}{n_1} + \frac{b_2}{n_2}\right) + b_1 b_2 \overline{e}\left(\frac{1}{n_2} - \frac{1}{n_1}\right)\right]\end{aligned}$$

And an integrated market increases welfare if and only if: $var(\theta_1 - \theta_2) \geq \frac{1}{((2+n_1)b_1 + (2+n_2)b_2)(b_1+b_2)}\left[(a_2 - a_1)(n_2 b_1 + n_1 b_2) + b_1 b_2 \overline{e}(n_1 - n_2)\right]$.

D-4 Proof and generalization of corollary 12

In this section I demonstrate the results of corollary 4 in the case of a continuum of demand states for the case of additive variations. Here variations are not random so the permit price is the same in all demand states. Demand states are represented by $\varepsilon \in [0,1]$, the inverse demand function in a demand state is $P(Q)+\varepsilon$ and $P'(x+y)+P''x \leq 0$ for all $x, y \geq 0$. I consider symmetric distribution of capacity (cf Grimm and Zoetl (2005) for a demonstration of existence and uniqueness of symmetric equilibrium).

1. Equilibrium capacity is lower and base production higher with Cournot competition:

For a permit price σ, in low demand states the capacity constraint is not binding and firms produce the Cournot production $Q^C(\sigma,\varepsilon)$ that satisfies the first order condition:$P + P'Q^C/n + \varepsilon = \sigma$. The capacity constraint of all firms is binding for $\varepsilon \geq \tilde{\varepsilon}$, the threshold depends on the quantity of capacity and satisfies $P(K)+P'(K)k-\sigma+\tilde{\varepsilon}=0$. At the symmetric equilibrium of the capacity game, aggregate and individual quantities of capacity satisfy:

$$c = \int_{\tilde{\varepsilon}}^{1} P(K^C)+\varepsilon+P'(K^C)k^C-\sigma dF(\varepsilon) \tag{D.3}$$

For variations are additive, injecting the expression of $\tilde{\varepsilon}$ gives:

$$c = \int_{\tilde{\varepsilon}}^{1} \varepsilon - \tilde{\varepsilon} dF(\varepsilon) \tag{D.4}$$

Hence on the long run the value of the threshold state $\tilde{\varepsilon}$ only depends of the distribution of the random variable and the cost of capacity. Particularly it is not influenced by the number of active firms n. So one can consider that states $\varepsilon < \tilde{\varepsilon}$ are base demand states and $\varepsilon > \tilde{\varepsilon}$ are peak ones. Similar calculations can be done for the first best - competitive- situation. It appears that the threshold state at the first best optimum also satisfies equation (D.4). With a slight abuse of notation the solution of equation (D.4) is also denoted $\tilde{\varepsilon}$.

The permit price clears the permits market so that aggregate production with Cournot competition is equal to the first best aggregate production:

$$\int\limits_{\varepsilon} Q^C(\sigma^C,\varepsilon) - Q^*(\sigma^*,\varepsilon) dF(\varepsilon) = 0$$

Both productions are increasing with respect to ε and I assume that:

$$0 < \frac{\partial Q^C}{\partial \varepsilon} = \frac{1}{-P' - \frac{1}{n}(P' + P''Q^C)} < \frac{\partial Q^*}{\partial \varepsilon} \text{ for } \varepsilon \leq \tilde{\varepsilon}$$

This assumption is satisfied with a linear demand and while apparently relatively innocuous seems difficult to avoid to obtain subsequent results. With only two demand states as in the main text, it could be avoided.

Hence the difference $Q^C - Q^*$ is decreasing with respect to ε. Capacity constraints are binding in both cases for $\varepsilon \geq \tilde{\varepsilon}$. Hence, the difference $Q^C - Q^*$is positive for small ε and negative for high ε i.e. $k^C(n) < k^*$.

2. An increase of the number of firms decreases base production and increases capacity and welfare:

Above reasoning can be reproduced to show that $Q^C(n,\varepsilon) > Q^C(n+1,\varepsilon)$ in low demand states and $k^C(n) < k^C(n+1)$.

Furthermore, the reallocation of the constraint due to an increase of the number of firms increases welfare:

I assume that $Q^C$and $k^C$are differentiable with respect to n and relaxed the integer constraint, the derivative of welfare with respect to n is:

$$\frac{\partial W}{\partial n} = \int\limits_{0}^{\tilde{\epsilon}} (P+\varepsilon) \frac{\partial Q^C}{\partial n} dF(\varepsilon) + \left[\int\limits_{\tilde{\varepsilon}}^{1} P + \varepsilon dF(\varepsilon) - c \right] \frac{\partial k^C}{\partial n} \tag{D.5}$$

Because aggregate quantities are equal whatever the number of firms the following relationship holds between the derivative of production in low demand states and capacity:

$$\frac{\partial k^C}{\partial n} = \frac{1}{1 - F(\tilde{\varepsilon})} . \int\limits_{0}^{\tilde{\epsilon}} \frac{\partial Q^C}{\partial n} dF(\varepsilon) \tag{D.6}$$

And, first order conditions imply that:

$$P(Q^C) + \varepsilon = \sigma^C - P'(Q^C)\tfrac{Q^C}{n}$$
$$\int_{\tilde{\varepsilon}}^{1} P + \varepsilon dF(\varepsilon) - c = \left[1 - F(\tilde{\varepsilon})\right]\left[\sigma^C(n) - P'(k^C)\right]$$

Injecting these three relations into the expression (D.6) of welfare derivative gives:

$$\frac{\partial W}{\partial n} = \int_0^{\tilde{\epsilon}} \left[-P'(Q^C)\frac{Q^C}{n} - P'(k^C)\frac{k^C}{n}\right]\frac{\partial Q^C}{\partial n} dF(\varepsilon)$$

And finally this expression is positive because (i) $-P'(Q^C)Q^C - P'(k^C)k^C$ is increasing with respect to ε(because $P'.Q$ is decreasing with respect to Q) and (ii) the sign of $\partial Q^C/\partial n$changes once to become positive.

3. A subsidy of capacity increases welfare:

Let's consider a subsidy s on capacity and denote $Q^C(s)$ and $k^C(s)$ the Cournot quantities with such a subsidy. I assume that these quantities can be differentiated with respect to s and welfare also. First, with a subsidy of capacity the equilibrium quantity of capacity increases and the quantity produced in low demand states decreases. The derivative of welfare with respect to s is similar to (D.6) and similar manipulations can be done with an additive term:

$$\frac{\partial W}{\partial s} = \int_0^{\tilde{\epsilon}} \left[-P'(Q^C)\frac{Q^C}{n} - P'(k^C)\frac{k^C}{n}\right]\frac{\partial Q^C}{\partial s} dF(\varepsilon) - s\frac{\partial k^C}{\partial s}$$

It is strictly positive at $s = 0$, hence a striclty positive subsidy increases welfare.

Bibliography

Adilov, N. (2006). Forward Markets, Market Power, and Capacity Investment, *Cornelle University, working paper* . 142

Aghion, P. and Bolton, P. (1987). Contracts as a Barrier to Entry, *American Economic Review* **77**(3): 388–401. 142

Allaz, B. and Vila, J. (1993). Cournot competition, futures markets and efficiency, *Journal of Economic Theory* **59**(1): 1–16. 142

Amir, R. (1996). Cournot Oligopoly and the Theory of Supermodular Games, *Games and Economic Behavior* **15**(2): 132–148. 70, 155

Amir, R. (2005). Ordinal versus cardinal complementarity: The case of Cournot oligopoly, *Games and Economic Behavior* **53**(1): 1–14. 70

Amir, R., Garcia, F. and Knauff, M. (2006). Endogenous Heterogeneity in Strategic Models: Symmetry-breaking via Strategic Substitutes and Nonconcavities, *CORE Discussion Paper No. 2006/8* . 68

Amir, R. and Nannerup, N. (2005). Asymmetric Regulation of Identical Polluters in Oligopoly Models, *Environmental and Resource Economics* **30**(1): 35–48. 117

Arellano, M. and Serra, P. (2007). A model of market power in electricity industries subject to peak load pricing, *Energy Policy* **35**(10): 5130–5135. 43

Arrow, K. (1975). Vertical Integration and Communication, *Bell Journal of Economics* **6**(1): 173–183. 142

Barnett, A. (1980). The Pigouvian Tax Rule under Monopoly, *American Economic Review* **70**(5): 1037–41. 112

Boiteux, M. (1960). Peak-Load Pricing, *The Journal of Business* **33**(2): 157–179. 23

Borenstein, S., Bushnell, J., Kahn, E. and Stoft, S. (1995). Market Power in California Electricity Markets, *Utilities Policy* **5**(3-4): 219–236. 27

Buchanan, J. (1969). External Diseconomies, Corrective Taxes, and Market Structure, *American Economic Review* **59**(1): 174–177. 112

Chao, H., Oren, S. and Wilson, R. (2005). Restructured Electricity Markets: Reevaluation of Vertical Integration and Unbundling, *EPRI Technical Paper, July* . 26

Corchón, L. (2008). Welfare losses under Cournot competition, *International Journal of Industrial Organization* . 60, 69

Corchón, L. and Fradera, I. (2002). Comparative statics in Cournot free entry equilibrium, *Mathematical Social Sciences* **44**(2): 155–168. 140

Cournot, A. A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses*, Dunod, 2001. 28

Crampes, C. and Creti, A. (2003). Capacity Competition in Electricity Markets, *Universite des Sciences Sociales de Toulouse, mimeo* . 27

Cramton, P. and Stoft, S. (2006). The Convergence of Market Designs for Adequate Generating Capacity, *A White Paper for the Electricity Oversight Board* . 41

Cremer, H., Marchand, M. and Thisse, J. (1989). The Public Firm as an Instrument for Regulating an Oligopolistic Market, *Oxford Economic Papers* **41**(1): 283–301. 90

Crew, M., Fernando, C. and Kleindorfer, P. (1995). The Theory of Peak-Load Pricing: A Survey, *Journal of Regulatory Economics* **8**(3): 215–248. 23

de Fraja, G. and Delbono, F. (1990). Game Theoretic Models of Mixed Oligopoly, *Journal of Economic Surveys* **4**(1): 1–17. 90

D.G.COMP (2007). *Structure and Performance of six European Wholesale Electricity Markets in 2003, 2004 and 2005*, European Commission, (Report prepared by London Economics). 42

DGEMP (2003). *Coûts de référence de la production électrique*, Ministère de l'économie, des finances et de l'industrie. 16, 17

Dixit, A. (1979). A model of duopoly suggesting a theory of barriers to entry, *Bell Journal of Economics* **10**: 20–32. 90

Dixit, A. (1980). The Role of Investment in Entry Deterrence, *Economic Journal* **90**(357): 95–106. 67, 90

Ehrenmann, A. and Smeers, Y. (2008). Energy only, capacity market and security of supply. A stochastic equilibrium analysis, *CORE Université catholique de Louvain, working paper* . 141

Elberfeld, W. (2003). A note on technology choice, firm heterogeneity and welfare, *International Journal of Industrial Organization* **21**(4): 593–605. 66

Fabra, N., von der Fehr, N. and de Frutos, M. (2008). Investment Incentives and Auction Design in Electricity Markets, *CEPR Discussion Paper No. DP6626* . 28, 43

Fabra, N., Von Der Fehr, N. and Harbord, D. (2006). Designing Electricity Auctions, *The RAND Journal of Economics* **37**(1): 23–46. 27

Finon, D. (2008). Investment Risk Allocation in Decentralised Electricity Markets - The Need of Long-Term Contracts and Vertical Integration, *OPEC Energy Review* **32**(2): 150–183. 26, 41

Finon, D. and Pignon, V. (2008). Electricity and long-term capacity adequacy: The quest for regulatory mechanism compatible with electricity market, *Utilities Policy* **16**(3): 143–158. 21

Finon, D. and Roques, F. (2008). Contractual and financing arrangements for nuclear investment in liberalized markets : which efficient combination?, *Larsen Working Paper 13* . 26

Gabszewicz, J. and Poddar, S. (1997). Demand Fluctuations and Capacity Utilization under Duopoly, *Economic Theory* **10**(1): 131–146. 42

Green, R. (2006). Investment and Generation Capacity, *in F. Lévêque, Competitive Electricity Market and Sustainability, Edward Elgar* . 23

Green, R. J. and Newbery, D. M. (1992). Competition in the British Electricity Spot Market, *The Journal of Political Economy* **100**(5): 929–953. 27

Gromb, D., Ponssard, J. and Sevy, D. (1997). Selection in Dynamic Entry Games, *Games and Economic Behavior* **21**(1-2): 62–84. 141

Hirshleifer, D. (1988). Risk, Futures Pricing, and the Organization of Production in Commodity Markets, *The Journal of Political Economy* **96**(6): 1206. 141

Hung, N. and Sartzetakis, E. (1998). Cross-Industry Emission Permits Trading, *Journal of Regulatory Economics* **13**(1): 37–46. 113

Joskow, P. and Kahn, E. (2001). A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000, *NBER Working Paper* . 27

Joskow, P. L. (2006). Competitive Electricity Markets and Investment in New Generating Capacity, *MIT, W.P. 06-009* . 26, 41, 90

Joskow, P. L. and Tirole, J. (2006). Procurement by the ISO: Supplementary Material for "Reliability and Competitive Electricity Markets" , *Mimeo MIT and IDEI* . 26, 90

Joskow, P. L. and Tirole, J. (2007). Reliability and Competitive Electricity Markets, *RAND journal of economics* **38**(1): 60–84. 23, 40, 56, 90

Klemperer, P. and Meyer, M. (1989). Supply Function Equilibria in Oligopoly under Uncertainty, *Econometrica* **57**(6): 1243–1277. 27

Lahiri, S. and Ono, Y. (1988). Helping Minor Firms Reduces Welfare, *Economic Journal* **98**(393): 1199–1202. 69, 83

Lu, Y. and Poddar, S. (2005). Mixed oligopoly and the choice of capacity, *Research in Economics* **59**(4): 365–374. 91

Mankiw, N. and Whinston, M. (1986). Free Entry and Social Inefficiency, *Rand Journal of Economics* **17**(1): 48–58. 66, 84, 140

Maskin, E. (1999). Uncertainty and entry deterrence, *Economic Theory* **14**(2): 429–437. 67

Merrill, W. and Schneider, N. (1966). Government Firms in Oligopoly Industries: A Short-Run Analysis, *Quarterly Journal of Economics* **80**(3): 400–12. 90

Mills, D. and Smith, W. (1996). It pays to be different: Endogenous heterogeneity of firms in an oligopoly, *International Journal of Industrial Organization* **14**(3): 317–329. 66, 67

Montgomery, W. (1972). Markets in Licenses and Efficient Pollution Control Programs, *Journal of Economic Theory* **5**(3): 395–418. 118

Murphy, F. and Smeers, Y. (2005). Generation Capacity Expansion in Imperfectly Competitive Restructured Electricity Markets, *Operations research* **53**(4): 646. 43, 44

Neuhoff, K., Grubb, M. and Keats, K. (2005). Impact of the allowance allocation on prices and efficiency, *University of Cambridge Working Papers in Economics* (0552). 114

Newbery, D. M. (1998). Competition, Contracts, and Entry in the Electricity Spot Market, *Rand Journal of Economics* **29**(4): 726–749. 19, 42

Newbery, D. and Stiglitz, J. (1981). *Theory of Commodity Price Stabilization*, Oxford University Press. 141

Novshek, W. (1980). Cournot Equilibrium with Free Entry, *The Review of Economic Studies* **47**(3): 473–486. 67

Novshek, W. (1985). On the Existence of Cournot Equilibrium, *The Review of Economic Studies* **52**(1): 85–98. 70, 159

Pal, D. (1998). Endogenous timing in a mixed oligopoly, *Economics Letters* **61**(2): 181–185. 90

Ponssard, J. (1979). The Strategic Role of Information on the Demand Function in an Oligopolistic Market, *Management Science* **25**(3): 243–250. 141

Ponssard, J. (1991). Forward Induction and Sunk Costs Give Average Cost Pricing, *Games and Economic Behavior* **3**: 221–236. 141

Requate, T. (2005). Environmental policy under imperfect competition– a survey, *Christian-Albrechts-Universität Kiel, Department of Economics: Economics Working Paper* (2005-12). 112

Reynolds, S. and Wilson, B. (2000). Bertrand–Edgeworth Competition, Demand Uncertainty, and Asymmetric Outcomes, *Journal of Economic Theory* **92**(1): 122–141. 28, 43

Roques, F., Nuttal, W., Newbery, D., de Neufville, R. and Connors, S. (2006). Nuclear power: A hedge against uncertain gas and carbon prices?, *The Energy journal* **27**(4): 1–23. 41, 141

Sartzetakis, E. (1997). Tradeable Emission Permits Regulations in the Presence of Imperfectly Competitive Product Markets: Welfare Implications, *Environmental and Resource Economics* **9**(1): 65–81. 113

Sartzetakis, E. (2004). On the Efficiency of Competitive Markets for Emission Permits, *Environmental and Resource Economics* **27**(1): 1–19. 113

Schmalensee, R. (1981). Economies of Scale and Barriers to Entry, *The Journal of Political Economy* **89**(6): 1228–1238. 67

Segal, I. and Whinston, M. (2000). Exclusive Contracts and Protection of Investments, *The RAND Journal of Economics* **31**(4): 603–633. 142

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games, *International Journal of Game Theory* **4**(1): 25–55. 93

Simpson, R. (1995). Optimal pollution taxation in a Cournot duopoly, *Environmental and Resource Economics* **6**(4): 359–369. 112

Spence, A. (1977). Entry, Capacity, Investment and Oligopolistic Pricing, *Bell Journal of Economics* **8**(2): 534–544. 90

Stoffaës, C. (1994). *Entre monopole et concurrence: la régulation de l'énergie en perspective historique*, Editions PAU. 17

Vives, X. (1999). *Oligopoly Pricing: Old Ideas and New Tools*, MIT Press. 93, 116, 155

von der Fehr, N.-H. and Harbord, D. (1997). Capacity Investment and Competition in Decentralised Electricity Markets, *Department of Economics University of Oslo, memorandum 27/97* . 43, 56

von der Fehr, N.-H. M. and Harbord, D. (1993). Spot Market Competition in the UK Electricity Industry, *The Economic Journal* **103**(418): 531–546. 27

Ware, R. (1984). Sunk Costs and Strategic Commitment: A Proposed Three-Stage Equilibrium, *The Economic Journal* **94**(374): 370–378. 90

Weitzman, M. (1974). Prices vs. Quantities, *Review of Economic Studies* **41**(4): 477–491. 124

Wilson, R. (2002). Architecture of Power Markets, *Econometrica* **70**(4): 1299–1340. 19

Wolak, F. and Patrick, R. H. (2001). The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market, *NBER Working Paper* . 27

Wolfram, C. (1999). Measuring Duopoly Power in the British Electricity Spot Market, *The American Economic Review* **89**(4): 805–826. 27, 44

Zoetl, K. (2008). *Investment Incentives under Uncertainty*, PhD thesis, CORE, Louvain la Neuve. 28, 43, 142