

La nouvelle théorie de la régulation des monopoles naturels

Philippe Gagnepain*

*Department of Economics
Universidad Carlos III Madrid*

Janvier 2001

Résumé

L'inefficacité d'une entreprise résulte d'un ensemble de caractéristiques et de décisions en partie non-identifiables par un observateur extérieur au processus de production. L'analyse économique récente a cherché à tirer partie de ces situations et a permis de les appréhender en traitant les problèmes d'asymétrie d'information entre une autorité et un agent. Ces problèmes intègrent des situations d'aléa moral où les agents prennent des actions non-observables par l'autorité et des situations d'anti-sélection où les agents disposent d'informations privées sur leur environnement technologique. La régulation des monopoles naturels a ainsi fait l'objet d'un réexamen à la lumière de la théorie des incitations à partir du début des années 80. Cet article propose un aperçu de la nouvelle théorie de la régulation des monopoles naturels et tente de fournir un état presque exhaustif des travaux empiriques réalisés sur le sujet.

The inefficiency of a firm depends on several characteristics and decisions which are non-observable by an external individual. The economic theory has recently taken into account these situations through asymmetric information problems between principals and agents. Such problems consider moral hazard and adverse selection issues. The regulation of natural monopolies has thus been re-examined through the window of the theory of incentives. This article proposes an overview of the new theory of regulation and tries to provide an almost exhaustive state of empirical works realized so far on the topic.

* Correspondance : Department of Economics, Universidad Carlos III de Madrid, Calle Madrid, 126, 28903 Getafe, Madrid – Espagne. Tél : (34) 91 624 5732 ; Fax: (34) 91 624 9329. E-mail : philippe@eco.uc3m.es

1. Introduction

Une activité de production caractérisée par des rendements d'échelle croissants est plus coûteuse lorsqu'elle est assurée par plusieurs unités plutôt que par une seule. Le fait qu'une entreprise bénéficie d'économies d'échelle au cours du processus de production justifie sa position de monopole sur le marché. Ce principe s'applique par exemple à des industries aussi diverses que l'électricité, le gaz, les télécommunications, le transport sous toutes ses formes. A un degré moindre, le recours à un monopole est également motivé par des soucis de coordination des services. Ce dernier point est particulièrement vrai dans le cadre du transport des passagers. Les monopoles en place peuvent être publics ou privés et font l'objet d'une régulation par une autorité centrale ou locale qui souhaite restreindre leur pouvoir monopolistique.

Au cours des vingt dernières années, le paysage industriel mondial a fait l'objet de nombreux bouleversements dans le domaine de la régulation des entreprises et de l'organisation des marchés des biens de consommation et des services. Le débat entre les adeptes des systèmes capitalistes et socialistes a laissé la place à un autre débat axé sur l'éventuelle nécessité de l'intervention de l'Etat dans la régulation des monopoles publics ou privés. Ce changement a eu plusieurs conséquences : dans un premier temps, les autorités publiques ont fait usage de schémas réglementaires incitatifs en matière de productivité ; dans un second temps, ces mêmes autorités ont songé à la possibilité de déréglementer les marchés, de privatiser les entreprises et de faire appel à la concurrence.

Avant les années 80, la théorie de la régulation des monopoles naturels a largement ignoré les problèmes informationnels susceptibles d'interférer dans les rapports entre une entreprise régulée et une autorité.¹ La théorie proposée jusqu'au début des années 80 s'est principalement concentrée sur la partie demande de la régulation, c'est à dire la tarification proposée aux consommateurs. La régulation à la Ramsey-Boiteux (Boiteux, 1956) fut par exemple extrêmement populaire dans la théorie économique. L'importance des effets de la régulation de l'offre sur les coûts n'a été prise en considération que plus tard, lorsqu'on a pris conscience que la portée du principe des mécanismes incitatifs sur les gains potentiels en termes de bien-être pour la société était bien plus importante.

Ce n'est donc qu'à partir des années 80 que le principe de la régulation a été pensé en termes de relations principal-agent avec les travaux de Loeb et Magat (1979), Baron et Myerson (1982) et Laffont et Tirole (1986).² Les relations étudiées supposent un déficit informationnel chez l'autorité (le principal). Généralement, l'opérateur (l'agent) dispose d'une connaissance privée de la technologie utilisée, de l'état des coûts d'exploitation ou de la demande du marché. La nouvelle théorie de la régulation propose donc de considérer systématiquement les rapports entre régulateur et opérateur à travers le cadre normatif principal-agent. S'inspirant des techniques de la conception de mécanismes, cette approche élabore les mécanismes régulateurs optimaux et requiert pour cela une définition précise des objectifs de l'autorité et de la firme et une prise en compte rigoureuse des contraintes économiques et informationnelles. Les contraintes informationnelles constituent l'enjeu majeur de cette théorie. Leur existence gêne le contrôle de l'autorité et empêche celle-ci de mettre en place la politique réglementaire qui s'avère être la meilleure pour la société.

On distingue généralement deux types de contraintes informationnelles, l'anti-sélection et l'aléa moral.

L'aléa moral se rapporte à l'ensemble des variables endogènes qui ne sont pas observables par le régulateur mais sont connues par l'opérateur. Ces variables synthétisent les

actions de l'entreprise qui affectent sa production et/ou ses coûts. L'effort de productivité est le terme généralement considéré. Il consiste par exemple à investir dans des activités de recherche et de développement mais il peut également être interprété comme le soin avec lequel un manager surveille le travail de ses employés, gère une grève, achète des facteurs de production à un prix raisonnable ou affecte à des tâches précises des employés aux formations adéquates.

L'anti-sélection se rapporte à l'ensemble des variables exogènes qui ne sont pas observables par le régulateur mais sont connues par l'opérateur. Elles sont caractérisées par exemple par les capacités productives de l'opérateur, c'est à dire la technologie dont il dispose et qu'il utilise pour assurer la prestation d'un service ou la production d'une quantité de biens. Le déficit informationnel de l'autorité peut également être lié à la demande des consommateurs et plus particulièrement à son élasticité-prix. La présence d'une telle incertitude sur l'environnement productif d'un opérateur autorise celui-ci à disposer de rentes. En effet, une entreprise efficace, c'est à dire bénéficiant par exemple de coûts de production faibles, peut prétendre supporter des coûts élevés. Ainsi, elle reçoit une rente positive puisqu'elle touche un remboursement supérieur aux coûts anticipés. Elle peut également exercer un niveau d'effort faible de façon à ce que ses coûts de production soient juste remboursés par le montant fixé par le contrat, ce qui lui permet également de bénéficier d'une rente positive.

Cet article propose principalement un aperçu du cadre théorique de la nouvelle théorie de la régulation et des tests de cette théorie effectués par la littérature empirique. Nous nous concentrons sur le cadre normatif particulier où l'autorité fait face simultanément dans un environnement statique à des problèmes d'anti-sélection et d'aléa moral et peut observer les coûts agrégés réels de l'opérateur en fin de période. Le modèle présenté s'inspire donc essentiellement des travaux de Jean-Jacques Laffont et Jean Tirole. L'étude considère des opérateurs offrant des services sur des marchés organisés en réseaux. Ces services répondent le plus souvent à des règles de service universel.

Comme une présentation pertinente de la théorie de la régulation des monopoles naturels ne peut se limiter à la seule description d'un cadre normatif statique entre deux agents uniques, nous proposons d'étendre la discussion sur plusieurs points. Tout d'abord, pour aller au-delà de la régulation des monopoles naturels dans un cadre statique (l'exploitation d'un service ou la production d'un bien peut être reconduite d'une période à l'autre), il convient de présenter également une discussion des modèles dynamiques de la régulation. Ensuite, la reconduction tacite des contrats d'exploitation d'une période à l'autre est de plus en plus contestée ; les procédures d'appels d'offre, qui consistent à délivrer une licence de monopole à l'entreprise la plus efficace parmi plusieurs candidats, font l'objet d'intérêts grandissant de la part des autorités publiques. Enfin, un nombre croissant d'industries est soumis au principe de l'ouverture à la concurrence. Ainsi, l'opérateur principal, fait face à de nouveaux entrants qui proposent des services complémentaires ou substituables au service initial. Ces différentes perspectives ont un impact important sur le comportement des agents productifs et plus particulièrement sur l'efficacité du processus de production considéré.

La section 2 présente les schémas réglementaires habituellement utilisés mais qui ne sont pas nécessairement optimaux car ils négligent les contraintes informationnelles considérées par la théorie. Ces schémas particuliers s'inscrivent à l'intérieur du support théorique optimal présenté dans la section 3. La section 4 considère le cas des systèmes d'appels d'offre et suggère que le mécanisme optimal qui en résulte est le même que celui présenté dans la section précédente. La section 5 propose de considérer le cas de la dynamique de la régulation. La section 6 évoque l'ouverture à la concurrence des marchés de monopole, la section 7 présente les rares approches empiriques qui ont pour cadre général la régulation des monopoles en

situation d'information asymétrique. La section 8 propose une courte discussion sur l'utilisation des mécanismes optimaux en pratique et conclut.

2. Mécanismes de régulation usuels

Le programme de régulation d'un monopole définit un ensemble d'objectifs économiques que le monopole doit respecter. Le contenu de ces objectifs varie en fonction du schéma considéré. La rémunération de l'opérateur dépend du cadre normatif qui fixe les objectifs de recettes et de coûts.

Les mécanismes de régulation actuels se distinguent les uns des autres selon deux critères. Tout d'abord, l'entreprise régulée est ou n'est pas autorisée à percevoir un transfert de la part de l'autorité afin de pallier un déficit budgétaire anticipé ou réel. Ensuite, on distingue couramment les mécanismes à caractère incitatif et non incitatif en termes d'effort de productivité, c'est à dire, en termes de réduction des coûts.

Régimes de type coût du service

Le cas où l'entreprise touche un transfert est généralement particulier aux contrats qui lient une entreprise à une autorité de régulation centrale ou locale. Dans le premier cas, l'acheteur et l'autorité sont généralement les mêmes personnes. La Défense Nationale est un exemple d'activité qui correspond bien à la première définition tandis que le transport urbain est exemplaire de la seconde. Le régulateur fixe généralement les niveaux d'offre de service à assurer par l'entreprise et les niveaux de prix que celle-ci doit appliquer. Ces contrats sont très peu incitatifs dans la mesure où l'opérateur bénéficie d'un remboursement total de ses coûts réels. Le transfert obtenu est indépendant de la performance réalisée. L'entreprise régulée dans ce cas là n'est pas nécessairement disposée à fournir le niveau d'effort optimal. D'une part, elle ne prend aucun risque financier, d'autre part elle n'a pas intérêt à améliorer sa productivité puisqu'elle ne touche pas les dividendes de l'effort consenti (elle ne peut espérer aucun profit).

Régimes de type prix-fixe

A l'opposé des schémas précédents, on trouve les contrats de type *prix-fixe*. L'autorité fixe les niveaux d'offre et de prix. Dans ce cas là, l'entreprise supporte entièrement ses coûts d'exploitation. Elle reçoit en contrepartie un montant forfaitaire anticipé qui est supposé garantir l'équilibre financier de l'exploitation. On comprend bien que tout dépassement des coûts anticipés soit alors à la charge de l'opérateur, ce qui lui fournit de très fortes incitations pour exercer le niveau d'effort qui garantit une réduction des coûts optimale. Les contrats de type *prix-fixe*, tout comme les contrats de type *coût du service*, sont utilisés par exemple dans l'industrie du transport urbain en France.

Les mécanismes incitatifs

Les contrats incitatifs sont les schémas intermédiaires entre le principe du *coût du service* et celui du *prix-fixe*. Ces contrats définissent un cadre où l'autorité et l'opérateur partagent les risques liés aux coûts selon un cadre bien défini.

Un mécanisme incitatif a pour support un contrat linéaire qui définit la part des coûts d'exploitation supportée par l'opérateur. On considère ici un cadre réglementaire qui suppose que les recettes commerciales sont versées à l'autorité et que les coûts d'exploitation sont remboursés ex-post. La rémunération perçue par l'exploitant est alors caractérisée par un transfert net composé d'une partie fixe et d'une partie variable qui ampute la base de

remboursement forfaitaire qui est fixée au départ. L'entreprise supporte une fraction $b \in [0,1]$ de ses coûts d'exploitation telle que son transfert serait

$$t = a - b(C - C_0),$$

où C_0 désigne le montant anticipé des coûts d'exploitation, C est le coût observé ex post, a est la partie fixe du transfert et b caractérise la "puissance" du schéma incitatif. Les contrats linéaires dits incitatifs avec $b \in]0,1[$ sont encadrés par deux cas extrêmes. D'une part, les contrats de type *coût du service* ont une pente $b = 0$. D'autre part, les contrats linéaires de pente $b = 1$ caractérisent les schémas de type *prix-fixe*.

La définition et la mise en place des contrats incitatifs optimaux constituent le centre d'intérêt de la nouvelle théorie de la régulation. Un de ses enjeux majeurs est le compromis à réaliser entre une réduction des coûts d'exploitation par une augmentation de l'effort de productivité et une diminution de la rente abandonnée à l'opérateur. Lorsque le financement de l'activité est assuré par une taxation des consommateurs coûteuse car distordant les prix, l'autorité souhaite à la fois proposer à la firme un contrat fortement incitatif en matière de réduction de coût et abandonner une rente minimale. Or, ces deux objectifs sont précisément contradictoires. L'exemple des schémas de type *prix-fixe* et *coût du service* est à ce propos particulièrement illustrateur. En terme d'incitation aux efforts de productivité et dans une situation d'asymétrie d'information, les contrats de type *prix-fixe* sont de bons outils dans la mesure où l'opérateur récupère entièrement les gains dus à ses efforts. Les contrats de type *coût du service* exercent un effet minimum en matière d'incitation à l'effort. Inversement, les contrats de type *prix-fixe* nécessitent l'abandon d'une rente élevée pour les firmes les plus efficaces et présentent des risques pour les moins efficaces tandis que les régimes de type *coût du service* permettent au régulateur de capturer entièrement la rente et garantissent les revenus des opérateurs.

La régulation par le taux de rendement

Dans le cas où l'autorité ne verse pas de transfert, l'opérateur doit couvrir entièrement ses charges par le biais de la vente du produit ou du service aux consommateurs. La forme de régulation par les prix la plus ancienne est le principe du *taux de rendement*. La règle fondamentale dans ce cas est que le prix moyen fixé par l'autorité que l'opérateur est autorisé à appliquer permet aux recettes commerciales d'égaliser en théorie les coûts totaux.

La détermination du *taux de rendement* raisonnable fait l'objet de deux étapes successives. Au cours d'une première étape, le régulateur détermine le stock de capital exact de l'opérateur à partir de l'observation des investissements effectués au cours des périodes précédentes. Ce calcul tient naturellement compte de la dépréciation du capital à chaque période. Le coût futur lié à la quantité de capital et de facteurs de production consommés par le processus de production est alors estimé en tenant compte de l'inflation et d'éventuels chocs cycliques dans le futur. Le *taux de rendement* fixé correspond alors au coût du capital investi et le niveau de revenu requis par l'opérateur est la somme des coûts d'exploitation et du *taux de rendement* du capital. La seconde étape consiste simplement à établir le prix moyen associé qui égalise le revenu effectif et le revenu requis. Naturellement, le prix moyen déterminé n'est que le reflet d'une tarification complexe qui varie avec le type du produit ou du consommateur considéré. Cette dernière étape peut s'avérer particulièrement délicate dans la mesure où il est difficile de dissocier les coûts individuels liés à la production de biens différents.

La régulation par le *taux de rendement* permet un contrôle significatif de la tarification monopolistique de l'entreprise grâce à une surveillance soutenue de ses profits. En revanche, le principe présente deux inconvénients majeurs. D'une part, la détermination d'un *taux de rendement* du capital équitable n'est pas aisée. D'autre part, le pouvoir incitatif d'un tel mécanisme est forcément limité. Certains auteurs, comme Joskow et Schmalensee (1986), considèrent cependant que la distinction entre les régimes de type *prix-plafond*, définis ci-dessous, et les régimes de type *taux de rendement* ne peut en aucun cas se résumer à leurs puissances incitatives respectives, les délais de révisions dans le cas du *taux de rendement* ne permettant pas une continuité dans la garantie de couverture des dépenses engagées.

Un effet indésirable lié à la régulation par le *taux de rendement* est l'incitation au surinvestissement dont profite l'entreprise. Le taux de rémunération du capital étant supérieur à celui du marché, l'entreprise substitue de plus en plus du capital au travail dans son processus de production. Cet effet est connu sous le nom d'effet d'Averch et Johnson (voir Averch et Johnson (1962)). Le choix d'un taux de rendement supérieur à celui du marché dans ce cas peut être perçu comme une rente abandonnée par le régulateur. Cette rente ne correspond cependant à aucun arbitrage de la part du régulateur à qui Averch et Johnson ne prêtent aucun objectif implicite.

Les exemples de régulation par le principe du *taux de rendement* sont nombreux : on peut citer celui de la régulation d'*AT&T* aux Etats-Unis (en 1980, ce système fut partiellement remplacé par un *prix-plafond*).

La régulation par le prix-plafond

La régulation par *prix-plafond* est apparue au cours des années 80 lors des opérations de privatisation effectuées principalement dans les industries britanniques du gaz, de l'électricité et des télécommunications et les industries américaines de télécommunication.³ La régulation par le principe du *prix-plafond* consiste à fixer un prix de vente maximal pour chaque bien ou un sous-ensemble pondéré de biens. L'entreprise est alors libre de fixer son prix à la limite autorisée ou au-dessous de cette limite. L'avantage majeur de ce type de régulation est qu'il possède les mêmes propriétés incitatives que les schémas de type *prix-fixe* dans la mesure où le *prix-plafond* fixé ne garantit pas des recettes commerciales suffisantes pour couvrir entièrement les coûts si ceux-ci sont exagérément importants. De plus, ce principe permet à l'opérateur d'utiliser une tarification de type Ramsey plus consistante avec l'idée d'efficacité économique. Le risque est évidemment d'abandonner des rentes importantes aux entreprises. De plus, il a été souvent reproché à ce type de mécanisme d'affecter la qualité du service ou du bien offert. Par exemple, les services de *British Telecom* se sont considérablement dégradés après l'introduction d'un *prix-plafond*.

Bien qu'il ne soit pas fixé à partir d'un examen précis de sa comptabilité, le prix plafond garantit au mieux l'équilibre du budget de l'entreprise. Un plafond trop élevé confère à l'entreprise un pouvoir de monopole tandis qu'un prix trop bas l'empêche d'équilibrer son budget, quel que soit le niveau d'effort fourni. Néanmoins, l'opérateur est généralement autorisé à augmenter régulièrement le plafond en fonction de l'évolution des conditions productives. Le taux de croissance autorisé du plafond est fixé par la formule *RPI-X* qui stipule que le taux ne doit pas dépasser le taux de croissance de l'indice du prix de détail moins un taux de croissance anticipé de progrès technique.

Parmi les exemples célèbres de régulation par le principe du *prix-plafond* et de la formule *RPI-X*, on peut citer celui des compagnies régionales distributrices d'électricité au Royaume-Uni. Chaque compagnie a, sur le territoire qu'elle couvre, un pouvoir de monopole. Le régulateur doit alors fixer des plafonds que les prix pratiqués par les compagnies ne doivent

pas dépasser. Dans cet exemple particulier, les plafonds sont fixés pour une durée de 5 ans. L'entreprise assurant le transport est toujours en position de monopole naturel et est également régulée par un *prix-plafond*. Un nouveau calcul ($RPI - 3\%$) fut réalisé en 1994 afin de capturer les bénéfices substantiels dégagés depuis 1990.⁴ Dans le cadre des télécommunications britannique, le tableau ci-dessous donne un exemple d'une révision régulière du plafond.

Tableau 1 : Encadrement tarifaire de *British Telecom*

Date	Prix-plafond
1984-1989	$RPI-3,00\%$
1989-1991	$RPI-4,50\%$
1991-1993	$RPI-6,25\%$
1993-1997	$RPI-7,50\%$

Source : Thonier (1995)

Partage des coûts ou des profits

Le principe du *partage des coûts et des profits* est un cas intermédiaire des régulations par *taux de rendements* ou *prix-plafond*. Lorsque le taux de rendement réel du capital excède celui anticipé, l'autorité peut diminuer légèrement le prix moyen de façon à allouer à l'opérateur une partie des profits au lieu de leur totalité. Le partage des coûts est effectué en révisant ponctuellement le prix moyen à partir de l'examen de l'évolution de certains facteurs de production seulement. Ceci a pour effet de responsabiliser en partie l'opérateur puisque le prix appliqué ne reflète pas totalement les coûts d'exploitation et le taux de rendement requis. La régulation par le partage des coûts s'apparente donc aux mécanismes incitatifs en terme d'incitation à la productivité.

3. Le modèle théorique

Avant le début des années 80, les autorités centrales ou locales qui ont utilisé les schémas réglementaires présentés dans la section précédente n'ont pas intégré les problèmes informationnels identifiés par la nouvelle théorie de la régulation. Les régulations par *prix-plafond* ou *coût du service* ont été couramment employées. Elles ont été cependant proposées de façon aveugle dans la mesure où elles étaient offertes simultanément à plusieurs opérateurs d'une même industrie sans distinction de leur niveau d'efficacité. La nouvelle théorie de la régulation propose une régulation théorique optimale car chaque schéma réglementaire tient compte des caractéristiques réelles du monopole régulé.

3.1. Le cadre général

On considère une industrie particulière où un ensemble de services ou de biens est proposé et fourni par une seule firme. Des économies d'échelle significatives assurent à cette firme la qualité de monopole naturel et justifient ainsi sa position d'offreur unique sur le marché. Les coûts de production induits par l'activité sont décrits par

$$C(\tilde{Y}, e, \theta), \quad (1)$$

où $\tilde{Y} = (Y^1, \dots, Y^n)$ est un vecteur de biens de production ou de services proposés par l'entreprise, θ un paramètre d'anti-sélection qui caractérise son inefficacité technique uniquement et e est un paramètre d'aléa moral qui indique le niveau d'effort que celle-ci exerce en réponse à son inefficacité productive. Ainsi, à niveau de production donné, une firme plus inefficace fait face

à des coûts d'exploitation plus élevés tandis qu'un niveau d'effort plus important permet une réduction des coûts. Bien évidemment, à niveaux d'inefficacité et d'effort constants, une augmentation des quantités produites conduit à des coûts plus élevés.

Produire un effort est coûteux. En effet, les activités de recherche et de développement, de formation de la main d'œuvre, de sélection de facteurs de production de meilleure qualité conduisent à des charges supplémentaires. L'effort est donc la source d'une "désutilité" notée $\psi(e)$, fonction croissante et convexe.

L'entreprise connaît ses capacités productives et par là son inefficacité et elle choisit le niveau d'effort de productivité qu'elle est disposée à fournir.

La vente du panier de biens produits ou de services offerts génère une recette commerciale $R(\tilde{Y})$ qui, associée à un transfert t_0 versé par l'autorité, assure un revenu global $t_0 + R(\tilde{Y})$ à l'entreprise. Comme le montant du transfert est fixé par le principal, on adopte, sans perte de généralité, la convention comptable qui stipule que l'autorité reçoit les recettes commerciales et paie les coûts d'exploitation. La rémunération de l'entreprise est alors garantie par le transfert net t où

$$t = t_0 + R(\tilde{Y}) - C(\tilde{Y}, e, \theta), \quad (2)$$

et son utilité est

$$U = t - \psi(e). \quad (3)$$

On suppose ici que la firme est neutre face au risque. La prise en compte de l'aversion au risque ne change pas fondamentalement la méthodologie. Si l'observation des coûts est sans aléa, le mécanisme incitatif proposé n'est pas affecté. En revanche, dans le cas d'une incertitude liée à l'observation des coûts, la puissance du mécanisme incitatif est affaiblie, la part des coûts supportée par l'entreprise étant alors plus faible (voir Laffont et Tirole, 1986).

L'entreprise considérée est régulée par une autorité supérieure qui ne bénéficie pas d'information sur les niveaux d'effort et d'inefficacité de l'entreprise. L'autorité dispose simplement de croyances sur la distribution de θ sur l'intervalle $[\underline{\theta}, \bar{\theta}]$. La distribution admet une fonction de répartition $F(\theta)$ et une densité $f(\theta)$ positive et continue sur l'intervalle. L'autorité observe les coûts d'exploitation agrégés. Ainsi, elle n'est pas en mesure de distinguer les différentes charges liées aux différentes activités de production. En contrepartie, l'autorité est supposée avoir une connaissance acceptable des coûts internes liés à l'effort de productivité. Cette dernière hypothèse est fortement discutable mais peut être abandonnée lorsque des formes fonctionnelles adéquates sont utilisées pour décrire la technologie.

La consommation de l'ensemble des services ou des biens procure un surplus social brut $V(\tilde{Y})$ dont la définition dépend du type de bien ou de service considéré. Le surplus social inclut un paramètre représentatif de la distorsion issue de l'activité de taxation des consommateurs. En principe, la problématique présentée ici considère une activité monopolistique déficitaire où les recettes commerciales ne suffisent pas à financer l'ensemble des opérations. Le financement de l'activité est assuré par des fonds publics prélevés au moyen d'un système de taxation des consommateurs. Toute taxation est coûteuse car sa réalisation entraîne des charges et des distorsions dans l'allocation des ressources. Le coût des fonds publics ainsi défini par λ

implique que chaque euro prélevé pour financer une activité coûte en fait à la société $1 + \lambda$ euros. L'existence du coût des fonds publics est primordiale dans l'analyse de la régulation des opérateurs dans un contexte d'asymétrie d'information puisqu'elle donne un poids considérable au coût généré par l'activité de production et supporté par la société. Elle nécessite que l'abandon de rente à l'entreprise régulée soit minimal.⁵ Loeb et Magat (1979) suggèrent que lorsque le coût des fonds publics est nul, la présence d'asymétrie d'information n'est pas dommageable pour le bien-être de la société dans le sens où l'autorité peut atteindre le bien-être de premier rang généralement accessible en présence d'information parfaite.

A partir de la définition des différents paramètres et agrégats économiques intervenant dans la problématique, nous définissons le bien-être social par la somme

$$W(\tilde{Y}) = V(\tilde{Y}) - (1 + \lambda)[t + C(\tilde{Y}, e, \theta)] + U, \quad (4)$$

soit, en remplaçant U par son expression dans (3),

$$W(\tilde{Y}) = V(\tilde{Y}) - Z(\tilde{Y}), \quad (5)$$

où

$$Z(\tilde{Y}) = (1 + \lambda)[C(\tilde{Y}, e, \theta) + \psi(e)] + \lambda U \quad (6)$$

représente le coût social engendré par la consommation des services et des biens. On suppose alors que l'autorité a pour objectif la maximisation de l'espérance du bien-être social. Les outils dont dispose l'autorité pour la maximisation du bien-être social sont définis par l'ensemble $\{\tilde{Y}, e, U\}$. Les situations d'asymétrie d'information et d'information parfaite diffèrent par la possibilité qu'a l'autorité d'observer l'inefficacité θ de l'opérateur. Dans les deux cas, la méthodologie considère en fait un problème d'anti-sélection pur dans la mesure où les variables de contrôle définies dépendent uniquement de l'inefficacité, observable ou non, de l'opérateur. A ce propos, on définit par

$$E(\tilde{Y}, \theta, C), \quad (7)$$

le niveau d'effort requis pour un opérateur d'inefficacité θ pour produire la quantité \tilde{Y} à un coût C . Ainsi, les coûts d'exploitation et le coût de l'effort ne contiennent plus qu'une seule inconnue, l'inefficacité θ . La présence d'asymétrie d'information empêche l'autorité de mettre en place la politique de premier rang qui garantit à la société un bien-être maximal. On qualifiera le bien-être atteint de bien-être de second rang. Les deux sous-sections suivantes exposent les régulations optimales de premier et de second rang.

3.2. Régulation optimale dans le cadre d'information parfaite

Dans le cas où l'autorité a connaissance des capacités productives de l'exploitant, elle lui impose des niveaux de production, d'effort et de rente de façon à maximiser le bien-être social sous la contrainte $U \geq 0$. Les conditions de premier ordre issues de ce programme impliquent tout d'abord que le niveau de production requis est tel que la valeur sociale marginale de la production égale le coût marginal supporté par les consommateurs soumis à la taxation. Ensuite, il apparaît que le niveau d'effort qui assure un coût d'exploitation et un coût interne minimaux

égalise le coût marginal de l'effort et la désutilité marginale de l'effort. Enfin, les conditions de premier ordre impliquent que le transfert versé à l'exploitant assure un simple remboursement du coût interne de l'effort. Ainsi, aucune rente informationnelle coûteuse pour la société n'est versée à l'opérateur puisque la situation idéale décrite ici la rend caduque.

Le régulateur dispose alors de plusieurs types de schémas pour mettre en place le mécanisme optimal. Très simplement on peut imaginer qu'il verse à l'opérateur un transfert $t^* = \psi(e^*)$ et qu'il fixe le montant des coûts d'exploitation $C(\tilde{Y}^*, \theta, e^*)$ à atteindre par l'exploitant. Tout écart par rapport aux objectifs fixés fait alors l'objet de sanctions. Plus rigoureusement, l'autorité peut imposer un contrat de type prix-fixe

$$T = \psi(e^*) + [C(\tilde{Y}^*, \theta, e^*) - C(\tilde{Y}^*, \theta, e)], \quad (8)$$

où $C(\tilde{Y}^*, \theta, e)$ est le montant des coûts effectivement réalisés à partir du niveau de production requis par l'autorité. Le transfert perçu par l'opérateur se décompose alors en une partie fixe et une partie variable qui dépend de l'effort réalisé. Ce type de contrat est parfaitement incitatif dans la mesure où tout dépassement des coûts fixés au départ entraîne une sanction financière de l'opérateur qui a intérêt à fournir le niveau d'effort requis.

3.3. Régulation optimale en présence d'information asymétrique

On suppose maintenant que l'autorité ne connaît pas parfaitement le degré d'efficacité de l'opérateur. La maximisation de l'espérance du bien-être social relève alors des techniques de la conception de mécanismes d'abord développées par Mirrlees (1971) et notamment mises en pratique par Baron et Myerson (1982) dans les modèles de régulation des monopoles.

Le principe de la conception de mécanismes s'applique aux problèmes de discrimination monopolistique par les prix, aux systèmes d'enchères et de taxation, de fourniture de biens publics et à la régulation des monopoles dans des contextes d'information asymétrique. Au départ, toutes ces applications opposent un principal et un ou plusieurs agents. Le principal souhaite déterminer des actions qui dépendent de certaines caractéristiques de l'agent seulement connues de ce dernier. Dans le cas de la discrimination de second degré par les prix, par exemple, le monopole en place ne dispose pas au départ de l'information nécessaire sur la disponibilité à payer des consommateurs. Un autre exemple est celui de la taxation optimale dans lequel le gouvernement ne dispose pas de l'information nécessaire sur l'habilité des consommateurs à percevoir un salaire élevé.

Le principal ne peut directement demander aux agents de révéler leur information car ceux-ci peuvent mentir si leur intérêt en dépend. Il doit alors élaborer un mécanisme qui incite les agents à révéler leurs caractéristiques authentiques au moyen de transferts monétaires ou d'autres instruments. Le mécanisme en question étant coûteux, le principal fait face à un compromis entre efficacité et rente et le schéma optimal propose alors le compromis idéal. Le lecteur peut se reporter à Fudenberg et Tirole (1991, chapitre 7) ou à Salanié (1994), entre autres, pour un descriptif complet de la méthodologie.

Plus particulièrement, le mécanisme incitatif optimal dans notre problème est obtenu en faisant appel au principe de révélation.⁶ Le principe de révélation établit que s'il existe une stratégie optimale pour une firme à qui on propose un contrat particulier, alors ce contrat peut être répliqué par un mécanisme de révélation direct.

Définissons le jeu plus en détail : il met en présence un principal et un agent caractérisé par son inefficacité, inconnue du principal. De manière générale, le principal souhaite mettre en

œuvre une allocation de ressource $w(\cdot)$ aussi bonne que possible qui dépend bien entendu de l'inefficacité. Pour parvenir à ses fins, le principal élabore et impose à l'agent un mécanisme $[w(\cdot), M]$ qui se compose d'un espace de messages M pour l'agent et d'une fonction $w(\cdot)$ de M dans l'espace des allocations. Les agents qui connaissent la fonction $w(\cdot)$ jouent un jeu d'annonces où l'espace M constitue leur ensemble de stratégies et la fonction $w(\cdot)$ détermine leur allocation et donc leur utilité. Compte tenu du mécanisme $[w(\cdot), M]$, l'agent choisit l'annonce m^* qui maximise son utilité $U(w, \theta)$, soit

$$m^* \in \arg \max_{m \in M} U(w(m), \theta), \quad (9)$$

et il obtient l'allocation

$$w^*(\theta) = w(m^*(\theta)). \quad (10)$$

Le principe de révélation stipule alors que le principal peut se limiter à des mécanismes directs et révélateurs, c'est à dire des mécanismes où l'agent annonce son information et a intérêt à dire la vérité. Si l'allocation $w^*(\theta)$ peut être mise en œuvre par un mécanisme quelconque, le principal peut la mettre en œuvre par un mécanisme direct révélateur où l'annonce de l'agent $\hat{\theta}$ coïncide avec son vrai type θ .

Dans le modèle de régulation exposé dans cette section, le mécanisme de révélation est décrit par l'ensemble $\{\tilde{Y}(\hat{\theta}), C(\hat{\theta}), t(\hat{\theta})\}$ qui dépend de l'annonce de l'agent $\hat{\theta}$ et qui est tel que

$$U(\theta, \theta) \geq U(\hat{\theta}, \theta), \quad \forall (\hat{\theta}, \theta) \in [\underline{\theta}, \bar{\theta}], \quad (11)$$

$$U(\theta, \theta) \geq 0, \quad \forall \theta \in [\underline{\theta}, \bar{\theta}], \quad (12)$$

où $U(\hat{\theta}, \theta)$ est l'utilité perçue par l'entreprise de type θ lorsqu'elle annonce $\hat{\theta}$. La contrainte incitative (11) stipule que l'agent ne doit pas être incité à sousestimer voire à surestimer son efficacité au moment de l'annonce pour que le mécanisme révèle sa vraie inefficacité. Ainsi, l'utilité perçue par l'agent lorsqu'il annonce la vérité doit être supérieure à l'utilité perçue lorsqu'il ment sur son type. En outre, la contrainte de rationalité individuelle (12) impose une utilité non-négative pour l'opérateur afin que ce dernier ne refuse pas de participer au mécanisme régulateur. Ainsi, lorsque l'entreprise annonce un type, le principal lui impose un niveau de production $\tilde{Y}(\hat{\theta})$ au coût $C(\hat{\theta})$ et lui verse un transfert $t(\hat{\theta})$.

A partir de l'expression de l'effort définie dans (7) et de l'expression préalable de U dans (3), l'utilité de la firme est en définitive

$$U(\hat{\theta}, \theta) = t(\hat{\theta}) - \psi \left[E(\tilde{Y}(\hat{\theta}), C(\hat{\theta}), \theta) \right]. \quad (13)$$

La contrainte incitative s'obtient alors à l'issue de deux étapes successives. Dans un premier temps, pour que le mécanisme proposé incite l'opérateur à annoncer sa vraie inefficacité, il est

nécessaire que son annonce maximise son utilité lorsque celui-ci dit la vérité. En notant $E(.) = E(\tilde{Y}(\hat{\theta}), C(\hat{\theta}), \theta)$, la contrainte de premier ordre correspondante est

$$\left. \frac{\partial U(\hat{\theta}, \theta)}{\partial \hat{\theta}} \right|_{\hat{\theta}=\theta} = \left. \frac{\partial t(\hat{\theta})}{\partial \hat{\theta}} \right|_{\hat{\theta}=\theta} - \psi'(E(.)) \left[\sum_{i=1}^n \frac{\partial E(.)}{\partial Y_i(\hat{\theta})} \frac{\partial Y_i(\hat{\theta})}{\partial \hat{\theta}} + \frac{\partial E(.)}{\partial C(.)} \frac{\partial C(.)}{\partial \hat{\theta}} \right] \Bigg|_{\hat{\theta}=\theta} = 0 \quad (14)$$

L'autorité définit dans un second temps la règle d'abandon de la rente $\varphi(\theta) = U(\theta, \theta)$ aux firmes en fonction du type de l'agent à qui elle propose le contrat. Ainsi, d'après (14) ou à partir du théorème de l'enveloppe directement appliqué à l'expression (13), on obtient

$$\varphi'(\theta) = -\psi' \left[E(\tilde{Y}(\theta), C(\theta), \theta) \right] \frac{\partial E(.)}{\partial \theta}. \quad (15)$$

La rente ainsi définie est le prix que le principal doit payer pour obtenir que les types efficaces révèlent leur information. Pour que le mécanisme incite les agents à révéler leur vrai type, il doit abandonner les rentes les plus élevées aux opérateurs les plus efficaces et donc les rentes les plus faibles aux opérateurs les moins efficaces. Une firme de type θ peut, par intérêt, se déclarer de type $\theta' > \theta$, et ainsi produire un effort moindre pour atteindre l'objectif $\{C(\theta'), \tilde{Y}(\theta')\}$ fixé par l'autorité. Consécutivement à une diminution de l'effort $\partial E(.)/\partial \theta$, le gain de l'opérateur en terme d'utilité est alors égal à $\psi'[E(.)] \partial E(.)/\partial \theta$. On comprend bien que l'autorité doit alors garantir à l'agent le même gain en lui versant une rente afin que celui-ci annonce son vrai type.

A partir de la règle énoncée par la contrainte incitative, la contrainte de rationalité individuelle devient

$$\varphi(\bar{\theta}) = 0. \quad (16)$$

En effet, il apparaît que la contrainte initiale est saturée lorsque $\theta = \bar{\theta}$. Sachant que l'objectif de l'autorité est d'abandonner une rente minimale dans tous les cas de figure, la rente versée à l'entreprise la plus inefficace est nulle.

Pour déterminer l'allocation du contrat incitatif optimal, l'autorité choisit indifféremment le mécanisme direct et révélateur $\{\tilde{Y}^*(\theta), C^*(\theta), t^*(\theta)\}$ ou $\{\tilde{Y}^*(\theta), e^*(\theta), \varphi^*(\theta)\}$ qui maximise le bien-être social. Le programme à résoudre par l'autorité consiste donc à déterminer les niveaux de production, d'effort et de rente optimaux de façon à maximiser l'espérance du bien-être social sous les contraintes (15) et (16).

Les variables $e(\theta)$ et $\tilde{Y}(\theta)$ sont les variables de contrôle du problème et $\varphi(\theta)$ est considéré comme la variable d'état. Les conditions de premier ordre sont suffisantes si en plus, $\partial C/\partial \theta \geq 0$ et $\partial Y_i/\partial \theta \leq 0$, ce qui implique que le mécanisme autorise les firmes les plus inefficaces à avoir les coûts les plus élevés et à produire moins. Le processus de résolution habituel ignore ces contraintes dans un premier temps mais les confronte ensuite aux solutions des conditions de premier ordre. Les conditions de premier ordre sont

$$\frac{\partial V(\tilde{Y})}{\partial Y_i} = (1 + \lambda) \frac{\partial C(\tilde{Y}, e, \theta)}{\partial \tilde{Y}_i} + \lambda \frac{F(\theta)}{f(\theta)} \psi'(e) \frac{d(\partial E(\cdot)/\partial \theta)}{dY_i}, \quad (17)$$

et

$$-\frac{\partial C(\tilde{Y}, e, \theta)}{\partial e} - \frac{\lambda}{1 + \lambda} \frac{F(\theta)}{f(\theta)} \left[\psi''(e) \frac{\partial E(\cdot)}{\partial \theta} + \psi'(e) \frac{\partial^2 E(\cdot)}{\partial \theta \partial C} \frac{\partial C(\tilde{Y}, e, \theta)}{\partial e} \right] = \psi'(e). \quad (18)$$

Par rapport à la politique mise en place par l'autorité dans un contexte d'information parfaite, on remarque l'apparition de deux distorsions dans (17) et (18). Ces distorsions résultent de la volonté qu'a l'autorité de réduire au minimum la rente abandonnée à l'entreprise. Par exemple, le niveau d'effort requis par (18) est inférieur à l'effort exigé par une autorité qui dispose d'une information parfaite sur l'efficacité de l'entreprise.

Pour comprendre les implications de la condition (18), posons

$$I = \psi''(e) \frac{\partial E(\cdot)}{\partial \theta} + \psi'(e) \frac{\partial^2 E(\cdot)}{\partial \theta \partial C} \frac{\partial C(\tilde{Y}, e, \theta)}{\partial e}, \quad (19)$$

soit

$$\left[-\psi'(e) - \frac{\partial C(\tilde{Y}, e, \theta)}{\partial e} \right] (1 + \lambda) f(\theta) = \lambda F(\theta) I, \quad (20)$$

d'après (18) et (19). Notons que $I = d|\partial \varphi(\theta)/\partial \theta|/de$. Ainsi, I mesure la hausse de la rente perçue par les individus $[\underline{\theta}, \theta]$ en quantité $F(\theta)$ lorsque le niveau d'effort exercé par l'entreprise de type θ augmente d'une unité. Le coût correspondant pour la société est représenté par le côté droit de (20). La réduction de l'effort requise diminue la rente mais s'avère coûteuse parce qu'elle se répercute sur les coûts d'exploitation. Ainsi, en contrepartie, le coût social consécutif à une distorsion de l'effort vers le bas est donné par le côté gauche de (20). Ce coût est amplifié par la distorsion issue de la gestion des fonds publics $1 + \lambda$ et survient avec une probabilité $f(\theta)$. L'équation (18) montre clairement le compromis entre abandon de rente et efficacité. On notera que $F(\underline{\theta}) = 0$, ainsi, d'après (18), l'effort requis lorsque l'entreprise est la plus efficace possible est équivalent au niveau de premier ordre en cas d'information parfaite. Ainsi, il n'y a pas de distorsion pour les firmes les plus efficaces.

L'équation (17) inclut une correction incitative au principe similaire qui suggère que l'autorité réalise un compromis entre le surplus créé par la production et la rente associée. Cette correction n'est cependant pas nécessaire dans tous les cas de figure puisque son existence dépend de la forme fonctionnelle décrivant la technologie, comme le suggère la règle qui suit.

Une règle fondamentale est la dichotomie entre incitation à la productivité et tarification. Une telle dichotomie existe lorsque l'incitation à la productivité est entièrement fixée par l'aspect "offre" de la régulation, c'est à dire par la condition (18). Dans ce cas, la correction incitative dans la partie "demande" de la régulation est nulle et la tarification pratiquée est complètement indépendante des problèmes informationnels qui peuvent survenir. D'après (17), la dichotomie s'applique si (Voir Laffont et Tirole, 1993, page 179)

$$\frac{d(\partial E(\cdot)/\partial \theta)}{dY_i} = 0. \quad (21)$$

Cette condition technique est vérifiée pour une certaine classe de fonctions de coût.

3.4. Mise en œuvre des mécanismes optimaux par des contrats linéaires

La section 2 a évoqué la mise en place de mécanismes incitatifs par l'intermédiaire de contrats linéaires. L'ensemble des contrats linéaires constitue un menu proposé à l'opérateur par le principal. Ce menu est optimal si les paramètres du contrat sont déterminés par les conditions de premier ordre (17) et (18). Ainsi, l'opérateur qui fait face au menu choisit forcément le contrat qui correspond à son type. L'entreprise la plus efficace choisit ainsi le contrat où $b = 1$, c'est à dire celui qui requiert un niveau d'effort de premier ordre. Inversement, la firme la moins efficace choisit le contrat qui exige le niveau d'effort le plus faible, c'est à dire, celui pour lequel b est minimal. Les opérateurs dont l'inefficacité est définie dans l'intervalle $]\underline{\theta}, \bar{\theta}[$ choisissent un contrat à la puissance incitative intermédiaire. Une condition technique est requise pour la validité de cette proposition. Il est nécessaire que le transfert $t(C(\tilde{Y}^*, e^*, \theta))$ issu de l'allocation optimale soit une fonction convexe des coûts effectivement réalisés. Le menu des contrats est alors défini comme la famille des tangentes au transfert.

Les mécanismes régulateurs évoqués dans la section 2 sont couramment utilisés dans la réalité. Les contrats de type *prix-fixe* ou *coût du service*, par exemple, lorsqu'ils sont issus d'un menu de contrat, sont optimaux. En revanche, ces mêmes contrats, ainsi que les schémas de type *prix-plafond* ou *taux de rendement*, ne sont pas optimaux lorsqu'ils sont proposés simultanément à plusieurs entreprises opérant à l'intérieur d'une même industrie, sans distinguer l'efficacité individuelle des agents.

4. Les systèmes d'enchères

Les schémas réglementaires optimaux présentés dans la section précédente considèrent des relations entre une autorité et un opérateur unique. En présence d'information asymétrique, l'opérateur est soumis au principe de révélation, ce qui permet aux autorités d'éviter l'abandon de rentes importantes. Cependant, même si l'opérateur révèle sa véritable inefficacité, le bien-être social de second rang associé à ces schémas reste dépendant de l'efficacité de l'opérateur. En effet, un opérateur inefficace, même s'il révèle ses véritables capacités, reste un opérateur inefficace. En revanche, si l'autorité fait face à plusieurs opérateurs, elle a la possibilité de choisir l'opérateur le plus efficace : il s'agit du principe de l'enchère ou de l'appel d'offre.

Pendant plusieurs décennies, les services qui bénéficiaient du statut de monopole naturel ont été assurés par des firmes appartenant aux autorités publiques. L'avènement des procédures de privatisation et l'apparition de nouveaux candidats pour l'exploitation des services ont incité les gouvernements à mettre en place des systèmes d'enchères pour l'attribution des licences de monopole. C'est par exemple le cas actuellement en France pour le transport urbain et plus particulièrement la construction de l'infrastructure lourde de type métro ou tramway. Le principe de l'enchère ou de l'appel d'offre permet aux régulateurs de tester les coûts de production des entreprises et de réduire dans certains cas l'abandon de rentes. Lorsqu'un régulateur fait face à plusieurs candidats, le système d'enchère permet de sélectionner l'opérateur le plus efficace.

Considérons l'exemple suivant : une autorité publique organise une enchère afin de délivrer une licence de monopole pour l'exploitation d'un projet \tilde{Y} . N opérateurs participent à l'enchère. La technologie de chaque opérateur i est décrite par la fonction de coût

$$C_i(\tilde{Y}_i, e_i, \theta_i), \quad i = 1, \dots, N. \quad (22)$$

L'organisateur de l'enchère ne connaît ni θ_i ni e_i . Il dispose simplement de croyances quant à la distribution des θ_i . Ces croyances sont définies par l'intervalle $[\underline{\theta}, \bar{\theta}]$, la fonction de densité $f(\theta)$ et la fonction de distribution $F(\theta)$. Les candidats participant à l'enchère annoncent simultanément leur efficacité. L'organisateur maximise le bien être espéré de la société et utilise un mécanisme révélateur. La contrainte incitative est obtenue sous l'hypothèse de comportement de type Nash-Bayésien. L'entreprise qui emporte l'enchère est celle qui annonce la meilleure efficacité. Le contrat optimal impose des objectifs $\{\tilde{Y}^*(\theta), e^*(\theta), \varphi^*(\theta)\}$ identiques à ceux définis dans la section (3.3.) Ainsi, le problème est identique à celui où l'autorité fait face à un opérateur unique et offre un menu de contrats. L'intérêt de l'enchère est bien entendu de sélectionner la firme la plus efficace et de payer des rentes plus faibles.

Bien que l'utilisation des mécanismes d'enchère améliorent a priori l'efficacité du processus de production, son utilisation par les autorités publiques n'est pas systématique. La raison est double. Tout d'abord, l'organisation d'une enchère est un processus lent et extrêmement coûteux pour la société. Il convient donc de comparer les coûts liés à son organisation aux gains nés d'une meilleure productivité. Ensuite, une enchère doit être répétée afin de pouvoir toujours sélectionner le meilleur candidat (l'opérateur le plus efficace aujourd'hui n'est pas forcément l'opérateur le plus efficace demain.) Si les opérateurs savent que leur pouvoir de monopole est à durée limitée, ils ont tendance à sous-investir car ils savent qu'ils ne récupéreront pas le rendement du capital investi (Williamson, 1976).

5. La dynamique de la régulation

Le modèle présenté dans la section précédente (3.3) est un modèle théorique statique. Il tente de représenter des situations où le contrat proposé par l'autorité organisatrice à l'opérateur est un contrat de court terme qui porte sur une période unique.⁷ Dans la pratique cependant, l'exploitation d'un service est souvent une activité de long terme qui porte sur plusieurs périodes. On observe donc une succession de contrats et le cadre étudié est dynamique.

Lorsque le problème de la régulation d'un monopole est projeté dans un cadre dynamique, le régulateur est confronté à un problème majeur appelé « ratchet effect » (« effet de cliquet ».) Le « ratchet effect » est la résultante de deux effets : d'une part, les relations entre le régulateur et l'opérateur sont caractérisées par des asymétries d'information ; d'autre part, le cadre d'analyse est dynamique et toute information révélée à la période t par l'opérateur peut être utilisée par le régulateur à la période suivante. Intuitivement, si un opérateur révèle en t qu'il est très efficace, l'autorité considère que l'agent est capable de produire à très faibles coûts. Ainsi, l'autorité fixera en $t+1$ des objectifs élevés à l'opérateur, ce qui annule toute possibilité de rentes pour ce dernier. Si on se réfère à la théorie présentée dans la section (3.3), ceci revient à proposer un contrat de second rand avec mécanisme révélateur à l'opérateur en t puis un contrat de type *prix-fixe* avec information parfaite. Malheureusement, l'opérateur, qui est parfaitement conscient du risque de capture par l'autorité, sous-estime volontairement son efficacité productive en t afin de bénéficier de rentes au cours des périodes suivantes.

Afin de déterminer quels sont les modes de régulation optimaux dans un cadre dynamique, il convient de distinguer deux cas : dans un premier cas, le régulateur s'engage à ne pas utiliser l'information dont il dispose afin de capturer la rente de l'opérateur. Dans le second cas, le régulateur ne s'engage pas.

Lorsque le régulateur s'engage à ne pas utiliser en seconde période l'information acquise en première période, le contrat optimal de long terme est simplement la répétition des contrats statiques de second rang présentés dans la section (3.3). A chaque période, le régulateur soumet l'opérateur au principe du mécanisme révélateur afin que celui-ci révèle sa véritable efficacité productive en échange d'une rente informationnelle. Il semble cependant que cette situation soit difficilement applicable à la réalité dans la mesure où l'engagement du régulateur est peu crédible.

L'hypothèse selon laquelle le régulateur ne peut pas ou ne veut pas s'engager à utiliser l'information acquise au cours de la période précédente est sans doute la plus crédible. En première période, le régulateur propose le contrat $T_1(C_1)$. Sur la base de l'observation des coûts de première période, le régulateur offre en seconde période le contrat $T_2(C_1, C_2)$. Laffont et Tirole (1988) ont montré que dans ce cas, en raison de la présence du « ratchet effect », il n'existe pas d'équilibre séparateur pour le contrat de première période. Ceci est vrai dans le cas d'un continuum $[\underline{\theta}, \bar{\theta}]$ d'opérateurs. Lorsque le problème est restreint à deux types d'opérateurs ($\underline{\theta}$ et $\bar{\theta}$), il est possible de trouver une solution en incitant l'opérateur le plus inefficace à fournir encore plus d'effort dans le cadre statique.

6. Concurrence

La nouvelle théorie de la régulation s'intéresse aux relations contractuelles entre une autorité publique et un opérateur unique. Le monopole était l'organisation industrielle la plus fréquemment utilisée par les autorités publiques jusqu'au début des années 80. Ceci est de moins en moins vrai aujourd'hui. La raison principale est que le statut de monopole naturel de certaines industries est de plus en plus remis en cause. Cette section propose un aperçu des principes d'ouverture d'un marché à la concurrence. Pour plus de détails, le lecteur peut par exemple se reporter au numéro 119 d'Economie et Prévision (1995), à Armstrong, Cowan et Vickers (1995), Martin (1998), Helm et Jenkinson (1998) et Laffont et Tirole (2000.)

6.1. Ouverture à la concurrence

A la fin des années 70 s'est amorcée aux Etats-Unis avec le secteur du transport aérien une des premières ouvertures à la concurrence. Cette tendance s'est poursuivie aux cours des années 80 et 90. L'ouverture à la concurrence constitue aujourd'hui un enjeu économique majeur et concerne aussi bien le transport aérien que les télécommunications, l'électricité, les réseaux postaux, le transport ferroviaires et le transport urbain (voir Tableau 2 pour un état de la concurrence dans les pays de la Communauté Européenne.)

Tableau 2 : Degré de libéralisation dans l'industrie européenne

	Allemagne	France	Italie	Espagne	Suède	Grande-Bretagne
Transport aérien	Elevé	Moyen	Elevé	Moyen	Elevé	Elevé
Electricité	Moyen	Faible	Faible	Moyen	Elevé	Très élevé
Services postaux	Moyen	Faible	Très faible	Moyen	Très élevé	Moyen
Transport ferroviaire	Faible	Faible	Très faible	Très faible	Elevé	Très élevé
Transport urbain	Très faible	Faible	Faible	Moyen	Elevé	Très élevé
Télécommunications	Elevé	Elevé	Moyen	Moyen	Très élevé	Très élevé

Source : Bergman et alii (1998) et IDEI (1999).

L'introduction de la compétition est née de la volonté des gouvernements de réorganiser les tarifications, d'améliorer la productivité, de réduire l'abandon de rentes et de réduire l'interférence entre le pouvoir politique et la gestion de l'entreprise. Celle-ci réclame un recensement pertinent des propriétés technico-économiques d'une industrie : économie d'échelle et progrès technique sont les deux points principaux à considérer.

En théorie, la concurrence entre les entreprises permet la création de marchés plus « efficaces » dans le sens où ils améliorent le bien-être des sociétés. De manière générale, la concurrence est plus facile à introduire dans les industries où les coûts fixes sont faibles (transport aérien ou routier par exemple.) Si une industrie est un monopole naturel, l'ouverture à la concurrence n'est pas pertinente. Les années récentes ont fait l'objet de nombreuses études tentant de déterminer quelles industries étaient ou non des monopoles naturels.⁸

Parfois, à l'intérieur même d'une industrie, certains éléments du processus de production peuvent être taxés de monopoles naturels tandis que d'autres sont potentiellement concurrentiels. Le cas de l'électricité est particulièrement représentatif : les parties « offre » et « production » sont toutes deux potentiellement concurrentielles tandis que la partie « transmission » par lignes à haute tension et la distribution en basse tension sont des monopoles naturels. D'autres exemples existent : les services longues distances pour les télécommunications aux Etats-Unis sont ouverts à la concurrence, de même que la communication par mobiles aux Etats-Unis et en Europe, la production de gaz, les systèmes de gaz naturel en Grande-Bretagne et les services postaux avec l'apparition des services express.

Il existe deux principaux types de concurrence. Tout d'abord, la concurrence pour l'obtention d'un marché de monopole ; dans ce cas, l'adjudication des droits de franchise est réalisée à l'aide de systèmes d'enchères (voir section 4.) Ensuite, la concurrence à l'intérieur même d'un marché (c'est le cas qui nous intéresse principalement dans cette section) ; l'autorité politique autorise alors plusieurs opérateurs à fournir le même service.

La transition d'un système monopolistique vers un système concurrentiel est complexe. Il existe de multiples contraintes qui retardent sa mise en place. Tout d'abord, des problèmes d'ordres contractuels. Lorsqu'une activité est verticalement intégrée, il existe une multitude de contrats explicites ou implicites à l'intérieur même du secteur. L'instauration de concurrence et l'éclatement de l'activité entraîne une incertitude si ces contrats disparaissent. Le second problème est lié à l'existence du principe de service universel.⁹ La distribution de l'eau, de l'énergie, le transport doivent être accessibles à tous les individus même si ce droit rend l'activité économiquement déficitaire. L'introduction de la concurrence est a priori contraire à l'esprit de service universel, à moins que l'éventuelle réduction des coûts d'exploitation de l'activité soit suffisamment importante.

6.2. Tarification et charge d'accès

Un point clé de l'ouverture d'un marché à la concurrence est la définition d'une « vraie » politique d'interconnectabilité entre les différents services proposés et d'une tarification d'accès au réseau adéquate. Le réseau principal est généralement un monopole naturel et est considéré comme un facteur de production indispensable pour le nouvel entrant.¹⁰ Le service final proposé par un nouvel entrant combine ainsi l'usage de l'infrastructure en place et celui d'autres biens et services situés en aval. Il est logique que les concurrents du secteur aval participent au financement de l'infrastructure amont. L'autorité régulatrice doit alors définir un prix d'entrée qui autorise une allocation des ressources efficace et permette d'éviter les problèmes de contournement (le contournement, ou *by-pass*, consiste à utiliser une technologie de substitution si les charges d'accès sont trop lourdes ; dans le secteurs des

télécommunications par exemple, un exemple classique est celui où l'entrant préfère l'usage de la radio ou du satellite à celui du réseau en place). Cette charge d'accès doit respecter plusieurs exigences. Les coûts de l'infrastructure amont doivent être couverts ; elle peut également permettre à l'opérateur en place de financer les charges liées au principe de service universel ; enfin, cette charge doit inciter l'opérateur en place et l'entrant à investir dans des technologies de pointe et doit fournir de bonnes incitations en matière de réduction des coûts.

Il existe un débat plutôt actif quant à la nature du prix d'entrée. Une première école¹¹ préconise la *Efficient Component Pricing Rule* (ECPR), c'est à dire une charge d'accès optimale incluant le coût incrémental moyen lié à l'accès d'un nouvel opérateur au réseau et le coût d'opportunité subi par l'opérateur en place.¹² Deux critiques principales peuvent être adressées à cette approche : d'une part, elle ne considère pas les coûts fixes à recouvrer par le propriétaire de l'infrastructure, d'autre part, le prix d'accès optimal est totalement déconnecté de la demande. La seconde approche (voir Laffont et Tirole, 1994, 1996 et 2000) s'appuie sur le principe de la tarification à la Ramsey. Le paragraphe suivant propose une présentation succincte de cette dernière :

On considère deux opérateurs. L'un d'eux exploite un réseau local. Le second est un opérateur qui propose un service « longue distance » et se connecte au réseau local. On note Y_0 la quantité de service local offerte au prix p_0 , Y_1 la quantité de service « longue distance » offerte par l'opérateur local au prix p_1 et Y_2 la quantité de service « longue distance » offerte par l'opérateur entrant au prix p_2 . La quantité de service qui transite par le réseau local est donc $Y = Y_0 + Y_1 + Y_2$. Soient respectivement c_0, c_1 et c_2 , les coûts marginaux sur le service local, sur le service longue distance pour l'opérateur de l'infrastructure locale et sur le service longue distance pour l'entrant. La consommation du service local assure aux usagers le surplus net $S(p_0)$ tandis que la consommation des services « longue distance » assure aux usagers le surplus net $S(p_1, p_2)$.

Le propriétaire de l'infrastructure locale propose un service d'accès au prix a à son concurrent. Ce dernier revend le service au prix $p_2 = a + c_2$. On suppose que l'entrant ne réalise pas de profit au départ. Afin de déterminer le prix d'accès optimal, il s'agit de définir dans un premier temps la tarification optimal pour les trois services. Le programme à résoudre s'écrit

$$\max_{p_0, p_1, p_2} S(p_0) + S(p_1, p_2) + \pi(p_0, p_1, p_2), \quad (23)$$

sous la contrainte

$$\pi(p_0, p_1, p_2) = (p_0 - c_0)Y_0 + (p_1 - c_1 - c_0)Y_1 + (p_2 - c_2 - c_0)Y_2 - k_0 \geq 0, \quad (24)$$

où $\pi(\cdot)$ représente le profit engendré par les trois services et k_0 est un coût fixe lié à la construction du réseau local. Le prix des services ainsi que le prix d'accès au réseau doivent permettre le remboursement des coûts fixes k_0 . Si ces coûts fixes étaient nuls, les prix optimaux seraient respectivement les coûts marginaux c_0 , $c_1 + c_0$ et $c_2 + c_0$. Les conditions de premier ordre associées au programme (23)-(24) sont

$$\frac{p_0 - c_0}{p_0} = \frac{\lambda}{1 + \lambda} \frac{1}{\hat{\eta}_0}, \quad (25)$$

$$\frac{p_1 - c_1 - c_0}{p_1} = \frac{\lambda}{1 + \lambda \hat{\eta}_1} \frac{1}{p_1}, \quad (26)$$

$$\frac{p_2 - c_2 - c_0}{p_2} = \frac{\lambda}{1 + \lambda \hat{\eta}_2} \frac{1}{p_2}, \quad (27)$$

où $\hat{\eta}_i$ est la superélasticité-prix qui intègre les possibles substituabilités ou complémentarités entre les deux services « longue distance » proposés par les deux opérateurs. Les prix sont ainsi fixés au-dessus des coûts marginaux afin de financer le remboursements des coûts fixes k_0 . La tarification est de type Ramsey. Plus la demande de service est inélastique et plus la substituabilité entre deux services est faible, plus le prix du service est élevé. Le prix d'accès optimal est alors facilement dérivé :

$$a = c_0 + \frac{\lambda}{1 + \lambda \hat{\eta}_2} \frac{p_2}{p_2}. \quad (28)$$

Ce prix dépend du coût marginal local c_0 et de la superélasticité de la demande $\hat{\eta}_2$. Rendre l'opérateur local seul responsable du remboursement des coûts fixes (par l'intermédiaire de p_0 et de p_1) conduirait à une importante distorsion de la demande en faveur de l'entrant.

Cette approche nécessite une bonne connaissance de la demande de biens et services. Les principaux détracteurs de cette méthode ont avancé qu'une autorité organisatrice ne possédait certainement pas cette information. Laffont et Tirole ont alors répondu qu'une tarification à la Ramsey était potentiellement décentralisable en soumettant les opérateurs à une tarification de type *prix-plafond*.

6.3. Un premier bilan

Si les expériences de déréglementation, notamment celles engagées au début des années 80 aux Etats-Unis et en Grande-Bretagne, ont fait l'objet de controverses au moment de leurs mises en œuvre, il semble maintenant que le bilan soit relativement favorable, notamment pour le consommateur. Certaines initiatives ont cependant été un échec.

La réussite de l'ouverture des marchés à la concurrence passe apparemment par la mise en place de règles d'interconnexion qui permettent aux consommateurs de bénéficier d'externalités de réseau et par l'instauration de règles de service universel précises. Nous terminons cette section sur la concurrence en évoquant quelques expériences représentatives et en présentant les enjeux majeurs dans les principales industries de réseau pour les années à venir. Cette sous-section suggère également qu'il existe des industries de réseau (l'électricité, le transport aérien et les télécommunications) pour lesquelles la notion de monopole naturel est presque caduque aujourd'hui.

Transport urbain

L'exemple de l'ouverture à la concurrence britannique (Londres et Irlande du Nord exceptés) en 1985 suggère que le transport urbain n'a sans doute pas intérêt à être déréglementé.¹³ Les entreprises municipales et nationales furent privatisées au début des années 80. Jusque en 1994, l'industrie britannique a vu ses coûts baisser de 30% malgré un accroissement de l'offre de 12.9%. En parallèle, on a pu observer une détérioration importante

de la qualité du service, une augmentation des prix de 31.8% et une diminution de la demande de 26.2%.

L'échec de la déréglementation dans ce secteur provient sans doute du fait qu'aucune règle de service universel n'a été définie par une autorité (les opérateurs ont une liberté totale en matière d'itinéraire, d'horaires et de prix) et que la qualité du service a été totalement sacrifiée au profit de la rentabilité. En outre, le service réclame une interconnectabilité difficilement réalisable par plusieurs opérateurs. Un système d'appel d'offre pour l'attribution des concessions de transport commence à se développer. C'est le cas en Grande-Bretagne pour Londres et les régions mal desservies par les opérateurs du marché déréglementé ; c'est également le cas pour la France et l'infrastructure lourde de type métro et tramway.

Transport aérien

La déréglementation du transport aérien américain s'opéra progressivement entre 1978 et 1983. Elle a conduit à une augmentation significative du nombre de transporteurs (61 entreprises en 1986 contre 28 en 1976) et de l'éventail des tarifs disponibles. Un des événements les plus spectaculaires auxquels a conduit la libéralisation est sans doute le développement rapide des vols domestiques avec correspondance et la création d'un réseau en étoile composé de plaques tournantes (*hubs*). Cette modification de la structure du réseau est sans doute à l'origine de la réussite de la déréglementation de l'industrie. Toutes les études concluent que la qualité du service s'est améliorée. Les prix se sont alignés sur les coûts, en particulier parce que la pratique des subventions croisées a peu à peu disparu. Les consommateurs ont vu les choix d'itinéraires et de fréquences s'élargir considérablement. Il semble donc que l'on puisse conclure ici que la dérégulation du transport aérien américain fut globalement une réussite.

Il apparaît nécessaire que la libéralisation du transport aérien s'accompagne d'un système de mise aux enchères des créneaux aéroportuaires dans les aéroports, sans quoi la concurrence sur les routes principales est difficile à mettre en œuvre pleinement. De plus, l'ouverture à la concurrence doit s'accompagner de règles précises dans le but de réduire le pouvoir de monopole de certaines compagnies qui dominent un *hub* particulier, à la manière des grosses compagnies américaines aux Etats-Unis.

Electricité

L'exemple britannique est particulièrement intéressant. La libéralisation de l'électricité britannique est issue de la mise en application de l'Electricity Act en 1990. La concurrence a entraîné l'entrée de nouveaux producteurs malgré l'existence d'un duopole dominant (*National Power* et *Power Gen*) dans l'électricité d'origine fossile. La distribution est maintenant assurée par 12 *Regional Electricity Companies* (REC) privatisées. Le transport proprement dit n'est pas touché par la nouvelle réglementation. Le recul permet de constater aujourd'hui que l'expérience fut satisfaisante : elle permit des gains de productivité importants dans la production et des prix de consommation modérés, même si des problèmes sont apparus au niveau de la planification à long terme des investissements.

De manière générale, il semble que l'ouverture à la concurrence de la production de l'électricité soit une voie à suivre, afin notamment d'éviter des prix trop élevés comme ce fut le cas en Angleterre. Un autre enjeu important est l'exercice de pressions concurrentielles au niveau de la distribution en permettant par exemple aux gros consommateurs que sont les entreprises de traiter directement avec les producteurs d'électricité.

Les télécommunications

En Grande-Bretagne, le service fut assuré au départ par un duopole (*British Telecom* et *Mercury*). 51% du capital de British Telecom fut ensuite privatisé et un *prix-plafond* fut

instauré. Le secteur fut ensuite complètement libéralisé en 1991 avec notamment l'entrée massive des câblo-opérateurs sur le marché de la téléphonie locale. Le bilan après trois années de pratique fait ressortir d'importants gains de productivité et une baisse des tarifs plus orientés vers les coûts grâce notamment à la suppression des subventions croisées par *British Telecom*. Les consommateurs ont par ailleurs bénéficié d'un élargissement de l'offre de services téléphoniques. Le seul point négatif est le relatif échec de la tentative de modification de la structure concurrentielle de l'industrie. *Mercury* s'est révélé un concurrent peu dangereux et *British Telecom* bénéficie toujours d'une position largement dominante (75% de part de marché en 1998 pour la téléphonie nationale).

Dans ce domaine, il semble que les enjeux principaux soient de déterminer les modes et les prix d'accès pour les nouveaux entrants. Il s'agit d'utiliser chaque segment le plus possible et d'éviter les contournements coûteux et inutiles. Un prix d'accès pour les nouveaux entrants basé sur les coûts marginaux empêche le remboursement des coûts fixes et motive d'autres voies de recherche telles que celle présentée précédemment avec une mise en œuvre par le principe du prix-plafond.

Services postaux

Il s'agit là d'un service aux caractéristiques particulières. Tout d'abord, la demande de courrier postal ne pourra au mieux que se maintenir à son niveau actuel au cours des années futures ; ce phénomène est naturellement dû au développement de la messagerie électronique. Ensuite, le facteur travail est largement majoritaire dans le processus de production. Ainsi, toute réforme telle que l'ouverture à la concurrence, toute perte de part de marché auraient des répercussions sociales et politiques importantes. Il semble que seule la distribution du courrier réponde aux propriétés de monopole naturel, ce qui laisse envisager une possible ouverture à la concurrence des autres activités du service. Cependant, hormis les services express, il existe peu d'exemples d'ouverture à la concurrence.

Eau

Dans le secteur de l'eau, des coûts fixes très importants, notamment au niveau de la distribution, imposent l'existence de monopoles naturels locaux. On revient alors au classique problème de régulation avec compromis entre efficacité et abandon de rente. Il ne semble donc pas qu'il existe un intérêt quelconque à l'introduction de la concurrence dans ce secteur.

Transport ferroviaire

L'industrie du transport de voyageurs est à ce jour fortement réglementée. Le service est généralement offert par un opérateur public unique. Les contraintes d'interopérabilité et de réduction des coûts rendent difficilement envisageable l'ouverture à la concurrence des grandes lignes et des lignes régionales. Un système d'appel d'offre pour l'attribution des concessions à des entreprises privées permettrait cependant une baisse significative des coûts d'exploitation d'une industrie très fortement déficitaire.

7. Tests empiriques

Nombreux sont les travaux de recherche économique sur la fourniture de services publics s'inspirant de la nouvelle théorie de la régulation. En revanche, peu d'études empiriques ont été réalisées en comparaison à ce jour. Plusieurs raisons peuvent expliquer ce constat. Tout d'abord, tester la nouvelle théorie de la régulation nécessite des bases de données qui permettent l'observation de plusieurs relations contractuelles entre plusieurs agents et plusieurs principaux. Ceci est possible si l'industrie considérée est constituée d'une multitude de monopoles locaux. Ensuite, les principes d'anti-sélection et d'aléa moral utilisent, par définition, des paramètres et

des agrégats non observables. Ainsi, l'économètre supporte les mêmes déficits informationnels qu'un responsable politique chargé de la régulation d'un monopole. Les choix des formes fonctionnelles relatives à la représentation des coûts d'exploitation, des coûts internes, l'intégration du coût des fonds publics, de la distribution des types des agents, la méconnaissance de la demande du produit ou du service considéré sont autant d'obstacles à une analyse empirique.

Différents types de tests empiriques ont cependant été proposés. Devant la difficulté technique, les estimations par les techniques du calibrage et la simulation ont primé. Ces techniques ont pour objet de comparer les différents modes régulateurs existants et théoriques en terme de bien-être apporté aux consommateurs. Les modèles structurels, moins nombreux, ont proposé des méthodes originales pour la prise en compte et le traitement de l'asymétrie d'information. Ces modèles élaborent les outils économiques nécessaires à la mise en place de mécanismes réglementaires optimaux dans la réalité. L'objet de cette section est de donner un recensement le plus exhaustif possible des approches proposées. On présente d'abord le principe des estimations des modèles à formes réduites, puis des modèles basés sur la simulation et enfin des modèles structurels.

Modèles à formes réduites

Avant d'entreprendre une modélisation structurelle ou par simulation, il peut être utile d'effectuer une analyse empirique préliminaire par un modèle économétrique à forme réduite qui suffit parfois à tester une théorie. Un modèle à forme réduite peut fournir une intuition sur les effets de certaines variables entre elles sans nécessiter une modélisation complète de l'environnement qui caractérise l'activité de production.

Mathios et Rogers (1989) s'intéressent à l'industrie du téléphone aux Etats-Unis et plus particulièrement aux appels à longue distance à l'intérieur d'un Etat. L'industrie a longtemps fait l'objet d'une régulation par le *taux de rendement*. A partir de 1984, plusieurs Etats ont adopté une régulation par le *prix plafond* afin d'autoriser une plus grande flexibilité à AT&T en matière de fixation des prix. Les auteurs proposent une approche empirique qui compare le niveau des prix appliqués dans les Etats où les schémas de type *taux de rendement* sont utilisés aux Etats où une régulation par le *prix plafond* est employée.¹⁴ La variable expliquée est le ou les prix pratiqués par les différents Etats. Les variables explicatives incluent différentes caractéristiques du réseau et une variable muette qui prend la valeur 1 si l'Etat autorise une flexibilité des prix à travers un régime de *prix-plafond*.

L'analyse suggère que les prix pratiqués sont généralement plus faibles de 7 à 13% dans les Etats où une plus grande flexibilité tarifaire est autorisée. Les auteurs justifient leurs résultats par une incitation à l'innovation plus importante dans le cas de la régulation par le principe du *prix-plafond*. Ainsi, la diminution correspondante des coûts autorise une diminution des prix.

A la différence de Mathios et Rogers, Dalen et Gomez Lobo (1995) observent l'impact d'un mode particulier de régulation sur les coûts d'exploitation et non sur les prix. Ils utilisent pour cela une forme fonctionnelle de type translog et incorporent au modèle une variable muette qui prend la valeur 1 si l'opérateur est régulé selon un mode standard. L'application porte sur L'industrie du transport urbain norvégien qui est fortement déficitaire.

En 1983, la règle du remboursement intégral du déficit a laissé la place à des procédures de négociation des objectifs financiers en début d'exercice. Au départ, les négociations sur les prix et les coûts étaient individuelles. En 1986, plusieurs autorités locales adoptèrent un système standard qui imposa notamment un certain nombre de critères universels pour la détermination des coûts d'exploitation. Ce procédé, comparable au principe de la *Yardstick competition*, s'est

avéré beaucoup plus incitatif en matière de réduction des coûts que celui basé sur la négociation individuelle. L'estimation suggère que les entreprises régulées sur la base d'un schéma standard ont en moyenne des coûts inférieurs de 3,6% par rapport aux entreprises régulées sur la base de critères individuels.

Gagnepain (1998) propose une procédure d'estimation similaire dans le cadre de la régulation du transport urbain en France. A partir de l'estimation d'une fonction de coût de type translog, il montre que les coûts d'exploitation sont en moyenne inférieurs de 2,05% pour les opérateurs régulés par des régimes de type *prix-fixe*.

Le choix des régimes réglementaires par les autorités de cette industries fait l'objet de l'analyse de Caillaud et Quinet (1993). Les paramètres caractérisant les puissances incitatives des mécanismes et les risques encourus par les opérateurs et les autorités sont supposés dépendre d'un ensemble de facteurs exogènes et endogènes comme l'aversion au risque de l'exploitant, la volatilité des recettes et des coûts, le coût des fonds publics, le niveau de la demande, l'efficacité des politiques de commercialisation ou de productivité, l'incitation à la commercialisation et à la productivité désirée par le principal. Afin d'établir un lien théorique entre ces différents facteurs et la puissance incitative du contrat, les auteurs utilisent la théorie des contrats et des incitations. De l'analyse théorique sont tirés plusieurs enseignements qui sont testés empiriquement à l'aide d'une forme réduite et un modèle économétrique de choix discrets. Les variables expliquées sont les paramètres des régimes contractuels. Plusieurs variables explicatives sont introduites afin de simuler les effets des paramètres exogènes et endogènes évoqués ci-dessus. Les résultats empiriques confirment globalement les intuitions de départ.

Techniques de simulation et de calibrage

Les techniques de simulation et de calibrage n'ont pas à proprement parler de fondements économétriques. Les modèles qui les utilisent ne permettent pas d'estimer par exemple une technologie particulière. A partir de la définition préalable d'un ensemble de paramètres et de valeurs injectées en guise de données exogènes, ces techniques permettent plutôt de comparer le bien-être procuré par différents régimes réglementaires ou d'observer la trajectoire d'agrégats économiques dans des situations particulières. Elles permettent une validation ou un rejet de la théorie économique à laquelle elles se rapportent.

La première approche empirique utilisant de telles techniques est sans doute celle de Schmalensee (1989). Celui-ci compare les propriétés d'un ensemble de "bons" régimes réglementaires en terme de bien-être procuré à la société. Par ce terme, il désigne les mécanismes utilisés dans la réalité, qui sont particulièrement bien compris et administrés par leurs utilisateurs. Il s'intéresse plus particulièrement aux régimes linéaires de type *prix plafond* et *coût du service* (sans transfert) et à des cas intermédiaires de partage des risques liés aux coûts entre l'entreprise régulée et le régulateur. Le modèle considère les coûts d'exploitation marginaux d'un monopole. Ceux-ci contiennent une part d'incertitude liée à des chocs exogènes et ils dépendent d'un effort inconnu du régulateur et de l'économètre. L'exercice de l'effort conduit à un coût interne qui dépend de l'effort et d'un paramètre d'anti-sélection inconnu du régulateur. Un contrat linéaire proposé au monopole détermine le prix qu'il applique. Ce prix est une fonction des coûts réalisés.

La distribution de l'incertitude liée aux coûts et celle du terme d'anti-sélection, l'élasticité de la demande dépendent d'un ensemble de paramètres positifs. La simulation consiste à déterminer le régime qui maximise le bien-être de la société à partir d'un ensemble de valeurs pour ces paramètres. L'auteur conclut qu'en l'absence d'incertitude sur les coûts, les

régimes de type *prix-plafond* sont toujours optimaux pour la société. En revanche, un accroissement de l'incertitude privilégie plutôt les mécanismes moins incitatifs.

Gasmi, Ivaldi et Laffont (1994) se basent au départ sur le même modèle structurel et la même grille de valeurs que Schmalensee. L'incertitude concerne néanmoins les coûts totaux et non plus simplement les coûts marginaux et le mode de financement de l'activité autorise cette fois le versement de transferts par l'autorité. Le modèle propose une comparaison, en terme de bien-être, des mécanismes de régulation optimaux en information parfaite ou asymétrique aux "bons" régimes réglementaires de Schmalensee et un régime de type *prix plafond*.

L'analyse montre qu'un mécanisme optimal de second rang atteint le bien-être de premier rang lorsque la pondération du profit de la firme dans le bien-être est maximale. De manière générale, tous les mécanismes considérés sont proches du bien-être de premier rang lorsque cette pondération augmente. La régulation par le prix-plafond abandonne des rentes plus importantes que les autres régimes. Néanmoins, en y incorporant des principes de partage des profits, le régulateur peut restreindre l'abandon de la rente au monopole et atteindre un niveau de bien-être comparable à celui issu d'une régulation optimale.

Les travaux de Gasmi, Laffont et Sharkey (1997, a et b) proposent de construire un menu de contrats optimaux de second rang à partir de données simulées et générées par un programme d'optimisation.¹⁵ Cette construction permet de comparer dans un second temps différents modes de régulation en terme d'effets redistributifs sur la société. Le modèle s'intéresse aux entreprises de télécommunication locales aux Etats-Unis. Le cadre général de l'analyse est basé sur une estimation préalable des coûts de distribution des entreprises à l'aide d'une fonction de type translog. L'estimation des coûts permet aux auteurs de simuler l'allocation optimale de second rang issue et de proposer notamment un menu de contrats linéaires.

Cette réalisation nécessite un calibrage préalable des paramètres ou agrégats a priori non observables et non quantifiables que sont l'efficacité technique, l'effort, la désutilité de l'effort, le coût des fonds publics et le surplus des consommateurs. La méthodologie générale utilisée est proche de l'esprit de celle de Wunsch (1994) décrite ci-après. L'efficacité technique et l'effort sont mesurés respectivement à partir d'indices dérivés des prix du capital et du travail. La fonction de demande individuelle, nécessaire pour le calibrage du surplus du consommateur, est caractérisée par une élasticité calibrée pour l'industrie considérée. Une forme quadratique est utilisée pour le calibrage de la désutilité de l'effort dont les paramètres s'inspirent de l'observation des effets des réductions de masses salariales après la dérégulation ou la privatisation d'entreprises de télécommunication. Les croyances du régulateur suivent une loi uniforme. Le coût des fonds publics est arbitrairement fixé à 0.3. Cette valeur semble appropriée pour les pays développés à systèmes de taxation efficaces.

Les résultats vérifient tout d'abord les propriétés classiques de l'allocation optimale développées dans cet article. L'effort optimal, la rente informationnelle, la désutilité de l'effort et le surplus du consommateur sont des fonctions décroissantes de l'inefficacité. L'estimation laisse apparaître une correction incitative sur les prix faible. L'hypothèse de dichotomie des préoccupations incitatives et de la tarification est testée et validée.

On analyse ensuite les conséquences redistributives de différentes formes de régulation. Le mécanisme incitatif optimal avec observation des coûts ex-post (Laffont et Tirole, 1986) est comparé à celui sans observation possible des coûts ex-post (Baron et Myerson, 1982), à la régulation par *prix plafond* avec et sans partage des profits, à la régulation par le *coût du service* et au cas où l'information entre monopole et régulateur est dite parfaite. Il apparaît que la situation d'information parfaite est bien entendu la meilleure et permet d'atteindre le bien-être social de premier rang. La régulation par *prix plafond* est la plus incitative en matière

d'incitation à la productivité et donc abandonne les rentes informationnelles les plus élevées. Le gain d'efficacité se transforme ainsi en un important coût social. Le bien-être général associé est proche des niveaux atteints par les mécanismes optimaux à la Laffont-Tirole et Baron-Myerson lorsque le régulateur associe au *prix plafond* un schéma de partage des profits. Cette conclusion est similaire à celle évoquée dans Gasmi, Ivaldi et Laffont (1994). La régulation par le coût du service est la plus préjudiciable pour la société en terme de bien-être.

L'approche proposée par Wunsch (1994) se distingue clairement des modèles de simulation précédents. Tout d'abord, son objectif n'est pas de comparer les effets sur le bien-être de différents modes de régulation mais de construire un menu de mécanismes incitatifs optimaux. Ensuite, à la différence de Gasmi, Laffont et Sharkey, la procédure d'élaboration du mécanisme prend en compte l'environnement réglementaire et économique réel des opérateurs qui constituent la base de donnée utilisée. En ce sens, l'auteur considère que les mécanismes existant ne sont pas optimaux. L'observation de la comptabilité individuelle des firmes suggère plutôt des schémas peu incitatifs en matière d'effort à la productivité. Ce postulat de départ est évidemment primordial.

L'auteur propose alors de construire par la simulation les mécanismes optimaux de second ordre et de calculer les bénéfices associés en termes de réduction des coûts. Une telle construction nécessite, comme dans Gasmi, Laffont et Sharkey, un calibrage ou une estimation préalable de la distribution du terme d'inefficacité, de la désutilité de l'effort et du coût des fonds publics. Le champ d'application est l'industrie du transport urbain à l'échelle de l'Europe. Une étude préalable (Wunsch, 1996) fournit à l'auteur l'estimation d'une fonction de coût du secteur.

L'idée est alors d'utiliser les écarts-types des intervalles de confiance autour des coûts estimés pour dériver une distribution plausible de l'inefficacité des opérateurs. En effet, Wunsch suggère que l'asymétrie d'information entre entreprise régulée et régulateur se limite à la variance inexplicite de l'estimation des coûts. L'écart type estimé représente 15% des coûts. La distribution de l'inefficacité des opérateurs est alors approximée par une distribution normale de même écart-type. La désutilité de l'effort est supposée avoir une forme quadratique. La détermination des paramètres de la fonction est dérivée de l'observation de la comptabilité des entreprises et des expériences de dérégulation en Europe. Le coût des fonds publics est supposé nul. Au vu des résultats, il semble que d'importants gains en terme de réduction des coûts d'exploitation pourraient être atteints par la mise en place d'un mécanisme optimal.

Les modèles économétriques structurels

Les modèles précédents, basés sur des formes réduites ou utilisant des techniques de simulation ne s'imprègnent en aucun cas de l'environnement réglementaire, institutionnel ou économique qui règle le comportement d'une entreprise. Les formes fonctionnelles qui servent de support à l'estimation ne tirent leur identité d'aucune approche économique préalable.

Le propre des approches structurelles est justement qu'elles prennent en compte au mieux les interactions qui génèrent les données auxquelles le modèle se rapporte. La description du comportement du producteur dans le cadre de son activité est nécessaire pour une représentation fidèle de la réalité. Ainsi, une forme fonctionnelle structurelle est forcément dérivée d'un modèle théorique, pertinent dans la mesure du possible, approprié à l'industrie considérée.

Il semble que Wolak (1994) soit le premier à proposer un modèle structurel qui traite de la régulation des monopoles en présence d'asymétrie d'information. Il considère un problème d'anti-sélection pure.

Wolak s'intéresse à l'industrie de la distribution d'eau aux Etats-Unis et propose ainsi une estimation des coûts d'opérations des entreprises chargées du service. La structure des coûts opérationnels et de la demande considérée tient compte du cadre réglementaire qui influence le comportement des opérateurs. Le mode de régulation en place utilise des mécanismes optimaux de second rang du type de ceux proposés par Baron et Myerson (1982). De manière générale, l'opérateur choisit le stock de capital nécessaire qui assure le niveau de production requis par l'autorité. Après avoir observé le choix de l'opérateur, l'autorité détermine le prix fixe à payer par les consommateurs pour accéder au réseau et le prix marginal que ceux-ci doivent supporter pour chaque unité supplémentaire de service consommée. Deux structures informationnelles sont considérées. Une première suppose que le régulateur connaît l'efficacité productive de l'entreprise et peut ainsi fixer les prix du service en fonction de cette efficacité. Un second schéma suppose une situation d'asymétrie d'information. Dans ce cas, le régulateur propose des combinaisons incitatives de prix d'accès et de prix marginaux qui incitent l'opérateur à révéler sa véritable identité par la sélection de son stock de capital. En situation d'information parfaite, le régulateur égalise le prix marginal au coût marginal du service et le taux de rémunération du capital à sa productivité marginale. En information imparfaite, prix et taux subissent une distorsion.

L'estimation est basée sur un système composé d'une fonction de coût d'exploitation et d'une fonction de demande du service dans les deux cas d'information complète ou incomplète sur la productivité du travail. La fonction de coût intègre l'inefficacité inobservable de l'opérateur. Wolak distingue, à cet effet, la quantité de travail observable qui conduit aux coûts d'exploitation, de la quantité de travail efficace qui permet à l'opérateur d'atteindre le niveau de production requis. Les deux quantités de travail convergent l'une vers l'autre lorsque l'opérateur réduit son inefficacité.

La fonction de demande intègre les conditions de premier ordre de la régulation optimale. La procédure d'estimation elle-même associe des techniques non-paramétriques pour la densité de l'inefficacité et des techniques paramétriques pour l'estimation de la technologie. Un test non-emboîté conclut à la supériorité du modèle d'information incomplète. Les résultats suggèrent également que les approches standard préalables qui ignorent les asymétries d'information éventuelles surestiment généralement les économies d'échelles des monopoles.

Thomas (1995) propose également une application qui intègre des mécanismes optimaux de second rang. Il considère une estimation structurelle dans le cadre de la régulation de la pollution de l'eau en France. Une entreprise, par son activité de production, peut dégager des effets externes négatifs en polluant par exemple les sources d'eau environnantes. L'intervention d'une autorité peut alors être nécessaire pour assurer une canalisation satisfaisante des émissions polluantes.

L'instauration d'une taxe assure un revenu unitaire à chaque entreprise pour chaque réduction unitaire du niveau de pollution. En présence d'une taxe fixée à un niveau imparfait, les entreprises polluantes réduisent leurs activités de contrôle des émissions. L'autorité locale, en l'occurrence l'Agence française du contrôle de la pollution, doit alors intervenir et proposer un schéma régulateur aux entreprises de façon à leur fournir les incitations nécessaires pour investir dans une capacité de traitement des émissions satisfaisante. Thomas suppose que les mécanismes contractuels sont optimaux et assurent à la société un bien-être de second rang. Le bien-être de premier rang ne peut être atteint car l'autorité locale ne possède pas d'information sur le coût marginal de l'activité de réduction des émissions. Le mécanisme optimal définit les niveaux optimaux de réduction des émissions polluantes et de capital nécessaire à l'exploitation de l'activité. Bien évidemment, une rente est abandonnée, qui devient de plus en plus coûteuse au fur et à mesure que des opérateurs plus inefficaces sont régulés.

L'analyse empirique présente l'estimation indirecte des paramètres d'une fonction de coût liée à l'activité de réduction des émissions. Le terme dénotant l'inefficacité est supposé avoir une densité de type log-normale. L'estimation suit une procédure à deux étapes. La densité de l'inefficacité des opérateurs est obtenue dans un premier temps puis les paramètres structurels sont estimés dans un second temps en intégrant les conditions de premier ordre inhérentes aux niveaux de production et de capital optimaux.

Deux résultats principaux sont obtenus. D'une part, la politique de régulation optimale consiste à contracter uniquement avec les entreprises les plus efficaces. Les entreprises les plus inefficaces ne sont pas régulées. Ensuite, il apparaît que la taxe actuellement en place dans l'industrie devrait être deux fois plus élevée si l'autorité locale voulait fournir aux entreprises polluantes les incitations nécessaires à une réduction de la pollution socialement satisfaisante sans avoir recours à la régulation directe.

L'industrie du transport urbain norvégien décrite dans Dalen et Gomez-Lobo (1995) sert de support à l'analyse structurelle proposée dans Dalen et Gomez-Lobo (1996). Les opérateurs de transport font face à deux types de régulation aux puissances incitatives distinctes en matière d'effort de productivité. Un premier type est basé sur des pratiques de négociation individuelle. L'autre fixe des objectifs de production, de prix et de coût sur la base d'un schéma standard et s'inspire ainsi des principes de la *yardstick competition*.

Dalen et Gomez-Lobo proposent dans un premier temps l'estimation d'une fonction de coût. La structure des coûts intègre le comportement de maximisation de l'utilité des firmes. Les auteurs supposent que chaque schéma est décrit par un transfert du type de celui décrit dans la section 3.4. L'utilité des firmes dépend du transfert obtenu, des coûts internes, des coûts d'exploitation et des recettes commerciales. Les modes réglementaires sont supposés exogènes au problème, c'est à dire que l'estimation n'intègre pas le programme du principal à la manière de l'approche structurelle de Wolak (1994). Les opérateurs maximisent leur utilité et allouent ainsi un niveau d'effort particulier à l'activité.

A l'instar de Wolak (1994), le facteur travail constitue ici la source de l'information asymétrique. Une forme fonctionnelle particulière est calibrée pour les coûts internes. L'estimation de la technologie permet aux auteurs de proposer une distribution des efficacités à partir des écarts types des intervalles de confiances autour des coûts dans l'esprit de la méthode suggérée par Wunsch (1994).

En considérant plusieurs valeurs possibles pour le coût des fonds publics, Dalen et Gomez-Lobo calculent les gains en bien-être induits par la mise en place d'un mécanisme optimal de second-rang du type de celui proposé par Laffont et Tirole (1986).¹⁶ Afin de pallier un abandon de rente élevé, associé à ce type de mécanismes, les auteurs proposent également une évaluation empirique des gains induits par l'instauration d'un système d'appel d'offres. Le support théorique est le modèle d'enchère de Laffont et Tirole (1987) qui montre que le principe de la meilleure enchère permet de réduire l'abandon de rentes tout en conservant un caractère incitatif comparable à celui du mécanisme optimal de second rang. Le gain dépend bien entendu du nombre d'agents participant à l'appel d'offre.¹⁷ Dans le modèle de Dalen et de Gomez-Lobo, une simple enchère avec deux participants permet de réduire de moitié la rente versée à l'opérateur.

A l'inverse de Wolak (1994), Thomas (1995) et dans une moindre mesure de Dalen et Gomez-Lobo (1996), Gagnepain et Ivaldi (1998 a et b) considèrent que, dans le cadre particulier de l'industrie du transport urbain français, les contrats en place ne sont pas optimaux au sens défini par la nouvelle théorie de la régulation et donc que ceux-ci sont à élaborer. Cette hypothèse de départ est primordiale. Il s'agit, à partir de l'observation des contrats effectivement utilisés dans l'industrie, d'estimer les caractéristiques des différents opérateurs de transport

(efficacité et effort) et des zones de transport elles-mêmes (coût des fonds publics et élasticité de la demande). Ceci passe par la prise en compte simultanée de l'offre, de la capacité et de la demande de transport et donc par la construction d'un système à trois niveaux. A cet effet, la nouvelle théorie de la régulation est considérée comme un outil et ne fait pas l'objet d'une tentative de validation systématique.

Le cadre d'analyse considéré est celui du modèle principal-agent avec asymétrie d'information où l'inefficacité globale liée au processus de production se décompose en une inefficacité technique et un effort endogène de productivité. Au cours du processus, un ou plusieurs facteurs de production sont affectés. On distingue alors la quantité de facteur "efficace" de la quantité de facteur utilisée, le rapport entre ces deux quantités dépendant de l'inefficacité technique et de l'effort de chaque opérateur. Cette distinction permet de construire dans une première étape des frontières de production et de coût conditionnelles à un niveau d'effort endogène. Cet effort est endogène dans le sens où il dépend du cadre réglementaire en place. Si l'opérateur maximise son utilité à régulation donnée, l'effort endogène optimal dépend des prix des facteurs, du niveau de production, de l'inefficacité technique et de la technologie. En réintroduisant l'effort optimal dans la frontière de production ou de coût préliminaire, on obtient la forme structurelle finale à estimer.

L'estimation des caractéristiques "réelles" des opérateurs et du site de production associé permet alors d'élaborer des politiques de régulation optimales qui améliorent le bien-être des consommateurs de transport urbain.

8. Conclusion

Cet article a rappelé comment la prise en compte de l'asymétrie d'information dans les relations entre le régulateur et l'entreprise régulée peut conduire à un mécanisme de régulation optimal de second rang. Lorsque les coûts d'exploitation sont observables en fin de période, les mécanismes de régulation optimaux peuvent être décentralisés au moyen d'un menu de contrats linéaires. Un opérateur qui fait face au menu de contrats choisit le mécanisme approprié qui correspond à son type d'inefficacité parce que tout autre choix lui procurerait une utilité moindre. Par opposition au niveau de premier ordre accessible en situation d'information parfaite, l'allocation définie par ce mécanisme conduit au bien-être social de second ordre parce que les contraintes incitatives induites nécessitent un abandon de rente par le principal et parce que le niveau d'effort requis par l'autorité est inférieur au niveau de premier ordre.

La nouvelle théorie de la régulation des monopoles naturels propose un cadre théorique qui utilise des mécanismes révélateurs. Dans la réalité, les autorités publiques se servent-elles de tels mécanismes ? Observe-t-on des menus de contrats ? Si oui, de tels menus sont-ils optimaux au sens définis par la théorie ?

A notre connaissance, il n'existe que très peu, voire pas d'exemples d'industries qui utilisent des menus de contrats linéaires strictement inspirés des mécanismes présentés en section 3. On peut imaginer que l'observateur extérieur au processus de sélection ne peut observer que le contrat final sélectionné et non le menu de départ. Cependant, il existe plusieurs facteurs qui expliquent la difficulté de mise en place d'un mécanisme révélateur.

La première raison est essentiellement d'ordre technique : la théorie suppose que l'autorité, lorsqu'elle construit le mécanisme, connaît le coût interne engendré par l'effort de l'entreprise. Cette hypothèse est assez forte dans la mesure où on a supposé par ailleurs que la clairvoyance comptable de l'autorité est limitée. Ainsi, dans le cadre de la régulation d'une firme multi-produits, par exemple, on a supposé que l'autorité soit incapable de distinguer les coûts issus des différentes activités de production. De même, la connaissance subjective que peut

avoir l'autorité de la distribution du type des agents, est nécessairement limitée, voire imprécise. L'impact de la génération de l'effort sur les coûts est également arbitraire. L'autorité ne sait pas, a priori, si l'effort affecte tous les facteurs de production ou si son effet est concentré sur certains d'entre eux uniquement.

La seconde raison est liée au caractère dynamique des relations contractuelles entre les autorités et les opérateurs. Le « ratchet effect » empêche l'existence d'équilibres séparateurs pour les contrats de première période. Ainsi, il n'est pas possible de mettre en place des mécanismes révélateurs dans ce contexte.

Enfin, et plus simplement, la détermination et la mise en place d'un mécanisme optimal peuvent s'avérer extrêmement coûteuses et les motivations de l'autorité à cet effet peuvent être limitées, vu que, comme exposé dans la section 7, certaines combinaisons de schémas couramment utilisés, *prix-plafond* et *partage des profits*, assurent à la société un bien-être proche du bien-être de second rang.

Dans la réalité, il semble que l'on observe essentiellement des menus composés de deux contrats (voir Mathios et Rogers, 1989, Lewis et Sappington, 1989, Dalen et Gomez-Lobo, 1995 et Gagnepain, 1998 entre autres) de types polaires comme ceux présentés dans la section 2 : ce sont des contrats très incitatifs (de type *prix-fixe* ou *prix-plafond*) ou des contrats très peu incitatifs (de type *coût du service* ou *taux de rendement*) en matière d'effort à la productivité. Comment interpréter ces contrats ? Deux réponses sont sans doute possibles :

Si l'on se réfère strictement à la méthodologie présentée dans cet article, de tels contrats ne sont pas optimaux. Cependant, si l'on considère que face aux difficultés techniques et informationnelles énoncées ci-dessus, les autorités publiques restreignent l'espace des capacités productives des opérateurs à deux types (« plutôt efficace » et « plutôt inefficace » par exemple), il est peut-être possible d'accepter l'hypothèse selon laquelle ces contrats polaires constituent un menu optimal où un opérateur plutôt inefficace choisit un contrat de type *coût du service* et un opérateur plutôt efficace choisit un contrat de type *prix-fixe*.

A l'opposé, on peut également imaginer que face aux difficultés techniques et informationnelles, les autorités publiques (très certainement averses au risque) se limitent à proposer des contrats peu incitatifs lorsqu'elles souhaitent s'impliquer financièrement dans le service régulé ou des contrats incitatifs lorsqu'elles souhaitent de désengager financièrement. Cette hypothèse est extrêmement plausible, elle expliquerait pourquoi, au cours des 20 dernières années et face à l'accroissement de leur endettement, les autorités centrales ou locales ont peu à peu abandonné les principes du *taux de rendement* et du *coût du service* pour ceux du *prix-plafond* et du *prix-fixe*.

En définitive, comme suggéré par Armstrong, Cowan et Vickers (1994), convient-il de considérer les modèles de la nouvelle théorie de la régulation comme des guides pour l'intuition des économistes (l'implication centrale est que plus le régulateur est informé, meilleur sera le mécanisme réglementaire pour le bien-être de la société), voire comme une norme à laquelle toute régulation de monopole devrait se référer plutôt qu'une représentation de la réalité ? Si tel est le cas, les mécanismes optimaux de premier et de second rang sont à élaborer et à mettre en place. Pour ce faire, les travaux économétriques structurels sont certainement d'une grande importance. Ils peuvent en effet, à partir de l'observation des schémas réglementaires pratiqués, permettre d'identifier les outils nécessaires à la construction des mécanismes de premier et de second-rang. Le champ de réflexion dont Leibenstein (1966) fut le principal initiateur, qui combine les notions de frontière de production et de coût d'une part, les incitations d'autre part, fournira certainement à la nouvelle théorie de la régulation les outils indispensables à sa mise en pratique.

Bibliographie

- Allais, M. Le problème de la coordination des transports et la théorie économique. *Bulletin des Ponts et Chaussées et des Mines*, 1947.
- Armstrong, M., Cowan, S., et J. Vickers. Regulatory reform: Economic analysis and British experience. *Cambridge: MIT Press*, 1994.
- Averch, H., et L. Johnson. Behavior of the firm under regulatory constraint. *American Economic Review*, 1962, Vol. 52 : 1052-1069.
- Baron, D., et R. Myerson. Regulating a monopolist with unknown costs. *Econometrica*, 1982, Vol. 50 : 911-930.
- Boiteux, M. Sur la gestion des monopoles publics astreints à l'équilibre budgétaire. *Econometrica*, 1956, Vol. 24 : 22-40.
- Caillaud, B., et E. Quinet. Analyse du caractère incitatif des contrats de transport urbain. Mimeo, 1993, *CERAS*.
- Crew, M. A., et P.R. Kleindorfer. Incentive regulation in the United Kingdom and the United States: Some lessons. *Journal of Regulatory Analysis*, 1996, Vol. 9 : 211-225.
- Dalen, D.M., et A. Gomez-Lobo. Regulation and incentive contracts: An empirical investigation of the Norwegian bus transport industry. *Working paper series N°W96/8, IFS*, 1996.
- Dalen, D.M., et A. Gomez-Lobo. Incentive contracts: An empirical investigation of the Norwegian bus transport industry. Mimeo, 1995.
- Donald, S.G., et D.E.M. Sappington. Choosing among regulatory options in the United States Telecommunication industry. *Journal of Regulatory Analysis*, 1998, Vol. 12 : 225-340.
- Fudenberg, D., et J. Tirole. Game theory. *Cambridge, MIT Press*, 1991.
- Gabel, D., et M. Kennet. Estimating the cost structure of the local telephone exchange network. *National Regulatory Research Institute Report N°91-16*, 1991.
- Gagnepain, P. Structures productives de l'industrie du transport urbain et effets des schémas réglementaires. *Economie et Prévision*, 1998, Vol. 35 : 95-107.
- Gagnepain, P., et M. Ivaldi. Stochastic frontiers and asymmetric information models. *Document de travail GREMAQ*, 1998a.
- Gagnepain, P., et M. Ivaldi. Current versus optimal regulatory policies: Evidence from the French urban transport industry", *Document de travail IDEI*, 1998b.
- Gasmi, F., Ivaldi, M., et J.J. Laffont. Rent extraction and incentives for efficiency in recent regulatory proposals. *Journal of Regulatory Economics*, 1994, Vol. 6 : 151-176.
- Gasmi, F., Laffont, J.J., et W.W. Sharkey. Incentive regulation and the cost structure of the local telephone exchange network. *Journal of Regulatory Analysis*, 1997a, Vol.12 : 5-25.
- Gasmi, F., Laffont, J.J., et W.W. Sharkey. Empirical evaluation of regulatory regimes in local telecommunications markets. *Document de travail IDEI*, N°68, 1997b.
- Gasmi, F., Laffont, J.J., et W.W. Sharkey. The natural monopoly test reconsidered. *Mimeo*, 1999.
- Gibbard, A. Manipulation for voting schemes. *Econometrica*, 1973, Vol. 41 : 587-601.
- Helm, D., et T. Jenkinson (Eds). Competition in regulated industries. *Oxford University Press*. 1998.
- IDEI. "Network Industries and Public Service", [European Economy, n°4, 1999](#)
- Joskow, P., et R. Schmalensee. Incentive regulation for electric utilities. *Yale Journal of Regulation*, 1986, Vol. 4 : 1-49.
- Laffont J.J. A propos de l'émergence de la théorie des incitations. *Revue Française de Gestion*, 1993, N°96 : 13-19.

- Laffont, J.J. The new economics of regulation ten years after. *Econometrica*, 1994, Vol. 62 : 507-537.
- Laffont, J.J., et J. Tirole. Using cost information to regulate firms. *Journal of Political Economy*, 1986, Vol. 64 : 614-641.
- Laffont, J.J., et J. Tirole. Auctioning incentive contracts. *Journal of Political Economy*, 1987, Vol. 95 : 931-937.
- Laffont, J.J., et J. Tirole. The dynamics of incentive contracts. *Econometrica*, 1988, Vol. 56 : 1153-1175.
- Laffont, J.J., et J. Tirole. A theory of incentives in procurement and regulation. *Cambridge: MIT Press*. 1993.
- Laffont, J.J., et J. Tirole. Competition in telecommunications. *Cambridge: MIT Press*. 2000.
- Leibenstein, H. Allocative efficiency versus "X-efficiency". *American Economic Review*, 1966, Vol. 56 : 392-415.
- Lewis, T., et D. Sappington. Countervailing incentives in agency problems. *Journal of Economic Theory*, 1989, Vol. 20 : 405-416.
- Loeb, M., et W. Magat. A decentralized method of utility regulation. *Journal of Law and Economics*, 1979, Vol. 22 : 399-404.
- Martin, S. (Ed). Competition Policies in Europe. *North Holland*. 1998.
- Mathios, A.D., et R.P. Rogers. The impact of alternative forms of state regulation of AT&T on direct-dial, long distance telephone rates. *RAND Journal of Economics*, 1989, Vol. 20 : 437-453.
- Mirrlees, J. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 1971, Vol. 38 : 175-208.
- Reichelstein, S. Constructing incentive schemes for government contracts, an application of agency theory. *The Accounting Review*, 1992, Vol. 67 : 712-731.
- Salanié, B. Théorie des contrats. *Economica*, 1994.
- Sappington, D.E.M., et D.L. Weisman. Designing incentive regulation for the telecommunications industry. *Cambridge: MIT Press*. 1996.
- Schmalensee, R. Good regulatory regimes. *Rand Journal of Economics*, 1989, Vol. 20 : 417-436.
- Thomas, A. Regulating pollution under asymmetric information: The case of industrial waste water treatment. *Document de travail GREMAQ*, 1995.
- Thonier, A. Dix ans de réformes des télécommunications au Royaume-Uni : échec du pilotage de la structure de marché et libéralisation complète. *Economie et Prévision*, 1995, Vol. 119 : 107-119.
- Vickers, J., et G. Yarrow. Privatization: An economic analysis. *Cambridge: MIT Press*. 1988.
- Williamson, O. Franchise bidding for natural monopolies. *Bell Journal of Economics*, 1976, Vol. 7 : 73-104.
- Wolak, F.A. An econometric analysis of the asymmetric information, regulator-utility interaction. *Annales d'Economie et de Statistique*, 1994, Vol. 34 : 13-69.
- Wunsch, P. Estimating menus of linear contracts for mass transit firms. Mimeo, 1994, CORE, Louvain-La-Neuve, Belgique.
- Wunsch, P. Cost and productivity of major urban transit systems in Europe. *Journal of Transport Economics and Policy*, Mai 1996 : 171-186.

¹ Allais (1947) et Averch et Johnson (1962) font vraisemblablement partie des rares exceptions à ce sujet.

² Voir Laffont (1993) pour une revue de la théorie des incitations et Laffont (1994) pour une revue de la nouvelle théorie de la régulation.

³ Voir Vickers et Yarrow (1988) et Armstrong, Cowan et Vickers (1994) pour plus de détails sur l'expérience Britannique qui concerna d'autres industries que celles évoquées dans cette section. Voir Lyon et Toman (1991) pour un aperçu de l'expérience américaine dans le cadre des industries du gaz et du transport et Crew et Kleindorfer (1996) pour un aperçu général sur l'application des régimes inspirés du principe du taux de rendement et du *prix-plafond* en Grande Bretagne et aux Etats-Unis. Les approches de Joskow et Schmalensee (1986) sur l'industrie de l'électricité et de Donald et Sappington (1998) sur celle des Télécommunications aux Etats-Unis sont plus théoriques.

⁴ Lorsqu'une entreprise est régulée par un *prix-plafond* et que celle-ci dégage des profits, le régulateur peut procéder à une révision du prix afin de capturer la rente du producteur. Cette procédure peut cependant avoir des effets inattendus dans la mesure où l'opérateur peut anticiper cette révision et décider de ne pas réaliser le niveau d'effort optimal. Ce principe est connu sous la dénomination *ratchet effect*. Cet effet peut transformer une régulation théoriquement très incitative en une régulation très peu incitative (*taux de rendement*).

⁵ Dans le mécanisme de régulation optimale de Baron et Myerson (1982), λ est un paramètre de pondération qui définit le poids du profit de l'entreprise dans le bien-être de l'ensemble de la société et non comme un paramètre représentatif d'une quelconque distorsion issue de la taxation des consommateurs. Le modèle, bien que adoptant une méthodologie sensiblement différente conduit cependant aux mêmes types de mécanismes incitatifs.

⁶ Gibbard (1973) est à l'origine de ce principe.

⁷ Par court terme, on entend généralement des périodes de 5 années environ.

⁸ Voir Gasmi et alii (1999) par exemple.

⁹ Les industries en réseau ont en commun l'existence des obligations de service universel. Les obligations de service universel sont motivées par l'instauration d'une certaine équité chez les consommateurs tout comme leur existence est motivées par la création d'externalités de réseau. Il s'agit pour un opérateur d'offrir le service à tous les consommateurs avec des niveaux de qualité et de prix satisfaisants. Le service universel peut prendre différentes formes ; obligation de continuité, de régularité de service, de capacité et de qualité minimale, obligation de prix en particulier pour les groupes à faibles revenus. En pratique, ces obligations sont financées au moyen de subventions croisées par les prix et de subventions versées par les autorités publiques.

¹⁰ Dans le cas d'un accès à sens unique, l'opérateur en place n'a pas besoin d'utiliser le réseau de l'entrant. L'accès bilatéral est une situation où deux concurrents font chacun usage de leur propre réseau et du réseau de leur adversaire.

¹¹ Voir Baumol et Sidak (1994, chapitre 7).

¹² Par coût d'opportunité, on entend ici perte de profit lié au détournement d'une partie de la demande par le nouvel entrant.

¹³ L'exemple de la déréglementation chilienne conduit aux mêmes conclusions.

¹⁴ Dans le domaine des télécommunications aux Etats-Unis, il existe une importante littérature empirique, essentiellement américaine, qui analyse les effets de l'introduction des régimes de *prix-plafond* sur les prix, les coûts, la productivité et la qualité du service, l'investissement et les profits des opérateurs. Voir Sappington et Weisman (1996) pour les références.

¹⁵ Il s'agit du modèle d'optimisation LECOM (Local Exchange Cost Optimization Model) développé par Gabel et Kennet (1991).

¹⁶ Pour être plus précis, ils ne considèrent que la partie coût du bien-être. Ils supposent en effet que les mécanismes optimaux affectent faiblement le surplus du consommateur et les recettes commerciales. L'analyse empirique propre au transport urbain présentée dans cette thèse tend à confirmer cette hypothèse.

¹⁷ Lorsque le nombre de participants à l'enchère est très élevé, la rente cédée au vainqueur devient nulle et le bien-être de premier rang peut alors être atteint