

UNIVERSITE DE NEUCHATEL
FACULTE DE DROIT ET DES SCIENCES ECONOMIQUES

**La fourniture quasi-privée
des services publics locaux
en réseau**

THESE

Présentée à la Faculté de droit et des sciences économiques pour obtenir le
grade de docteur ès sciences économiques

par

Laurent Christe

Imprimerie de l'Évêque, Neuchâtel

1992

Monsieur Laurent Christe est autorisé à imprimer sa thèse de doctorat ès sciences économiques intitulée "La fourniture quasi-privée des services publics locaux en réseau".

Il assume seul la responsabilité des opinions énoncées.

Neuchâtel, le 3 mars 1992

Le Doyen
de la Faculté de droit
et des sciences économiques

Henri-Robert Schüpbach

AVANT PROPOS

Cette recherche n'aurait pu être réalisée sans les encouragements et la collaboration de plusieurs personnes. J'aimerais leur exprimer ici toute ma gratitude et ma reconnaissance.

Je tiens à remercier le Professeur Claude Jeanrenaud, directeur de thèse. Ses encouragements et conseils m'ont été extrêmement précieux et m'ont permis d'acquérir la rigueur indispensable à toute recherche scientifique.

Grâce à son soutien, j'ai pu effectuer en 1990 un séjour à l'University of California, Los Angeles avec l'aide d'une bourse de relève accordé par le Fonds national de la recherche scientifique. A ce propos, je remercie également le Professeur Werner Z. Hirsch qui m'a chaleureusement accueilli à UCLA et m'a permis de compléter ma formation au sein d'un environnement de très haut niveau scientifique.

J'aimerais encore remercier Nathalie Schwab et Jean-Paul Debrot pour leurs remarques et critiques constructives.

Enfin, je tiens à exprimer toute ma reconnaissance à ma famille pour ses encouragements ainsi qu'à Sandrine pour les travaux dactylographiques, son soutien moral et son immense patience tout au long de cette étude.

Cette étude a pu être publiée grâce au soutien financier du siège de Neuchâtel de la société KPMG Fides.

TABLE DES MATIÈRES

	page
1. Introduction	1
Première partie: Revue de la littérature	3
2. La fourniture des services collectifs	3
2.1. Justification de l'intervention publique: les échecs du marché	3
2.1.1. Externalités	4
2.1.2. Biens collectifs	5
2.1.3. Rendements d'échelle croissants	6
2.1.4. Risque et incertitude	8
2.1.5. Équité distributive	8
2.2. Échecs de l'intervention publique	9
2.2.1. Caractéristiques de la demande de biens collectifs	9
2.2.2. Caractéristiques de l'offre de biens collectifs	10
2.2.3. Typologie des échecs de l'intervention publique	11
3. Efficacité de la fourniture des services collectifs	13
3.1. La productivité et sa mesure	13
3.2. Fonction et frontière de production	14
3.3. L'efficacité productive et sa mesure	15
3.4. Estimation de la frontière de production ou de coût	18
3.4.1. Frontière non statistique et non paramétrique	19
3.4.2. Estimation non statistique d'une frontière paramétrique déterministe	22
3.4.3. Estimation statistique d'une frontière non paramétrique	23
3.4.4. Estimation statistique d'une frontière paramétrique déterministe	23
3.4.5. Estimation d'une frontière paramétrique stochastique	24
3.4.6. Critique de la littérature	27
4. Efficacité des services publics en réseau: études empiriques et comparatives	29
Deuxième partie: Fourniture efficace des services collectifs organisés en réseau	33
5. Les modes de fourniture des services collectifs	33

5.1. Introduction	33
5.2. Approche théorique	35
5.3. Modes de fourniture	37
5.3.1. Gestion directe	38
5.3.2. Coopération entre collectivités publiques	38
5.3.3. Sous-traitance	39
5.3.4. Concession	40
5.3.5. Subvention	41
5.3.6. Remise de bons	42
5.3.7. Volontariat	43
5.3.8. Marché	43
6. Evaluation des modes de fourniture des services collectifs	45
6.1. Nature des services et choix du mode de fourniture	45
6.2. Efficacité des modes de fourniture	46
6.2.1. Gestion directe	47
6.2.2. Coopération entre collectivités publiques	48
6.2.3. Sous-traitance	49
6.2.3. Concession	50
6.2.5. Subvention	52
6.2.6. Remise de bons	53
6.2.7. Volontariat	54
6.2.8. Marché	54
7. Services collectifs organisés en réseau: problématique	57
7.1. Caractéristiques des services en réseau	57
7.2. Notions de monopole naturel	58
7.3. Réglementation du monopole naturel: les raisons	61
7.4. Justification de la réglementation	64
7.5. Objectifs de l'intervention publique	68
7.6. Effets de la réduction des coûts de production sur le bien-être collectif	70
7.7. Les différentes réglementations des entreprises en réseau	73
7.8. Réglementation par le taux de rendement du capital	73
7.9. Réglementation par les prix	75
7.10. Réglementation par la qualité	77
8. Alternatives à la réglementation	81
8.1. Propriété publique ou gestion directe	81
8.1.1. Objectifs	81
8.1.2. Raisons	82

8.2. La concession	82
8.2.1. Soumission de la concession	83
8.2.2. Nombre d'offres	84
8.2.3. Efficacité des prix	84
8.2.4. Evaluation des soumissions	85
8.2.5. Application du contrat	85
8.2.6. Durée du contrat	86

**Troisième partie: Efficacité des services collectifs: Application
aux entreprises suisses de transports urbains** 89

9. Introduction	89
10. Modélisation de la structure de production	91
10.1. Dualité entre fonctions de coût et de production	91
10.2. La fonction de coût translog	92
11. Le modèle économétrique	99
11.1. Estimation de l'efficacité productive	102
11.2. Estimation d'une frontière déterministe	103
11.3. Estimation d'une frontière stochastique	103
12. Procédures d'estimation	105
12.1. Choix des variables de qualité et d'environnement	105
12.2. Le modèle	106
13. Description des données	109
13.1. L'output	110
13.2. Les inputs	110
13.3. Les coûts	111
13.4. Les variables de qualité et d'environnement	113
14. Les résultats	115
14.1. Productivité partielle des inputs	115
14.2. Mesure de l'efficacité productive	118
14.3. Estimation de l'efficacité productive à partir d'une frontière de coût déterministe	121
14.4. Estimation de l'efficacité productive à partir d'une frontière de coût stochastique	124
14.5. Elasticités de substitution et rendements d'échelle	134
15. Performance des entreprises de transports urbains en fonction du mode de fourniture	139
16. Conclusions	145

Références	149
-------------------	------------

1. Introduction

La mesure de l'efficacité des entreprises en réseau constitue le thème principal de ce travail. Les services tels que la distribution d'eau et d'électricité, les transports publics urbains, le téléphone sont disponibles grâce à des réseaux. Or l'existence d'un réseau conduit souvent le marché à adopter une structure de monopole que l'on appelle monopole naturel ou technique. Contrairement à une situation de concurrence parfaite, le monopole ne permet pas d'atteindre un optimum économique et cela pour deux raisons. Premièrement, le monopoleur tend à profiter de sa situation pour fournir aux usagers une quantité insuffisante de services et à vendre ces derniers à un prix trop élevé. Deuxièmement, l'absence de concurrence n'incite pas le producteur à une gestion efficace de son entreprise ce qui mène à une situation d'inefficacité productive. Ces raisons ont poussé l'Etat à intervenir afin de corriger les comportements et de tenter de maximiser le bien-être collectif. Nous nous sommes intéressés aux différents moyens d'intervention à la disposition des autorités publiques et à leur impact sur l'efficacité dans la production des services. Il s'agit alors de trouver quelles sont les possibilités d'accroître la concurrence en choisissant le mode de fourniture le mieux adapté aux caractéristiques du service. Nous verrons qu'il n'y a pas uniquement une alternative entre production publique et production privée mais qu'il existe entre ces deux extrêmes de nombreuses possibilités. Nous nous sommes alors penchés sur la mesure de l'efficacité des entreprises de transports publics urbains afin de déterminer si une situation de concurrence amène effectivement le producteur à agir efficacement.

La première partie de cette étude est consacrée d'une part à l'analyse des modes de fourniture des services publics dans la littérature économique et, d'autre part, aux récents développements des méthodes de mesure de la performance dans la production de biens et services. Un survol des précédentes études empiriques permettant de comparer la performance du secteur public à celle du secteur privé est également présenté.

Dans la seconde partie, nous avons examiné plus attentivement les modes de fournitures des services publics, en particulier les conditions et les circonstances dans lesquelles ils peuvent être appliqués. Nous verrons que leur domaine d'application est très vaste (chapitre 5) et examinerons la relation entre la manière dont la prestation est offerte et son efficacité aussi bien du côté de la production que de celui de la consommation (chapitre 6). Le chapitre 7 traite des

services en réseau proprement dit et des monopoles naturels qui leur sont associés. Les différents moyens d'intervention des autorités publiques locales sont considérés ainsi que leur impact sur le comportement des producteurs. Cette analyse porte, d'une part, sur les différents instruments permettant de réglementer une entreprise en réseau et, d'autre part, sur les alternatives à la réglementation dont dispose l'Etat (chapitre 8).

La troisième partie est consacrée entièrement à la mesure de l'efficacité des entreprises suisses de transports publics urbains. Nous avons effectué une mesure de l'efficacité productive à partir d'une modélisation de la structure de production en nous basant sur une théorie récente, celle de la dualité (chapitre 10). En plus d'une évaluation de la productivité partielle des facteurs de production, nous avons retenu deux méthodes de mesure de l'efficacité, toutes deux obtenues à partir d'une fonction de coût. La première s'appuie sur une méthode déterministe, la seconde sur un modèle stochastique; toutes deux permettent de construire une frontière de coût, c'est-à-dire celle qui détermine l'ensemble des productions possibles. L'intérêt d'utiliser deux mesures différentes de l'efficacité réside dans la possibilité de tester la cohérence des résultats. Ces derniers ont d'ailleurs montré une concordance révélatrice de la convergence des deux méthodes (chapitre 14).

L'utilisation des fonctions de coût permet en outre d'obtenir un certain nombre de caractéristiques de la structure de production comme par exemple les rendements d'échelle. Ceux-ci constituent une information extrêmement précieuse dans le contexte de l'analyse de la production de services en réseaux. Les résultats révèlent que les transports urbains bénéficient de rendements d'échelles croissants quel que soit leur niveau de production.

Une comparaison des performances des entreprises de transports urbains en fonction de leur mode de fourniture a été finalement réalisée afin d'étudier l'efficacité relative de chaque mode (chapitre 15). Les résultats obtenus montrent que la performance n'est pas liée au mode de fourniture.

PREMIERE PARTIE
REVUE DE LA LITTERATURE

2. La fourniture des services collectifs

La théorie économique a toujours admis la nécessité d'une intervention de l'Etat dans la fourniture de services tels que la justice, la protection intérieure et extérieure du pays ou encore l'éducation scolaire de base. Ces prestations ont la caractéristique de pouvoir être consommées simultanément par plusieurs individus, c'est une des raisons pour lesquelles on les appelle services collectifs. Nous savons néanmoins que le secteur public comprend beaucoup d'autres activités. Il convient par conséquent d'examiner, d'une part, les raisons qui ont amené l'Etat à produire certains services et, d'autre part, les problèmes générés par cette intervention.

2.1. Justification de l'intervention publique: les échecs du marché

On juge habituellement les succès et échecs des mécanismes de marché grâce à deux critères:

- l'efficacité;
- l'équité distributive.

Un marché est défini comme efficient si la production du même niveau de prestations ne peut être obtenu à un coût plus faible ou, autrement dit, si des prestations plus importantes ne peuvent être générées à un coût identique.

Si les mécanismes du marché permettent d'accomplir une tâche à un coût inférieur à celui d'un arrangement institutionnel alors le marché est relativement efficace. Dans le cas contraire, on dit que le marché est relativement inefficace.

Le second critère permet de juger les résultats des mécanismes de marché dépasse largement les frontières de la micro-économie. Il est surtout lié à la formulation, l'évaluation et l'application des différentes politiques publiques (Wolf 1986). De plus, les problèmes de redistribution sont souvent soumis à des

jugements subjectifs au sujet des succès et échecs présumés des mécanismes de marché. Dans cette perspective, l'évaluation de l'équité distributive est un problème difficile à résoudre. En effet, selon l'interprétation donnée à l'équité (meilleure répartition du revenu, équité *verticale* ou *horizontale*, etc.), l'objectif visé peut varier énormément.

La plupart des auteurs associent les échecs du marché à une allocation inefficace des ressources; ils excluent par conséquent les problèmes de distribution sauf lorsque ces derniers affectent l'efficacité. Les causes d'échecs du marché sont diverses, nous commenceront par celles qui déterminent l'efficacité. Elles dépendent d'un certain nombre de considérations normatives qui permettent d'établir si l'allocation des ressources est optimale au sens de Pareto.

2.1.1. Externalités

Selon Buchanan et Stubblebine (1962), une externalité est présente lorsque l'utilité d'un individu A dépend non seulement des activités (X_1, X_2, \dots, X_n) qui sont exclusivement sous son contrôle mais également de l'activité Y_j qui est hors de son contrôle. Un individu peut donc consommer les biens X_1, \dots, X_n et être affecté par l'activité Y_j générée par quelqu'un d'autre. Cette définition souligne qu'une externalité implique non seulement une interdépendance - l'action d'un agent influence la situation d'un autre agent - mais aussi une absence de prix, l'échange se produisant directement sans l'intermédiaire des mécanismes du marché, c'est-à-dire sans compensation.

Les externalités peuvent avoir un impact positif (économie externe) ou négatif (déséconomie externe). Elles peuvent avoir lieu au sein du processus de production (une entreprise améliore la formation de son personnel dont une partie ira travailler pour d'autres firmes), entre producteurs et consommateurs (la pollution de l'air par les cheminées de l'entreprise) ou entre consommateurs (le fumeur de cigarettes qui impose une déséconomie externe aux non-fumeurs).

Si l'internalisation de l'externalité ne peut être réalisée par négociation entre les agents concernés, c'est à l'Etat qu'il appartient d'intervenir. Les principales stratégies utilisées par les autorités publiques sont les suivantes (Cullis et Jones 1987):

- imposer le bien générant l'externalité;
- subventionner le producteur de l'externalité ou l'individu qui la subit;

- réglementer l'activité, en fixant une norme à ne pas dépasser par exemple;
- rassembler l'agent générant l'externalité et celui qui la supporte au sein d'une seule unité de décision.

Ces différentes alternatives donnent une idée des possibilités d'intervention de l'Etat, il faut toutefois garder à l'esprit que ces dernières ne sont pas sans coût. Arrow (1971) estime que les coûts de transaction supportés par l'Etat dans sa tentative de résoudre les échecs du marché doivent être comparés aux coûts engendrés par les externalités. Cela signifie donc que l'intervention de la collectivité ne devrait pas être automatique.

2.1.2. Biens collectifs

C'est principalement à partir des contributions de Samuelson (1954, 1955) que les fondements théoriques des biens collectifs ont été véritablement mis en place. Les biens et services produits et consommés ont été classés entre deux pôles où les extrêmes sont représentés par les biens collectifs purs et les biens privés purs. Entre ces deux extrêmes, on trouve les biens et services mixtes. Musgrave et Musgrave (1980) définissent deux caractéristiques principales permettant de classer les services dans chacune des catégories énoncées ci-dessus :

- a) **la possibilité d'exclusion.** Elle s'applique lorsque le fournisseur d'un bien peut exclure de la jouissance de celui-ci tout individu qui n'accepterait pas d'en payer le prix. La possibilité d'exclusion permet donc le transfert des droits de propriétés d'un bien contre un prix.
- b) **la rivalité dans la consommation.** Elle désigne une situation dans laquelle la consommation d'un bien par un individu empêche tout autre personne de consommer ce même bien.

Les biens collectifs purs sont ceux dont la consommation n'est pas rivale et l'exclusion impossible ou très difficile. La seconde caractéristique implique que les consommateurs ne seront pas incités à révéler le montant qu'ils seraient disposés à payer pour obtenir une unité du bien. Les économistes parlent à ce propos du problème du passager clandestin (*free rider*). Lorsqu'on rencontre ce type de comportement chez les consommateurs, le marché est voué à l'échec: les préférences n'étant pas révélées, il n'y a aucune incitation à produire le bien. Ici

encore, l'intervention de l'Etat s'avère nécessaire car il peut instaurer un impôt afin de financer la production du bien en question.

2.1.3. Rendements d'échelle croissants

Lorsqu'il y a des économies d'échelle dans la production, le coût moyen d'un bien est décroissant si la quantité produite augmente sur un certain intervalle de l'échelle de production. Les économies d'échelle sont très souvent liées à un coût fixe important et/ou à une offre totalement ou partiellement indivisible comme c'est le cas pour les transports en commun, les télécommunications ou la distribution d'eau et d'électricité. A l'équilibre, c'est-à-dire au point où la courbe de demande coupe le coût marginal, ce dernier se situe alors au-dessous du coût moyen. La fixation du prix égal au coût marginal - condition pour atteindre un optimum social - entraîne des pertes pour le producteur, ses recettes étant plus faibles que ses coûts dans l'intervalle de production réalisable. Aucune entreprise n'est incitée à produire ce bien en quantités requises par la maximisation du bien-être collectif. Ce bien devra par conséquent être fourni au travers d'une intervention publique: soit une production publique, soit une production par une entreprise privée réglementée ou subventionnée.

La figure 2.1. illustre plus clairement ce problème. S'il n'y a qu'un seul producteur sur le marché, le prix se fixera en p_2 et l'output sera alors au niveau q_2 , point auquel la recette marginale (RMa) coupe le coût marginal (CMA). Dans cette situation, le profit sera maximisé. Il y aura par contre une perte allocative égale à la surface abd puisque les consommateurs paieront plus que le coût marginal. Même si l'équilibre se situait en q_1 avec un prix p_1 , niveau pour lequel les producteurs ne feraient pas de profit, l'optimum ne serait pas atteint car l'avantage marginal est supérieur au coût marginal.

L'intervention publique peut prendre différentes formes:

- le bien est vendu au coût marginal par une entreprise en mains publiques;
- une réglementation du marché est imposée;
- les prix sont déterminés par l'Etat;
- le taux de rendement du capital est limité par un plafond;

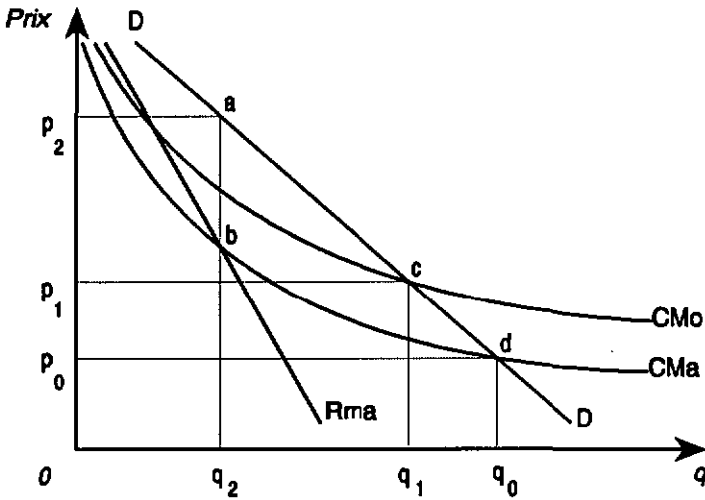


Figure 2.1 Economies d'échelle dans la production

De telles interventions permettent rarement d'atteindre une situation efficace au sens de Pareto bien qu'elles cherchent à s'en approcher.

Si le producteur est autorisé à fournir le bien de manière à couvrir ses coûts sans faire de profits, il fixera la quantité produite à un niveau où la courbe de demande croise la courbe du coût moyen. L'équilibre obtenu n'est alors plus optimal du point de vue du bien-être social.

En l'absence de toute intervention publique, le producteur choisira de maximiser son profit. S'il se trouve dans une situation de monopole ce qui est fréquent lorsque l'offre est totalement ou partiellement indivisible, il choisira de vendre une quantité telle que son offre se situe à l'intersection de son coût marginal et sa recette marginale. Cette solution est la moins satisfaisante pour la collectivité car elle engendre la perte la plus importante de bien-être social. Dans ces deux derniers cas, l'intervention publique devient souhaitable si la société désire maximiser le bien-être social ou du moins s'en approcher. Rappelons toutefois que le coût de cette intervention doit être au moins compensé par l'accroissement du bien-être sinon celle-ci ne se justifie pas.

2.1.4. *Risque et incertitude*

La théorie économique est très souvent basée sur des modèles dans lesquels producteurs et consommateurs sont supposés connaître avec certitude les prix actuels et futurs des biens et des facteurs. En réalité, l'avenir est incertain et, par conséquent, les prix futurs également. Arrow (1970) a démontré que, dans un monde incertain, il y a un ensemble de circonstances dans lesquelles le mécanisme de prix concurrentiels produit une allocation optimale des ressources. Cependant, bien que sa proposition permette aux individus d'éviter la charge d'une couverture du risque (en achetant une assurance sur la vie par exemple), de tels marchés sont incomplets en raison des coûts élevés d'organisation que ces derniers génèrent par rapport aux avantages assurés. Par conséquent, le risque ne pouvant être totalement assuré sur un marché, une intervention publique se révèle nécessaire pour s'approcher d'un optimum social.

2.1.5. *Equité distributive*

Nous savons qu'un optimum de Pareto n'est pas unique. En supposant qu'il soit atteint, il s'agit de déterminer quelle situation optimale est préférable puisqu'à chacune d'elle correspond une distribution différente de l'utilité parmi les membres de la société (Boadway 1979). Même un marché fonctionnant parfaitement peut s'écarter des préférences de la société en ce qui concerne la distribution du revenu et de la richesse. Il convient donc d'examiner l'efficacité relative de différentes mesures de redistribution. D'un certain point de vue, il est adéquat de considérer l'inéquité comme un échec du marché. Dans cette perspective, les mécanismes du marché produisent une redistribution du revenu socialement indésirable.

2.2 *Echecs de l'intervention publique*

Nous avons vu que les échecs du marché peuvent conduire à une production insuffisante de services. A l'opposé, le résultat des échecs de l'intervention de l'Etat est une production trop importante d'activités. La littérature économique traitant de ce problème se base sur le point fondamental suivant: l'intervention publique dans une économie génère ses propres coûts qui peuvent surpasser les coûts résultant d'un marché imparfait.

L'analyse des échecs de l'intervention de l'Etat (*government failures* ou *nonmarket failures*) se base essentiellement sur les travaux de Wolf (1986). Les

échecs s'expliquent à la fois par les caractéristiques de l'offre et de la demande de biens publics. Nous examinerons celles-ci et présenteront ensuite la typologie des échecs selon Wolf .

2.2.1. Caractéristiques de la demande de biens collectifs

La demande de biens collectifs se distingue de celle des biens privés par un certain nombre de caractéristiques:

- a) **La prise de conscience des imperfections dans le fonctionnement du marché.** Au fur et à mesure du développement des activités économiques, les échecs du marché sont devenus plus fréquents et plus importants. Il en résulte un éloignement d'un équilibre socialement optimal (pollution croissante, pouvoir monopolistique des entreprises, effets de congestion dus à l'augmentation de la densité de la population, etc.).
- b) **L'émergence des organisations politiques.** A partir des années 60, les mouvements politiques et les groupes de pression (mouvement féministes, écologistes, groupes de consommateurs, etc.) ont favorisé la prise de conscience des échecs du marché. Ils ont, de plus, exercé une pression sur les autorités publiques afin que ces dernières interviennent pour remédier à ces échecs.
- c) **Le court horizon (ou myopie) des politiciens.** Etant donné le court laps de temps séparant deux législatures (souvent quatre ans), un décalage important apparaît entre le court horizon des politiciens et le temps relativement long nécessaire à la compréhension d'un problème particulier dans le but d'appliquer des remèdes pour le résoudre. Par conséquent, les coûts et avantages futurs tendent à être ignorés alors que les coûts et avantages actuels sont amplifiés. Feldstein (1980) appelle cet effet *the inherent myopia of the political process*.
- d) **Le décalage entre coûts et avantages.** Une distorsion de la demande de biens collectifs provient de la "rupture" entre ceux qui bénéficient des avantages des programmes et ceux qui en paient le coût. Ce phénomène survient lorsque les avantages d'une action publique sont concentrés au sein d'un groupe particulier tandis que les coûts sont dispersés parmi tous les contribuables ou consommateurs. Les bénéficiaires du programme sont donc beaucoup plus fortement incités à soutenir celui-ci que l'ensemble des

contribuables à s'y opposer. Il en résulte un programme ou une réglementation inefficace ou inéquitable.

2.2.2. Caractéristiques de l'offre de biens collectifs

L'offre de biens collectifs se distingue par deux particularités très importantes. La première concerne la difficulté de définir et de mesurer la production. En effet, les prestations publiques sont souvent malaisées à définir et extrêmement difficiles à mesurer aussi bien en termes de quantité qu'en termes de qualité. Ces biens sont fréquemment des produits intermédiaires qui sont au mieux des approximations de l'output final désiré. Par exemple, les restrictions sur la distribution de l'alcool par la Régie fédérale des alcools font partie d'un but plus vaste qui est celui de la préservation de la santé publique. La qualité des services collectifs n'est pas facile à assurer en raison du manque d'information à son sujet. Dans le cas de biens privés, cette information est transmise aux producteurs par le choix des consommateurs. Il est par exemple assez ardu de déterminer si la qualité des programmes d'enseignement est meilleure actuellement qu'il y a cinq ans.

La seconde caractéristique porte sur le monopole de la production. Les biens et services fournis par l'Etat sont produits habituellement par un seul office qui a, de ce fait, un monopole souvent basé sur une loi. Cette absence de concurrence empêche toute comparaison et rend d'autant plus pénible l'évaluation de la qualité du service produit.

2.2.3. Typologie des échecs de l'intervention publique

A) ABSENCE DE LIAISON ENTRE COÛTS ET RECETTES

Sur le marché, le lien entre le coût de production d'un bien et les recettes nécessaires à le couvrir est fourni par le prix. Dans le cas des activités de l'Etat, ce lien est absent puisque les recettes permettant d'assurer la production du bien proviennent de sources autres que les prix. De plus, lorsqu'une politique publique prend la forme d'une réglementation, d'un transfert ou d'une production directe de biens, il y a tendance à ce que cette activité présente des coûts inutiles, c'est-à-dire une inefficacité-X. Si des possibilités technologiques permettaient de réduire les coûts, d'accroître la productivité ou de réaliser des économies d'échelle, ces opportunités risquent d'être tout simplement ignorées

ou moins bien exploitées que dans le cas d'une activité soumise aux lois du marché.

Enfin, les objectifs de certaines politiques publiques peuvent être inaccessibles. Wolf (1986) cite l'exemple d'une politique sociale qui financerait un projet de recherche irréaliste visant à transformer des individus, avec un Q.I. inférieur à 70, en journalistes.

B) INTERNALITES ET BUTS ORGANISATIONNELS

Pour mener à bien leurs activités, les offices publics ont besoin de se fixer un certain nombre de normes puisque les indicateurs directs de leurs performances sont absents, indicateurs qui sont évidemment disponibles sur le marché. Wolf (1986) appelle ces normes des *internalités*, c'est-à-dire "the goals that apply with non market organization to guide, regulate, and evaluate agency performances and the performance of agency personnel".

Etant donné la difficulté de la mesure de l'output et l'absence de feed-back de la part des consommateurs, le secteur public développe des internalités n'ayant qu'un lien très vague avec les objectifs collectifs qu'il est sensé poursuivre. Autrement dit, les internalités deviennent les éléments des fonctions d'utilité que les fonctionnaires publics cherchent à maximiser.

L'existence d'internalités signifie donc que les coûts et avantages privés dominent les jugements des décideurs publics. Elles ont donc pour effet d'augmenter l'offre du secteur public. L'objectif de croissance du budget de l'office est une des internalités les plus souvent rencontrées car les incitations à proposer de nouvelles dépenses sont plus fortes que celles visant à les réduire (augmentation des responsabilités et donc des salaires, accroissement du pouvoir de l'office, etc.).

C) EXTERNALITES DERIVEES

L'intervention de l'Etat, en vue de corriger les échecs du marché, peut produire des effets secondaires non prévus dans d'autres domaines que celui visé. En raison du caractère imprévisible de ces externalités dérivées, l'office responsable de leur création n'en tient pas compte dans ses prises de décisions. Ce genre d'effets secondaires est difficile à identifier parce qu'ils agissent avec un certain retard par rapport à l'action entreprise mais également à cause de politiques publiques mal définies. L'exemple le plus frappant concerne la réglementation des monopoles: la réglementation du taux de rendement du capital induit une

substitution inefficace du capital au travail car c'est le meilleur moyen pour l'entreprise d'accroître son profit¹. Le coût social d'une telle externalité peut même excéder celui que la réglementation tentait de réduire.

D) INEQUITE DISTRIBUTIVE

L'activité de l'Etat va parfois dans le sens d'une plus grande inégalité. Cette activité confère à certaines personnes un pouvoir qui n'est pas toujours exercé avec compétence et honnêteté. Bien que ce pouvoir soit généralement soumis à des contrôles et dépende de lois, de procédures administratives ou d'institutions politiques ou sociales, des inégalités peuvent survenir. Cet arbitraire provient souvent de la complexité des lois ce qui rend la réalisation de l'équité horizontale et verticale très difficile avec les instruments dont dispose l'Etat.

¹ La littérature économique se réfère à cet effet sous l'appellation *Averch-Johnson effect*. Nous verrons ce problème de façon plus détaillée dans le chapitre 7.

3. Efficacité de la fourniture des services collectifs

Au cours de ses développements, l'analyse économique de la production a utilisé différents concepts afin de décrire la performance des entreprises. Les notions de productivité, de rentabilité, de progrès technique et d'efficacité ont été largement utilisées. Nous allons examiner ces différentes approches en commençant par la plus simple et en terminant par les développements les plus récents.

3.1. La productivité et sa mesure

La notion de productivité, c'est-à-dire l'aptitude des facteurs de production à réaliser celle-ci est très souvent employée en raison de sa simplicité de mise en oeuvre et de l'interprétation rapide qui en résulte. Lorsqu'un seul facteur de production (travail, capital ou autre) est pris en considération, on parle de productivité partielle. Celle-ci se mesure par le rapport du volume de la production (mesurée en quantité ou en valeur) à celui du facteur utilisé. L'impact de la productivité d'un seul input sur la performance d'une entreprise est cependant difficile à établir en raison de l'intervention de plusieurs facteurs dans le processus productif. Il convient alors de considérer simultanément l'ensemble des facteurs de production si l'on veut pouvoir porter un diagnostic sur le comportement d'une entreprise ou d'une branche économique. Il s'agit donc d'obtenir un indicateur de la productivité globale des inputs. Se pose par conséquent le problème de leur agrégation en une seule unité et souvent celui de l'agrégation des outputs dans les cas fréquents où le processus productif analysé génère plusieurs biens différents.

L'approche des *nombres indices* cherche à résoudre cette difficulté en se basant sur la mesure de la *productivité totale des facteurs* (PTF). Cette méthode de mesure de la productivité a été utilisée par Denison (1974) et Kendrick (1973) dans leurs travaux sur l'évolution de la productivité globale de l'économie américaine. Les indices de la méthode PTF consistent en un rapport de deux indices distincts, l'un pour les outputs et l'autre représentant les inputs¹. L'indice de l'output peut être une simple mesure non pondérée d'outputs homogènes (kWh d'électricité produite) ou une mesure pondérée d'outputs hétérogènes. Au

¹ L'agrégation séparée des outputs et des inputs suppose un progrès technique neutre et une fonction de transformation séparable en outputs et en inputs (Cowing et Stevenson 1981).

niveau de l'unité de production, la mesure des inputs doit inclure tout ceux qui sont utilisés dans le processus de production (capital, travail, matériel, énergie, etc) et doit considérer toutes les phases du processus de production (génération, transmission et distribution de l'électricité par exemple). La mesure de la performance grâce à la méthode PTF est sensiblement améliorée par rapport à celle obtenue à partir de la productivité partielle des inputs, car elle permet de tenir compte de l'ensemble des facteurs de production. Des difficultés subsistent toutefois pour distinguer les effets des rendements d'échelle et de la substitution des facteurs sur le progrès technique.

3.2. *Fonction et frontière de production*

L'utilisation du concept de fonction de production permet d'entrevoir la performance d'une entreprise de manière plus globale. Une fonction de production est généralement définie comme la relation existant entre les quantités d'inputs utilisées et les quantités d'outputs réalisées. Cette relation peut s'écrire sous forme mathématique comme une fonction

$$Y(y,x) = 0, \quad y \geq 0, \quad x \geq 0, \quad (3.1)$$

où x est un vecteur d'inputs non négatifs alors que y est un vecteur d'outputs également non négatifs. Dans le cas d'un seul et unique output, il est plus aisé de représenter l'expression (3.1) sous la forme

$$y = f(x) \quad (3.2)$$

où pour n'importe quelle combinaison d'inputs correspond un niveau unique d'output. Selon cette expression, nous supposons que l'entrepreneur choisit toujours le niveau d'output disponible le plus élevé possible. La fonction de production définie ci-dessus exclut donc toute possibilité d'inefficacité technique et se confond par conséquent avec la frontière de production délimitant l'espace de l'output techniquement réalisable pour l'entreprise étant donné un vecteur d'inputs (Chambers 1988). Elle spécifie ainsi les quantités maximales d'outputs accessibles pour tout niveau d'input et à l'inverse, pour tout niveau d'output, les quantités minimales d'inputs nécessaires à son obtention.

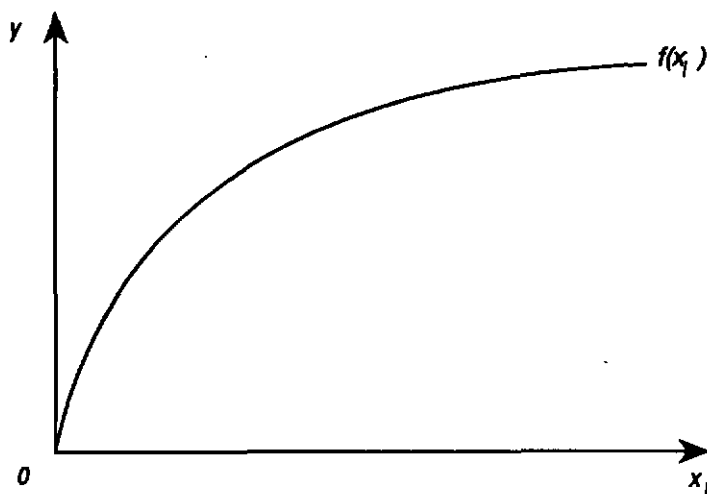


Figure 3.1 Frontière de production

Sur la figure 3.1 où l'output est représenté en fonction d'un seul input, l'aire comprise entre la courbe $f(x_i)$ et l'axe horizontal correspond à l'ensemble des productions techniquement possibles. La fonction $f(x_i)$, en délimitant l'espace supérieur, correspond donc à la frontière de production.

On suppose encore que cette fonction de production est monotone croissante, quasi-concave, continue, différentiable et comprend le vecteur nul. Toutefois plusieurs auteurs ne s'appuient que sur un nombre limité de ces propriétés.

3.3. L'efficacité productive et sa mesure

De la définition précédente de la fonction de production, il est facile d'en déduire l'idée d'efficacité productive. A partir de l'ensemble des productions possibles (cf. figure 3.1), toute entreprise ne se situant pas sur la frontière est considérée comme inefficace, c'est-à-dire qu'elle pourrait fournir un output supérieur étant donné les quantités de facteurs de production utilisées. Ainsi, une observation est déclarée soit efficace, soit inefficace. Si nous reprenons l'expression (3.2), nous pouvons maintenant exprimer les deux situations sous la forme

$$y \leq f(x) \quad (3.3)$$

où l'égalité stricte représente l'ensemble des points efficaces se trouvant sur la frontière, alors que l'inégalité stricte correspond à l'ensemble des points inefficaces situés entre l'abscisse et la frontière, mais n'appartenant à celle-ci.

La détermination de l'efficacité productive peut également être obtenue à partir de la fonction de coût total minimum, puisque celle-ci est définie comme la relation qui lie à la quantité produite la valeur minimale des inputs permettant cette production (Malinvaud 1968). L'ensemble des observations inefficaces correspond alors à l'aire située au-dessus de la courbe du coût total. Etant donné que la fonction du coût total minimum est le dual de la fonction de production, l'efficacité obtenue à partir de la première implique toujours l'efficacité dérivée de la seconde. Cependant, le contraire n'est pas forcément vrai. En effet, la fonction de coût permet de considérer aussi bien l'inefficacité technique dans la production, c'est-à-dire une utilisation non optimale des inputs, que l'inefficacité allocative dans la production², soit une utilisation des inputs dans de mauvaises proportions par rapport à leur prix. La fonction de production ne permet que de mesurer l'efficacité technique puisque l'efficacité allocative nécessite la connaissance du prix des facteurs.

Afin de présenter le problème de la mesure de l'efficacité, supposons que la frontière de production soit connue. L'article de Farrell (1957) fait figure de pionnier dans la mesure de l'efficacité et puisqu'une grande partie de la littérature économique traitant de ce sujet s'y réfère, nous allons brièvement présenter sa démarche.

Considérons un seul output (y) produit à l'aide de deux inputs (x_1 et x_2); la frontière de production prend donc la forme

$$y = f(x_1, x_2) \quad (3.4)$$

Supposons encore que cette fonction soit caractérisée par des rendements d'échelle constants de façon à pouvoir l'écrire

$$I = f(x_1/y, x_2/y) \quad (3.5)$$

² Le concept d'efficacité allocative a le plus souvent un sens différent, voir Bös (1986).

Considérons une firme A utilisant (x_1^0, x_2^0) pour produire l'output y^0 . Sur la figure 3.2, A correspond au point $(x_1^0/y, x_2^0/y)$. L'isoquant UU' représente les différentes combinaisons des deux facteurs qu'une entreprise efficace pourrait utiliser pour produire l'output unitaire.

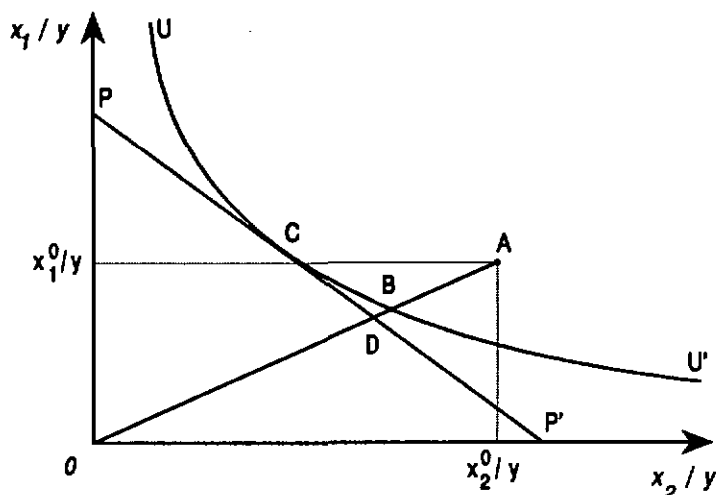


Figure 3.2 Mesure de l'efficacité productive

L'**efficacité technique** de l'entreprise A est donc OB/OA et mesure la proportion de (x_1^0, x_2^0) réellement nécessaire pour produire y^0 . Ce rapport prend la valeur unitaire pour les firmes efficaces et devient infiniment petit lorsque le montant des inputs par unité d'output devient infiniment grand.

Si PP' représente le rapport des prix des inputs, le point C constitue le point de minimisation du coût total. Puisque le coût en D est le même qu'en C , l'**efficacité allocative** de la firme A peut être définie par le rapport OD/OB . Ce dernier mesure la réduction possible du coût total par l'utilisation appropriée des proportions d'inputs en fonction de leur prix.

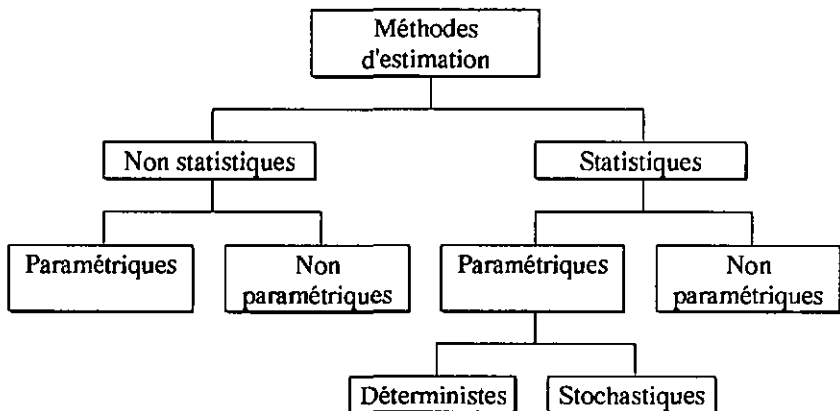
Finalement, l'efficacité totale de l'entreprise est définie par le rapport OD/OA qui est égal au produit des efficacités technique et allocative. L'inefficacité totale représente la réduction du coût total résultant d'un mouvement du point A observé vers le point C , celui qui minimise le coût total.

Si l'hypothèse des rendements d'échelle constants n'est plus imposée, la démarche de Farrell peut toujours être utilisée. L'isoquant UU' devient l'isoquant spécifique au niveau d'output y^0 (Farrell & Fieldhouse 1962, Forsund & Hjalmarsson 1974).

3.4. Estimation de la frontière de production ou de coût

Dans le chapitre précédent, nous avons supposé que la frontière de production était connue. Il s'agit maintenant de discuter des méthodes permettant d'estimer cette frontière. Celles-ci sont classées selon qu'elles se basent sur un modèle statistique ou non. Chacune de ces deux approches se divise encore en méthodes paramétriques et non paramétriques selon que l'on fait l'hypothèse ou non que la frontière peut être représentée par une fonction dont il faut estimer les paramètres. De plus, il faut également ajouter que si les méthodes non statistiques sont toutes déterministes, c'est-à-dire que la frontière ne comporte pas de terme d'erreur, les méthodes statistiques paramétriques peuvent être soit déterministes, soit stochastiques.

En suivant Thiry & Tulkens (1988), nous pouvons résumer ces méthodes d'estimation de la manière suivante:



Nous n'utiliserons, dans la partie empirique, que des méthodes statistiques paramétriques³. Cependant, nous présenterons brièvement les autres méthodes afin de mieux expliquer les raisons de notre choix.

3.4.1. *Frontière non statistique et non paramétrique*

Afin d'estimer l'efficacité technique et allocative dans la production, Farrell (1957) a utilisé une méthode en construisant une enveloppe convexe des rapports inputs-outputs observés, enveloppe obtenue par des techniques de programmation linéaire. La frontière est alors estimée avec les observations les plus efficaces qui représentent un sous-ensemble de l'échantillon. Cette procédure n'est basée sur aucun modèle explicite de la frontière, la seule condition étant que les observations ne peuvent se situer au-dessous de celle-ci, comme nous pouvons le constater sur la figure 3.3.

On remarque alors que ce sont les observations les plus efficaces qui constituent la frontière. L'efficacité technique est mesurée comme décrit précédemment par le rapport OB/OA où B représente une observation efficace hypothétique utilisant la même proportion d'inputs que A .

L'avantage de cette approche est qu'elle n'impose aucune forme fonctionnelle à la frontière. Ses principaux inconvénients résident d'une part, dans l'hypothèse restrictive des rendements d'échelle constants et d'autre part, dans l'estimation de la frontière à partir d'un sous-ensemble d'observations de l'échantillon, ce qui la rend particulièrement sensible aux observations extrêmes ainsi qu'aux erreurs de mesure.

La méthode proposée par Deprins, Simar et Tulkens (1984) constitue le prolongement des travaux de Farrell. Basée uniquement sur l'hypothèse de libre disposition des inputs et des outputs (appelée ci-après méthode FDH pour *Free Disposal of Hull*), cette approche n'impose pas la convexité de l'ensemble de production et permet de construire une frontière quelconque.

³ Notons que les termes déterministes et stochastiques s'appliquent uniquement à l'estimation de la frontière. Il est bien clair que toute méthode statistique est basée sur l'aléatoire.

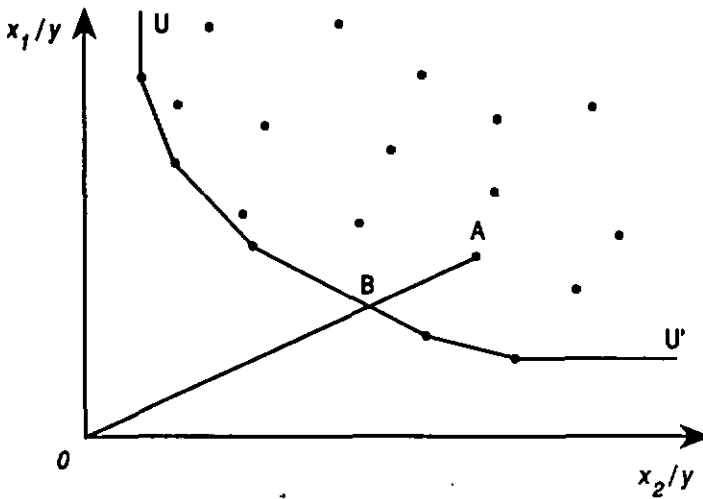


Figure 3.3 Mesure de l'efficacité selon Farrell

La méthode FDH consiste à déclarer inefficace toute observation qui est dominée par au moins une autre pour laquelle les quantités d'outputs produites sont supérieures et les quantités d'inputs utilisées sont inférieures. Les observations déclarées efficaces sont logiquement appelées non dominées. Dans le cas d'un seul output y et d'un seul input x , on distingue facilement sur la figure 3.4 les entreprises dominées, donc inefficaces, telles que E et F de celles qui sont efficaces, telles que A, B, C et D . L'hypothèse ci-dessus implique que la frontière de production ait une forme en escalier.

La mesure numérique de l'efficacité de la production est effectuée de deux façons (Thiry et Tulkens 1988):

- a) l'**inefficacité en output** (distance verticale sur le graphique) est estimée par le rapport entre l'output de l'observation en question et l'output le plus élevé des observations qui la domine en output;
- b) l'**inefficacité en input** (distance horizontale) est obtenue par le rapport dont le dénominateur est l'input de l'observation et le numérateur, le plus petit des inputs qui la dominant en input. L'observation G est ainsi déclarée inefficace en input puisque l'entreprise B utilise moins d'inputs pour une même production. Néanmoins, G est efficace en output.

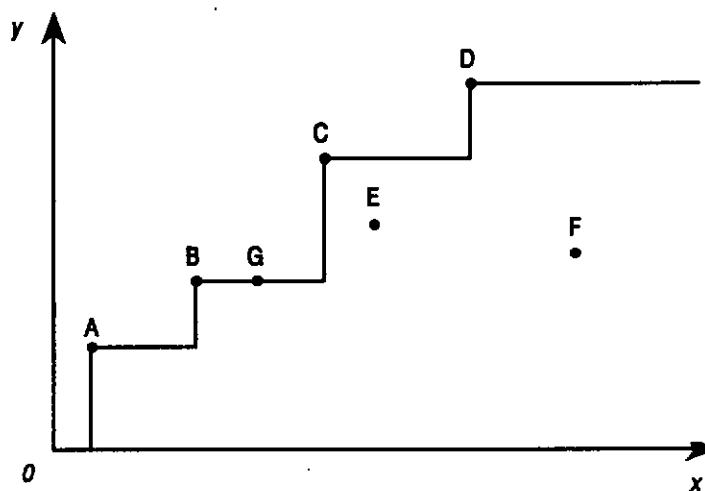


Figure 3.4 Mesure de l'efficacité selon la méthode FDH

L'avantage de cette approche réside dans sa très grande flexibilité pour déterminer la frontière de production ainsi que dans sa simplicité. Par contre, en plus de l'inconvénient de la méthode de Farrell concernant l'estimation de la frontière à partir d'un sous-ensemble d'observations, deux mesures de l'efficacité doivent être effectuées. Ce second inconvénient rend l'analyse problématique puisque certaines observations, bien que se situant sur la frontière, peuvent être déclarées efficaces en output et inefficace en input ou réciproquement. De plus, un autre problème apparaît lorsqu'on désire généraliser la méthode au cas de plusieurs inputs et outputs. Les conditions pour qu'une observation soit déclarée inefficace deviennent si exigeantes lorsque le nombre d'inputs et/ou d'outputs s'accroît, que les observations déclarées inefficaces deviennent extrêmement rares, voire inexistantes. Enfin, lorsqu'une observation se situe dans un intervalle de production où il n'y a aucune autre entreprise, elle sera automatiquement déclarée efficace, car aucune comparaison n'est alors possible, ce qui la rend efficace par défaut.

Une autre approche non paramétrique, la méthode DEA (*Data Envelopment Analysis*) utilise des techniques de recherche opérationnelle. Elle a été développée par Charnes, Cooper & Rhodes (1978) afin de mesurer l'efficacité technique. L'efficacité d'une entreprise particulière est donnée par le programme

$$\text{Maximiser } h_0 = \frac{\sum_{r=1}^s u_r y_{r0}}{\sum_{i=1}^m v_i x_{ij}} \quad (3.6)$$

$$\text{sous la contrainte: } \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1, \quad j=1, \dots, n$$

$$u_r, v_i > 0, \quad r = 1, \dots, s \quad i = 1, \dots, m$$

où les y_{rj} et les x_{ij} représentent respectivement les r outputs et i inputs observés de la j ème entreprise. Les u_r et v_i sont les variables de pondération positives qui doivent être déterminées en résolvant le programme ci-dessus. Une observation est efficace si et seulement si $h_0=1$ (voir aussi Chames, Cooper et Rhodes 1981, Färe et Hunsaker 1986).

3.4.2. Estimation non statistique d'une frontière paramétrique déterministe

Aigner et Chu (1968) ont été les premiers à utiliser une frontière paramétrique déterministe à l'aide d'une méthode non statistique.

Ils ont spécifié une frontière de production homogène de type Cobb-Douglas – ce pourrait être n'importe quel autre type de fonction – de telle sorte que toutes les observations se trouvent sur ou au-dessous de la frontière $f(x, \beta)$ où β représente les paramètres à estimer. La forme générale du modèle s'écrit donc en notation logarithmique

$$\ln y = \ln f(x, \beta) - u, \quad u \geq 0 \quad (3.7)$$

Pour une fonction de production de type Cobb-Douglas, cela donne

$$\ln y = \beta_0 + \sum_{i=1}^m \beta_i \ln x_i - u, \quad u \geq 0 \quad (3.8)$$

L'estimation du vecteur β est la solution du programme

$$\text{Minimiser } \sum_{k=1}^n |\ln y_k - \beta_0 - \sum_{i=1}^m \beta_i \ln x_{ik}| \quad (3.9)$$

$$\text{sous la contrainte: } \ln y_k \leq \beta_0 + \sum_{i=1}^m \beta_i \ln x_{ik} \quad \text{pour tout } k$$

où $k=1, \dots, n$ représente les n observations (entreprises).

Aigner et Chu proposent d'effectuer cette estimation soit par une technique de programmation linéaire en minimisant la somme des valeurs absolues des résidus u , soit par programmation quadratique en minimisant la somme des carrés des résidus u . Bien que Aigner et Chu ne l'aient pas réalisée, l'efficacité technique des observations peut être obtenue directement à partir des résidus puisque u représente l'inefficacité technique.

3.4.3. Estimation statistique d'une frontière non paramétrique

Avec les méthodes non paramétriques précédentes, il subsiste le problème de la présence d'une perturbation statistique. Les données sont supposées être exemptes de toute erreur de mesure et la liste des inputs est exhaustive.

Une tentative de traitement statistique d'un modèle non paramétrique a été réalisée par Banker et Maindiratta (1985). Ils imposent uniquement à la frontière d'être monotone et concave par rapport aux inputs, ce qui nécessite un certain nombre de contraintes sous formes d'inégalités. Bien que les propriétés statistiques des estimateurs soient incertaines, cette stratégie d'un traitement statistique de modèles non paramétriques semble prometteuse car elle permet de combiner le traitement des perturbations statistiques avec la flexibilité des méthodes non paramétriques.

3.4.4. Estimation statistique d'une frontière paramétrique déterministe

Le modèle non statistique proposé par Aigner et Chu peut aisément être soumis à une analyse statistique en faisant certaines hypothèses au sujet de x et de u . Supposons maintenant une technologie du type Cobb-Douglas, ce qui donne une fonction

$$\ln y_k = (\beta_0 - \mu) + \sum_{i=1}^m \beta_i \ln x_{ik} - (u_k - \mu) \quad (3.10)$$

qui a des résidus de moyenne zéro et satisfait aux conditions pour qu'un traitement par la méthode des moindres carrés ordinaires (MCO) soit convergent.

Richmond (1974) note que la méthode des MCO donne des estimations convergentes de $(\beta_0 - \mu)$ et des β_i , c'est-à-dire de tous les paramètres de la fonction de production, excepté la constante. En supposant une distribution particulière pour le terme u , sa moyenne peut être estimée de façon non biaisée et convergente à partir des moments centraux (deuxième et troisième ordre par exemple) des résidus des moindres carrés. L'estimation de μ peut alors être utilisée pour corriger le terme constant $(\beta_0 - \mu)$ de façon à obtenir des estimations convergentes de tous les paramètres de la frontière.

La difficulté de cette technique appelée moindres carrés ordinaires corrigés (MCOC) réside dans le fait que même après correction du terme constant, certains résidus peuvent encore se situer au-dessus de la frontière, ce qui rend la mesure de l'efficacité de chaque observation quelque peu délicate. Une autre difficulté de la méthode MCOC se situe dans la correction du terme constant, car celle-ci n'est pas indépendante de la distribution choisie pour u . Ainsi deux distributions distinctes donneront systématiquement des estimations différentes de l'efficacité technique.

Une solution à ces problèmes consiste à estimer la frontière de production par MCO et à corriger le terme constant de façon à ce qu'aucun résidu ne soit positif, un seul étant nul (Gabrielson 1975, Greene 1980). Autrement dit, on ajoute au terme constant le résidu positif le plus élevé obtenu par MCO.

3.4.5. Estimation d'une frontière paramétrique stochastique

Dans les modèles exposés précédemment, la frontière est considérée comme déterministe. Toute la variation de la performance d'une entreprise est attribuée à la variation de son efficacité par rapport à une frontière déterminée commune à toutes les firmes. Cette notion ignore malheureusement que la performance d'une entreprise peut être affectée aussi bien par des facteurs n'étant pas sous son contrôle (mauvaises conditions météorologiques, manque de chance, rupture dans l'offre des inputs, etc) que par ceux sous son entière responsabilité (l'inefficacité proprement dite). L'objet du modèle de frontière stochastique est de considérer séparément les effets non contrôlables par les firmes des autres effets, c'est-à-dire de l'inefficacité.

Ces arguments ont amené Aigner et al. (1977) et Meeusen et van den Broek (1977) à construire un modèle avec un terme d'erreur composé de deux parties:

- un élément symétrique permettant des variations aléatoires de la frontière entre les observations et qui incorpore les effets des erreurs de mesure, des variables explicatives non observées et des chocs aléatoires hors du contrôle des gestionnaires;
- un élément asymétrique représentant les effets de l'inefficacité par rapport à la frontière stochastique (Forsund et al. 1980).

Une frontière de production de type Cobb-Douglas peut être obtenue à partir du modèle suivant:

$$\ln y_k = \beta_0 + \sum_{i=1}^m \beta_i \ln x_{ik} + (v_k - u_k) \quad (3.11)$$

$k=1, \dots, n$ observations, $u \geq 0$.

Dans tous les modèles de frontière stochastique, le terme d'erreur aléatoire v est supposé suivre une distribution normale de moyenne nulle et de variance σ_v^2 . On suppose alors que le terme d'inefficacité u suit une distribution normale tronquée (en zéro ou non) ou une distribution exponentielle (Stevenson 1980).

En introduisant des distributions de probabilité spécifiques pour u et v (par exemple semi-normal et normal) et en supposant qu'ils sont indépendants l'un de

l'autre ainsi que par rapport aux inputs, la fonction de vraisemblance peut être définie et les estimateurs du maximum de vraisemblance peuvent alors être calculés.

Bien qu'il soit moins efficace asymptotiquement, le modèle peut aussi être estimé par les moindres carrés ordinaires corrigés en ajoutant au terme constant un estimateur convergent de $E(u)$ obtenu à partir des moments centraux des résidus des moindres carrés ordinaires.

La faiblesse de ces méthodes réside dans leur incapacité à décomposer les résidus $(v_k - u_k)$ individuellement afin d'obtenir leurs deux composants séparément. L'inefficacité de chaque observation ne peut être obtenue; seule une estimation de l'inefficacité moyenne de l'échantillon ($E(u)$) peut être calculée.

Ce problème a été résolu par Jondrow et al. (1982) qui ont proposé une procédure permettant d'estimer l'inefficacité technique pour chaque observation. Ils proposent d'évaluer $E(u_k | v_k - u_k)$ à partir des valeurs ajustées de $(v_k - u_k)$ et des paramètres de la fonction. Ils donnent ensuite une formule pour $E(u | v - u)$ dans le cas où v suit une distribution normale et u une distribution semi-normale. Il est évident que diverses distributions de l'efficacité mèneront à des résultats empiriques fort différents en ce qui concerne l'estimation des inefficacités individuelles (Schmidt 1986). D'un point de vue empirique, il est donc important de tester ces hypothèses et de considérer différentes distributions afin de voir laquelle convient le mieux.

L'inconvénient des modèles de frontière stochastique est leur dépendance vis-à-vis de la distribution arbitraire choisie pour le terme d'inefficacité u .

Grâce aux développements de la théorie de la dualité (Shephard 1953), les modèles de frontières stochastiques ne s'appliquent pas uniquement aux frontières de production, mais également aux frontières de coût. Nous l'avons déjà mentionné, les fonctions de production et de coût sont deux façons alternatives de définir la même technologie. L'estimation statistique des frontières de coût se fait de la même manière que pour la frontière de production, en gardant à l'esprit que les observations se situent maintenant au-dessus de celle-ci. Etant donné que l'efficacité peut être obtenue à partir de l'une ou l'autre frontière, le choix d'une frontière de coût est plus appropriée si l'output est exogène, alors qu'une frontière de production convient mieux lorsque seuls les inputs sont exogènes.

Schmidt et Lovell (1979, 1980) ont considéré un système d'équations composé d'une frontière de production de type Cobb-Douglas et d'un ensemble d'équations représentant les conditions de premier ordre que les firmes sont supposées respecter afin de minimiser le coût total. Cette démarche leur permet non seulement d'accroître l'efficacité des paramètres en exploitant les restrictions entre les équations (les paramètres de la fonction de production apparaissent dans les conditions de premier ordre), mais encore d'obtenir des estimations des efficacités techniques et allocatives des observations.

Malheureusement, une analyse similaire à celle de Schmidt et Lovell devient beaucoup plus difficile avec des technologies plus complexes que celles du type Cobb-Douglas. Greene (1980) a introduit un système d'équations consistant en une frontière de coût de type translog et des équations de part optimale des facteurs dans le coût total (*share equations*). L'estimation d'un tel modèle présente l'avantage de la flexibilité de la fonction translog et permet de tester différentes hypothèses relatives aux propriétés de la fonction de coût (homogénéité, homothéticité, etc). Il subsiste cependant le délicat problème de la relation entre le terme d'erreur de la frontière de coût et ceux des équations de part optimale des inputs. Greene suppose l'indépendance entre ceux-ci, ce qui ne semble pas tout à fait exact puisque toute proportion incorrecte d'un facteur accroît le coût total. Cela signifie que pour une équation de part optimale d'un facteur donné, une erreur positive ou négative accroît le coût total. Il existe donc certainement une distribution jointe entre les termes d'erreurs de la frontière de coût et les équations de part optimale qui devrait permettre de tenir compte de cette dépendance. Elle n'est toutefois pas triviale et reste à être découverte.

3.4.6. Critique de la littérature

Le principal avantage des méthodes paramétriques réside dans la possibilité de déterminer une frontière de production dans une forme mathématique simple et dans leur capacité à s'accommoder des rendements d'échelle non constants. Elles permettent en outre d'obtenir des informations sur les caractéristiques de la fonction de production telles que l'élasticité d'échelle et les élasticités de substitution entre les facteurs.

Cependant, la forme mathématique, déterminée a priori, peut imposer une structure inadaptée à la frontière. Enfin, l'approche paramétrique limite très sévèrement le nombre d'observations qui peuvent être déclarées techniquement efficaces.

4. Efficacité des services publics en réseau: études empiriques et comparatives

Plusieurs auteurs ont estimé l'efficacité relative des entreprises publiques et privées dans le domaine des services en réseau. La production et la distribution de l'électricité est certainement le secteur qui a été le plus souvent analysé jusqu'ici. Nous mentionnerons ci-après les études les plus récentes.

Yunker (1975) a effectué une comparaison entre les secteurs public et privé aux Etats-Unis en considérant les coûts de la production, de la transmission et de la distribution d'électricité comme une fonction de l'output et du nombre de clients. Il ne considère pas le prix des inputs dans son analyse. L'échantillon comprend 24 entreprises publiques et 49 privées produisant entre 0.5 et 4 millions de mega-watt heures (MWh) par années et desservant au moins 1000 clients. Ses résultats montrent que les coûts sont légèrement plus faibles pour les firmes publiques, la différence avec les entreprises privées n'étant toutefois pas significative.

Meyer (1975) a considéré des compagnies américaines produisant et distribuant l'électricité parmi lesquelles 30 sont publiques et 30 privées. Il a estimé une fonction de coût différente pour la production, la transmission, la distribution et la maintenance. Son analyse montre que, pour chacune des années observées – 1967 à 1969 –, la structure des coûts des deux types d'entreprises est différente. Il trouve que les coûts de production sont décroissants par rapport à l'output et sont généralement plus faibles pour les firmes publiques. Les coûts de transmission et ceux de la maintenance sont aussi plus faibles pour le secteur public. Bien que ses résultats soient significatifs, Meyer n'a pu déterminer la source de la différence de performance entre entreprises publiques et privées. De plus, il n'a ni tenu compte des différentes technologies utilisées ni du fait que les firmes publiques avaient un coût du capital plus faible.

Afin de remédier aux problèmes liés au prix des inputs, Neuberger (1977) a inclus ces prix dans son analyse ainsi que d'autres caractéristiques telles que la longueur des lignes de distribution et la surface du territoire desservi. Son étude effectuée aux Etats-Unis comprend 90 entreprises privées et 75 publiques. Il incorpore directement dans sa fonction de coût par client une variable dichotomique tenant compte du mode de propriété. Cette dernière variable

montre que les entreprises publiques sont 9% moins chères que les firmes privées avec, cependant, un niveau de signification de 10% seulement.

Une autre étude comparative a été effectuée par Pescatrice et Trapani (1980). Ils n'ont examiné que des firmes produisant l'électricité à l'aide d'une énergie fossile et exclu toutes celles réalisant également la transmission et la distribution. Leur échantillon contient 56 observations dont 33 sont des firmes privées. Leur analyse tient compte de l'âge des équipements et d'un coût du capital spécifique à chaque entreprise. Ces éléments leur ont permis d'estimer une fonction de coût de type translog ainsi que les fonctions de demande des inputs. Leurs résultats montrent que les coûts sont entre 25% et 43% plus élevés pour les compagnies privées durant la période 1965-70. De plus, Pescatrice et Trapani ont également trouvé que le taux du progrès technique dans les entreprises publiques est considérablement plus élevé que dans les firmes privées. Ils concluent que la meilleure efficacité du secteur public provient du processus de réglementation imposé aux entreprises privées.

Nous citerons encore l'étude de Foreman-Peck et Waterson (1985) qui se réfère à un échantillon d'entreprises britanniques municipales et privées. Basée sur des données de 1937, leurs résultats indiquent qu'en moyenne les compagnies publiques sont moins efficaces. Ils observent néanmoins que les grandes entreprises publiques sont aussi performantes que les compagnies privées, alors que de nombreuses petites firmes publiques sont moins efficaces.

D'autres services en réseau ont également reçu une certaine attention. Dans le domaine de la distribution de l'eau, Crain et Zardkoohi (1978) ont examiné des données de 1970 provenant de 88 compagnies de taille homogène dont 24 sont privées. Au contraire de la plupart des études précédentes, ils trouvent que les entreprises privées ont des coûts 25% plus faibles. Cependant, la structure des coûts pour les secteurs public et privé est significativement différente, indiquant que l'ampleur des économies d'échelle est plus importante pour les firmes publiques. En restreignant leur échantillon à 78 observations, Crain et Zardkoohi (1980) trouvent que les entreprises privées sont surcapitalisées alors que le secteur public utilise trop de main-d'oeuvre. Feigenbaum et Teeple (1983) obtiennent des résultats similaires alors que l'étude de Bruggink (1982) montre que les firmes publiques sont plus efficaces.

Dans le domaine des transports publics, Caves et Christensen (1980) ont étudié la performance relative de deux compagnies canadiennes de chemin de fer, l'une privée et l'autre publique. Basés sur la méthode de la productivité totale des

facteurs, leurs résultats montrent que l'entreprise publique n'est pas moins efficace que celle du secteur privé. Ils observent encore que les deux compagnies examinées agissent sur un marché concurrentiel et que leur relativement bonne performance est très certainement due à cet environnement.

Une étude menée au Royaume-Uni par Pryke (1982) compare l'efficacité des compagnies aériennes. Il conclut que les entreprises privées sont, d'une part, plus profitables et, d'autre part, plus efficaces que leurs concurrentes du secteur public. Des résultats allant dans le même sens ont été obtenus plus récemment par Forsyth (1984). L'étude de Manzini (1990) parvient à des conclusions différentes de celles de Pryke et Forsyth. Partant d'un échantillon de 50 compagnies de transport aérien, il a cherché à déterminer si un mode de propriété (public/privé) était techniquement plus efficace que l'autre à partir de l'estimation d'une fonction de production de type translog. Bien qu'aucune tendance ne soit apparue en faveur de l'un ou l'autre des deux secteurs, Manzini relève que le cadre réglementaire associé au mode de propriété affecte l'efficacité technique des compagnies. Dans ce contexte, ses résultats lui permettent d'affirmer que l'expérience américaine de déréglementation a exercé une influence positive sur la performance des transporteurs aériens.

DEUXIEME PARTIE
FOURNITURE EFFICACE DES SERVICES COLLECTIFS
ORGANISES EN RESEAU

5. Les modes de fournitures des services collectifs

5.1. Introduction

L'éventail des alternatives imaginées par les collectivités publiques locales pour fournir des services à leur population se révèle d'autant plus large que les pressions extérieures (résistance des contribuables, volonté collective de "privatiser" certains services) sont importantes. Cependant le choix du mode de fourniture à disposition de chaque collectivité dépend fortement de la nature du service à fournir. Si le service a toutes les caractéristiques d'un bien collectif pur – impossibilité ou difficulté d'exclusion, non rivalité dans la consommation – une action collective est alors indispensable pour le produire et couvrir les coûts générés par sa production. Nous avons déjà vu que même pour les biens mixtes, l'intervention collective peut s'avérer nécessaire pour plusieurs raisons. On peut y ajouter la volonté que la prestation soit absolument offerte en quantité suffisante à la population.

L'action collective n'est toutefois pas forcément synonyme d'une intervention du secteur public. Des individus peuvent coopérer en vue de fournir une prestation de nature collective sans qu'il y ait nécessairement participation du secteur public. On trouve, par exemple, des équipements de loisirs tels que des courts de tennis, des piscines, des salles d'exposition ou des réseaux câblés qui sont aussi bien en mains privées que publiques. On pourrait penser que la coopération volontaire n'est réalisable que dans le cas des services pour lesquels l'exclusion est possible. La fourniture de services dont l'exclusion est difficile peut toutefois être effectuée par une association privée. L'entretien des rues d'un quartier (nettoyage, ramassage de la neige) pourrait être effectué par les habitants eux-mêmes ou par une association des résidents. En fait, lorsque la collectivité est relativement petite et les préférences de ses membres homogènes, les pressions sociales peuvent être suffisantes pour assurer une contribution de chacun et

éviter les passagers clandestins. Les services du feu assurés par des volontaires est un exemple bien connu.

Les actions volontaires privées pour fournir un bien collectif conviennent cependant moins bien lorsque la communauté devient relativement grande et que les intérêts de ses membres divergent. Dans ce cas, l'administration publique peut se donner les moyens d'obliger les individus à financer, par une contribution monétaire involontaire, la fourniture des services collectifs. Elle décide également quels services doivent être délivrés à la population, quels types de financement adopter et comment allouer les biens entre les membres de la société.

Ainsi, tout service collectif exige une action collective. Toutefois, cela ne signifie pas que si la collectivité publique assume la responsabilité de la fourniture du service, ce dernier soit forcément produit par les employés de l'administration. Plusieurs alternatives sont possibles. Par exemple, le service peut être sous-traité ou concédé à une entreprise privée.

A travers ces exemples, on remarque que la fourniture d'un service collectif peut être séparée de sa production proprement dite. L'exemple le plus frappant est la réparation des routes. La commune est responsable du maintien des conditions de circulation mais la réparation elle-même est très souvent confiée à une entreprise privée.

Bien que les services collectifs en réseau ne représentent qu'un sous-ensemble des prestations publiques, leur fourniture n'en prend pas moins plusieurs formes. C'est pourquoi nous traiterons dans ce chapitre l'ensemble des modes de fourniture des services collectifs.

5.2. Approche théorique

Nous avons vu que les façons de fournir un service collectif sont diverses. Elles dépendent de plusieurs facteurs :

- caractéristiques économiques du service (bien collectif pur, bien mixte ou bien privé pur);
- caractéristiques et structure économique du marché;
- taille et homogénéité de la collectivité;

- préférences des autorités publiques et de la population;
- efficacité de la production et de la fourniture du service.

En fonction de ces éléments, la fourniture du service prendra une forme différente avec une intervention publique plus ou moins importante. Les deux formes extrêmes sont, d'une part, une gestion et une production effectuées totalement par le secteur public et, d'autre part, une libre fourniture du service par un système de concurrence entre plusieurs entreprises privées.

Avant de décrire chaque mode de fourniture, il convient de prendre en considération un certain nombre d'aspects.

Il s'agit en effet de distinguer :

- quel est le consommateur du service;
- quel est le producteur de la prestation;
- quel agent détermine les caractéristiques du service à offrir et sélectionne le producteur;
- quel est l'éventail de choix du consommateur;
- quelle est la situation du ou des producteurs;
- de quelle manière le service est financé.

Le consommateur est celui qui reçoit le service et en jouit. Ce peut être un individu, un ménage, un résident d'une aire géographique donnée ou un groupe spécifique de personnes ayant des caractéristiques communes (étudiants, handicapés).

Le producteur du service est un agent qui réalise directement le travail ou délivre le service au consommateur. Un producteur peut être une unité de l'administration publique, un groupement de collectivités publiques, une entreprise privée, une association de volontaires, une organisation sans but lucratif ou, dans certains cas, le consommateur lui-même.

L'agent qui détermine les caractéristiques du service et met en relation le producteur et le consommateur est souvent une unité de l'administration publique. Nous l'appellerons *superviseur*, Savas (1982) le nomme *arranger*. Cet agent décide également si le service doit être fourni et détermine la population

qui va en bénéficier. Le superviseur peut être l'administration communale ou cantonale, une association de volontaires ou le consommateur lui-même.

La distinction entre fournir et produire un service devient alors importante. Pour une grande quantité de prestations, l'administration publique est essentiellement le superviseur, c'est-à-dire un instrument à la disposition de la société lui permettant de décider ce qui devra être réalisé collectivement, les individus qui pourront en bénéficier, le niveau d'offre du service (quantité et qualité) et de quelle manière celui-ci devra être financé (par l'impôt, par une taxe ou par un prix). Dans cette perspective, la production d'un service peut être totalement indépendante des décisions prises de façon collective. Si, par exemple, une administration communale décide de fournir un service à sa population et de le financer collectivement – par l'impôt – cela ne signifie pas que la production sera automatiquement effectuée par les propres employés de la commune. Bien que la responsabilité de la fourniture du service incombe à l'administration publique, il peut y avoir différents producteurs.

L'éventail de choix du consommateur doit être considéré comme sa liberté à sélectionner un service sur le marché en fonction de ses préférences. Une fois le service mis à sa disposition, ses possibilités de choix dépendent :

- d'une éventuelle obligation de consommer (service de police par exemple);
- de l'unicité du service, chacun consommant exactement les mêmes quantité et qualité de la prestation;
- de la détermination par l'individu de la quantité consommée en fonction du prix du bien et de l'utilité qu'il lui attache;
- de la diversité du service mis à disposition de la population – différentes qualités par exemple;
- de l'adaptation de l'offre à la demande globale de la population desservie.

Lorsque le consommateur agit seul en tant que superviseur, on remarque alors que son éventail de choix par rapport au service est le plus large.

La situation du ou des producteur(s) se réfère au type de marché sur lequel le service est offert. La situation de concurrence parfaite ne se rencontre que rarement en raison de la nature des biens et services collectifs. A l'inverse, le monopole est plus fréquent. On verra par la suite qu'il est possible de créer une

situation de concurrence même dans les cas où une seule entreprise produisant le service est plus efficace que plusieurs (*monopole naturel*). Pour la collectivité, l'intérêt de connaître le type de marché réside avant tout dans la volonté de s'approcher le plus possible de l'optimum social.

Enfin, le type de financement du service a des conséquences sur le comportement des consommateurs et sur celui des producteurs. Si un service fournit à la population est financé entièrement de manière collective, le consommateur aura tendance à en consommer le plus possible et le producteur ne sera pas incité à se comporter efficacement puisqu'il n'est pas rétribué en fonction de la quantité vendue.

5.3. *Modes de fourniture*

Les différentes possibilités à la disposition des collectivités publiques ont été répertoriées principalement par Savas (1982, 1987) et par Hatry (1983). Nous suivrons en grande partie la méthodologie de Savas et classerons les modes de fourniture en gestion directe, coopération entre collectivités publiques, sous-traitance, concession, subvention à une organisation privée, remise de bons, association de volontaires et enfin marché. A remarquer que ces modes de fourniture ne tiennent pas compte de l'intervention publique sous forme de réglementation. Celle-ci s'applique généralement au dernier mode mais peut être également être combinée avec un autre, en particulier avec la concession comme nous le verrons par la suite.

5.3.1. *Gestion directe*

Description: La collectivité publique (administration communale ou cantonale) gère et produit elle-même le service à l'aide de ses propres équipements et de son personnel. Elle fixe le niveau et la qualité de la prestation offerte. De même, dans certains cas, elle détermine la population cible. Le financement est assuré soit par les ressources générales du budget (impôt), soit par un tarif ou encore par une combinaison des deux.

Parmi les services les plus couramment offerts à la population par la gestion directe en Suisse, on trouve l'administration générale, les services de justice et

de police, l'enseignement de base, l'éclairage public, les services sociaux ou encore la distribution et l'assainissement de l'eau¹.

Buts: Les raisons pour lesquelles la collectivité publique décide d'offrir un service en utilisant la gestion directe sont nombreuses. Tout d'abord, l'accès au service est aisé et chaque individu devrait pouvoir en bénéficier librement, quel que soit son revenu. Un certain niveau de qualité de la prestation est assuré car la production de celle-ci reste sous le contrôle de la collectivité publique. La continuité et la régularité de l'approvisionnement sont garanties puisque l'Etat ne peut faire faillite. Un des objectifs principaux de la gestion directe réside dans le fait que la décision de produire un service ne tient pas uniquement compte des avantages privés – profitant directement au consommateur – mais prend également en considération les avantages et désavantages externes. Enfin, la quantité offerte correspond à un niveau qui permet de maximiser le bien-être collectif indépendamment des contraintes financières auxquelles le producteur privé doit faire face, c'est-à-dire équilibrer les coûts et les recettes.

5.3.2. Coopération entre collectivités publiques

Description: Plusieurs collectivités publiques (généralement des communes) ayant les mêmes besoins se regroupent au sein d'un organisme (syndicat, association) pour fournir un service en commun. Cet organisme prend la forme juridique d'une personne morale de droit public. Il offre le service à la population de ses membres soit directement (personnel et équipements propres), soit indirectement en mandatant une entreprise privée pour produire la prestation.

En Suisse, la coopération entre communes est fréquemment utilisée en raison du grand nombre de communes de petite taille. On peut citer les équipements sportifs tels que patinoires et piscines, le traitement des ordures ménagères, les hôpitaux, les transports régionaux, l'enseignement secondaire.

Buts: La coopération entre collectivités publiques permet à celles-ci d'atteindre un seuil critique en deçà duquel il leur serait impossible d'offrir un service faute

¹ Ces tendances sont basées sur les résultats d'une enquête menée en 1988 auprès des communes suisses de plus de 5'000 habitants. Cette enquête a été réalisée dans le cadre d'un mandat confié par le Fonds national de la recherche scientifique et portant sur l'efficacité dans la fourniture des services publics locaux.

de moyens financiers d'une part et de consommateurs potentiels d'autre part. C'est très souvent le cas avec un service exigeant un investissement important. La coopération permet alors de réaliser des économies d'échelle et de réduire les coûts d'exploitation. L'utilisation de ce mode de fourniture vise non seulement à éliminer la duplication des équipements et du personnel mais aussi à améliorer la qualité de la prestation. Tous ces éléments ont évidemment une influence positive sur la baisse des coûts engendrés par le service offert.

L'autre but de la coopération intercommunale est de résoudre les problèmes provenant de déséconomies externes liées à la consommation de certains biens ou services (Schwab et Christe 1990). En se mettant d'accord pour réaliser un investissement commun, les collectivités publiques contribuent à réduire les effets négatifs qu'elles exercent les unes sur les autres. La construction d'une station d'épuration des eaux usées desservant plusieurs communes en est un exemple.

5.3.3. *Sous-traitance*

Description: La collectivité publique confie à un tiers – entreprise privée ou autre collectivité publique – le soin de produire la totalité ou une partie du service selon ses directives et sous sa responsabilité. Le sous-traitant est payé directement par la collectivité qui finance la prestation, soit par les ressources générales du budget, soit par un tarif ou par une combinaison des deux.

La forme la plus fréquemment rencontrée est un contrat entre la collectivité publique et une entreprise privée. Une seconde solution consiste en un contrat entre deux collectivités publiques, la première achetant le service à la seconde.

Parmi les prestations le plus souvent sous-traitées en Suisse, on trouve le ramassage des ordures ménagères, l'enlèvement de la neige, l'entretien des routes, les service d'ambulances ou encore les transports d'écoliers.

Buts: L'objectif principal de la seconde forme de sous-traitance est la réalisation d'économies d'échelle, bien que la première puisse également y parvenir. En achetant le service à une autre collectivité, généralement plus grande, celui-ci pourra être produit avec des équipements de taille optimale.

L'autre forme poursuit principalement deux buts. Le premier est de créer une situation de concurrence afin que le service soit produit le plus efficacement possible. Pour l'obtention du contrat, la collectivité publique met alors en

concurrence plusieurs entreprises privées. Le second but est d'éviter l'engagement de main-d'oeuvre spécialisée qui ne pourrait être utilisée rationnellement (architectes, juristes, ingénieurs, etc). Il est en effet préférable de sous-traiter ces activités si les besoins sont trop faibles ou irréguliers (Fisk et al. 1984).

L'objectif principal de la sous-traitance – réduire les coûts de la fourniture d'un service – est basé sur les hypothèses suivantes: par rapport à une administration publique, l'entreprise privée est plus flexible dans la gestion de son personnel, évite les comportements bureaucratiques et, grâce à la concurrence, est fortement incitée à produire plus efficacement ainsi qu'à innover plus rapidement (Fitts & Lindveit 1984).

5.3.4. Concession

Description: La collectivité publique accorde à une entreprise privée ou semi-publique le droit et la responsabilité de fournir un service collectif dont elle a le monopole. Le consommateur paie le service directement au producteur. La concession est un acte juridique qui fait l'objet d'un contrat, limité dans le temps, fixant les droits et devoirs du concessionnaire. La collectivité peut ainsi réglementer le prix, la qualité et le niveau du service. De ce fait, l'aire géographique desservie par l'entreprise peut également être déterminée.

La concession peut être exclusive – une seule entreprise produit et vend le service – ou non exclusive. Le premier cas s'applique surtout aux situations de monopole naturel pour lesquelles il est préférable, du point de vue de l'efficacité économique, qu'une firme unique produise une quantité donnée du service plutôt que plusieurs entreprises (sous-additivité de la fonction de coût de l'industrie). Dans le second cas, la collectivité publique attribue la concession à plusieurs firmes se faisant concurrence. Le consommateur peut alors effectuer un choix entre différents fournisseurs en fonction de ses préférences. Quelle que soit la forme de concession accordée, il est évidemment indispensable que le service puisse être offert avec la possibilité d'exclure les individus qui ne désiraient pas en payer le prix.

La production des services collectifs en réseaux conduit à une situation de monopole naturel pour laquelle la concession s'adapte particulièrement bien. Ce mode de fourniture offre en effet le moyen de créer une situation de concurrence dans laquelle l'entreprise n'est pas incitée à profiter de sa position de

monopoleur. Les services tels que la distribution d'eau, de gaz ou d'énergie, les transports publics, les télécommunications sont souvent attribués au moyen d'une concession exclusive. Par contre, les services d'ambulances ou les taxis peuvent être accordés à plusieurs entreprises bénéficiant de concessions non exclusives.

Buts: La concession permet à la collectivité publique d'exercer un contrôle sur un service collectif sans pour autant le fournir elle-même. Cela lui permet de n'engager ni ses ressources financières, ni son personnel tout en garantissant la disponibilité du service.

La plupart des objectifs de la concession sont identiques à ceux de la sous-traitance. En faisant jouer la concurrence entre plusieurs entreprises pour l'obtention de la concession, celle qui aura finalement le droit de produire le service sera incitée à agir efficacement ainsi qu'à offrir aux consommateurs une prestation de qualité.

De plus, le consommateur étant contraint à payer le prix du service, le producteur peut aisément ajuster son offre à la quantité demandée. Lorsque la concession est non exclusive, les consommateurs exercent une pression sur les entreprises qui ne satisferaient pas leurs exigences en matière de qualité ou de prix.

5.3.5. Subventions

Description: La collectivité publique verse une subvention à une entreprise privée, à une organisation sans but lucratif ou à une association de volontaires afin de l'inciter à accomplir une prestation d'intérêt général. Lorsque la consommation d'un service doit être encouragée, la subvention permet d'en réduire le prix, d'en améliorer la qualité ou d'en accroître le niveau. La subvention peut prendre non seulement la forme d'une contribution monétaire mais également celle d'une exemption fiscale ou de prêts sans intérêts.

Le système des subventions est largement utilisé dans le domaine de la santé afin de rendre plus accessibles les services médicaux à l'ensemble de la population (subventions à des hôpitaux ou à des compagnies d'assurance maladie). Les services culturels tels que théâtres, opéras, musées ou orchestres symphoniques sont souvent subventionnés, tout simplement afin de rendre leur consommation possible car les prix nécessaires à la couverture de leurs coûts de production seraient dissuasifs.

Les collectivités publiques emploient aussi ce système lorsque le marché est un monopole naturel global et désirent que la firme produise à un niveau pour lequel la demande est égale au coût marginal. A ce point, le coût moyen étant supérieur au coût marginal, le producteur ne couvre pas ses coûts. Etant donné la quantité offerte, la collectivité publique octroie alors à ce dernier une subvention lui permettant d'équilibrer son budget .

Buts: Nous l'avons déjà mentionné, l'objectif principal de ce mode de fourniture est d'encourager la consommation d'un service. Mais pourquoi la collectivité publique ne le produirait pas elle-même? Il y a deux raisons à cela. La première provient de l'accès qu'ont certaines organisations à des ressources telles que travail de volontaires, donations et fonds privés indisponibles pour les collectivités publiques; cela permet à ces organisations de fournir le service de façon moins coûteuse. La seconde, puisque la subvention profite aux usagers du service, la collectivité publique peut s'assurer que certains groupes de la population y ont également accès.

5.3.6. Remise de bons

Description: La collectivité publique remet à différentes catégories de consommateurs des bons représentant une certaine valeur monétaire. Ceux-ci sont à faire valoir sur le marché pour l'achat de biens ou services d'intérêt général. Les bons permettent à leurs bénéficiaires de se procurer un service déterminé qu'ils ne pourraient se permettre d'acheter au prix courant du marché. Ce système donne au consommateur la liberté d'effectuer un choix en sélectionnant la quantité et la qualité désirées du service ainsi que son producteur. A condition que ce dernier soit agréé par la collectivité publique, il reçoit alors de celle-ci la contre-valeur du bon.

Bien que rarement utilisé, on trouve ce mode de fourniture dans le système des logements sociaux ou dans l'enseignement. Dans les transports, la remise de bons permet aux personnes âgées et aux handicapés d'utiliser les taxis plutôt que les transports en commun.

Buts: Par rapport aux subventions, la remise de bons donne aux consommateurs une plus grande liberté d'action quant au choix du service. En outre, la concurrence entre les fournisseurs peut mener à une amélioration de la qualité de la prestation et à une meilleure efficacité. Enfin, les bons permettent d'adapter

assez précisément l'offre d'un service aux besoins réels de la population puisque celle-ci agit en fonction de ses préférences.

5.3.7. *Volontariat*

Description: Une organisation de volontaires se substitue ou s'associe à la collectivité publique afin d'accomplir une prestation d'intérêt public. Aucun contrôle n'est en principe exercé par le secteur public sur la fourniture du service.

On trouve des activités bénévoles dans les domaines des services sociaux (repas à domicile, transports de personnes âgées ou d'écoliers) et paramédicaux (premiers secours, soins à domicile).

Buts: Le volontariat permet d'éviter une intervention publique. Le service est offert sans financement par le budget public et sans utiliser les employés municipaux. De plus, aucune réglementation particulière n'est nécessaire. Finalement, l'organisation de volontaires obtiendra une plus grande confiance de la part des consommateurs parce qu'ils estiment que celle-ci se sent beaucoup plus concernée par son activité que n'importe quel autre producteur.

5.3.8. *Marché*

Description: La collectivité publique laisse les entreprises privées fournir la prestation. Ce sont donc les lois du marché qui déterminent la quantité et la qualité de la production. La collectivité publique n'intervient plus, sinon pour imposer certaines normes. Le consommateur peut donc ainsi choisir librement le service qu'il désire en fonction de ses préférences et de son budget. Certains biens mixtes sont offerts par le marché, tels les services de sécurité, les zoos et même le ramassage des ordures (Savas, 1987).

Buts: Pour les prestations possédant les caractéristiques de bien privé pur – possibilité d'exclusion et rivalité – le marché est le système qui convient le mieux car aucune ressource publique n'est requise. De plus, sous la condition d'un marché concurrentiel, la situation d'équilibre permet d'atteindre un optimum économique. Les exemples ci-dessus montrent cependant que même certains services dont la consommation n'est pas rivale, peuvent être fournis par le marché.

Tableau 5.1 Modes de fourniture

MODE DE FOURNITURE	PRODUCTEUR	SUPERVISEUR	FINANCEMENT
GESTION DIRECTE	Collectivité publique	Collectivité publique	Collectivité publique et/ou consommateur
COOPERATION ENTRE COMMUNES	Association de collectivités publiques	Collectivités publiques	Collectivités publiques et/ou consommateur
SOUS-TRAITANCE	Entreprise privée ou autre collectivité publique	Collectivité publique	Collectivité publique et/ou consommateur
CONCESSION	Entreprise privée ou semi-publique	Collectivité publique et/ou consommateur ²	Consommateur
SUBVENTIONS	Entreprise ou association privée	Collectivité publique et consommateur	Collectivité publique et consommateur
REMISE DE BONS	Entreprise privée	Collectivité publique et consommateur	Collectivité publique et consommateur
VOLONTARIAT	Association de volontaires	Association de volontaires	—
MARCHE	Entreprise privée	Consommateur	Consommateur

² Lorsque la concession est non exclusive le consommateur peut effectuer un choix.

6. Evaluation des modes de fourniture de services publics

Le but principal d'une évaluation est la recherche d'une fourniture économiquement efficace des services publics. Celle-ci dépend de nombreux facteurs ayant un impact aussi bien sur la production que sur la consommation de la prestation. Il faut souligner que le choix d'un mode de fourniture dépend avant tout des caractéristiques du bien à fournir. C'est pourquoi nous allons examiner brièvement cet aspect de la fourniture des services publics.

6.1. Nature des services et choix du mode de fourniture

Nous avons déjà constaté que plusieurs modes de fourniture ne pouvaient pas s'appliquer à certains types de biens. Il est clair que les biens privés purs peuvent être fournis par n'importe quel système, quoique certains soient mieux adaptés que d'autres. A l'opposé, les biens collectifs purs ne peuvent être délivrés qu'à travers une intervention publique - gestion publique, coopération entre collectivités publiques, sous-traitance - ou par une collaboration volontaire. Les autres alternatives ne peuvent s'appliquer puisqu'elles nécessitent la fixation d'un prix et donc la possibilité d'exclusion.

Les biens mixtes se divisent en deux catégories. La première comprend ceux qui possèdent la propriété d'exclusion mais dont la consommation n'est pas rivale. A l'instar des biens privés purs, ces services peuvent être offerts quel que soit le mode de fourniture. Les services de la seconde catégorie possèdent les caractéristiques opposées et ne peuvent être offerts au travers des systèmes de bons, de concessions ou du marché. Dans quelques cas, ils peuvent être fournis en subventionnant le producteur. On pourrait en effet imaginer que la subvention soit attribuée aux entreprises qui polluent un cours d'eau afin de les inciter à traiter leurs déchets, le bien mixte étant alors le cours d'eau.

Tableau 6.1 Modes de fournitures et natures des services

	RIVALITE	NON-RIVALITE
EXCLUSION POSSIBLE	Biens privés purs gestion directe coopération sous-traitance concession subvention bons volontariat marché	Biens mixtes gestion directe coopération sous-traitance concession subvention bons volontariat marché
EXCLUSION IMPOSSIBLE	Biens mixtes gestion directe coopération sous-traitance subvention volontariat	Biens publics purs gestion directe coopération sous-traitance volontariat

6.2. *Efficacité des modes de fourniture*

Etant donné les restrictions précédentes, nous pouvons maintenant nous interroger sur l'efficacité de chaque mode de fourniture. Elle se divise en deux parties, la première concerne la production, la seconde porte sur la consommation du service.

Pour juger de l'efficacité de la production, on utilise un certain nombre de critères:

- coût du service pour la collectivité publique;
- existence d'une situation de concurrence;
- contrôle du producteur par la collectivité publique;
- qualité du service fourni;
- optimalité de l'échelle de production;
- stabilité et continuité de la fourniture;
- existence de fraude de la part du producteur.

Les critères suivants permettent d'apprécier l'efficacité de la consommation:

- coût du service pour le consommateur;
- liberté de choix;
- liaison entre coûts et avantages;
- équité (tarification selon le principe d'équivalence ou selon celui de la capacité contributive);
- effets redistributifs;
- prise en compte des groupes de population défavorisés.

6.2.1. Gestion directe

A) EFFICACITE DE LA PRODUCTION

Lorsqu'on recherche à tout prix la stabilité et la continuité de la fourniture d'un service, la gestion directe convient parfaitement. En effet, les risques inhérents à une production privée (faillite, cessation d'activité, arrêt de la production d'un service non rentable, etc) sont pratiquement inexistantes. Cet argument peut avoir un poids considérable lorsqu'il s'agit de fournir des services essentiels tels que la santé ou la distribution d'eau et d'électricité. La gestion directe permet d'éviter non seulement les risques d'interruption de la production mais d'assurer également une qualité minimale sur laquelle les consommateurs peuvent compter.

En outre, il semble que le contrôle de la production par l'autorité de surveillance soit meilleure lorsque ce sont les propres employés de la collectivité publique qui gèrent l'unité productrice que dans les cas où ce sont des firmes privées qui assurent l'offre du service. Cependant, les gestionnaires peuvent chercher à maximiser leur propre utilité plutôt que le bien-être social, il s'agit alors de trouver un système les incitant à rechercher l'intérêt général.

Bien que la gestion directe permette de satisfaire plusieurs objectifs économiques et sociaux, de nombreuses critiques lui sont adressées concernant son inaptitude à maintenir les coûts de production à leur niveau minimum. Cela s'explique souvent par:

- l'absence de concurrence, d'où manque de pression sur les coûts et impossibilité de comparaison avec une autre unité productrice;
- le comportement bureaucratique des gestionnaires qui visent à maximiser le budget du service et ainsi à accroître leur pouvoir (Niskanen 1971);
- le biais existant entre la taille effective des équipements et leur taille optimale d'où l'alternative entre produire un service à un coût trop élevé ou ne pas le produire du tout.

B) EFFICACITE DANS LA CONSOMMATION

Selon le mode de financement (impôt ou tarif) la situation du consommateur varie sensiblement. Toutefois, quel que soit le type de financement, le coût du service pour le consommateur dépend surtout de l'efficacité de la production puisque, d'une manière ou d'une autre, c'est lui qui le supporte. Le financement par un tarif permet cependant au consommateur d'exercer un meilleur contrôle sur la fourniture de la prestation. Il peut mieux ajuster sa consommation à ses besoins et donne ainsi une indication à l'administration publique sur la demande du service. Il ne dispose en revanche d'aucune liberté de choix car l'offre est unique.

La gestion directe permet d'offrir un service à l'ensemble de la population et de mieux prendre en considération les besoins particuliers de certains groupes, tels les personnes à bas revenu, les personnes âgées ou encore les enfants.

En raison des effets redistributifs substantiels dus au financement par l'impôt, ce mode de fourniture permet de satisfaire plus fréquemment le principe de la capacité contributive (*ability to pay*) plutôt que celui de l'équivalence (*benefit principle*).

6.2.2. *Coopération entre collectivités publiques*

A) EFFICACITE DE LA PRODUCTION

Par rapport à la gestion directe, la coopération intercommunale permet de produire un service avec des équipements de taille mieux adaptée. L'impact sur les coûts de production peut donc être important. En outre, la qualité du service peut être améliorée grâce à l'élargissement du financement et de la population desservie. Lorsque la production du service est confiée à une entreprise privée,

la situation est la même que dans le cas de la sous-traitance et les conditions d'application sont identiques.

B) EFFICACITE DANS LA CONSOMMATION

En général, la coopération entre collectivités publiques a les mêmes effets sur la consommation du service que la gestion directe. Dans quelques cas cependant des effets géographiques peuvent réduire la qualité du service pour les habitants éloignés des équipements et les inciter à ne pas consommer la prestation.

6.2.3. *Sous-traitance*

A) EFFICACITE DE LA PRODUCTION

La principale raison de l'utilisation de la sous-traitance est la diminution du coût pour la collectivité publique. Savas (1987) mentionne les conditions préalables suivantes pour y parvenir:

- le cahier des charges est défini sans ambiguïté ce qui implique une définition précise des caractéristiques de la qualité et du niveau du service;
- plusieurs producteurs potentiels sont disponibles ce qui permet d'instaurer un climat de concurrence entre les soumissionnaires;
- la collectivité publique peut contrôler la performance du sous-traitant;
- le contrat a une durée limitée et contient des clauses de pénalités au cas où le cahier des charges n'est pas respecté.

Si la sous-traitance répond à ces critères, le coût du service a de grandes chances d'être inférieur à celui de la gestion directe. De plus, l'échelle de production s'approche de son niveau optimal car l'entreprise sous-traitante a la possibilité de produire la même prestation pour différentes communes. Rappelons néanmoins que le coût total du service pour la collectivité publique ne correspond pas seulement à la rétribution du sous-traitant. Les coûts de contrôle et d'établissement du contrat par l'administration publique doivent être également comptabilisés. Le principal inconvénient de ce mode de fourniture, par rapport aux deux précédents, réside dans le risque de rupture de la production en raison de la faillite ou de la cessation d'activités de l'entreprise privée.

B) L'EFFICACITE DANS LA CONSOMMATION

Les effets sont les mêmes que pour la gestion directe et la coopération intercommunale; il convient toutefois de souligner que si le coût peut diminuer pour le consommateur - la sous-traitance permet alors de réduire réellement le coût global -, ce coût peut aussi augmenter si un climat de concurrence n'a pas été établi entre les entreprises pour l'obtention du contrat ou s'il y a entente entre elles.

6.2.4. Concession

A) EFFICACITE DE LA PRODUCTION

Dans la plupart des cas - situation de monopole naturel, les services fournis à travers le système de concession nécessitent un seul producteur afin de minimiser les coûts. Il est donc impossible de faire jouer la concurrence sur le marché. La collectivité publique doit, d'une part, simuler la situation de concurrence - condition d'efficacité de la concession - pour l'obtention du contrat et, d'autre part, établir une réglementation du marché.

L'attribution de concessions exclusives doit répondre à un certain nombre de critères afin de maximiser le bien-être social et d'obliger l'entreprise privée à produire au moindre coût:

- l'exclusion par le prix est possible à un coût raisonnable;
- le service doit pouvoir être défini précisément;
- le prix de la prestation doit pouvoir être estimé clairement;
- plusieurs entreprises potentielles se font concurrence pour obtenir le contrat de concession (procédures d'enchères);
- la durée de la concession est limitée.

La condition de concurrence est souvent difficile à remplir en raison de la nature du service fourni par la concession. En effet, les services, tels ceux en réseau, exigent des moyens financiers relativement importants (distribution de l'électricité par exemple), ce qui restreint significativement le nombre de soumissionnaires. C'est la raison pour laquelle la réglementation du monopole

s'avère très souvent indispensable afin de protéger les consommateurs contre des attitudes discriminatoires de la part du concessionnaire.

Les formes de réglementation ont fait l'objet de nombreuses études (Vickers et Yarrow 1985, Bailey 1973). L'une des formes le plus fréquemment utilisées aux Etats-Unis consiste à limiter le profit par l'établissement d'un prix maximum estimé afin de permettre au concessionnaire d'obtenir un taux de rendement "équitable" des capitaux investis. Cette réglementation du taux de rendement permet, certes, de protéger les consommateurs, néanmoins elle n'encourage pas l'entreprise à produire efficacement mais plutôt à trop investir pour accroître son profit global (*Averch-Johnson effect*).

Une autre forme rencontrée au Royaume-Uni vise à lier les variations du prix du service à l'indice des prix à la consommation. L'entreprise concessionnaire ne peut augmenter ses prix que jusqu'à concurrence du taux annuel de croissance de l'indice des prix à la consommation, moins un facteur X déterminé en fonction des gains de productivité attendus dans la branche. La grande difficulté de l'application de cette méthode est la détermination de la productivité future de la branche, car l'administration publique obtient cette information de l'entreprise qui, de son côté, a tout intérêt à sous-estimer ses gains d'efficacité.

La réglementation par les prix doit, en général, être liée à la fixation de normes de qualité, afin que la pression sur les coûts ne se fasse pas au détriment de celle-ci. Si la réglementation de la concession est effectuée correctement, l'entreprise sera alors incitée à choisir un équipement de taille optimale, c'est-à-dire proche du minimum de son coût moyen.

B) EFFICACITE DANS LA CONSOMMATION

L'accès au service est limité aux individus acceptant d'en payer le prix. Il y a donc une forte incitation, pour le consommateur, à établir un lien entre coûts et avantages reçus et à mesurer réellement ses propres besoins face à la prestation offerte par le concessionnaire, incitation inexistante lorsque le service est financé par l'impôt. Bien qu'il doive payer un prix, l'individu n'a pas la possibilité de choisir un fournisseur puisque celui-ci est unique. Par contre, lorsque la concession n'est pas exclusive, la présence de plusieurs entreprises sur le marché - généralement un nombre restreint - permet aux consommateurs d'accroître quelque peu leur liberté de choix. Etant donné que seul le consommateur finance le coût du service, il est évident que la concession n'entraîne aucun effet redistributif. De plus, ce mode de fourniture ne tient pas

compte des groupes défavorisés à moins que la collectivité publique n'applique des mesures complémentaires (subventions, bons).

Si les conditions d'efficacité de la production ne sont pas remplies, on risque de voir apparaître un coût élevé pour les consommateurs et une baisse substantielle de la qualité. A cet égard, plusieurs études empiriques effectuées aux Etats-Unis (Savas et Stevens 1976, Quigley et Kemper 1976) ont montré que la concession était plus coûteuse que la gestion directe et la sous-traitance dans les services de ramassage des ordures ménagères. Dans le domaine de l'assurance immobilière en Suisse, Buschor (1983) a établi que le coût était sensiblement plus faible dans les cantons ayant un monopole public que dans ceux où le service était fourni par des compagnies privées.

6.2.5. Subvention

A) EFFICACITE DE LA PRODUCTION

L'utilisation du système de subvention vise plus à encourager la consommation d'un service qu'à améliorer l'efficacité du ou des producteurs. Très souvent, la subvention bénéficie à un producteur unique (hôpital, théâtre, etc.) ce qui ne contribue évidemment pas à accroître la concurrence. Cependant, selon Hatry et Valente (1983), le coût pour la collectivité devrait être inférieur à celui d'une fourniture directe et ceci d'autant plus que le contrôle par le secteur public est relativement peu important, d'où un coût administratif moindre.

Etant donné le faible niveau de contrôle du producteur (absence de cahier des charges) et la nature des services concernés, la collectivité publique ne peut s'assurer d'une meilleure qualité du service par l'octroi de la subvention. En effet, si la subvention est supprimée, la prestation risque tout simplement de ne plus être fournie. La collectivité ne peut pas non plus complètement éviter le risque d'une interruption du service due à un problème (financier ou de main-d'oeuvre) de l'entreprise ou de l'organisation sans but lucratif.

B) EFFICACITE DANS LA CONSOMMATION

La subvention permet de réduire significativement le prix d'une prestation pour le consommateur, le coût direct diminue donc. Par contre, les non-utilisateurs du service participent à son financement par l'intermédiaire des impôts utilisés pour couvrir la subvention. Il y a donc un effet redistributif relativement important entre les non-usagers et les consommateurs du service au profit de ces derniers.

L'accès au service étant amélioré, la subvention permet de tenir compte, dans une certaine mesure, des groupes de population défavorisés.

6.2.6. Remise de bons

A) EFFICACITE DE LA PRODUCTION

Selon le service fourni, le système des bons peut permettre à plusieurs producteurs potentiels d'entrer sur le marché. La concurrence entre ceux-ci exerce une certaine pression sur leur coût de production. C'est d'ailleurs l'une des conditions pour que le système des bons fonctionne efficacement. Toutefois, il n'est pas évident que, pour la collectivité publique, les coûts soient plus faibles qu'une production directe par le secteur public. Le coût administratif peut effectivement s'avérer relativement important car la gestion de la remise des bons demande un personnel régulier. En revanche, la qualité du service sera probablement meilleure en raison de la diversité des fournisseurs. Le risque d'une rupture dans l'approvisionnement s'en trouve réduit.

B) EFFICACITE DANS LE CONSOMMATION

L'avantage principal de ce mode de fourniture est le choix offert aux consommateurs. En fonction de leurs préférences, ils peuvent librement déterminer leur fournisseur et le type de prestation désirée. Il est donc important que les individus soient bien informés au sujet des conditions du marché, en particulier sur le coût et la qualité du service. Cette dernière devrait d'ailleurs pouvoir être facilement estimée par le consommateur.

Si la réduction du coût global pour la population ne semble pas significative, les effets redistributifs sont au contraire évidents. Puisque la collectivité publique peut déterminer les groupes d'individus qui peuvent obtenir des bons, il y a redistribution entre les non-bénéficiaires qui contribuent, par l'impôt, au financement du service et les bénéficiaires des bons. Par ce système, la prise en considération des groupes défavorisés est très efficace.

6.2.7. Volontariat

A) EFFICACITE DE LA PRODUCTION

Une association de volontaires ne fournira une prestation que si le besoin s'en fait sentir et si le résultat de son action procure à ses membres une satisfaction

suffisante. L'intérêt ici n'est donc pas d'établir une concurrence entre fournisseurs, il s'agit avant tout d'inciter des individus à s'unir et à entreprendre une action collective pour laquelle ils se sentent motivés. L'intervention publique peut varier sensiblement selon la prestation fournie. Dans certains cas, l'association de volontaires agit en totale indépendance (association culturelles ou de loisirs, actions charitables ou sociales). Dans d'autres, l'administration publique intervient en fournissant un encadrement en personnel ou encore des équipements (service du feu, transports d'écoliers, bibliothèques). Le contrôle du producteur est alors directement proportionnel au degré d'intervention publique. La qualité du service dépend essentiellement de la qualification des volontaires et de leur entraînement ainsi que de leur motivation.

B) EFFICACITE DANS LA CONSOMMATION

Quel que soit le service, l'utilisation de volontaires permet, soit de réduire les impôts ou les tarifs, soit de fournir la prestation gratuitement ou à un prix symbolique. En général, l'action volontaire est limitée à une aire géographique déterminée (association de quartier) ou à un groupe spécifique (soins à domicile des personnes âgées). Cependant, pour les individus qui peuvent en bénéficier, l'accès au service est aisé.

6.2.8. *Marché*

Laisser le marché fournir un service suppose que ce dernier contienne une grande partie des caractéristiques d'un bien privé pur. Il faut en outre s'assurer que plusieurs entreprises produisent le service et se fassent concurrence sur le marché. Les conditions d'efficacité requises, aussi bien du point de vue de la production que de la consommation, sont celles qui prévalent dans la théorie classique. Nous n'y reviendrons donc pas.

En laissant le marché libre de fournir une prestation, la collectivité publique prend le risque que certains individus n'aient pas du tout accès au service en raison de son coût ou de l'éloignement de son lieu de production. La consommation de la prestation par l'ensemble de la population ne doit donc pas être considérée comme une condition nécessaire par la collectivité publique.

Tableau 6.2 Evaluation de l'impact des différents modes de fourniture

MODES DE FOURNITURE CARACTERISTIQUES	Gestion directe	Coopération	Sous-traitance	Concession	Subventions	Remise de bons	Volontariat	Marché
Exige une description précise du service	non	non	oui	oui	non	non	non	non
Exige plusieurs producteurs potentiels	non	non	oui	oui	non	oui	non	oui
Permet de réaliser des économies d'échelle	0	2	1	1	?	0	1	2
Améliore la qualité du service	0	2	1	1	2	3	3	3
Accroît le risque de discontinuité de la fourniture du service	0	0	3	2	2	2	1	1
Réduit le coût pour la collectivité publique	0	1	2	3	1	0	2	3
Réduit le coût pour le consommateur	3	3	3	0	2	2	2	0
Accroît le choix du consommateur	0	0	0	1 ¹	1	3	0	3
Lie coûts et avantages pour le consommateur	0	0	0	3	2	2	1	3
Exerce un effet redistributif	3	3	3	0	1	2	1	0

Remarque : l'effet peut être nul (0), faible (1), marqué (2), important (3)

Source: Hatry et Valente, 1983, Savas 1987

¹ Seulement lorsque la concession n'est pas exclusive.

7. Services publics organisés en réseau: problématique

7.1. Caractéristiques des services en réseau

Les services fournis au travers d'un réseau peuvent prendre des formes très différentes. Ils ont cependant un certain nombre de caractéristiques communes qui permettent de les identifier. Nous les examinerons tout d'abord sous l'angle de la production, puis du point de vue de la demande. Mais il convient tout d'abord de définir la notion de réseau.

Un réseau est un système, souvent complexe, de lignes ou de canaux interconnectés. Ces lignes ou canaux transportent des biens ou services aussi divers que des messages audio ou vidéo, des personnes, de l'énergie, de l'eau potable ou des eaux usées, des marchandises ou encore des informations.

Du point de vue de la production, les services en réseau se caractérisent par une forte intensité du facteur capital et par des coûts fixes importants. Ceux-ci sont souvent irrécupérables, c'est-à-dire que l'entreprise qui fournit le service perd ses investissements si elle décide de sortir du marché (réseaux téléphoniques, électriques, distribution d'eau ou de gaz). De plus, dans certains cas (téléphone, électricité), l'output est difficile ou impossible à stocker. Enfin, un point commun à une majorité de services en réseau est leur production par une firme unique en raison des caractéristiques particulières de la fonction de coût. Cette particularité est très importante tant du point de vue de la structure du marché et des relations qu'elle implique entre producteurs et consommateurs, que de celui de la réglementation et du contrôle du marché. L'importance considérable de cet aspect des services en réseau nécessite un examen détaillé de la problématique des monopoles naturels.

Sous l'angle de la consommation du bien ou service, nous pouvons dire qu'il y a non seulement une demande pour une utilisation réelle du réseau, mais également pour une utilisation potentielle de celui-ci. Sharkey (1988) parle de *demande d'accès au réseau*. D'autre part, à court terme, la demande varie au cours du temps ce qui oblige les producteurs de certains services à faire face à la totalité de celle-ci quel que soit son niveau. Les moyens de production se retrouvent donc souvent en état de surcapacité (télécommunications, transports urbains, électricité).

Finalement, les services en réseau sont souvent assimilés à des biens publics car leur consommation est généralement non rivale et leur production souvent indivisible (transports publics, télévision). De plus, les consommateurs peuvent bénéficier d'externalités positives. Par exemple, un nouvel abonné au téléphone, en élargissant la taille du système, permet d'accroître la qualité du service en augmentant le nombre de connections possibles et de réduire les coûts des communications entre les autres abonnés.

Les caractéristiques de l'offre et de la demande des entreprises organisées en réseau ont conduit les économistes à les considérer un peu rapidement comme des monopoles naturels. En raison de coûts fixes importants et de l'existence d'économies d'échelle, ce type d'entreprise s'est vu fréquemment réglementé par les administrations publiques de façon à ce qu'un seul producteur subsiste sur le marché.

Avant de discuter de la justification de l'intervention publique, il s'agit, en premier lieu, d'aborder la question du monopole naturel. Nous allons donc examiner, du point de vue de la théorie économique, les situations où il y a effectivement monopole naturel. Une grande partie de la littérature discutant des monopoles naturels (Baumol et al. 1982, Sharkey 1982, Encoua 1986, Encoua et Moreau 1987, Berg et Tschirhart 1988) s'appuie sur des activités multi-produits. Nous nous contenterons cependant de traiter le cas d'un seul output car notre objectif principal réside dans la mesure de l'efficacité.

7.2. Notions de monopole naturel

Une branche économique est un monopole naturel dans une aire géographique déterminée si et seulement si une entreprise unique peut produire l'output demandé à un coût inférieur à celui de deux ou plusieurs firmes ayant la même technologie (Vicker et Yarrow 1985). Le monopole naturel se définit donc en terme d'efficacité relative d'un producteur unique par rapport à l'efficacité de plusieurs.

Il faut remarquer toutefois que tous les monopoles ne sont pas forcément naturels mais peuvent résulter de circonstances particulières telles que la formation d'un cartel ou d'une guerre des prix entre une firme et ses concurrents afin d'éliminer ces derniers. Ils peuvent également provenir d'une intervention publique empêchant tout concurrent potentiel d'entrer sur le marché.

En supposant que plusieurs producteurs peuvent se faire concurrence, et en l'absence de toute réglementation, le monopole naturel apparaît lorsqu'une firme unique est capable de satisfaire la totalité de la demande à un coût inférieur à celui de plusieurs. Par conséquent, pour qu'un monopole soit naturel, il faut que la firme unique soit le producteur le plus efficace s'il n'y a aucune restriction à la concurrence (Sharkey 1982).

Avant de définir le monopole naturel en termes plus précis, il nous faut auparavant introduire la notion de fonction de coût. Nous nous limiterons à une branche économique produisant un output unique et dans laquelle chaque entreprise potentiellement capable de réaliser ce bien a accès à la même technologie. Nous avons déjà vu que celle-ci peut être exprimée de façon équivalente soit par une fonction de production, soit par une fonction de coût. Considérons la fonction de coût C de telle sorte que $C(q)$ représente le coût total de la production, par une seule firme, de q unités d'output. Etant donné l'hypothèse précédente, C exprime également la fonction de coût des autres entreprises de la branche. Supposons que l'output q soit produit par k entreprises,

chacune réalisant q_i ($\sum_{i=1}^k q_i = q$ et $q_i > 0$), alors le coût total de l'industrie est donné par $C(q_1) + C(q_2) + \dots + C(q_k)$.

Pour un niveau de production q donné, l'industrie est un monopole naturel si et seulement si

$$C(q) < \sum_{i=1}^k C(q_i) \quad (7.1)$$

Les fonctions de coût pour lesquelles l'inégalité (7.1) est vérifiée pour tout $k \geq 2$ sont dites **strictement sous-additives** en q . Cela signifie qu'il est moins coûteux de faire produire un output donné par une seule firme plutôt que par plusieurs. Cependant, il est plus important de souligner que la sous-additivité dépend de la forme de la fonction de coût et du volume de production. Une industrie peut être un monopole naturel pour de faibles quantités d'output mais ne plus l'être pour des quantités supérieures.

Jusqu'ici, les économistes avaient souvent identifié l'existence d'un monopole naturel à celle de rendements d'échelle croissants. Il est donc utile de faire une distinction entre les notions d'économies d'échelle et de sous-additivité.

Une fonction de coût est considérée comme ayant des rendements d'échelle croissants si

$$C(\lambda q) < \lambda C(q) \quad \text{pour tout } \lambda > 1 \text{ et } q \geq 0 \quad (7.2)$$

Divisons cette inégalité par λq , nous pouvons écrire

$$\frac{C(\lambda q)}{\lambda q} < \frac{C(q)}{q} \quad \text{pour tout } \lambda > 1 \text{ et } q \geq 0 \quad (7.3)$$

ce qui signifie que le coût moyen est une fonction décroissante de l'output.

Deux autres propriétés de la fonction de coût sont également liées au concept de monopole naturel, ce sont la concavité et la convexité. Une fonction de coût est dite strictement concave si pour deux productions possibles x et y

$$C[\theta x + (1-\theta)y] > \theta C(x) + (1-\theta)C(y) \quad \text{pour } 0 < \theta < 1 \quad (7.4)$$

et strictement convexe si

$$C[\theta x + (1-\theta)y] < \theta C(x) + (1-\theta)C(y) \quad \text{pour } 0 < \theta < 1 \quad (7.5)$$

Supposons que la fonction de coût soit différentiable, alors la concavité est équivalente à un coût marginal décroissant ($\partial^2 C / \partial q^2 < 0$) et la convexité à un coût marginal croissant ($\partial^2 C / \partial q^2 > 0$). Ces différentes définitions nous permettent de montrer que la sous-additivité est une notion plus générale que celle de coût moyen décroissant (rendements d'échelle croissants) ou de coût marginal décroissant (concavité). En suivant Sharkey (1982), nous pouvons alors dire que ni les rendements croissants, ni la concavité de la fonction de coût ne sont des conditions nécessaires pour obtenir la sous-additivité. La sous-additivité est donc compatible avec un coût marginal croissant (convexité) pour tout niveau d'output et avec un coût moyen croissant pour certains niveaux d'output mais pas pour tous.

La figure 7.1 montre bien que la sous-additivité n'implique pas forcément des rendements croissants. Le coût moyen, exprimé par la pente d'un segment entre l'origine et la fonction de coût, est décroissant jusqu'en q_0 , augmente entre q_0 et q_1 et décroît à nouveau pour des niveaux d'outputs supérieurs à q_1 . Une firme

unique produira q^* , où $q_0 < q^* < q_1$, plus efficacement que plusieurs entreprises bien que le coût moyen soit croissant dans le voisinage de q^* .

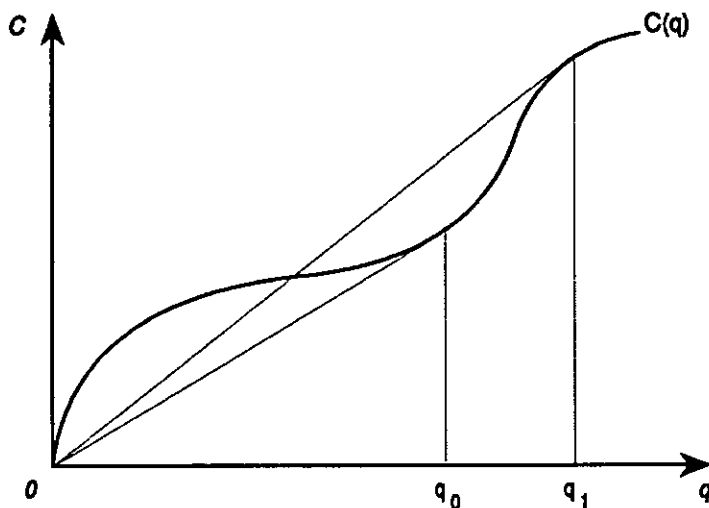


Figure 7.1 Fonction de coût et sous-additivité

Berg et Tschirhart (1988) font une différence entre monopole naturel local (*weak natural monopoly*) et global (*strong natural monopoly*). Si la fonction de coût satisfait la condition (7.3) pour tout niveau d'output, l'industrie est un monopole naturel global. Si, par contre, la fonction de coût satisfait la condition (7.1) mais pas (7.3) pour un niveau quelconque de l'output, alors la branche économique est un monopole naturel local. Cette distinction permettra de montrer que les effets d'une réglementation diffèrent selon le genre de monopole naturel.

7.3. Réglementation du monopole naturel: les raisons

La structure d'un marché qui possède les caractéristiques d'un monopole naturel, a très souvent justifié une réglementation par les pouvoirs publics. Le principe de l'imposition de contraintes s'articule autour de plusieurs arguments.

Le premier objectif poursuivi est de faire bénéficier la collectivité de conditions de production au coût minimum – une seule entreprise – et du surplus du consommateur en empêchant la firme de fixer un prix excessif. C'est l'argument

le plus connu. Un monopoleur non réglementé cherchera à offrir une quantité d'output lui permettant de maximiser son profit. L'égalité entre la recette marginale et le coût marginal résulte de ce processus. La réglementation permet alors d'améliorer le bien-être du consommateur et l'efficacité économique en fixant un prix égal au coût moyen.

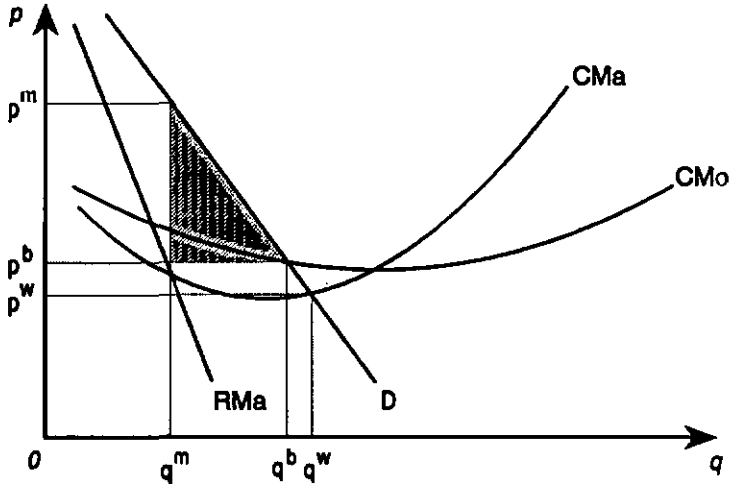


Figure 7.2 Monopole naturel et réglementation

La figure 7.2 illustre l'amélioration du bien-être collectif (surface hachurée) – traduit par l'augmentation du surplus du consommateur – résultant d'une action visant à contraindre le producteur à offrir le bien à un prix égal au coût moyen. Cependant, l'opportunité de ce type de réglementation a été fortement critiquée par plusieurs auteurs (Demsetz 1968, Baumol, Panzar et Willig 1982). Ils observent que sur un marché parfaitement contestable¹, une entreprise monopoliste ne peut obtenir des profits positifs. La réglementation ne doit alors s'appliquer qu'aux marchés non parfaitement contestables, c'est-à-dire lorsque les

¹ Un marché est dit parfaitement contestable si, (1) l'entrée est libre, (2) la sortie est possible sans coût, (3) les entreprises qui désirent entrer sur le marché disposent de la même technologie que celle(s) en place et (4) ont accès au même marché sans restriction (Curien et Gensollen 1987).

coûts d'entrée ne sont pas nuls ou lorsque la production exige des coûts fixes irrécupérables (*sunk costs*) qui freinent l'entrée en interdisant la sortie sans coût.

Le second objectif consiste à protéger un monopole naturel contre des comportements opportunistes de la part des consommateurs et des concurrents potentiels en face desquels le monopole peut parfois être vulnérable. Nous sommes alors en présence d'une configuration non soutenable². En l'absence de barrières à l'entrée et lorsque la production d'un bien ou service exige d'importants coûts fixes non récupérables, un monopole naturel n'est pas toujours assuré d'offrir le produit à un prix qui lui permette un équilibre budgétaire et empêche simultanément l'entrée temporaire de firmes sur certaines parts de marché. Sur un marché croissant où il y a des économies d'échelle ou dans une branche économique connaissant des progrès technologiques continus, de telles barrières peuvent s'avérer nécessaires.

Mais l'exemple le plus courant est celui où la nature du service fourni exige une capacité de production excédentaire afin d'anticiper les variations de la demande au cours du temps. Cette situation provoque alors un accroissement, temporaire du moins, des coûts de production du monopoleur. En l'absence de réglementation, d'autres firmes chercheront à entrer sur le marché afin d'exploiter des opportunités de profit à court terme, réduisant par là les avantages de long terme d'une production au coût minimum par une entreprise unique.

La volonté de stabiliser un marché constitue la troisième raison d'une intervention des pouvoirs publics. Sur un marché instable où l'on suppose que tous les producteurs se font concurrence sur une même base, l'absence de réglementation peut mener à une situation où aucune firme ne parvient à survivre. En effet, lorsque les coûts fixes sont peu élevés, une offre non réglementée peut facilement devenir excédentaire, entraînant l'instabilité du marché.

² Sur un marché quelconque, une configuration est soutenable lorsque (1) l'offre équilibre la demande, (2) aucune entreprise active ne fait de pertes, (3) aucun concurrent potentiel ne peut envisager une entrée profitable et (4) les projets d'entrée sont établis en supposant que les firmes en place ne modifieront pas leurs prix (Curien et Gensollen 1987, voir aussi Encoua et Moreau 1987).

7.4. Justification de la réglementation

La réglementation d'un monopole naturel ne se justifie que si son application permet d'accroître le bien-être collectif. Berg et Tschirhart (1988) estiment que les gains de la réglementation dépendent de trois facteurs:

- a) l'allocation des ressources en l'absence de réglementation;
- b) l'existence de barrières à l'entrée sur le marché;
- c) la présence d'un monopole local ou global.

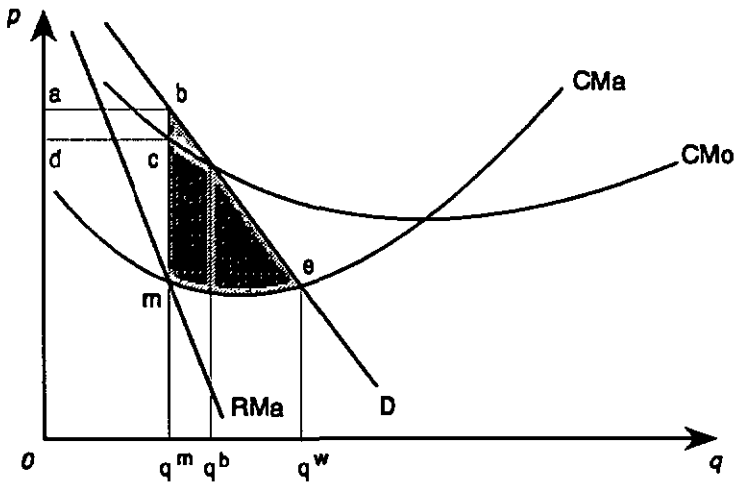


Figure 7.3 Avantage d'une tarification au coût marginal

Le premier facteur permet de justifier une intervention si la mauvaise allocation des ressources représente un coût social supérieur à celui de l'application de la réglementation. La figure 7.3 illustre une situation justifiant l'intervention publique. Etant donné le coût moyen CMo , l'excès de profit représenté par la surface $abcd$ est relativement faible par rapport à la perte pour la collectivité donnée par bem . Par contre, si bem est moins important que le coût de gestion de la réglementation, alors l'intervention ne se justifie pas.

La présence de barrières à l'entrée sur un marché peut avoir diverses origines. La plus courante est l'existence de coûts fixes irrécupérables. Quel que soit le type

de barrières, elle implique des interventions différentes selon que l'on se situe dans le contexte d'un monopole naturel global ou local.

Dans le cas d'un monopole naturel global, la firme non réglementée choisira un niveau d'output q^m alors que l'output socialement efficace est q^w (figure 7.2). Une intervention se justifie donc afin d'obliger le producteur à fournir q^w . Le problème provient du fait que p^w est alors inférieur au coût moyen, ce qui entraîne un déficit pour la firme. Le coût moyen étant décroissant, la collectivité peut avoir intérêt à laisser à l'entreprise son statut de monopole parce que la structure du marché permet de minimiser les coûts de production quel que soit le niveau d'output. Par contre, si la firme est contrainte de fixer un prix égal au coût marginal, elle doit être subventionnée, sinon elle cessera son activité, réduisant ainsi le bien-être collectif puisque tout le surplus de consommateur disparaît. Une autre solution consiste à s'éloigner d'un prix égal au coût marginal afin de laisser le producteur couvrir ses coûts sans qu'il puisse faire de profits. Cela mène à l'établissement d'un prix égal au coût moyen (q^b). Cette seconde démarche est moins efficace socialement puisque l'output est inférieur à q^w , mais reste cependant nettement préférable à q^m .

Lorsque nous sommes en présence de barrières à l'entrée, en conjonction avec un monopole naturel local, il existe un niveau d'output à partir duquel l'entrée d'une seconde firme sur le marché devient efficace. Sur la figure 7.4, le coût moyen de deux entreprises est représenté par $2C(q/2)/q$ afin de situer le point q^s qui est le niveau d'output permettant à la deuxième d'opérer. Dans l'intervalle $0 < q < q^s$, la fonction de coût est sous-additive. Une tarification efficace fixera le prix p^w (=coût marginal) et la quantité q^w . Rappelons qu'en ce point, la fonction de coût est sous-additive mais le coût moyen est croissant. Dans ce cas, aucun problème ne surgit puisque la firme peut établir un prix égal au coût marginal et obtenir un profit positif (=aire hachurée). Néanmoins, la réglementation s'avère nécessaire afin de protéger les consommateurs car, sans contrainte, la firme serait incitée à établir un prix pour lequel le coût marginal est égal à la recette marginale.

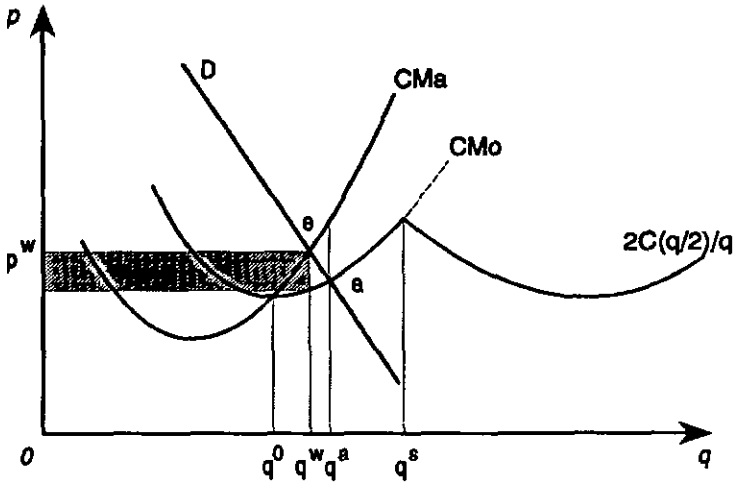


Figure 7.4 Barrière à l'entrée et sous-additivité

Supposons maintenant qu'il n'y ait aucune barrière à l'entrée. La démarche ci-après est basée sur deux hypothèses. La première stipule que les producteurs concurrents peuvent offrir un niveau d'output, soit identique, soit inférieur à celui du monopoleur en place. La seconde hypothèse considère que la sortie est sans coût, c'est-à-dire que le risque supporté par un entrant est très faible, voire inexistant. Prenons tout d'abord le cas du monopole naturel global (figure 7.3). Si la collectivité désire que l'output q^w soit produit, une réglementation est nécessaire sinon le monopoleur choisira un niveau inférieur. Celui-ci doit alors être subventionné, autrement il fera des pertes. D'autre part, si la collectivité préfère ne pas subventionner l'entreprise et la laisser couvrir ses coûts, une intervention ne se justifie plus. En effet, le monopoleur sera incité à fixer un prix qui ne lui rapporte aucun profit car une tarification générant un profit positif attirera de nouveaux concurrents sur le marché. L'entreprise en place étant indifférente à la solution choisie, la réglementation n'est pas nécessaire à moins que des considérations d'ordre social, politique ou d'environnement soient retenues.

Finalement, en présence d'une entrée libre et d'un monopole naturel local, la fonction de coût est sous-additive et le coût moyen croissant en q^w (figure 7.4). L'établissement de barrières à l'entrée s'avère nécessaire dans cette situation si l'on veut atteindre le prix p^w qui incite les firmes concurrentes à entrer sur le

marché dans la perspective de faire un profit positif représenté par l'aire hachurée. De plus, il n'existe pas de prix unique qui empêche toute entrée sur le marché. En l'absence de barrière, un concurrent pourrait toujours fixer un prix inférieur à celui du monopole, s'attribuer une part du marché et faire du profit. Supposons, par exemple, que le prix du monopoleur soit égal au coût moyen et que l'output se situe en q^a (figure 7.4). Un entrant pourrait ne satisfaire qu'une partie de la demande, ce qui lui permettrait de fixer un prix légèrement inférieur à celui du monopole et de faire un profit. On parle dans ce cas de monopole naturel mono-produit non soutenable (Curien et Gensollen 1987). Les autorités publiques sont maintenant obligées d'agir pour protéger l'entreprise en place contre de nouveaux concurrents. Si l'output maximisant le bien-être collectif se situe entre q^0 et q^a , une barrière à l'entrée doit être érigée afin d'attribuer à la firme un droit exclusif - une concession - d'exploiter le marché. Sans cet accord, aucun entrepreneur ne prendra le risque de voir d'autres producteurs entrer sur le marché.

Une forme de barrière à l'entrée pourrait être l'obligation pour tout entrant de satisfaire la totalité de la demande à un prix inférieur à celui du monopole. Si ce dernier produit à un prix égal au coût moyen et que le concurrent n'a, ni une meilleure technologie, ni un accès à des prix inférieurs pour ses inputs, alors cette solution permettra d'empêcher tout nouveau concurrent d'entrer sur le marché.

Tableau 7.1 Politique de réglementation des monopoles naturels

TYPE DE MONOPOLE	PRESENCE DE BARRIERES A L'ENTREE	ABSENCE DE BARRIERES A L'ENTREE
MONOPOLE NATUREL GLOBAL <i>(la tarification au coût marginal entraîne un déficit)</i>	Imposer $P=CMA$ et subventionner l'entreprise ou S'éloigner de la tarification au coût marginal afin d'éliminer le déficit	Imposer $P=CMA$ et subventionner l'entreprise ou Ne pas réglementer et laisser la menace d'une entrée
MONOPOLE NATUREL LOCAL <i>(la tarification au coût marginal permet un profit)</i>	Imposer $P=CMA$ et empêcher le monopole de faire un éventuel superprofit	Imposer $P=CMA$, empêcher l'entrée de concurrents sur le marché et le monopole de faire un superprofit

Source: Berg et Tschirhart, 1988

Nous résumons dans le tableau 7.1 ces différentes situations et les réponses possibles des autorités publiques. Ces interventions permettent à la collectivité d'accroître l'efficacité allocative (au sens de Pareto) grâce à une réglementation des prix et à une limitation des entrées sur le marché.

7.5. Objectifs de l'intervention publique

Lorsqu'une activité est un monopole naturel comme c'est le cas de la plupart des services en réseau, nous avons vu qu'il était préférable que le marché soit occupé par une seule firme tant qu'elle n'exploite pas sa position de monopole en fixant des prix excessifs par exemple. Nous en arrivons ainsi au centre du problème: dans une situation de monopole naturel, comment peut-on profiter d'une production efficace sans devoir supporter les désavantages d'un comportement monopolistique?

Nous ne considérerons, par souci de simplicité, que le cas d'une activité générant un seul output. L'objectif des autorités publiques est alors de maximiser le bien-être social (W) en incitant le monopole à choisir un prix qui permette de satisfaire le problème suivant:

$$\text{Maximiser } W = CS + \pi \quad (7.6)$$

$$\text{où } CS = \int_0^q p(q) dp - p(q)q \text{ et } \pi = p(q)q - C(q).$$

Le surplus du consommateur est représenté par CS , le profit du producteur par π et la fonction inverse de la demande par $p(q)$. Sur un marché de concurrence parfaite, la condition nécessaire de premier ordre de ce problème de maximisation est donnée par l'égalité entre le prix et le coût marginal. Les propriétés d'une telle situation et leurs conséquences sur le comportement des agents sont supposées être celles qui permettent d'atteindre un état économiquement efficace. Ce sont donc celles que toute société cherchera à satisfaire, même en présence d'un monopole naturel. Sans intervention, le monopoleur choisira un prix qui maximisera uniquement son profit, c'est-à-dire celui qui satisfait l'égalité entre la recette marginale et le coût marginal du producteur.

La réglementation doit alors s'adapter aux différentes situations: monopole naturel global ou seulement local, existence ou non de barrières à l'entrée. Comme nous l'avons vu, chaque cas requiert une action différente de la part des autorités publiques. Les objectifs poursuivis visent non seulement à protéger les consommateurs ou le monopole, mais également à assurer le maintien de la production.

Cependant, une intervention publique sur un marché en situation de monopole naturel ne se limite pas uniquement à rechercher un équilibre socialement optimal. Il s'agit aussi d'inciter la firme à produire efficacement le bien ou service. Puisque la fonction de coût du producteur intervient dans le problème de la maximisation du bien-être collectif, il est indispensable que le processus de production soit efficace, c'est-à-dire que la fonction de coût minimum soit satisfaite.

En situation de concurrence parfaite, les entreprises sont contraintes de produire au moindre coût sinon elles feraient des pertes et seraient obligées de quitter le marché ou de cesser leur activité. Dans le cas d'un monopole, le mécanisme de la concurrence ne fonctionne plus ou alors sous une forme atténuée bien que la théorie des marchés contestables affirme le contraire (voir Baumol, Panzar et Willig 1982). La pression sur les coûts est moins importante même si un office de contrôle est créé afin de surveiller la firme. En effet, les offices publics chargés de la réglementation ne pourront jamais être aussi motivés qu'un concurrent potentiel. Par conséquent, les managers ressentiront moins d'incitation à agir efficacement et les coûts risqueront de s'accroître. La firme produira au-dessus de son coût minimum. Leibenstein (1966) a identifié ce type d'inefficacité à l'inefficacité-X. De plus, l'entreprise jouissant d'une situation de monopole bénéficie d'une source d'informations (prix des facteurs, technologie) indisponible pour les autorités publiques. Cet avantage lui permet d'éviter des sanctions lorsqu'elle agit mal car, soit l'office de réglementation ne peut apprécier l'ampleur des écarts commis, soit il les pardonnera. On remarque donc que celui-ci ne peut évaluer toutes les décisions prises par le monopoleur. Une technologie inefficace peut être retenue et utilisée pendant plusieurs années, causant par exemple des prix plus élevés pour les consommateurs. D'autre part, sans une menace d'entrée de concurrents, l'incitation à innover est également absente d'où nouvelle source d'inefficacité, du moins à long terme.

Nous pouvons donc affirmer que les objectifs d'une réglementation ne devraient pas viser uniquement l'abaissement des prix vers les coûts de production mais

également s'assurer de l'efficacité interne de la firme. Plus cette dernière est performante, plus ses coûts seront faibles et, de ce fait, les prix pourront être réduits. L'ensemble de ces efforts contribue à l'accroissement du bien-être collectif.

7.6. Effets de la réduction des coûts de production sur le bien-être collectif

L'analyse économique du monopole naturel s'est essentiellement penchée sur les problèmes relatifs aux prix établis par le monopoleur ainsi qu'à la relation entre ceux-ci et la structure des coûts de la production. Quelles que soient les solutions proposées, les problèmes liés à une production non efficace n'ont suscité que peu d'intérêt. Pourtant, il apparaît que l'établissement de prix efficaces, c'est-à-dire ceux répondant à l'objectif de maximisation du bien-être social, ne peut être réalisé qu'à partir du moment où l'entreprise agit de manière à minimiser ses coûts. Ce point important a été mis en lumière par Waterson (1988). Prenons une firme en situation de monopole, en l'absence de réglementation, elle fixera un prix lui permettant de maximiser son profit. Supposons maintenant qu'un accroissement de l'efficacité technique dans son processus de production permette une réduction du coût moyen de AC_1 vers AC_2 (figure 7.5). D'une part, les consommateurs obtiennent une augmentation de leur surplus de $efkj$ dont une partie ($efnj$) représente un transfert de l'entreprise. D'autre part, le profit de la firme augmente également de $efgh$ en $jklm$. Le gain du bien-être collectif est donc égal à $efkj+jklm-efgh$, représenté par l'aire hachurée sur la figure 7.5.

Considérons maintenant une intervention des autorités publiques et supposons que les options à la disposition d'une agence de réglementation permettent d'assurer:

- soit une réduction de x pour-cent des prix pratiquée par le monopole,
- soit une réduction de x pour-cent des coûts de production,

avec un coût de mise en oeuvre identique, quelle que soit la solution choisie. Le problème consiste alors à déterminer quelle est la meilleure option.

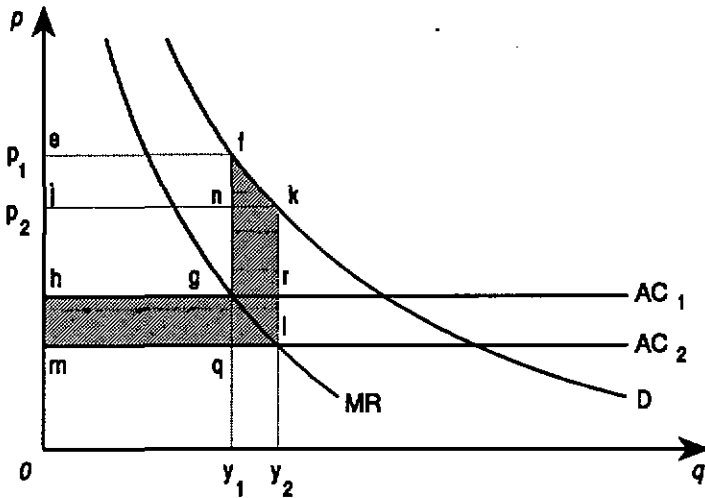


Figure 7.5 Effets de la réduction des coûts de production

Supposons que l'élasticité-prix de la demande du service q soit approximativement constante sur l'intervalle des productions possibles pour l'entreprise. Si elle n'est pas réglementée, cette dernière cherchera à égaliser son coût marginal et sa recette marginale afin de maximiser son profit, c'est-à-dire

$$p(q^*) + p'(q^*)q^* = C'(q^*) \quad (7.7)$$

où le membre de gauche de l'équation (7.7) représente la recette marginale du monopole, le membre de droite son coût marginal et q^* le niveau d'output permettant de maximiser le profit. Les conditions de premier ordre peuvent être exprimées en utilisant l'élasticité-prix de la demande qui est donnée par

$$\varepsilon(q) = \frac{p}{q(p)} \cdot \frac{dq(p)}{dp}, \quad \varepsilon(q) < 0 \quad (7.8)$$

où p est le prix du service q et $q(p)$ la fonction de demande.

En substituant l'expression (7.8) dans l'équation (7.7), nous obtenons (cf. Varian 1984, p. 81).

$$p(q^*) \left[1 + \frac{1}{\varepsilon(q^*)} \right] = C^*(q^*) \quad (7.9)$$

où $p(q^*)$ est la fonction inverse de demande. On remarque, grâce à (7.9), que le prix fixé par le monopole est directement lié à sa fonction de coût. Autrement dit, une variation des coûts entraîne un changement du prix dans les mêmes proportions. Pour la simplicité de la démonstration, supposons un coût marginal égal au coût moyen, ce dernier étant constant et prenons $C_1 > C_2$, deux niveaux de coût. Ainsi une réduction de x pour-cent $(= (C_1 - C_2)/C_1)$ du coût implique une diminution identique du prix; nous pouvons affirmer à partir de l'expression (7.9) que

$$\frac{C_1 - C_2}{C_1} = \frac{p_1^* - p_2^*}{p_1 - \eta} = \frac{p_1 - p_2}{p_1} \quad (7.10)$$

$$\text{où } \eta = \left| \frac{1}{1 + \varepsilon(q^*)} \right|, \quad p_1 = p_1(q^*) \text{ et } p_2 = p_2(q^*).$$

Par conséquent, une réduction du coût de la production entraîne une amélioration du bien-être collectif (aire *fkrjg* sur la figure 7.5) au travers de la diminution du prix de vente du service. Si cette compression des coûts est due à un accroissement de l'efficacité dans le processus de production, le bien-être social augmente d'un montant égal à la surface *hrlm*. Nous pouvons donc voir, grâce à cette démonstration, que la réduction des coûts a non seulement le même effet qu'une baisse du prix mais comporte encore d'autres conséquences positives. C'est la raison pour laquelle il faudrait chercher d'abord à inciter les entreprises à compresser leurs coûts de production avant de réglementer le marché.

Cette analyse montre bien l'importance du contrôle des coûts d'un monopole naturel. Etant donné une structure du marché économiquement efficace – une seule entreprise, le maintien d'un prix égal à un niveau de production inefficace peut avoir un effet moindre sur le bien-être collectif qu'une forme de contrôle où les prix seraient élevés mais où les coûts refléteraient une production performante. Cela signifie qu'une tarification excessive³ est préférable jusqu'au

³ C'est-à-dire l'établissement d'un prix de la part de l'entreprise unique qui lui permette de jouir d'une rente de monopole.

point où elle entraîne une perte de bien-être égale au gain obtenu grâce à la production au coût minimum.

En raison de ces conséquences importantes sur le bien-être collectif, nous examinerons en détail l'efficacité productive des entreprises en réseau dans l'analyse empirique. Nous chercherons également à expliquer pourquoi elles s'éloignent parfois du niveau idéal.

7.7. Les différentes réglementations des entreprises en réseau

La maximisation du bien-être collectif dans un secteur caractérisé par une situation de monopole naturel, tel que celui des entreprises en réseau, ne peut être réalisée qu'en conjonction avec une intervention publique. Celle-ci peut prendre de nombreuses formes et être plus ou moins contraignante. Nous examinerons dans ce chapitre une forme très répandue aux Etats-Unis depuis de nombreuses années et ayant gagné récemment une popularité croissante au Royaume-Uni, il s'agit de la réglementation d'entreprises privées par un office public. Waterson (1988) définit la réglementation comme "the control of an industrial activity by government in the sense of actions such as restrictions on firms entering the industry, constraints on firms actually in the industry, or both. We might add to these, government powers affecting substitutes and complements and direct subsidizing of activities."

L'objectif de la réglementation est principalement de garder sous contrôle le pouvoir du monopole. Les moyens utilisés – contrôle du taux de rendement du capital et contrôle direct des prix – visent essentiellement à maintenir les prix à un niveau équivalent à celui du point mort (*breakeven*), c'est-à-dire celui où l'entreprise ne fait ni pertes ni profits. Bien que considéré comme moins important, l'objectif d'un office de réglementation peut être d'assurer le maintien d'un certain niveau de qualité.

Nous commencerons par examiner la réglementation portant sur le taux de rendement du capital puisque c'est la forme qui a reçu la plus grande attention de la part des études aussi bien théoriques qu'empiriques.

7.8. Réglementation par le taux de rendement du capital

Considérons une branche économique dans laquelle il n'y a qu'une seule et unique entreprise réglementée. Celle-ci cherchera à maximiser son profit bien que soumise à une contrainte qui lui impose un taux de rendement spécifique s

sur son capital. s est établi de façon à se situer dans un intervalle où la borne inférieure (r) est représentée par le coût du capital de la firme et la borne supérieure (r_m) est le rendement maximum que le monopole obtiendrait en l'absence de réglementation, c'est-à-dire $r < s < r_m$. Averch et Johnson (1962) se sont penchés sur l'impact d'une telle contrainte dans le choix des inputs effectué par l'entreprise. Ils ont démontré analytiquement que celle-ci avait tendance à devenir surcapitalisée et inefficace. Ce résultat est connu sous le nom de "*Averch-Johnson effect*". Afin de montrer les effets de cette distorsion sur le bien-être collectif, il s'agit d'examiner brièvement le modèle.

Supposons qu'une entreprise réglementée produise un seul output (q) à partir de deux inputs x_1 et x_2 qui sont, respectivement, le travail et le capital. Les prix des facteurs de production sont représentés pour le travail par w et pour le capital par r . L'output est produit selon une fonction de production $q = f(x_1, x_2)$ et la fonction inverse de la demande de q est $p(q)$. Nous obtenons, à partir de ces deux dernières expressions, le revenu de la firme $R = p(q)q$. Celle-ci s'efforcera de maximiser son profit par rapport à x_1 et x_2

$$\pi = R - wx_1 - rx_2 \quad (7.11)$$

$$\text{sous la contrainte: } \frac{R - wx_1}{x_2} = s$$

Pour résoudre ce problème, nous pouvons utiliser la méthode du multiplicateur de Lagrange et obtenons:

$$L = R - wx_1 - rx_2 + \lambda(sx_2 + wx_1 - R) \quad (7.12)$$

où λ est le multiplicateur de Lagrange. En supposant que x_1 , x_2 et $\lambda > 0$, les conditions de premier ordre impliquent que

$$(1-\lambda)\partial R/\partial x_1 - (1-\lambda)w = 0 \quad (7.13)$$

$$(1-\lambda)\partial R/\partial x_2 - (1-\lambda)r = \lambda(r-s) \quad (7.14)$$

Sans contraintes, le rapport des prix des inputs est égal au rapport de leur productivité marginale sous les conditions de minimisation du coût. Dans le cas

d'une entreprise réglementée, nous obtenons, en divisant l'expression (7.14) par (7.13):

$$\frac{f_2}{f_1} = \frac{r}{w} - \frac{\lambda}{1-\lambda} \cdot \frac{s-r}{w} \quad (7.15)$$

où $f_i = \frac{\partial f(x_1, x_2)}{\partial x_i}$.

Etant donné que $0 < \lambda < 1$ et $s > r^4$, la firme ne minimise pas le coût de la production car $f_2/f_1 < r/w$. Cela démontre qu'il y a surcapitalisation. Ce biais ne signifie pas pour autant que l'entreprise réglementée choisira d'utiliser plus de capital ou de produire plus d'output que si elle n'était pas soumise à une réglementation. Cette remarque est très importante. En effet, si la réglementation du taux de rendement entraîne une augmentation significative de la quantité d'output par rapport à celle que produirait un monopole maximisant son profit sans contrainte, alors cette inefficacité de la production peut être tolérée. Toutefois, l'accroissement de l'output ne peut être garanti bien qu'il soit vraisemblable (Baumol et Klevorick 1970).

Le problème de ce type de réglementation est donc le suivant. En supposant qu'il y ait une augmentation de l'output, il s'agit de trouver le niveau pour lequel l'effet socialement désirable de cette variation n'est pas entièrement compensée par la perte de bien-être résultant de la distorsion dans les proportions des facteurs utilisés, ce qui implique un coût de production accru.

7.9. Réglementation par les prix

Les modèles de réglementation par les prix ont obtenu moins de succès que ceux impliquant une contrainte sur le taux de rendement du capital. Néanmoins, ils ont commencé à prendre une certaine importance depuis que le Royaume-Uni a décidé d'appliquer une variante de l'un de ces modèles à la réglementation de British Gas Corporation. Nous considérerons tout d'abord une méthode proche du modèle Averch-Johnson, ensuite nous examinerons la technique utilisée par l'administration britannique.

⁴ Pour la démonstration, voir Berg et Tschirhart 1989

La première technique consiste à réglementer le rapport entre le profit et les recettes en fixant une limite supérieure t à ce ratio. Dans ce cas, l'entreprise cherche à résoudre le problème suivant:

$$\text{Maximiser } \pi = R - wx_1 - rx_2 \quad (7.16)$$

$$\text{sous la contrainte: } \frac{R - wx_1 - rx_2}{R} \leq t$$

En résolvant ce système grâce à la méthode de Lagrange et en dérivant par rapport à chacun des deux inputs, on obtient les conditions du premier ordre:

$$\frac{\partial L}{\partial x_1} = MR_1 - w - \lambda[MR_1(1-t) - w] = 0 \quad (7.17)$$

$$\frac{\partial L}{\partial x_2} = MR_2 - r - \lambda[MR_2(1-t) - r] = 0 \quad (7.18)$$

En réarrangeant ces expressions et en divisant (7.18) par (7.17), cela donne:

$$\frac{f_2}{f_1} = \frac{w}{r} \quad (7.19)$$

L'entreprise n'a donc plus tendance à utiliser un rapport capital-travail qui ne minimise pas le coût de la production de q . Par conséquent, il n'y a pas de perte directe de bien-être collectif. Une réglementation de ce type, en contraignant les prix à se situer au-dessous de ce qu'ils seraient autrement, améliore le bien-être en réduisant les effets négatifs du monopole.

Il subsiste cependant un problème pratique important en ce qui concerne l'établissement de t . En effet, alors qu'il est relativement aisé, pour le régulateur, de se faire une idée sur un taux de rendement équitable du capital, la perception d'un rendement équitable en ce qui concerne le chiffre d'affaires est beaucoup plus vague. En raison de cette difficulté majeure, cette méthode n'a été que rarement utilisée.

L'autre possibilité est une réglementation grâce à laquelle les prix sont fixés de manière exogène. On laisse le monopole augmenter ses prix jusqu'à concurrence

de l'indice des prix à la consommation moins un certain pourcentage, x , qui représente les gains de productivité attendus dans l'activité en question. Cette méthode, appelée RPI-X (pour *retail price index minus x*), a pour but d'obliger la firme à adapter la combinaison de ses facteurs de production aux variations des prix de ces derniers.

Ce type de réglementation permet, d'une part, de contraindre la firme à produire plus efficacement car le facteur x , pour autant qu'il puisse être estimé correctement par les autorités publiques, va contraindre l'entreprise à réaliser les gains de productivité techniquement possibles. D'autre part, tant que x reste inchangé, la méthode RPI-X permet aux entreprises d'intérioriser les avantages provenant de progrès techniques supérieurs à ceux qui étaient attendus.

7.10. Réglementation par la qualité

Dans les modèles de réglementation examinés jusqu'ici, nous avons considéré la qualité du service fourni par un monopole naturel comme une variable exogène. En réalité, nous devrions tenir compte de la qualité offerte puisqu'elle est aussi importante que le prix dans la détermination du bien-être collectif. Qu'elle soit imposée par les autorités publiques ou fixée librement par le monopole, elle est influencée par toute activité réglementaire.

D'un point de vue théorique, Spence (1975) a montré que la prise en considération de la qualité dans un schéma de réglementation d'un monopole risquait d'introduire un élément difficile à estimer. Idéalement, l'office de réglementation devrait créer un lien entre le prix et la qualité en confrontant l'entreprise à une fonction de réaction qui refléterait le taux de substitution entre le prix et la qualité exprimé par la demande sur le marché. Le problème consiste à estimer ce taux de substitution car sa détermination exige la connaissance de la valeur attachée à la qualité par l'ensemble des consommateurs infra-marginaux. A moins d'entreprendre des enquêtes directement auprès de la population, il est peu vraisemblable que l'office de réglementation puisse réellement utiliser un tel schéma.

Toutefois, comme le démontre Spence, la réglementation du taux de rendement du capital pourrait être utilisée dans certaines circonstances comme une stratégie de l'optimum de second rang (*second best*), c'est-à-dire lorsque:

- la qualité maximisant le profit tend à être trop faible et

- la qualité est capital-intensive, c'est-à-dire qu'elle augmente lorsqu'on substitue du capital au travail.

La tendance pour une entreprise à surcapitaliser lorsqu'elle est soumise à une contrainte sur le taux de rendement compense alors, partiellement du moins, la propension à offrir une qualité trop faible. Si par contre, l'augmentation de la qualité exige un surcroît du facteur travail, ce qui est le cas pour de nombreux services, la réglementation du taux de rendement risque de détériorer encore le niveau de la qualité offerte.

Pazner (1975) a également examiné l'effet sur la qualité d'une réglementation par les prix. Il propose une réglementation qui ne modifie pas la qualité du service fourni par le monopole.

Lorsque la réglementation impose un prix $p=\bar{p}$, la qualité, représentée par t sur la figure 7.6, sera inférieure à son niveau optimal (t^m), c'est-à-dire à celui existant avant l'intervention publique. Le point t^m est le niveau de qualité pour lequel le monopole maximise son profit. Comme nous pouvons le constater, quelle que soit la quantité vendue, l'entreprise monopoliste réagit à une fonction de demande pour la qualité $-p(t)$ est l'inverse de cette fonction - qui est non pas *abc* mais *abd*. Elle choisit alors un niveau de qualité de $t=\bar{t}$ car elle ne peut obtenir un prix plus élevé en offrant une qualité supérieure à \bar{t} alors que, au-delà de ce point, son coût moyen $c(t)$ augmente.

Il est en revanche impossible de prévoir si la quantité demandée va augmenter ou diminuer à la suite d'une réglementation par les prix. Toute variation de la quantité demandée est influencée par deux facteurs. Premièrement, par la qualité du service. Une réduction du niveau de la qualité signifie, pour un prix inchangé, que la courbe de demande en fonction de la quantité va se déplacer vers la gauche. Autrement dit, les consommateurs seront moins incités à acheter un bien de qualité inférieure. Deuxièmement, par le prix réglementé (\bar{p}) qui sera plus faible que le prix p^m , fixé librement par le monopole. En conséquence, selon l'ampleur du déplacement de la courbe de demande, la quantité demandée pourra être supérieure ou inférieure à celle existant en régime de monopole. Deux situations peuvent donc se présenter. Si la quantité demandée diminue, alors la réglementation par les prix est inefficace puisque la qualité baisse également. En revanche, si la quantité augmente, il y aura une amélioration de la situation par rapport à celle qui prévalait avec un régime de monopole non réglementé.

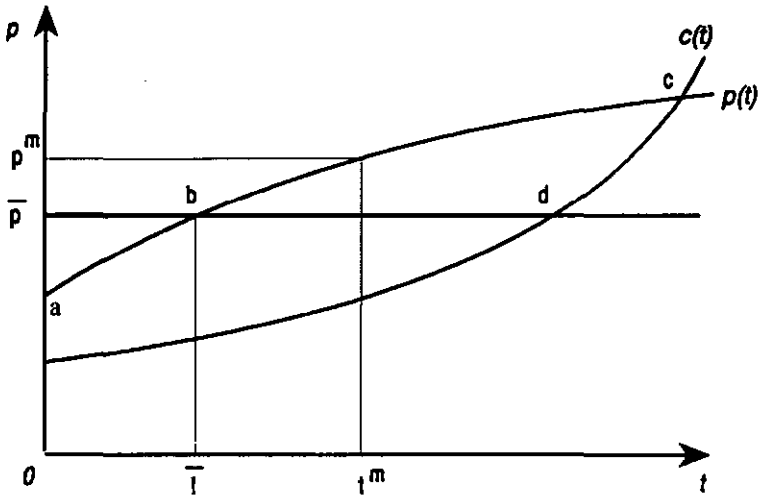


Figure 7.6 Effet sur la qualité d'une réglementation par les prix

Selon Pazner, la réglementation optimale consiste à proposer au monopole une subvention par unité d'output qui soit totalement indépendante de la qualité. Cette solution permettra d'inciter l'entreprise à produire un niveau égal à celui d'une situation de concurrence parfaite. Dans le cas où la firme n'atteindrait pas la production désirée, l'autorité de réglementation lui imposerait alors une taxe sur chaque unité produite. Puisque nous faisons l'hypothèse que le coût est uniquement une fonction de la qualité, il ne sera pas influencé par la subvention ou la taxe. L'objectif étant d'obtenir un prix identique à celui d'une concurrence parfaite, la subvention optimale doit donc être égale à la différence entre le prix et la recette marginale pour une quantité d'output correspondant à un marché concurrentiel. Cette proposition entraîne toutefois plusieurs inconvénients. Cela nécessite tout d'abord, pour l'autorité de contrôle, de connaître un certain nombre d'éléments, dont la fonction de demande du service – information que seule la firme connaît. De plus, Pazner suppose que la qualité n'est pas modifiée quelle que soit la situation (monopole, concurrence ou réglementation) et donc que le seul instrument de la subvention est suffisant pour parvenir à l'équilibre prix-quantité-qualité désiré. Cette hypothèse n'est certainement pas réaliste. Il faudrait alors multiplier les instruments – un pour la qualité, un autre pour la quantité – ce qui rend d'autant plus complexe l'application de la réglementation.

8. Alternatives à la réglementation

8.1. Propriété publique ou gestion directe

8.1.1. Objectifs

Le principal but poursuivi par la propriété publique concerne la maximisation du bien-être collectif (W) émanant du service en question. Rappelons que le bien-être collectif est, en général, défini comme la somme des surplus du consommateur et du producteur. Puisque n'importe quel accroissement des prix au-dessus du coût marginal entraîne une perte de bien-être, la solution la plus évidente consiste à fixer le prix au niveau du coût marginal.

Le problème devient:

$$\text{Maximiser } W = \int_0^q p dq - wx_1 - rx_2 \quad (8.1)$$

où $p = p(q)$

$q = q(x_1, x_2)$

$w = w_c$ (prix du travail)

$r = r_c$ (prix du capital)

La solution de ce problème implique l'efficacité technique et allocative dans la production, une vente efficace grâce à l'identification correcte de la fonction de demande et une allocation des ressources efficace par une juste détermination de l'output q_0 , c'est-à-dire où le prix égale le coût marginal.

Dans le contexte d'une concurrence parfaite, les contraintes sont supposées être satisfaites étant donné les effets de la "main invisible". Dans le cas d'un monopole, les contraintes ne sont pas automatiquement satisfaites et, par conséquent, la fonction objectif n'est pas maximisée. Les autorités publiques établissent alors des objectifs de gestion et imposent des contraintes afin d'atteindre le résultat souhaité.

8.1.2. Raisons

Une raison souvent invoquée dans les débats politiques est que la gestion directe assure la propriété publique des moyens de production. Un autre argument en faveur de la propriété publique est qu'elle facilite la planification économique de secteurs-clés de l'économie. Par exemple, la propriété publique des industries du charbon, du pétrole, du gaz et de l'électricité permet une planification plus aisée du secteur de l'énergie. Elle rend également plus aisée les mesures de redistribution du revenu et peut donc être désirable lorsque l'on veut donner un accès universel à certains services essentiels.

Lorsqu'une industrie crée des externalités, nous avons vu que la propriété publique permet souvent de résoudre plus directement les problèmes qui y sont liés. En effet, si l'objectif est plus large que la maximisation du profit, le contrôle d'une entreprise par les actionnaires est probablement moins approprié que le contrôle par l'Etat.

Enfin, certains affirment qu'au sein des firmes publiques, il y a moins de tensions entre gestionnaires et employés car les profits, s'ils existent, ne vont pas spécifiquement aux propriétaires du capital.

8.2. *La concession*

Une des manières de résoudre le problème du monopole naturel est d'établir une distinction entre le concept de *concurrence sur un marché* et celui de *concurrence pour un marché*. La concurrence pour obtenir le droit de desservir seul un marché permet de profiter aussi bien de la production par une seule firme, – minimisation des coûts de l'industrie – que du processus concurrentiel obligeant la firme à agir efficacement.

Les nombreuses situations où une entreprise unique fournit un service au sein d'une aire géographique déterminée proviennent souvent de l'octroi d'une concession par les autorités publiques locales. Quelles que soient les raisons pour lesquelles une firme se voit attribuer ce droit, il semble que ce processus ne réponde pas aux conditions économiques qui permettent de maximiser le bien-être collectif. C'est pourquoi nous allons examiner plus attentivement la proposition énoncée ci-dessous.

8.2.1. *Soumission de la concession*

L'idée de base du modèle de soumission de la concession est la suivante: mettre en concurrence au moment de la conclusion du contrat plusieurs fournisseurs potentiels. On utilise à cet effet une procédure d'enchères au travers de laquelle la concession est attribuée, pour un marché déterminé, à l'entreprise qui a proposé le prix de vente le plus bas (Berg et Tschirhart 1988). Ce processus devrait amener le producteur à fixer un prix égal au coût moyen de façon à éviter la mise en place d'une réglementation du monopole. Cette solution, très proche de celle proposée par Demsetz (1968), montre que la sous-additivité des coûts de production menant à une firme unique ne signifie pas forcément que le prix sera déterminé par un processus de maximisation du profit. Cependant, cette forme de concurrence n'est possible que si deux conditions sont satisfaites. Premièrement, les facteurs de production doivent être disponibles pour tous les soumissionnaires au travers de marchés ouverts où les prix sont établis de manière concurrentielle. Deuxièmement, le coût d'une entente entre entreprises rivales doit être dissuasif de façon à ce que la soumission soit le résultat d'un processus d'enchères (Braeutigam 1989). En créant une concurrence pour l'obtention du marché, la procédure de soumission rend ce marché plus contestable qu'il ne le serait autrement. Rappelons que la contestabilité signifie l'absence de coûts irrécupérables et la possibilité, pour les entrants potentiels, de s'introduire sur le marché en s'accaparant une part de celui-ci avant que la firme en place ne réagisse. Bien qu'en théorie le processus de mise aux enchères d'une concession semble très simple, son application soulève plusieurs questions:

- Combien d'entreprises faut-il inviter à soumissionner ?
- Quelle sera l'efficacité du vainqueur ?
- Comment évaluer les différentes soumissions?
- Comment imposer le respect du contrat?
- Comment déterminer la durée du contrat?

8.2.2. *Nombre d'offres*

Le but est d'obtenir autant d'offres que possible afin d'avoir le meilleur prix. Cependant, cela n'est pas aisé pour deux raisons:

- a) Même en supposant que les devis sont gratuits pour le client, ils entraînent malgré tout un certain coût: rendez-vous, discussion des exigences et évaluation des offres.
- b) Chaque firme peut, en ayant connaissance du nombre des autres soumissions, évaluer son rendement et proposer un prix en conséquence. Si par exemple une firme a une chance sur six de gagner le contrat, elle a une probabilité de $5/6$ de faire une perte égale aux coûts de préparation de la soumission. La décision de soumissionner ne se fera que si le gain résultant d'une offre gagnante dépasse ce coût. Cela signifie que, plus le nombre de soumissions augmente, plus l'offre gagnante sera élevée et ceci d'autant plus lorsque les offres sont proches les unes des autres et que le coût de leur préparation est élevé. Si les firmes doivent internaliser le coût d'un devis gratuit, il est donc préférable de solliciter peu de soumissions afin d'obtenir une offre gagnante plus avantageuse.

8.2.3. *Efficacité des prix*

Nous savons que le processus d'enchères mène dans le meilleur des cas à un prix égal au coût moyen. Cependant, une telle situation n'est pas idéale économiquement. Il serait préférable d'avoir un prix égal au coût marginal permettant de maximiser le bien-être collectif. Demsetz(1968) affirme que si la différence entre coût moyen et coût marginal est importante, il faut établir plusieurs prix le long de la courbe de demande. L'évaluation des tarifs devient alors problématique.. Deux solutions sont proposées:

- un contrat entre les soumissionnaires et les clients potentiels;
- les autorités publiques choisissent la structure du tarif.

La première est avant tout théorique et difficilement applicable car elle exigerait qu'un contrat soit conclu avec chaque consommateur. Avec la seconde solution, il existe le risque d'une mauvaise estimation de la fonction de demande d'où un tarif ne correspondant pas à la situation réelle du marché.

8.2.4. *Evaluation des soumissions*

Choisir l'entreprise la plus avantageuse n'est pas suffisant, il faut encore tenir compte de la qualité offerte. La tâche de l'évaluateur est plutôt de retenir la proposition offrant le meilleur rapport qualité/prix, ce qui requiert

inévitablement un jugement de valeur. Par exemple, dans le cas d'un service de transports urbains, comment faire son choix entre une soumission proposant un service meilleur marché mais peu fréquent et une autre plus chère proposant une fréquence plus élevée. En outre, il faut tenir compte du niveau de confort et de sécurité que l'on désire.

Le problème consiste donc à savoir ce que le régulateur doit exiger et ce qu'il doit laisser au jugement des soumissionnaires.

8.2.5. Application du contrat

Le contrat ne peut tenir compte de tous les éléments entrant dans la fourniture du service, particulièrement lorsqu'il s'agit de domaines tels que les télécommunications, la distribution d'électricité ou les transports publics. Certaines incertitudes persistent aussi bien pour les autorités publiques que pour les entreprises qui soumissionnent. Par conséquent, le respect rigoureux du contrat est difficile à exiger dans un tel environnement. Que peut faire l'administration publique dans ce cas? Une solution peu efficace serait de payer le concessionnaire le moins possible et de relancer la procédure de soumission. A ce moment le concessionnaire dispose d'un avantage considérable, celui de l'expérience. Il a donc des atouts pour faire une offre plus intéressante que celles des autres entreprises. Il serait alors préférable de renégocier le contrat avec la firme originale, d'autant plus que l'établissement d'une nouvelle entreprise requiert de nouveaux délais dans la mise en oeuvre du service et engendre des coûts supplémentaires.

Dans un environnement incertain, il y a donc une forte probabilité que les termes du contrat ne soient pas totalement respectés. Si les soumissionnaires savent qu'il existe une telle incertitude et si l'office de réglementation en est conscient, les offres auront tendance à être très optimistes en ce qui concerne les prix, la quantité, les délais de mise en oeuvre, etc., puisque les entreprises peuvent s'attendre à obtenir certains ajustements durant l'application du contrat, en particulier au niveau des prix.

La situation ressemble à un jeu avec une prise de décision séquentielle entre le principal – les autorités publiques – et l'agent – la firme sous contrat (Lewis 1986). En somme, il faut considérer les propositions initiales comme étant optimistes et les évaluer en conséquence.

Lorsque l'Etat constate que la performance du concessionnaire est vraiment trop faible par rapport à ce qui était prévu, il n'a parfois pas d'autre solution que de casser le contrat. Une telle décision peut s'avérer particulièrement difficile à prendre dans les cas relatifs à la distribution d'eau et d'électricité où la continuité de l'offre du service doit absolument être assurée. Le problème nous amène à la question de la durée du contrat.

8.2.6. Durée du contrat

Lorsque le concessionnaire possède la totalité des équipements nécessaires à l'exploitation du service, l'évaluation des actifs et les incertitudes concernant les prix futurs sont les deux principaux problèmes rencontrés dans l'établissement de la durée du contrat. Comme solution partielle, Demsetz (1968) propose un contrat incomplet à long terme alors que Posner (1974) suggère un contrat récurrent à court terme. Cette dernière solution permet de résoudre aisément le problème de l'incertitude. Par exemple, si le contrat portant sur la distribution de l'électricité est limité à deux ans, on peut s'attendre à ce que l'entreprise spécifie ses tarifs précisément ou, du moins, définisse exactement les circonstances entraînant une modification de ceux-ci. Par contre, un contrat de ce type ne résout pas le problème de l'évaluation des équipements. Si la firme prévoit un contrôle de ses actifs sur deux ans seulement, elle ne sera pas incitée à investir pour les améliorer. De plus, si le contrat n'est pas renouvelé, ce problème devient encore plus important. Le nouveau concessionnaire n'a pratiquement pas d'autre solution que de racheter les équipements. S'il est obligé d'en construire de nouveaux, d'une part, son coût marginal sera plus élevé que celui de l'ancien concessionnaire et, d'autre part, sera économiquement inefficace puisque la fourniture de ce service est caractérisée par des coûts irrécupérables. Les autorités publiques pourraient alors inclure dans le contrat l'obligation, pour l'ancien concessionnaire, de vendre le capital fixe à son successeur. Le prix de vente reste toutefois un problème difficile à résoudre. Le propriétaire cherchera à vendre à un prix lui permettant de couvrir le coût de la construction d'un système similaire, alors que l'acheteur désirera payer un prix proche de la valeur intrinsèque des équipements.

Le contrat incomplet à long terme cherche justement à apporter une réponse à cette question. Sur une période plus longue, dix ans par exemple, on peut estimer que des investissements effectués pendant la durée du contrat se révèlent valables pour l'entreprise. De cette façon, aussi bien le concessionnaire qu'un

entrant potentiel savent qu'ils auront à investir, ainsi la concurrence *ex ante* est plus étendue que dans le cas du contrat récurrent à court terme.

Cependant, Williamson (1976) remarque que l'incertitude s'accroît substantiellement avec le contrat incomplet à long terme. Premièrement, il n'est pas possible de fixer précisément un prix ni même de prévoir exactement son évolution durant la période du contrat car de nombreux éléments difficilement prévisibles entrent en considération:

- a) la demande pour le service peut varier, impliquant sa prévision afin de pouvoir établir un prix d'équilibre et déterminer la taille des équipements;
- b) la qualité étant liée aux décisions du concessionnaire en matière d'investissements, elle ne peut être établie que de façon peu précise.

La seconde difficulté majeure d'un contrat incomplet à long terme concerne l'application de sanctions lorsque le concessionnaire s'éloigne trop des exigences du contrat. La saisie des équipements par les autorités publiques est une mesure beaucoup trop sévère pour des négligences mineures. Aucune entreprise ne voudra prendre ce risque. De même, la mise à l'amende de la compagnie faisant une erreur peut sembler arbitraire lorsque l'on considère le nombre d'inconnues auxquelles doit faire face le concessionnaire.

Par conséquent, la "persuasion" devrait être la principale action à exercer dans les premières années du contrat.

TROISIEME PARTIE

L'EFFICACITE DES SERVICES PUBLICS EN RESEAU:

APPLICATION AUX ENTREPRISES SUISSSES DE TRANSPORTS URBAINS

9. Introduction

Nous avons vu jusqu'ici comment une entreprise en situation de monopole naturel se comportait. Les conséquences économiques d'une intervention publique ont également été examinées d'un point de vue normatif et positif et nous ont montré que la maximisation du bien-être collectif était, en réalité, un objectif difficile à atteindre. Il y a cependant des solutions mieux adaptées au cas particulier du monopole naturel.

Afin d'identifier les formes d'intervention des autorités publiques qui permettent d'atteindre la situation la plus favorable, nous nous sommes penchés sur la mesure de l'efficacité des entreprises de transports publics urbains. Ce service possède les caractéristiques d'un service en réseau et fait partie de la catégorie des biens collectifs. En se concentrant sur la mesure de l'efficacité productive de chaque firme, notre étude empirique cherche à identifier quelle est la situation la plus incitative pour les gestionnaires en ce qui concerne la minimisation des coûts de production. Le problème de la sous-additivité de la fonction de coût demande à être également examiné puisqu'il permet de justifier l'existence d'une firme unique au sein de l'aire géographique desservie. Cette question concerne plus particulièrement les grandes compagnies exploitant souvent plusieurs réseaux interconnectés. Il s'agit donc de déterminer si la structure du marché justifie l'existence d'un monopole ou s'il serait économiquement plus favorable de partager celui-ci entre plusieurs entreprises.

Notre démarche consistera, dans un premier temps, à estimer économétriquement l'efficacité productive de chaque firme observée, ce qui nous permettra de comparer sa performance en fonction du mode d'intervention publique. Nous nous pencherons ensuite sur les caractéristiques de la structure du marché afin de déterminer si la présence d'un monopole est justifiée ou non.

L'estimation de l'efficacité productive a été réalisée à partir d'une frontière de coût translog obtenue grâce aux propriétés de la théorie de la dualité. Le modèle économétrique repose quant à lui sur la méthode *seemingly unrelated regression* de Zellner (1962) qui est une technique des moindres carrés généralisés. Afin de tester leur convergence, deux types de frontières ont été retenues, l'une déterministe obtenue à partir de la technique du résidu minimum, l'autre stochastique estimée à l'aide de la méthode proposée par Jondrow et al. (1982). Nous avons introduit dans les frontières de coût des variables d'environnement et de qualité dans le but de mettre autant que possible toutes les observations dans une situation identique.

Enfin, nous avons supposé que les transports urbains étaient un monopole naturel lorsque nous étions en présence de rendements d'échelle croissants. Cette hypothèse simplificatrice par rapport à celle de sous-additivité des coûts de production a été retenue en raison de sa facilité à être estimée directement à partir de la frontière de coût.

10. Modélisation de la structure de la production

L'estimation de l'efficacité productive exige que l'on spécifie une fonction de production ou de coût (paramétrique ou non-paramétrique). Nous avons vu au chapitre 3 que cette dernière pouvait être évaluée de différentes façons. Puisque notre approche se base sur un modèle économétrique, cela exige la spécification d'une forme paramétrique. Les technologies traditionnelles telles que Leontieff, Cobb-Douglas et CES exigent qu'un certain nombre de restrictions soient placées a priori sur certains effets économiques du processus de production. Toutefois, les développements relativement récents de la théorie de la dualité ont rendu possible l'utilisation de formes imposant moins de restrictions. Vu l'hétérogénéité des compagnies de transports de notre échantillon, l'emploi de formes fonctionnelles plus flexibles permet à certaines caractéristiques du processus de production, telles que l'élasticité d'échelle ou de substitution, de dépendre des données. Ces caractéristiques peuvent varier à travers l'échantillon et ne doivent pas être paramétrisées comme c'est le cas pour les formes plus traditionnelles.

10.1. Dualité entre fonctions de coût et de production

Un résultat fondamental de la théorie de la dualité est qu'il existe des fonctions de coût et de production qui sont le dual l'une de l'autre, étant donné certaines conditions de régularité. La spécification d'une fonction de production implique ainsi une fonction de coût particulière et vice-versa. Par conséquent, la structure de production peut être étudiée empiriquement en utilisant soit une fonction de coût, soit une fonction de production. La dualité implique que la fonction dérivée de la demande d'un facteur quelconque puisse être obtenue en différenciant partiellement la fonction de coût par rapport au prix du facteur en question.

Nous supposons dans notre modèle que, en raison de l'intervention de l'Etat, l'objectif de chaque firme est de minimiser ses coûts de la production. Le prix du transport fixé par les autorités publiques est donc une donnée pour l'entreprise. A court terme, le profit, s'il est possible, ne peut être maximisé qu'en minimisant les coûts de production. De plus, il est raisonnable de considérer l'output comme étant exogène aux décisions des entreprises. En effet, ces dernières doivent satisfaire la demande aux prix établis et ne peuvent choisir le niveau d'output qui maximiserait leur profit. Enfin, on peut supposer également que les compagnies

de transports urbains s'approvisionnent en facteurs de production sur un marché concurrentiel; le prix de ceux-ci est donc exogène aux décisions de la firme. Ces différentes raisons ainsi que les possibilités offertes par la théorie de la dualité, nous ont amenés à estimer une fonction de coût car aussi bien le prix des inputs que l'output sont des variables exogènes. L'estimation économétrique d'une fonction de production entraînerait en effet un biais dans le cas des transports urbains puisqu'il est vraisemblable que les quantités d'inputs utilisées sont déterminées par les gestionnaires et sont donc endogènes au processus de production.

Etant donné l'hypothèse de minimisation du coût, une fonction de coût néo-classique peut être utilisée afin d'estimer les caractéristiques de la technologie de la production et le degré d'efficacité de l'industrie.

10.2. La fonction de coût translog

Nous avons retenu dans cette étude la fonction de coût translog (Christensen, Jorgenson et Lau 1973) qui fait partie des formes fonctionnelles flexibles et permet de représenter des structures de production très générales. Elle est une approximation du deuxième ordre d'une série de Taylor deux fois différentiable. Une de ses principales caractéristiques est de permettre aux rendements d'échelle de varier avec le niveau de l'output ce qui laisse la courbe du coût moyen avoir une forme en U classique.

En supposant que les coûts de production soient minimisés avec un prix des inputs donné, la spécification générale de la fonction de coût translog peut s'écrire sous forme logarithmique:

$$\ln C = \alpha_0 + \alpha_Y \ln Y + \frac{1}{2} \beta_{YY} (\ln Y)^2 + \sum_i \alpha_i \ln P_i + \frac{1}{2} \sum_i \sum_j \beta_{ij} \ln P_i \ln P_j + \sum_i \beta_{Yi} \ln Y \ln P_i \quad i, j = 1, \dots, n \quad (10.1)$$

où $\beta_{ij} = \beta_{ji}$,
 C = coût total,
 Y = output,
 P_i = prix de l'input i .

Pour que cette fonction de coût corresponde à une fonction de production se comportant "normalement" (*well-behaved*), elle doit être positive et homogène de degré un dans les prix des facteurs, c'est-à-dire que pour un niveau fixe d'output, le coût total doit augmenter proportionnellement lorsque tous les prix s'accroissent proportionnellement. Ceci implique les relations suivantes parmi les paramètres:

$$\sum_i \alpha_i = 1 \quad (10.2)$$

$$\sum_i \beta_{Yi} = 0 \quad (10.3)$$

$$\sum_i \beta_{ij} = \sum_i \beta_{ji} = \sum_i \sum_j \beta_{ij} = 0 \quad (10.4)$$

Nous l'avons déjà mentionné, une propriété intéressante de la fonction de coût est que les équations dérivées de la demande des inputs peuvent être aisément calculées en différenciant partiellement la fonction de coût par rapport aux prix des facteurs, c'est-à-dire

$$\frac{\partial C}{\partial P_i} = X_i \quad (10.5)$$

où X_i représente la quantité de l'input i .

Ce résultat connu sous le nom de lemme de Shephard (1953) peut être exprimé sous forme logarithmique pour la fonction de coût translog. Il s'écrit alors

$$\frac{\partial \ln C}{\partial \ln P_i} = \frac{\partial C}{\partial P_i} \cdot \frac{P_i}{C} = \frac{P_i X_i^*}{C} = S_i \quad (10.6)$$

où S_i indique la part du coût attribuée au facteur i et X_i^* est le niveau d'input minimisant le coût total. La fonction de coût translog nous donne alors les équations de part optimale des facteurs (*share equations*)

$$S_i = \alpha_i + \sum_j \beta_{ij} \ln P_j + \beta_{Yi} \ln Y \quad (10.7)$$

A noter que la somme des parts optimales des facteurs (S_i) est égale à l'unité puisque $\sum_i P_i X_i = C$.

Parmi les informations importantes que l'on peut tirer de la fonction de coût, les élasticités de substitution d'Allen (Allen 1938) et les rendements d'échelle sont des caractéristiques particulièrement intéressantes lorsque l'on traite des monopoles naturels. Le théorème d'Uzawa (1962) montre que toutes les élasticités de substitution d'Allen (σ_{ij}) peuvent être obtenues à partir de la fonction de coût grâce à la formule:

$$\sigma_{ij} = \frac{C C_{ij}}{C_i C_j} \quad (10.8)$$

où les indices de C indiquent les dérivés partielles de la fonction de coût par rapport aux prix des facteurs, c'est-à-dire

$$C_i = \frac{\partial C}{\partial P_i}, \quad C_j = \frac{\partial C}{\partial P_j} \quad \text{et} \quad C_{ij} = \frac{\partial^2 C}{\partial P_i \partial P_j}.$$

Pour la fonction de coût translog, nous obtenons:

$$\sigma_{ij} = \frac{(\beta_{ij} + S_i S_j)}{S_i S_j}, \quad i \neq j \quad (10.9)$$

$$\sigma_{ii} = \frac{[\beta_{ii} + S_i(S_i - 1)]}{S_i^2},$$

L'élasticité-prix de la demande pour le facteur de production i est alors

$$\eta_i = \sigma_{ii} S_i \quad (10.10)$$

Les rendements d'échelle ont été habituellement définis comme une mesure de la variation de l'output résultant d'une variation proportionnelle de tous les facteurs

de production. Hanoch (1975) remarque cependant qu'il serait préférable d'utiliser le concept d'*economies of size* qui exprime la relation entre le coût total et l'output le long du chemin d'expansion, c'est-à-dire où les prix des inputs sont constants et les coûts sont minimisés pour tout niveau d'output¹. Nous utiliserons par la suite ce dernier concept de rendement d'échelle qui nous donne l'élasticité du coût total par rapport à la quantité produite. Christensen et Greene (1976) définissent les rendements d'échelle (RE) de la manière suivante:

$$RE = 1 - \frac{\partial \ln C}{\partial \ln Y} \quad (10.11)$$

Cette variable prendra une valeur positive en présence de rendements d'échelle croissants et une valeur négative avec des rendements d'échelle décroissants.

Varian (1988) montre que l'élasticité d'échelle, selon la définition traditionnelle, est équivalente à l'élasticité que nous venons de définir uniquement dans le cas d'une fonction de production homothétique.

La raison pour laquelle nous utiliserons le concept d'*economies of size* pour mesurer les rendements d'échelle provient de l'application de la fonction translog. En effet, celle-ci n'oblige pas la structure de production à être homothétique, pas plus qu'elle ne contraint les élasticités d'échelle à être constantes sur tout l'intervalle de production comme ce serait le cas avec une fonction du type CES. Rappelons qu'une fonction de coût correspond à une structure de production homothétique si, et seulement si, la fonction de coût peut être considérée comme une fonction séparable entre l'output et les prix des facteurs, c'est-à-dire si

$$C(y, p_i) = h(y) \cdot c(p_i)$$

La restriction pour que la fonction translog soit homothétique est:

$$\beta \gamma_i = 0 \text{ pour tout } i.$$

Nous ne restreindrons pas, dans notre application, la fonction de production à être homothétique. Cependant, la description de fonctions homothétiques nous

¹ La justification et la définition des deux concepts d'*economies of scale* et d'*economies of size* sont données par Chambers (1988, p. 68-72).

permet d'étendre ce raisonnement à l'introduction, dans la fonction de coût, de variables exogènes au processus de production .

La raison de la prise en compte de telles variables est très simple. Nous savons que tout processus de production est influencé – positivement ou négativement – par l'environnement dans lequel il se situe. Une étude empirique ne peut tenter d'appréhender l'efficacité de ce processus sans tenir compte des contraintes qui agissent sur celui-ci. L'idée consiste donc à inclure un vecteur de variables exogènes directement dans la fonction de coût. S'agit-il encore de déterminer de quelle façon les introduire? Considérons la fonction de coût $C(y,p)$ et réécrivons-la sous la forme $C(y,p,z)$ de manière à ce que z corresponde au vecteur des variables exogènes au processus de production. Nous supposons alors que cette fonction puisse s'écrire:

$$C(y,p,z) = c(y,p)h(z) \quad (10.12)$$

ce qui implique l'homothéticité par rapport à z . Imposons encore à la fonction $h(z)$ l'homogénéité. Cela signifie que si l'on multiplie tous les arguments de la fonction par un même facteur t , la valeur de la fonction est multipliée par un facteur t^m . Cette hypothèse nous permet d'ajouter simplement le vecteur z et les paramètres qui lui sont liés, à la fonction de coût translog définie en (10.1) ce qui donne sous forme logarithmique:

$$\begin{aligned} \ln C = & \alpha_0 + \alpha_Y \ln Y + \frac{1}{2} \beta_{YY} (\ln Y)^2 + \sum_i \alpha_i \ln P_i + \frac{1}{2} \sum_i \sum_j \beta_{ij} \ln P_i \ln P_j \\ & + \sum_i \beta_{Yi} \ln Y \ln P_i + \sum_l \gamma_l \ln Z_l \end{aligned} \quad (10.13)$$

$$i, j = 1, \dots, n \text{ et } l = 1, \dots, m$$

où Z_l représente la variable d'environnement l . L'avantage d'une telle restriction sur le vecteur z est qu'il n'apparaît pas dans les équations de part optimale des facteurs (S_i). Cela semble compatible avec la théorie économique et le comportement des gestionnaires. En effet, ceux-ci sont supposés minimiser le coût de la production en considérant le prix des facteurs, l'output ainsi que

² Pour une description rigoureuse des fonctions homogènes, voir Silberberg (1990).

l'environnement comme des données exogènes à court terme. Par contre, la part optimale de chaque input dans le coût total ne devrait être déterminée que par des variables entrant directement dans le processus de production, soit le prix des facteurs et la quantité produite.

Bien que les gestionnaires puissent, dans une certaine mesure, agir sur la qualité du service offert, nous considérerons celle-ci comme étant imposée par les autorités publiques. Puisque chaque entreprise de transports urbains est tenue de respecter, d'une manière ou d'une autre, les conditions établies soit par un office de réglementation, soit par un organe de surveillance, nous pouvons aisément supposer que la qualité du service fait partie intégrante des exigences à suivre. Pour cette raison, nous introduirons la qualité dans la fonction de coût de la même façon que les autres variables exogènes au processus de production (Z_1). Cette hypothèse permet de maintenir une forme relativement simple de la fonction translog.

La fonction de coût translog utilisée dans cette méthode considère un output unique (Y) et trois facteurs de production: le capital (K), le travail (L) et l'énergie (E). De plus, cinq facteurs exogènes au processus de production sont introduits dont deux représentent la qualité du service³. Nous avons fait l'hypothèse d'un progrès technique neutre selon Hicks (1963), c'est-à-dire que le long du chemin d'expansion, le taux marginal de substitution technique est indépendant du temps. La fonction de coût est donc donnée par:

$$\begin{aligned} \ln C = & \alpha_0 + \alpha_Y \ln Y + \frac{1}{2} \beta_{YY} (\ln Y)^2 + \alpha_K \ln P_K + \alpha_L \ln P_L + \alpha_E \ln P_E + \frac{1}{2} \beta_{KK} (\ln P_K)^2 \\ & + \frac{1}{2} \beta_{LL} (\ln P_L)^2 + \frac{1}{2} \beta_{EE} (\ln P_E)^2 + \beta_{KL} \ln P_K \ln P_L + \beta_{KE} \ln P_K \ln P_E \\ & + \beta_{LE} \ln P_L \ln P_E + \beta_{YK} \ln Y \ln P_K + \beta_{YL} \ln Y \ln P_L + \beta_{YE} \ln Y \ln P_E \\ & + \gamma_1 \ln Z_1 + \gamma_2 \ln Z_2 + \gamma_3 \ln Z_3 + \gamma_4 \ln Z_4 + \gamma_5 \ln Z_5 + \gamma_6 t \end{aligned} \quad (10.14)$$

où C : coût total

³ Nous ne précisons pas encore quelles variables seront retenues dans notre application car celles-ci font l'objet d'une sélection particulière que nous décrivons au chapitre 12. Nous verrons également quelles raisons nous ont amenés à ne retenir que cinq facteurs exogènes et celles qui ont déterminé le choix d'une forme logarithmique.

Y : output

P_K : prix du capital

P_L : prix du travail

P_E : prix de l'énergie

Z_1 : facteur de qualité ou d'environnement l

t : temps

γ_t : taux du progrès technique

α_j , β_{ij} et γ_1 : paramètres à estimer.

Nous pouvons alors dériver les trois équations de part optimale des facteurs à partir du modèle ci-dessus et obtenons:

$$S_K = \alpha_K + \beta_{KK} \ln P_K + \beta_{KL} \ln P_L + \beta_{KE} \ln P_E + \beta_{YK} \ln Y \quad (10.15)$$

$$S_L = \alpha_L + \beta_{LL} \ln P_L + \beta_{KL} \ln P_K + \beta_{LE} \ln P_E + \beta_{YL} \ln Y \quad (10.16)$$

$$S_E = \alpha_E + \beta_{EE} \ln P_E + \beta_{KE} \ln P_K + \beta_{LE} \ln P_L + \beta_{YE} \ln Y \quad (10.17)$$

Nous remarquons que ni les facteurs exogènes au processus de production ni le progrès technique n'apparaissent dans les trois équations (10.15), (10.16) et (10.17).

11. Le modèle économétrique

Le modèle exposé dans le chapitre précédent suppose a priori que le coût est minimisé par rapport au coût des inputs. L'objectif de notre application sera donc de vérifier si les entreprises observées minimisent effectivement leurs coûts de production. La fonction de coût translog (10.14) doit donc être considérée comme une frontière, c'est-à-dire le lieu des points où le coût est réellement minimisé étant donné l'output, les prix des inputs, les variables exogènes autres processus de production et un certain niveau technologique. Sous forme analytique, nous pouvons écrire cette fonction sous la forme:

$$C \geq f(x, \theta)$$

où $f(x, \theta)$ représente la fonction de coût translog, x est le vecteur des variables indépendantes décrit ci-dessus et θ le vecteur des paramètres à estimer. Si l'on fait l'hypothèse que toutes les firmes se situent sur la frontière de coût, nous pouvons réécrire cette inégalité sous la forme:

$$C_k = f(x_k, \theta), \quad k=1, \dots, n \text{ observations.}$$

L'application d'un modèle économétrique nous oblige à introduire un terme d'erreur aléatoire au membre de droite de cette équation qui devient, sous forme logarithmique:

$$\ln C_k = \ln f(x_k, \theta) + \varepsilon_k \quad (11.1)$$

Le terme ε_k est une variable aléatoire tenant compte des erreurs de mesure et d'éventuelles variables manquantes dans le modèle. Ce terme explique pourquoi le coût ne se situe pas exactement sur la frontière. En abandonnant l'hypothèse que toute firme se situe effectivement sur la frontière de coût, nous pouvons considérer une possible inefficacité dans le processus de production.

Parmi les différentes approches décrites dans le chapitre 3, nous en retiendrons deux qui utilisent des méthodes paramétriques; il s'agit de la frontière déterministe et de la frontière stochastique. La première considère que le terme ε_k représente uniquement l'inefficacité productive. Dans l'expression (11.1), il est non négatif ($\varepsilon_k \geq 0$). La seconde approche consiste à considérer la frontière elle-même comme stochastique. Le terme ε_k de l'équation (11.1) se décompose

en deux éléments ($e_k = v_k + u_k$). Le premier, v_k est purement résiduel et est l'équivalent du terme d'erreur de la méthode des moindres carrés ordinaires. Le second élément, u_k est non négatif et représente l'inefficacité productive. Il reflète le fait que chaque observation doit se trouver, soit sur la frontière $\ln f(x_k, \theta) + v_k$, soit au-dessus de celle-ci.

Nous pourrions appliquer ces deux méthodes directement à la fonction de coût translog (10.14) et utiliser la méthode des moindres carrés ordinaires décalée proposée par Greene (1980) pour la première et la méthode de Jondrow et al. (1982) pour la seconde. Nous négligerions toutefois les informations supplémentaires contenues dans les équations de part optimale des facteurs (10.15), (10.16) et (10.17) qui peuvent aussi être estimées facilement. De plus, la fonction de coût translog contient de nombreux termes – le prix des inputs – variant peu d'une observation à l'autre ce qui peut entraîner un problème de multicollinéarité.

Berndt et Wood (1975) ont utilisé une procédure permettant d'estimer les équations de part optimale des facteurs comme un système de régression multivarié. Cette méthode est satisfaisante lorsqu'on fait l'hypothèse des rendements d'échelle constants, c'est-à-dire lorsque $\alpha_Y = 1$, $\beta_{YY} = 0$ et $\beta_{Yi} = 0$ pour tout i . De cette manière, les *share equations* contiennent tous les paramètres de la fonction de coût à l'exception de l'intercept (α_0) et toutes les caractéristiques économiques de la fonction de coût peuvent être obtenues. Toutefois, lorsqu'on n'impose pas des rendements d'échelle constants, leur méthode n'est plus valable puisque α_Y et β_{YY} n'apparaissent que dans la fonction de coût.

Puisqu'aucune de ces deux approches ne paraît satisfaisante, une meilleure procédure consiste à estimer simultanément la fonction de coût et les équations de part optimale des facteurs comme un système de régression multivarié. Le principal avantage provenant de l'introduction des *share equations* est qu'elles permettent d'accroître le degré de liberté – le nombre d'observations est multiplié par le nombre d'équations dans le système – sans ajouter de nouveaux paramètres. Nous obtenons ainsi des coefficients plus efficaces statistiquement que ceux résultant de la méthode des moindres carrés ordinaires appliquée à la fonction de coût uniquement. Berndt et Christensen (1973) sont, à notre connaissance, les premiers à avoir utilisé un tel système d'équations afin d'estimer une fonction de coût. Pour résoudre ce système, nous utiliserons la méthode SURE (*Seemingly Unrelated Regression Equations*) proposée par Zellner (1962) et ajouterons un terme d'erreur w à chacune des équations de part

optimale des facteurs. Puisqu'elles sont obtenues par différentiation, ces équations présentent l'avantage de ne pas contenir le terme d'erreur de la fonction de coût. Ecrivons maintenant le système d'équations (10.14) à (10.17) sous forme matricielle:

$$\begin{pmatrix} C_k \\ S_{ik} \end{pmatrix} = \begin{pmatrix} X'_k \\ Z'_{ik} \end{pmatrix} \theta + \begin{pmatrix} \varepsilon_k \\ w_{ik} \end{pmatrix}$$

où k représente l'indice de l'observations et i ($i=1,2,3$), la i ème *share equation*. A partir du système (10.14) à (10.17), il est clair que les variables apparaissant dans Z_{ik} sont également présentes dans X_k . Les vecteurs ($X_k : Z_{ik}$) sont arrangés de façon à ce que θ satisfasse les restrictions entre les équations sur les paramètres α_i , β_{ij} et β_{Yi} entrant dans le système. Les vecteurs ($\varepsilon_k : w_{ik}$) sont indépendamment et identiquement distribués. En suivant Zellner, nous supposons que deux termes d'erreur dans différentes équations sont corrélés seulement s'ils correspondent à la même observation. Ecrivons maintenant les vecteurs ($\varepsilon_k : w_{ik}$) sous une forme simplifiée e_i avec $i=1, \dots, 4$ équations. D'après l'hypothèse ci-dessus, la covariance entre deux termes d'erreur d'une observation k est alors

$$\text{cov}(e_{ik} e_{jk}) = \sigma_{ij} \quad i, j=1, \dots, 4 \quad \text{et} \quad k=1, \dots, n$$

Toutefois, cette procédure d'estimation ne peut pas encore s'appliquer à notre modèle car la matrice des covariances requise par la méthode SURE est singulière. Cela provient du fait que la somme des erreurs des *share equations* doit être égale à zéro pour chaque firme puisque $\sum S_i = 1$ pour chacune d'elle. Afin de rendre non singulière la matrice des covariances, il suffit d'effacer une des équations de part optimale des facteurs. Un problème persiste encore puisque l'estimation des paramètres ainsi obtenue n'est pas indépendante du choix des équations retenues dans le système.

Deux solutions sont alors possibles pour résoudre ce problème. La première consiste à appliquer la méthode du maximum de vraisemblance au système en omettant une des *share equations*. Barten (1969) a montré en effet que les estimateurs du maximum de vraisemblance ne varient pas en fonction de l'équation effacée. La seconde solution consiste à effectuer une itération de la

procédure SURE de Zellner¹. Kmenta et Gilbert (1968) ont montré que l'itération de la méthode SURE convergeait vers une estimation par la méthode du maximum de vraisemblance. Bien que les deux procédures soient aussi efficaces statistiquement l'une que l'autre, nos résultats sont basés sur la méthode du maximum de vraisemblance. Nous avons utilisé, dans notre application, le logiciel LIMDEP de W. Greene qui recommande la méthode du maximum de vraisemblance lorsqu'il y a des contraintes sur les paramètres. Afin de tester la convergence des deux procédures, les estimations des paramètres à partir de l'autre méthode seront néanmoins présentées.

11.1. Estimation de l'efficacité productive

Le modèle économétrique utilisé suppose que les entreprises de transports urbains cherchent à minimiser leur coût de production sous contrainte d'une frontière de production, soit déterministe, soit stochastique. Nous considérons que ces firmes peuvent être non seulement techniquement inefficaces – elles utilisent trop d'inputs – mais également allocativement inefficaces – elles n'opèrent pas sur leur chemin d'expansion au moindre coût. L'inefficacité allocative signifie que des erreurs sont faites dans le choix des proportions d'inputs utilisés en fonction de leur prix respectif. Les conditions de premier ordre de la minimisation du coût ne sont donc pas totalement satisfaites, ce qui justifie le terme d'erreur (w_k) dans chaque équation de part optimale des facteurs. Le terme d'erreur ε qui doit apparaître dans l'équation de coût (10.14) traduit les effets aussi bien de l'inefficacité technique qu'allocative.

Quel que soit le type de frontière employé – déterministe ou stochastique – nos résultats refléteront donc l'inefficacité globale des entreprises de transports urbains. Certains auteurs sont parvenus à décomposer celle-ci (voir Schmidt et Lovell 1979, 1980 et Kumbhakar 1988, 1989) et ont estimé les effets séparés de l'inefficacité technique et allocative sur les firmes observées. Cependant, leur méthode se restreint à des technologies bien déterminées (Cobb-Douglas,

¹ A partir d'une estimation S_0 de la matrice des covariances, les paramètres sont estimés et une nouvelle matrice est obtenue. On répète le processus jusqu'à ce que la variation de l'estimation des éléments de θ soit négligeable. Pour plus de détails, voir Srivastava et Giles 1987, p. 149-155.

Generalized production Function²) et ne peut être étendue aux formes fonctionnelles plus flexibles. Nous ne pourrions donc malheureusement pas déterminer les effets des contraintes institutionnelles et/ou sociales sur le choix des inputs effectué par les entreprises.

11.2. Estimation d'une frontière déterministe

L'estimation d'une frontière de coût déterministe demande que l'équation (10.14) soit quelque peu réarrangée afin de pouvoir toujours utiliser la méthode itérative SURE. En effet, celle-ci fait l'hypothèse que le terme d'erreur est distribué indépendamment et identiquement selon une loi normale de moyenne nulle. Cela exclut l'estimation d'une frontière avec une erreur toujours non négative. Afin de résoudre ce problème, nous utiliserons une adaptation de la méthode des moindres carrés ordinaires décalés proposée par Greene (1980). La technique est très simple: elle consiste à estimer le système d'équations (10.14) à (10.17) par la méthode itérative SURE puis à ajouter à l'intercept de la fonction de coût son résidu minimum. De cette façon, toutes les observations se trouveront au-dessus de la frontière, sauf une qui se situera sur celle-ci. Greene a montré que cette démarche menait à une estimation convergente de tous les paramètres de la frontière à l'exception de la constante.

Ainsi, la totalité de la distance séparant chaque observation de la frontière de coût est expliquée par l'inefficacité productive³.

11.3. Estimation d'une frontière stochastique

La méthode déterministe a reçu de nombreuses critiques surtout de la part des économètres. Ils lui reprochent d'éliminer toute incertitude due à des éléments qui ne sont pas contrôlables par les managers ou par les observateurs. De plus, cette technique considère, de par sa nature, qu'il n'y a pas d'erreur de mesure (Schmidt 1986). Dans ce contexte, le développement d'une frontière gardant son

² Développée par Zellner et Revankar (1969), cette seconde forme de fonction de production est une généralisation d'une Cobb-Douglas.

³ Cette technique a été appliquée par Burgat et Jeanrenaud (1990) à la mesure de l'efficacité-coût des entreprises de ramassage des déchets ménagers. Il n'ont toutefois estimé qu'une fonction de coût par la méthode des moindres carrés ordinaire sans considérer les *share equations* qui lui sont liées.

caractère stochastique est tout à fait compréhensible. Depuis Meeusen et van den Broek (1977) et Aigner, Lovell et Schmidt (1977) de nombreuses applications ont été réalisées.

L'application de la méthode stochastique à notre système d'équations (10.14) à (10.17) consiste à décomposer le terme d'erreur ϵ de la fonction de coût en deux éléments distincts. Le premier (v) représente les facteurs purement résiduels tels que les erreurs d'observation du coût, la chance, etc. C'est cet élément qui donne à la frontière de coût son caractère stochastique. La seconde composante (u) est non négative et reflète l'inefficacité technique et allocative dans la production.

Le terme d'erreur attaché aux équations de part optimale des facteurs (w_i) représente, quant à lui, aussi bien les effets purement résiduels que l'inefficacité allocative. Nous ne l'avons pas décomposé car cela n'est pas nécessaire. L'utilisation des inputs dans des proportions incorrectes peut, en effet, aller dans les deux sens, c'est-à-dire qu'il peut y avoir une sur-utilisation ou une sous-utilisation d'inputs. Les deux éléments de w_i seraient alors tous deux symétriques et aucune méthode ne serait capable de les isoler sans faire d'autres hypothèses à leur sujet.

Nous devons toutefois établir les relations qui existent entre les vecteurs v , u et w_i . Il semble raisonnable de supposer que v et u sont indépendants puisque le premier représente le caractère aléatoire de la frontière tandis que le second traduit l'éloignement des firmes de leur niveau efficace. Dans le même ordre d'idée, nous supposerons que v et les w_i sont indépendants. Enfin, les vecteurs u et w_i sont conjointement corrélés pour une même observation.

12. Procédures d'estimation

Nous avons maintenant deux modèles différents nous permettant d'estimer la performance des entreprises observées. Il nous reste cependant à déterminer quelles variables exogènes au processus de production nous allons introduire dans la fonction de coût. Nous montrerons ensuite comment nous obtenons le degré d'efficacité pour chaque firme selon le type de frontière employé.

12.1. Choix des variables de qualité et d'environnement

Le choix de ces variables est particulièrement délicat et important, c'est pourquoi il requiert une analyse particulière. Deux questions se posent en effet:

- combien de variables faut-il introduire dans le système?
- de quelle manière ces variables influencent-elles le coût de la production?

Bien que nous ayons fait l'hypothèse de séparabilité entre ces variables et celles de la fonction de coût, il s'agit de déterminer la forme fonctionnelle (linéaire, logarithmique, exponentielle, etc.) la plus appropriée. L'objectif étant d'obtenir, dans la mesure du possible, une relation statistiquement significative, nous avons testé plusieurs modèles et avons finalement retenu celui qui expliquait le mieux la variation du coût total. Cela signifie que nous avons effectué plusieurs régressions en changeant à chaque fois la forme fonctionnelle des variables de qualité et d'environnement. Le coefficient de corrélation le plus élevé a été obtenu en mettant celles-ci sous forme logarithmique. Nous avons donc retenu cette forme de liaison pour les introduire directement dans la fonction de coût.

Cependant, toutes ces variables n'expliquent pas, dans les mêmes proportions, les variations de l'efficacité entre les observations. Afin de ne garder que celles qui ont une influence significative, la procédure *stepwise* a été utilisée. Celle-ci a permis de retenir les cinq facteurs¹ suivants:

- la fréquence des arrêts (Z_1);

¹ En raison de la multicollinéarité entre les variables de qualité et celle représentant le type de compagnie (plusieurs moyens de transport ou non), cette dernière n'a pu être introduite dans le modèle bien qu'elle se soit avérée significative.

- la fréquence des courses (Z_2);
- le nombre de personnes transportées (Z_3);
- le nombre de personnes transportées par kilomètre parcouru (Z_4);
- le taux de couverture des charges (Z_5).

12.2. Le modèle

Nous connaissons maintenant l'ensemble des variables qui sont introduites dans le modèle. Il s'agit toutefois de savoir s'il convient de considérer l'existence d'un progrès technique. Bien que l'analyse porte sur une courte période (4 ans au maximum), nous avons fait l'hypothèse d'un progrès technique neutre selon Hicks et avons ajouté à l'équation du coût une variable de temps sous forme linéaire. Les résultats obtenus montrent qu'aucun progrès technique (coefficient non significatif) n'a été réalisé au cours de cette période. Nous avons donc considéré l'échantillon comme un nuage de points atemporel.

La totalité des variables de la fonction de coût et des équations de part optimale des facteurs sont donc connues (trois facteurs de production avec leur prix respectif, l'output ainsi que cinq variables de qualité et d'environnement. Nous pouvons alors procéder à l'estimation des paramètres du système d'équations.

Etant donné l'hypothèse d'une fonction de coût linéairement homogène dans les prix, celle-ci peut s'écrire:

$$\begin{aligned}
 \ln(C/P_K) = & \alpha_0 + \alpha_Y \ln Y + \frac{1}{2} \beta_{YY} (\ln Y)^2 + \alpha_L \ln(P_L/P_K) + \alpha_E \ln(P_E/P_K) \\
 & + \frac{1}{2} \beta_{LL} [(\ln P_L/P_K)]^2 + \frac{1}{2} \beta_{EE} [(\ln P_E/P_K)]^2 \\
 & + \beta_{LE} \ln(P_L/P_K) \ln(P_E/P_K) + \beta_{YL} \ln Y \ln(P_L/P_K) \\
 & + \beta_{YE} \ln Y \ln(P_E/P_K) + \gamma_1 \ln Z_1 + \gamma_2 \ln Z_2 + \gamma_3 \ln Z_3 \\
 & + \gamma_4 \ln Z_4 + \gamma_5 \ln Z_5 + \varepsilon
 \end{aligned} \tag{12.1}$$

Nous avons utilisé le prix du facteur capital pour normaliser l'équation car ce choix n'a aucune influence économique ou statistique sur les résultats (Schmidt et Lovell 1979). Puisque la méthode itérative SURE ne dépend pas de l'équation

omise dans le système pour rendre la matrice des covariances non singulière, nous avons éliminé l'équation de part optimale du capital. Le système se complète donc avec les deux équations suivantes:

$$S_L = \alpha_L + \beta_{LL} \ln(P_L/P_K) + \beta_{LE} \ln(P_E/P_K) + \beta_{YL} \ln Y + w_L \quad (12.2)$$

$$S_E = \alpha_E + \beta_{EE} \ln(P_E/P_K) + \beta_{LE} (\ln P_L/P_K) + \beta_{YE} \ln Y + w_E \quad (12.3)$$

Remarquons que les variables exogènes au processus de production (Z_1) n'apparaissent pas dans les deux dernières équations.

13. Description des données

Les entreprises suisses de transports publics urbains ne représentent pas une catégorie homogène dans les publications statistiques disponibles. Les informations les plus complètes concernent les compagnies les plus importantes, c'est-à-dire celles qui disposent d'au moins deux moyens de transports (tramway, trolleybus ou autobus). L'Office fédéral de la statistique en a recensé 15 qui desservent les 20 agglomérations suisses les plus importantes. Bien que moins détaillées, des données sont également disponibles pour les 178 entreprises de transports publics n'utilisant que des autobus. Toutefois, une grande partie de celles-ci effectuent principalement des transports régionaux et sortent donc de notre cadre d'analyse. Afin d'identifier les transports typiquement urbains, nous avons appliqué quelques critères empiriques tels que le nombre de lignes desservies, le nombre d'arrêts ou encore le nombre de véhicules. Toutes les informations concernant le processus de production proprement dit, c'est-à-dire le prix des facteurs, les caractéristiques du réseau et la qualité du service proviennent des éditions 1986 à 1990 de "*Les transports publics*" publié par l'Office fédéral de la statistique.

A partir des critères de sélection décrits précédemment, 69 entreprises répondant aux caractéristiques de transports urbains ont été retenues. Après élimination des observations pour lesquelles les données étaient insuffisantes, nous n'avons conservé que 43 firmes. Pour la période allant de 1985 à 1988, notre échantillon contient finalement 133 observations¹; il comprend aussi bien de très grands réseaux, tel Zurich avec 4,54 milliards de places-km en 1988, que des plus modestes, tel Langenthal avec 2,04 millions de places-km en 1986.

Les informations se rapportant au mode de fourniture du service proviennent de notre propre enquête réalisée en 1988 auprès des 69 firmes effectuant des transports publics urbains. Enfin, quelques données concernant les variables d'environnement nous ont été fournies directement par l'Office fédéral de la statistique.

¹ A noter que parmi les 43 firmes, les données des 14 "grandes" compagnies - celles utilisant plusieurs moyens de transports - couvrent les années 1985 à 1988. Les données des 29 autres entreprises ne concernent que la période de 1986 à 1988 et, pour certaines d'entre elles, une partie seulement de celle-ci.

13.1. L'output

La mesure de l'output la plus fréquemment utilisée dans les études sur les transports urbains est le nombre de places-kilomètres offertes car elle a l'avantage de ne considérer que l'offre de transport (Gathon 1989, Delis 1988). Viton (1981) utilise comme indicateur le nombre de véhicules-km et ne tient donc pas compte de la capacité des véhicules. Enfin, d'autres recherches se basent sur le nombre de voyageurs-km qui est une variable influencée aussi bien par des facteurs de l'offre que par ceux de la demande (Kumbhakar 1987). En raison des objections formulées à l'encontre des deux dernières mesures, nous avons opté pour la première. Celle-ci est le produit du nombre de places offertes par véhicule et du nombre de kilomètres parcourus par l'ensemble des véhicules en une année.

Etant donné notre hypothèse d'un output unique, la production a été agrégée lors que plusieurs moyens de transport étaient employés simultanément.

13.2. Les inputs

Nous l'avons déjà mentionné, trois facteurs de production ont été considérés, soit le travail, le capital et l'énergie. Le nombre de personnes employées par l'entreprise de transport est la mesure retenue pour le facteur travail. Le nombre d'heures effectuées par l'ensemble du personnel aurait été une mesure plus précise, malheureusement, cette variable n'était pas disponible. Puisque nous avons supposé que le capital était un input flexible, nous n'avons pas incorporé les éléments d'infrastructure car ceux-ci ne varient pas à court terme. C'est pourquoi seul le nombre total de places du parc de véhicules a été retenu. Lorsque plusieurs moyens de transport sont employés, l'agrégation a été effectuée en pondérant les nombres de places en parc de chaque moyen de transport par le nombre de véhicules-km parcourus par chacun d'eux. Enfin, étant donné que les différents moyens de transports utilisent deux sources d'énergie (électricité et carburant fossile), nous avons converti le carburant consommé par les autobus en équivalent kWh afin d'avoir une mesure homogène.

Plusieurs indications permettant de mesurer l'output et les inputs sont indiqués dans le tableau 13.1. On remarquera que la taille des entreprises varie considérablement. Cela s'explique non seulement par l'importance des localités

desservies, mais également par le fait que certaines firmes desservent les nombreuses communes d'une agglomération.

Tableau 13.1 Caractéristiques de la production des 133 observations (1985-1988)

Variables	Moyenne	Ecart-type	Minimum	Maximum
Places-km (en millions de km)	403.3	824.4	5.0	4542.2
Véhicules-km (en milliers de km)	3590.3	6612.5	65.0	33800.0
Véhicules	85.2	139.7	2.0	637.0
Places en parc	5027.7	8114.2	130.0	39850.0
Postes de travail	240.5	462.5	2.0	2434.0
Consommation d'énergie (en milliers de kWh)	12161.0	22611.0	233.0	114900.0

13.3. Les coûts

Les informations sur les coûts des facteurs de production disponibles dans la statistique des transports publics ont été parfois difficiles à interpréter. Les charges de personnel ont été obtenues à partir de deux rubriques: "*charges de personnel*" et "*frais de choses*". Bien que la seconde rubrique ne concerne pas entièrement le coût du travail, nous avons constaté qu'elle contenait en grande partie des frais de personnel auxiliaire. Le coût du capital est le résultat de l'agrégation des charges d'amortissement et d'intérêts. En raison des différentes pratiques comptables des entreprises, nous avons dû éliminer certaines observations pour lesquelles le coût du capital semblait totalement aberrant². La prise en compte des coûts de l'énergie n'a posé aucun problème car toutes les données étaient disponibles et sont apparues comme fiables. Le coût total d'exploitation représente finalement la somme du coût de chacun des trois inputs. Le prix du facteur travail correspond au coût annuel du personnel par personne occupée. Le prix du capital a été obtenu quant à lui en divisant son

² Une des raisons contribuant à un coût du capital "anormal" provient d'imputations partielles des frais financiers entre l'administration centrale et le département concerné par les transports urbains.

coût par le nombre de places en parc. Enfin, le prix du troisième facteur représente le coût annuel de l'énergie par kWh consommé.

Nous avons également inclus dans le tableau 13.2 le coût unitaire moyen qui donne une idée de l'intervalle dans lequel opère les entreprises de transports urbains de notre échantillon. On constate aussi de grandes différences dans les prix des facteurs, en particulier pour celui du capital. Ces écarts proviennent du mode de comptabilisation des coûts qui varie beaucoup d'une firme à l'autre et du mode de transport employé. Les différences concernant le prix de l'énergie sont dues au type d'énergie utilisée (carburant et/ou électricité) ainsi qu'aux prix locaux de l'électricité qui peuvent varier considérablement. Le prix du travail est celui qui fluctue le moins, cela s'explique par un niveau général des salaires relativement homogène entre les régions.

Tableau 13.2 Coûts et prix des facteurs

Prix et coûts	Moyenne	Ecart-type	Minimum	Maximum
Coût total (en milliers de frs)	24497.0	50231.0	265.0	269700.0
Coût du travail (en milliers de frs)	15839.0	32292.0	167.0	171100.0
Coût du capital (en milliers de frs)	3176.0	6681.0	25.0	37740.0
Coût de l'énergie (en milliers de frs)	896.0	1578.0	12.0	6728.0
Coût unitaire moyen (en cts par place-km)	5.8	2.0	3.0	17.2
Prix du travail (en frs par poste de travail)	78860.0	15726.0	44080.0	143600.0
Prix du capital (en frs par place en parc)	455.2	231.4	114.5	993.7
Prix de l'énergie (en cts par kWh)	6.8	2.6	2.8	14.0

Notons encore que les coûts sur l'ensemble de la période 1985-1988 ont été déflatés par rapport aux prix de 1985. Bien que cet ajustement ne soit pas nécessaire pour l'estimation de la fonction de coût, elle permet d'effectuer des comparaisons entre les différentes années.

13.4. Les variables de qualité et d'environnement

Nous avons regroupé ici l'ensemble des variables exogènes au processus de production ainsi que celles permettant de représenter la qualité du service offert. Vu l'importance de cette dernière, trois facteurs de qualité ont été envisagés. Ils s'agit de:

- la fréquence des arrêts, mesurée par la distance moyenne parcourue entre deux arrêts;
- la fréquence des courses, mesurée par le laps de temps s'écoulant entre deux passages successifs des véhicules durant les heures de trafic normal;
- l'importance en pourcentage des places assises par rapport à l'ensemble des places offertes.

La qualité du service varie en sens inverse des deux premiers indicateurs et dans le même sens que le troisième. Nous avons remarqué qu'en général, les firmes se situant dans les grandes agglomérations avaient une fréquence des arrêts et des courses nettement plus élevée que celle des compagnies exploitant des réseaux plus modestes. Le pourcentage des places assises est, par contre, plus élevé pour les petites entreprises, ce qui s'explique par l'utilisation d'autobus offrant plus de places assises que des trolleybus et tramways. Les caractéristiques du réseau proprement dit sont représentées par cinq variables. La première mesure la longueur totale des lignes du réseau, la seconde indique la densité du réseau (longueur totale des lignes divisée par l'aire desservie). L'importance de ce dernier est évaluée grâce à la population desservie, alors que la densité de la population permet d'obtenir un indicateur du degré d'urbanisation. La dernière variable est dichotomique et prend la valeur 1 lorsque plusieurs moyens de transports sont employés simultanément et la valeur 0 sinon.

Deux autres indicateurs donnent une idée de la demande de transport. Le premier indique le nombre de personnes transportées alors que le second représente le nombre de personnes transportées par kilomètre parcouru. Ce deuxième indicateur permet de connaître, d'une certaine manière, le taux de remplissage des véhicules.

Enfin, un dernier indicateur considère le taux de couverture des charges d'exploitation. Puisque la majorité des entreprises reçoivent des contributions de la part du secteur publique (prise en charge du déficit), cette variable indique

l'importance de la pression financière agissant sur l'entreprise et, par conséquent, sur l'incitation des gestionnaires à produire au coût minimum.

Tableau 13.3 Qualité des transports et variables d'environnement

Prix et coûts	Moyenne	Ecart-type	Minimum	Maximum
Distance moyenne entre deux arrêts (<i>en mètres</i>)	469.8	202.1	57.1	1362.0
Fréquence normale des courses (<i>en minutes</i>)	24.2	19.9	6.0	120.0
Importance des places assises (<i>en % du total</i>)	37.9	10.6	16.0	69.1
Longueur totale des lignes exploitées (<i>en mètres</i>)	64027.0	59100.0	2400.0	271400.0
Densité du réseau (<i>en mètres par km²</i>)	3509.6	3353.8	105.3	18020.0
Population desservie	50397.0	69826.0	1930.0	351100.0
Densité de la population	2287.6	2214.4	190.5	10920.0
Personnes transportées (<i>en milliers</i>)	23892.0	50470.0	139.0	282400.0
Personnes transportées par km parcouru	4.3	2.4	0.4	10.8
Taux de couverture des charges (<i>en %</i>)	68.9	18.0	12.0	106.0

Parmi les variables décrites ci-dessus, nous distinguerons plusieurs catégories:

- les éléments qui décrivent le réseau de transports (type de réseau, longueur et densité des lignes) et ne peuvent être modifiés, à court terme, ni par les autorités publiques, ni par les gestionnaires;
- les indicateurs de la qualité de l'offre de transport qui peuvent être soumis à certaines contraintes imposées par les autorités de surveillance;
- les variables d'environnement géographique qui représentent une donnée pour les compagnies et pour l'office de réglementation;
- un indicateur financier sous contrôle des autorités publiques.

Nous pouvons constater que les gestionnaires ont peu d'influence sur l'ensemble de ces facteurs, raison pour laquelle ces derniers ne sont pas directement intégrés dans le processus de production.

14. Résultats

14.1. Productivité partielle des inputs

Avant d'examiner l'efficacité productive des entreprises suisses de transports urbains, nous allons utiliser une mesure plus simple et plus courante: la productivité partielle des inputs. Cet indicateur est un ratio où le numérateur est représenté par l'output et le dénominateur par l'input considéré. Bien que les mesures de productivité partielles ne soient que peu satisfaisantes en raison de l'absence d'informations sur la fonction de production (rendements d'échelle, intensité relative des inputs, type de technologie, etc), elles donnent un aperçu de la performance des entreprises d'un échantillon.

L'indicateur le plus souvent utilisé concerne le facteur travail. Nous avons toutefois calculé également la productivité partielle du capital car cet input est relativement important dans les activités de transports urbains. Par ailleurs, la productivité partielle du facteur énergie devra être interprétée avec prudence étant donné les diverses sources énergétiques utilisées (électricité et carburant). Ces différences sont, en quelque sorte, conditionnées par les différents moyens de transports rencontrés (tramways, trolleybus et autobus).

Nous avons obtenu l'indicateur de la productivité partielle du travail en divisant l'output (nombre de places-kilomètres) par le facteur travail (nombre de personnes employées). Pour l'indicateur correspondant au capital, le dénominateur est représenté par le nombre de places du parc de véhicules. Enfin, pour déterminer la productivité partielle de l'énergie, nous avons divisé l'output par l'énergie consommée par les véhicules, exprimée en kWh. Ces trois indicateurs sont décrits dans le tableau 14.1 ci-dessous.

Tableau 14.1 Description de la productivité partielle des facteurs de production

Statistiques	Productivité partielle du travail	Productivité partielle du capital	Productivité partielle de l'énergie
Moyenne	1821.63	58.20	27.81
Ecart-type	666.36	24.16	6.46
Coefficient de variation	0.37	0.42	0.23
Minimum	491.18	10.52	12.11
Maximum	4139.57	113.99	46.46
Rapport Min/Max	8.43	10.83	3.84

Nous constatons que la productivité partielle du capital a la plus grande dispersion alors que celle de l'énergie est la plus concentrée. Les histogrammes de fréquence sont d'ailleurs révélateurs (figure 14.1, 14.2 et 14.3). En effet, pour le facteur travail, la productivité des observations est symétrique autour de leur moyenne. On remarque le même phénomène pour le facteur capital mais avec une dispersion plus grande. L'histogramme de fréquence du facteur énergie révèle une légère asymétrie due probablement à l'utilisation de deux sources d'énergie différentes.

En comparant la productivité partielle des entreprises de notre échantillon, nous avons constaté de fortes différences entre les compagnies exploitant plusieurs moyens de transports et celles n'utilisant que des autobus. En effet, la productivité partielle du travail est significativement plus élevée dans les entreprises de la seconde catégorie (l'hypothèse d'égalité des moyennes des deux sous-ensembles est rejetée à 95%). Par contre, nous trouvons la situation inverse pour les deux autres indicateurs de productivité partielle. Ces écarts s'expliquent, en partie du moins, par la structure différente des moyens de transport entre les deux catégories, différence dont est incapable de tenir compte un indicateur tel que la productivité partielle des inputs. D'une part, les véhicules de la première catégorie parcourent de plus grandes distances – donc produisent un output plus important – en raison de la fréquence élevée des trajets dans les villes ayant des tramways ou des trolleybus. D'autre part, les véhicules utilisant la force électrique s'avèrent moins gourmands en énergie que ceux fonctionnant à partir de carburants fossiles (meilleur rendement énergétique du moteur électrique).

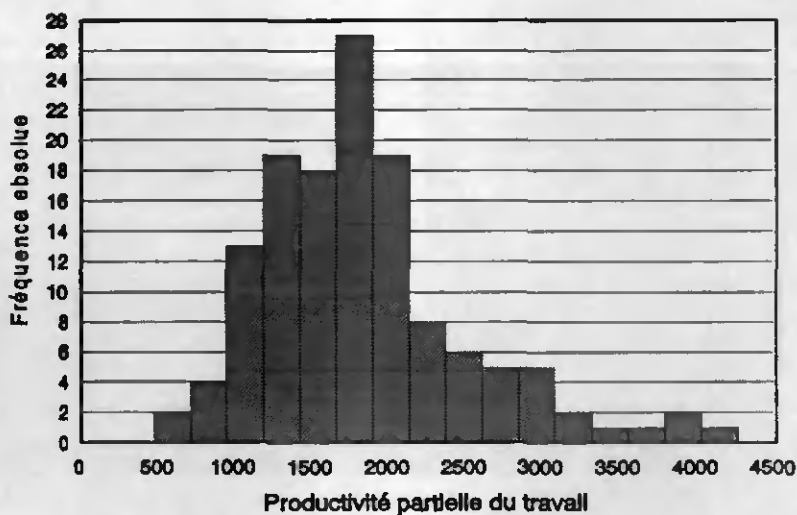


Figure 14.1 Histogramme de fréquence de la productivité partielle du travail

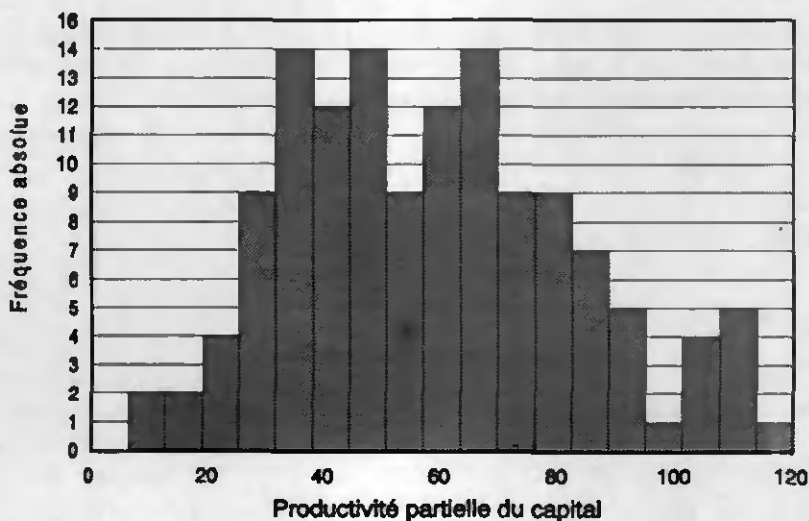


Figure 14.2 Histogramme de fréquence de la productivité partielle du capital

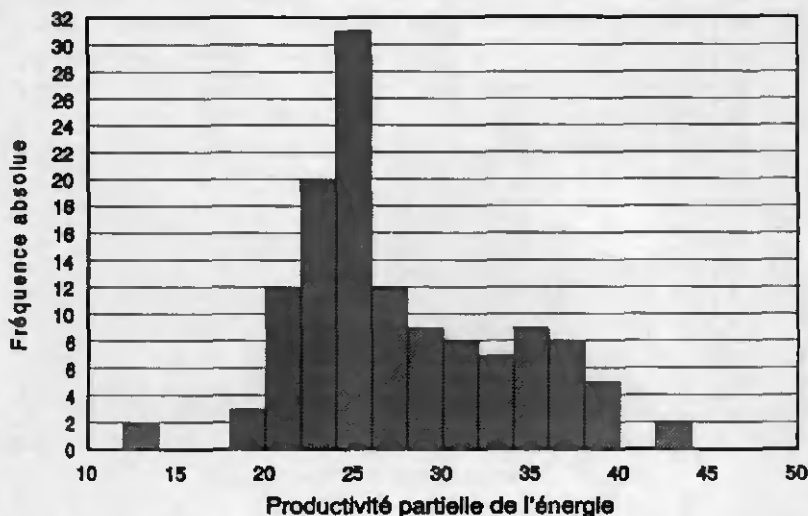


Figure 14.3 Histogramme de fréquence de la productivité partielle de l'énergie

Il semble par conséquent que la productivité partielle des inputs soit avant tout conditionnée par le choix des moyens de transport et que la performance proprement dite n'ait qu'un impact limité sur ce type d'indicateur. C'est la raison pour laquelle il convient d'employer une mesure de l'efficacité plus précise et tenant compte de l'ensemble des facteurs. Nous allons donc examiner plus attentivement l'efficacité productive des entreprises de transport urbain, efficacité obtenue à partir d'une frontière de coût.

14.2. Mesure de l'efficacité productive

A partir du système d'équations (12.1) à (12.3), nous avons appliqué la méthode itérative SURE à deux modèles différents: le premier étant strictement celui du système décrit ci-dessus. Dans le second, nous avons éliminé les facteurs exogènes au processus de production, c'est-à-dire les variables Z_1 dans l'expression (12.1). D'autre part, l'estimation des paramètres a été effectuée selon les méthodes disponibles sur le logiciel LIMDEP, soit:

- *Iterated Generalized Least Squares* qui est équivalente à
- *Maximum Likelihood Estimation of Constrained Multivariate Regression*.

Tableau 14.2 Paramètres de la fonction de coût¹

Coefficients	Modèle (1)	Modèle (2)	Modèle (3)	Modèle (4)
α_0	-11.1293 (-6.007)	-10.8570 (-6.289)	2.52848 (-1.469)	-0.87041 (-0.561)
α_Y	1.03574 (5.245)	1.00708 (5.479)	0.23660 (1.304)	0.11690 (0.717)
β_{YY}	-0.00796 (-0.756)	-0.00641 (-0.655)	-0.00533 (-0.618)	-0.00512 (-0.661)
α_L	0.09731 (1.492)	0.09688 (1.544)	0.14507 (2.058)	0.16835 (2.493)
β_{LL}	0.06858 (10.689)	0.06859 (10.897)	0.06335 (8.768)	0.06065 (8.674)
α_E	0.27260 (18.854)	0.27850 (19.497)	0.25605 (16.122)	0.25333 (16.089)
β_{EE}	0.03138 (24.851)	0.03219 (25.185)	0.02968 (22.006)	0.02984 (22.051)
α_K	0.63009 (10.457)	0.62463 (10.733)	0.59881 (9.391)	0.57832 (9.464)
β_{KK}	0.05070 (8.626)	0.05007 (8.711)	0.04779 (7.681)	0.04591 (7.670)
β_{LE}	-0.02463 (-17.492)	-0.02535 (-17.793)	-0.02263 (-13.799)	-0.02229 (-13.587)
β_{LK}	-0.04395 (-7.274)	-0.04324 (-7.314)	-0.04073 (-6.202)	-0.03836 (-6.057)
β_{EK}	-0.00675 (-5.903)	-0.00684 (-5.918)	-0.00706 (-5.801)	-0.00755 (-6.193)
β_{YL}	0.01672 (7.167)	0.01667 (7.645)	0.01583 (6.723)	0.01538 (6.942)
β_{YE}	-0.00221 (-4.305)	-0.00224 (-4.581)	-0.00205 (-4.031)	-0.00198 (-4.004)
β_{YK}	-0.01452 (-6.921)	-0.01443 (-7.272)	-0.01378 (-6.499)	-0.01339 (-6.678)

¹ Les chiffres entre parenthèses indiquent les valeurs du *t* de Student.

Tableau 14.2 Paramètres de la fonction de coût (suite)

Coefficients	Modèle (1)	Modèle (2)	Modèle (3)	Modèle (4)
γ_1			0.06942 (2.324)	0.08604 (3.208)
γ_2			-0.05002 (-1.207)	-0.10515 (-2.826)
γ_3			-0.07034 (-1.413)	-0.13123 (-2.943)
γ_4			0.77615 (8.772)	0.88717 (11.175)
γ_5			-0.64724 (-7.406)	-0.74850 (-9.537)
$\log \Sigma $	-19.55218	-19.55296	-20.05895	-20.08108

L'utilisation de ces deux méthodes a pour but essentiel de montrer que la première converge vers la seconde et que nous pouvons être indifférents quant à l'emploi de l'une plutôt que l'autre. Le tableau 14.2 confirme la similarité des deux méthodes. Les modèles (1) et (2) n'incorporent pas les facteurs d'environnement et de qualité dans la fonction de coût alors qu'ils sont présents dans (3) et (4). Les modèles (1) et (3) sont estimés à partir du maximum de vraisemblance tandis que les modèles (2) et (4) le sont à partir de la méthode des moindres carrés généralisés.

Les paramètres α_K , β_{KK} , β_{LK} , β_{EK} et β_{YK} ont été obtenus à partir des contraintes 10.2, 10.3 et 10.4 imposées à la fonction de coût. Les erreurs-types de ces paramètres ont été calculées à partir de la formule de la variance d'une somme de variables aléatoires².

Dans les deux premiers modèles, tous les paramètres sont significatifs à 99%, à l'exception de β_{YY} et de α_L . Les coefficients α_Y et β_{YY} ne sont plus statistiquement significatifs dans les deux derniers modèles. La raison est certainement une multicollinéarité entre les variables de qualité et l'output. Le log-déterminant de la matrice des covariances (Σ) montre que celle-ci ne varie pas selon la méthode, vérifiant ainsi leur convergence.

² Pour de plus amples précisions sur la façon d'obtenir la variance d'une somme de variables aléatoires, voir Mood, Graybill et Boes (1974), p. 178-179.

14.3. Estimation de l'efficacité productive à partir d'une frontière de coût déterministe

Nous avons calculé l'efficacité productive de chaque observation en prenant les modèles (1) et (3) comme base d'estimation. Le terme constant α_0^* correspondant à la frontière est obtenu en ajoutant à α_0 la valeur du résidu minimal de la fonction de coût qui est égal à -0,7739 pour le modèle (1) et à -0,5279 pour le modèle (3). Le degré d'efficacité productive (DEP) s'obtient alors en divisant le coût minimal C_k^* de chaque observation par le coût réel correspondant (C_k), c'est-à-dire

$$\text{DEP}_k = \frac{C_k^*}{C_k} = \exp(\varepsilon_{\min} - \varepsilon_k).$$

La valeur de $(\varepsilon_{\min} - \varepsilon_k)$ étant toujours non négative pour une seule observation, le degré d'efficacité productive n'est donc jamais supérieur à 1. L'examen de celui-ci sur l'ensemble des observations montre que la prise en compte des variables de qualité et d'environnement accroît substantiellement l'efficacité pour la plupart des entreprises (voir figure 14.6). En outre, le tableau 14.3 montre non seulement une augmentation de plus de 50% de l'efficacité moyenne de l'échantillon, mais également que l'introduction des facteurs exogènes au processus de production améliore également l'homogénéité de la série puisque l'écart-type passe de 0.148 à 0.129. Nous devons tout de même mentionner le très faible score de l'entreprise la moins performante puisque son coût total devrait être, selon le modèle, d'environ 84% et 72% plus faible si elle produisait au coût minimum.

Enfin, nous remarquons d'une part, que les compagnies les plus importantes ne sont pas les moins performantes bien que l'une des plus modestes soit la plus efficace (voir figure 14.7). D'autre part, l'introduction de variables de qualité et d'environnement est particulièrement importante pour les grandes entreprises comme le montre la différence entre les deux mesures de l'efficacité sur la figure 14.7.

Tableau 14.3 Degré d'efficacité productive d'après une frontière déterministe

	Degré d'efficacité productive	
	sans facteurs exogènes ³	avec facteurs exogènes
Moyenne	0.483	0.604
Ecart-type	0.149	0.131
Asymétrie	0.751	0.077
Minimum	0.164	0.283

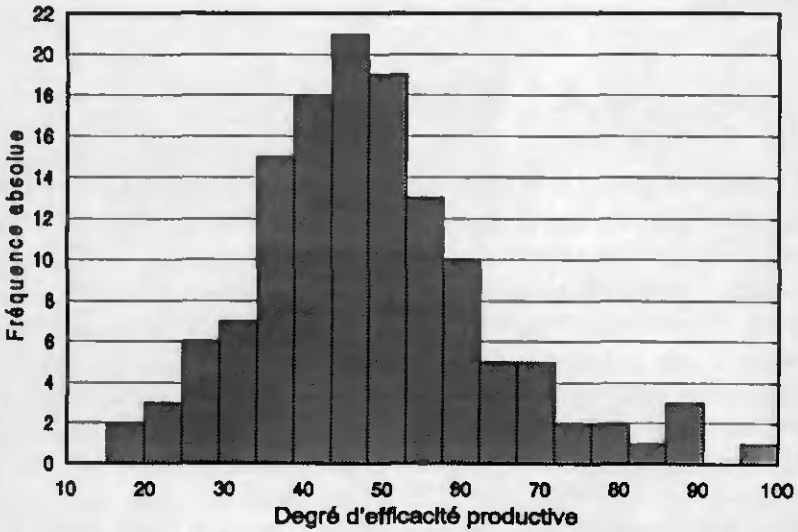


Figure 14.4 Histogramme de fréquence de l'efficacité productive – Frontière déterministe sans facteurs exogènes

³ Pour simplifier, le terme *facteurs exogènes* représente ci-après les variables de qualité et d'environnement.

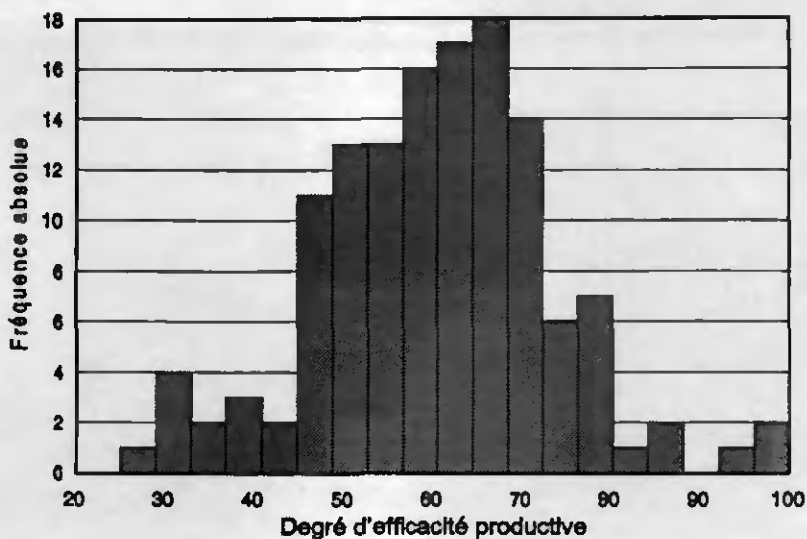


Figure 14.5 Histogramme de fréquence de l'efficacité productive – Frontière déterministe avec facteurs exogènes

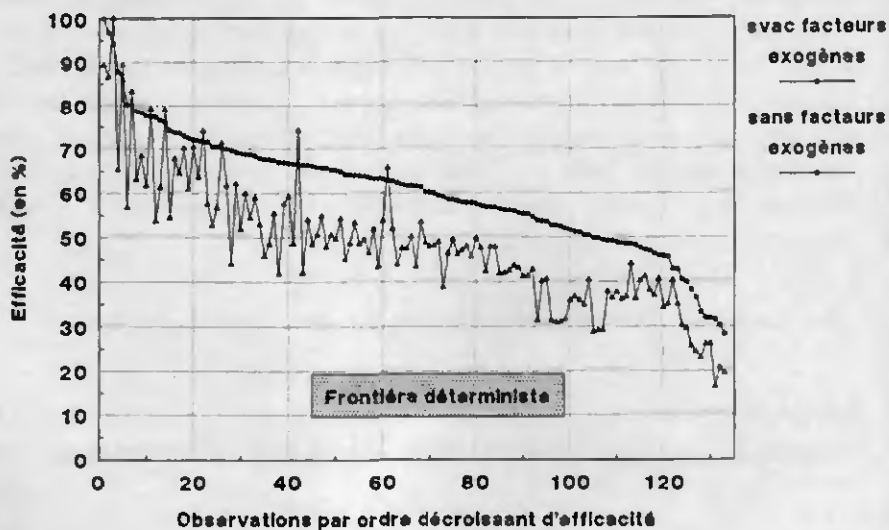


Figure 14.6 Degré d'efficacité avec et sans facteurs exogènes

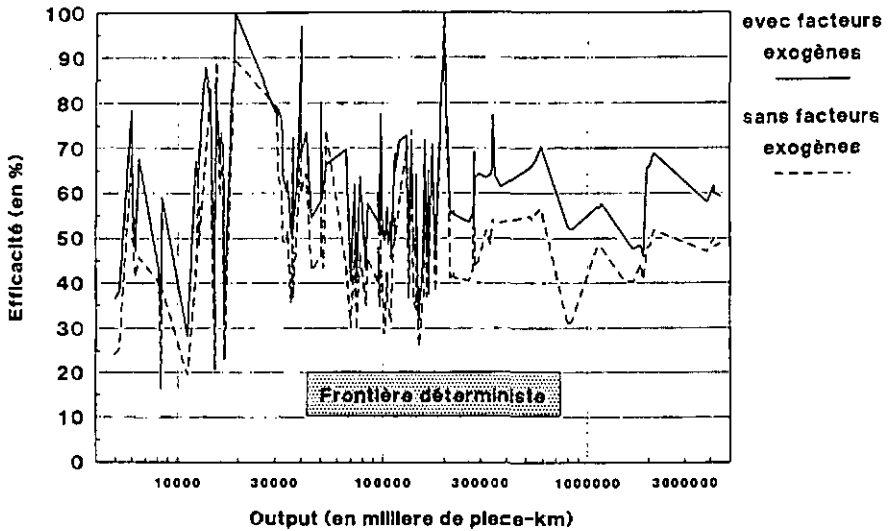


Figure 14.7 Degré d'efficacité productive en fonction de l'output.

L'estimation de l'efficacité à partir de la méthode déterministe révèle, dans notre application, une performance moyenne relativement faible. Cette constatation nous amène à poser la question de la fiabilité de cette méthode. En effet, ne doit-on pas examiner si d'autres facteurs influencent la fonction de coût minimum. De plus, la présence d'éventuels *outliers* risque de biaiser considérablement les résultats puisque la frontière peut très bien être obtenue à partir d'une observation extrême. Afin d'éliminer ces deux inconvénients de la méthode déterministe, nous allons, dans le chapitre suivant, nous baser sur l'estimation d'une frontière de coût stochastique.

14.4. Estimation de l'efficacité productive à partir d'une frontière de coût stochastique

Puisque la frontière est stochastique, le terme d'erreur de la fonction de coût se compose des deux éléments u et v décrits précédemment. Pour une observation particulière, le terme d'erreur s'écrit:

$$\varepsilon_k = v_k + u_k$$

où $v_k \sim N(0, \sigma_v^2)$ est un élément symétrique représentant le résidu habituel de toute relation statistique et $u_k \geq 0$ est un terme représentant l'inefficacité productive. A noter que ce dernier mesure l'éloignement d'une observation de sa frontière de coût minimum, il est donc l'équivalent de $(\varepsilon_{\min} - \varepsilon_k)$ dans le modèle déterministe⁴. L'estimation du système d'équation (10.14) à (10.17) nous permet d'obtenir les résidus $\hat{\varepsilon}$ qui peuvent être considérés comme des estimateurs du terme d'erreur ε . Le problème consiste alors à décomposer ces estimateurs afin d'isoler leurs composants, v et u . Schmidt et Lovell (1979) ont estimé l'inefficacité moyenne à partir d'une estimation de σ_u . Toutefois, notre objectif étant de mesurer l'inefficacité productive u_k pour chaque observation k , nous reprenons l'idée de Jondrow et al. (1982) en l'appliquant au cas d'une frontière de coût. Il s'agit alors de considérer la distribution conditionnelle de u étant donné ε . La moyenne de cette distribution sera utilisée pour obtenir une estimation de u . Nous supposons par la suite que u suit une distribution semi-normale, c'est-à-dire $u \sim |N(0, \sigma_u^2)|$.

Notons encore que la fonction de coût (10.14) doit être transformée puisque $E(\varepsilon) \geq 0$ étant donné que $E(u) > 0$. Afin de satisfaire les conditions habituelles des moindres carrés généralisés, la moyenne du terme d'erreur doit être égale à zéro. En ajoutant $E(\varepsilon)$ au terme constant et en soustrayant $E(\varepsilon)$ à ε , la moyenne du terme d'erreur $\varepsilon - E(\varepsilon)$ est égale à zéro dans le nouveau modèle. L'estimation du terme constant est alors donné par $\hat{\alpha}_0$ et puisque

$$E(\hat{\alpha}_0) = \text{plim} \hat{\alpha}_0 = \alpha_0 + E(u) = \alpha_0 + \sqrt{(\pi/2)} \sigma_u$$

nous pouvons former une estimation convergente de α_0 :

$$\tilde{\alpha}_0 = \hat{\alpha}_0 - \sqrt{(\pi/2)} \hat{\sigma}_u$$

où $\hat{\sigma}_u$ est une estimation de σ_u .

L'estimation de l'inefficacité demande quelques développements. Nous reprendrons la démarche de Jondrow et al. afin de montrer que la distribution

⁴ Afin de ne pas alourdir la notation, nous avons éliminé l'indice k dans la suite du chapitre.

conditionnelle de u , étant donné $\varepsilon = v + u$, est celle d'une variable tronquée en zéro et suivant une loi normale de moyenne μ_* et de variance σ_*^2 , où

$$\mu_* = \varepsilon \sigma_u^2 / \sigma^2$$

$$\sigma_*^2 = \sigma_u^2 \sigma_v^2 / \sigma^2$$

$$\sigma^2 = \sigma_u^2 + \sigma_v^2$$

En rappelant que v et u sont indépendants et que $\varepsilon = v + u$, il s'agit de trouver la distribution conditionnelle de u , étant donné ε .

La densité jointe de u et v est le produit de leurs densités respectives puisqu'ils sont indépendants, c'est-à-dire

$$f(u, v) = \frac{1}{\pi \sigma_u \sigma_v} \exp \left[\frac{-1}{2\sigma_u^2} u^2 - \frac{1}{2\sigma_v^2} v^2 \right], \quad u \geq 0 \quad (14.1)$$

En sachant que $v = \varepsilon - u$, la densité jointe de u et ε est donnée par

$$f(u, \varepsilon) = \frac{1}{\pi \sigma_u \sigma_v} \exp \left[\frac{-1}{2\sigma_u^2} u^2 - \frac{1}{2\sigma_v^2} (\varepsilon - u)^2 \right] \quad (14.2)$$

La fonction de densité de ε peut s'écrire, en suivant Aigner, Lovell et Schmidt (1977), comme

$$g(\varepsilon) = \frac{2}{\sigma \sqrt{2\pi}} \exp \left[\frac{-1}{2\sigma^2} \varepsilon^2 \right] \left[1 - \Phi \left(\frac{\varepsilon \sigma_u}{\sigma_v \sigma} \right) \right] \quad (14.3)$$

où Φ est la fonction de répartition d'une loi normale.

Nous savons en outre que la probabilité conditionnelle d'une variable aléatoire X , étant donnée la variable aléatoire Y , est donnée par (Judge et al. 1982):

$$f(x|y) = \frac{f(x,y)}{g(y)}, \quad g(y) \neq 0$$

Par conséquent, la densité conditionnelle de u , étant donné ε , est le rapport de (14.2) et de (14.3) qui s'écrit:

$$f(u|\varepsilon) = \frac{1}{1-\Phi(\varepsilon\sigma_u/\sigma_v\sigma)} \frac{1}{\sigma_*\sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma_*^2} + \frac{1}{\sigma_v^2}\mu\varepsilon - \frac{\sigma_u^2}{2\sigma_v^2\sigma^2}\varepsilon^2\right] \quad (14.4)$$

Cette expression peut se simplifier et s'écrire:

$$f(u|\varepsilon) = \frac{1}{1-\Phi(\varepsilon\sigma_u/\sigma_v\sigma)} \frac{1}{\sigma_*\sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma_*^2} (u - \sigma_u^2\varepsilon/\sigma^2)^2\right] \quad (14.5)$$

L'expression $1-\Phi(\cdot)$ est la probabilité qu'une variable aléatoire suivant une loi normale $N(\mu_*, \sigma_*^2)$ soit positive. Cette expression représente donc la densité de probabilité d'une variable aléatoire ($u|\varepsilon$) distribuée selon une loi normale σ_*^2 tronquée en zéro (Jondrow et al. 1982). Nous pouvons maintenant estimer u par sa moyenne ou par son mode. Nous ne retiendrons que la moyenne qui s'écrit (Kumbhakar 1989):

$$E(u|\varepsilon) = \mu_* + \sigma_* \left[\frac{\phi(\mu_*/\sigma_*)}{\Phi(\mu_*/\sigma_*)} \right] \quad (14.6)$$

où ϕ est la fonction de densité d'une loi normale.

Preuve: Nous savons que la moyenne d'une variable aléatoire suivant une loi normale tronquée en a avec des paramètres μ et σ^2 s'écrit (Greene 1990):

$$E(u|\varepsilon > \alpha) = \mu + \sigma \left[\frac{\phi(\alpha)}{1-\Phi(\alpha)} \right] \quad (14.7)$$

où $\alpha = (a-\mu)/\sigma$

Dans notre cas, $a=0$ et $u|\varepsilon \sim N(\mu_*, \sigma_*^2)$, ce qui donne l'expression suivante:

$$E(u|\epsilon) = \mu_* + \sigma_* \left[\frac{\phi(-\mu_*/\sigma_*)}{1-\Phi(-\mu_*/\sigma_*)} \right] \quad (14.8)$$

Puisque $\phi(-x) = \phi(x)$ du fait de la symétrie de la loi normale, nous obtenons l'expression (14.6). C.Q.F.D.

Les paramètres μ_* et σ_*^2 étant inconnus, ils doivent être remplacés par leur estimation, $\hat{\mu}_*$ et $\hat{\sigma}_*^2$. De plus, en calculant les second et troisième moments centraux à partir des résidus de la fonction de coût, nous obtenons des estimations convergentes de σ_u^2 et de σ_v^2 . Les second et troisième moments, μ_2 et μ_3 , du terme d'erreur sont donnés par (Aigner, Lovell et Schmidt, 1977, Schmidt et Lovell 1979):

$$\mu_2 = \sigma_v^2 + \left[\frac{\pi-2}{\pi} \right] \sigma_u^2$$

$$\mu_3 = \sqrt{\frac{2}{\pi}} \left[\frac{4-\pi}{\pi} \right] \sigma_u^2$$

Nous obtenons ainsi facilement les expressions suivantes de σ_u^2 et σ_v^2 :

$$\sigma_u^2 = \left[\sqrt{\frac{\pi}{2}} \left[\frac{\pi}{4-\pi} \mu_3 \right] \right]^{2/3} \quad (14.9)$$

$$\sigma_v^2 = \mu_2 - \left[\frac{\pi-2}{\pi} \right] \sigma_u^2 \quad (14.10)$$

En utilisant les estimateurs convergents $\hat{\mu}_2$ et $\hat{\mu}_3$ à la place de μ_2 et μ_3 dans (14.9) et (14.10), nous avons des estimateurs convergents $\hat{\sigma}_u^2$ et $\hat{\sigma}_v^2$ de σ_u^2 et σ_v^2 .

A partir des expressions (14.9) et (14.10), nous obtenons une estimation des u_k donnée par $E(u_k|\epsilon_k)$. Une autre méthode que celle des moments centraux peut également être utilisée. Il s'agit d'une estimation directe par la méthode du maximum de vraisemblance de l'équation (14.8) en remplaçant dans celle-ci le vecteur ϵ par son équivalent $(y-X\beta)$. Bien que cette seconde méthode permette d'éviter le problème dû à un éventuel terme négatif pour μ_3 , nous ne l'avons pas

employée étant donné que ce terme ne s'est jamais avéré négatif quel que soit le modèle retenu.

Après correction du terme constant (α_0), celui-ci est égal à -11.727 , -11.4525 , 1.972 et -1.4188 respectivement pour les modèles (1) à (4). Nous remarquons en outre que la variabilité du terme d'inefficacité (u) est beaucoup plus importante que celle du terme d'erreur (v) comme le montre le rapport $\lambda = \sigma_u / \sigma_v$ dans le tableau 14.4. De plus, l'introduction des variables d'environnement et de qualité réduit considérablement la variance de v ce qui montre bien leur importance sur le processus de production des entreprises de transports publics urbains.

En raison de l'hypothèse retenue pour la distribution du terme d'inefficacité u – loi normale tronquée en zéro – un grand nombre d'observations apparaissent comme étant relativement efficaces. Cet aspect de la distribution des observations selon leur efficacité (voir figures 14.8 et 14.9) souligne en quelque sorte l'importance de cette hypothèse. Il semble alors qu'un autre modèle pourrait être retenu. Stevenson (1980) suggère une distribution normale tronquée en $a \neq 0$ et ajoute: "It is not clear, however, why the mode of u should be expected to occur at $u=0$. If the error term u represents the level of inefficiency, the ALS (Aigner, Lovell and Schmidt) specifications are based on an implicit assumption that the 'likelihood' of inefficient behavior monotonically decreases for increasing levels of inefficiency. However, characteristics such as degree of educational training, intelligence, persuasiveness, etc. (factors which relate to managerial efficiency) are not likely distributed with such a monotonically declining density function over the population. Since the economic agents are humans or human institutions, the possibility of a non-zero mode for the density function of u would seem a more tenable presumption". Cependant, l'estimation du modèle de Stevenson entraîne de substantielles complications mathématiques et statistiques tout en maintenant des hypothèses fortes (loi normale tronquée ou loi gamma) qui demandent elles-mêmes à être vérifiées. De plus, son application à un système d'équations tel que le nôtre n'est pas forcément possible.

Tableau 14.4 Paramètres obtenus à partir d'une distribution normale tronquée en zéro

Paramètres	Modèle (1)	Modèle (2)	Modèle (3)	Modèle (4)
σ_u^2	0.1448	0.1437	0.1255	0.1219
σ_v^2	0.0462	0.0465	0.0078	0.0021
λ	1.7708	1.7589	4.0091	7.5955
σ	0.4370	0.4361	0.3652	0.3521

L'examen de la performance des entreprises de transports urbains montre un sensible accroissement de l'efficacité moyenne par rapport à celle obtenue par la méthode déterministe. En effet, l'efficacité globale s'accroît d'environ 90% avec la méthode stochastique lorsque les variables exogènes au processus de production ne sont pas considérées et d'environ 40% lorsque ces dernières sont introduites dans les modèles. Cette augmentation de l'efficacité moyenne provient du fait qu'avec la seconde méthode, la frontière ne dépend plus d'une observation unique qui peut être détachée de l'ensemble des autres observations (cf. figure 14.4) comme le montrent Burgat et Jeanrenaud (1990) dans leur analyse de l'efficacité des entreprises de ramassage des déchets ménagers. Cependant, quelle que soit la méthode utilisée pour mesurer l'inefficacité, on remarque clairement sur les figures 14.12 et 14.13 que les observations les moins efficaces sont les mêmes et ont des valeurs très proches. La figure 14.11 nous montre encore que les firmes les moins performantes sont de petite taille et que les grandes sont relativement bien classées. Nous ne pouvons toutefois en conclure que ce type d'industrie est un monopole naturel. Il faudrait effectuer pour cela des mesures complémentaires permettant de déterminer si la fonction de coût globale de l'industrie est réellement sous-additive⁵.

⁵ Voir à ce propos l'application de Evans et Heckman (1984) à l'entreprise américaine Bell Telephone.

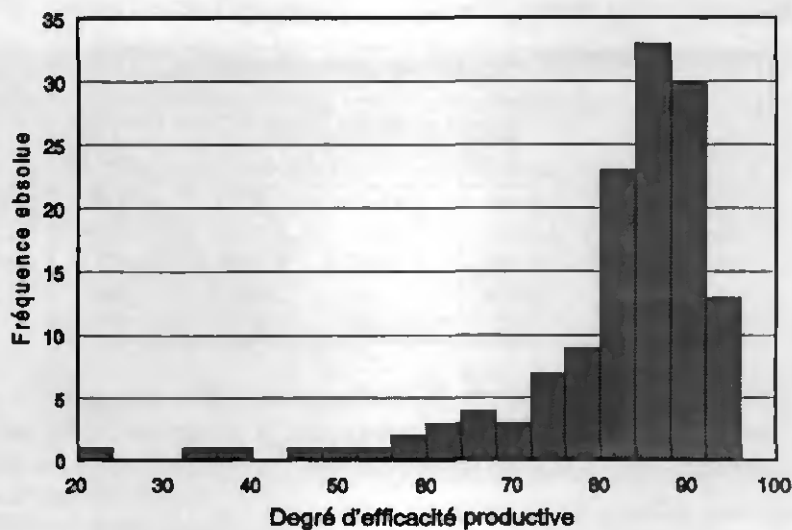


Figure 14.8 Histogramme de fréquence de l'efficacité productive – Frontière stochastique sans facteurs exogènes

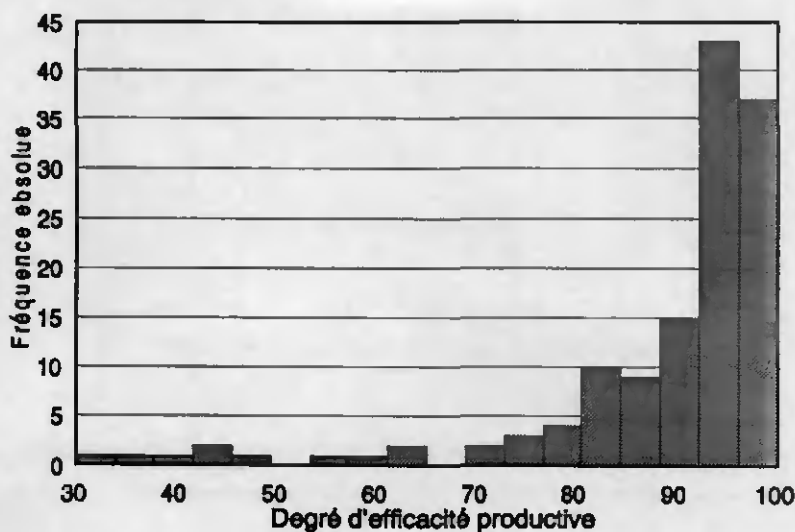


Figure 14.9 Histogramme de fréquence de l'efficacité productive – Frontière stochastique avec facteurs exogènes

Tableau 14.5 Degré d'efficacité productive d'après une frontière stochastique

	Degré d'efficacité productive ($1-u_k$)			
	sans facteurs exogènes		avec facteurs exogènes	
	Modèle (1)	Modèle (2)	Modèle (3)	Modèle (4)
Moyenne	0.8188	0.8188	0.8863	0.9104
Ecart-type	0.1207	0.1199	0.1353	0.1448
Minimum	0.2161	0.2215	0.3086	0.3313
Maximum	0.9488	0.9485	0.9876	1.1170
Asymétrie	-2.276	-2.260	-2.452	-2.330

Le facteur d'asymétrie montre que la distribution est fortement concentrée aux environs de zéro. Ce phénomène est pas surprenant puisque la distribution du terme d'inefficacité est basée sur l'hypothèse d'une loi normale tronquée en zéro.

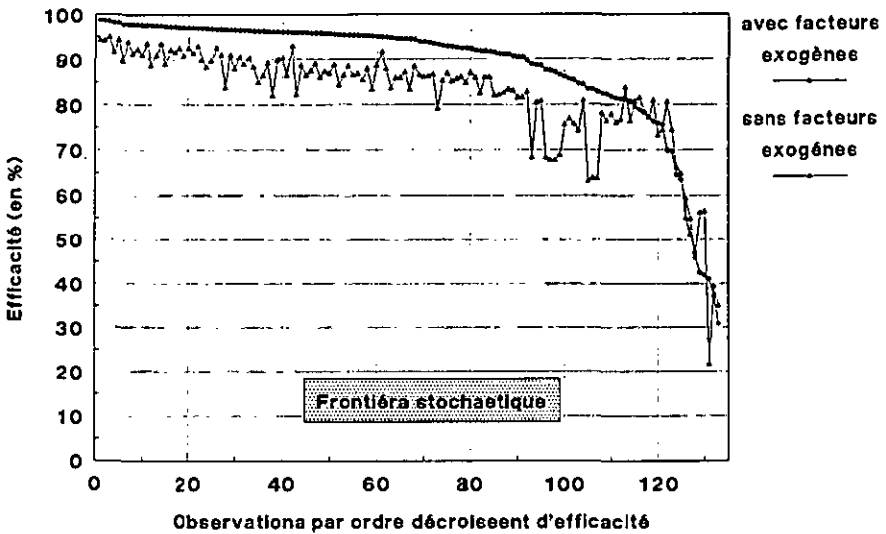


Figure 14.10 Degré d'efficacité productive avec et sans facteurs exogènes

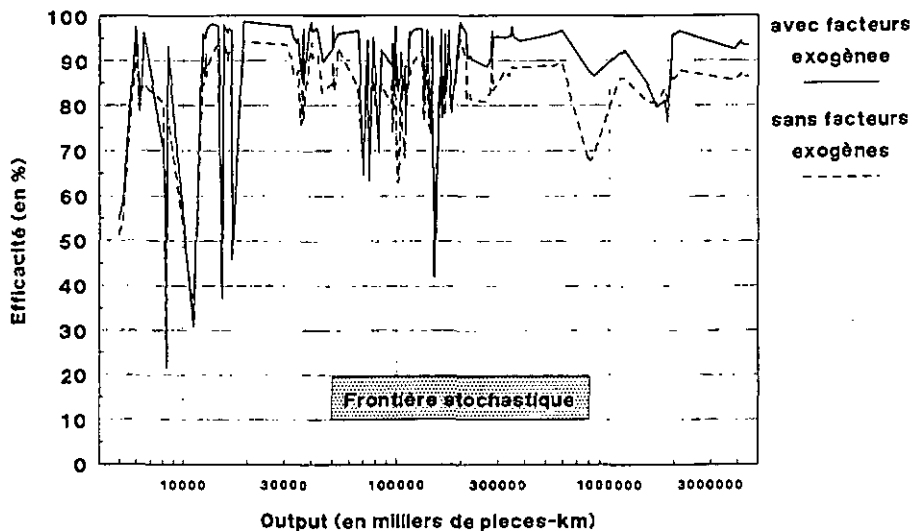


Figure 14.11 Degré d'efficacité productive en fonction de l'output

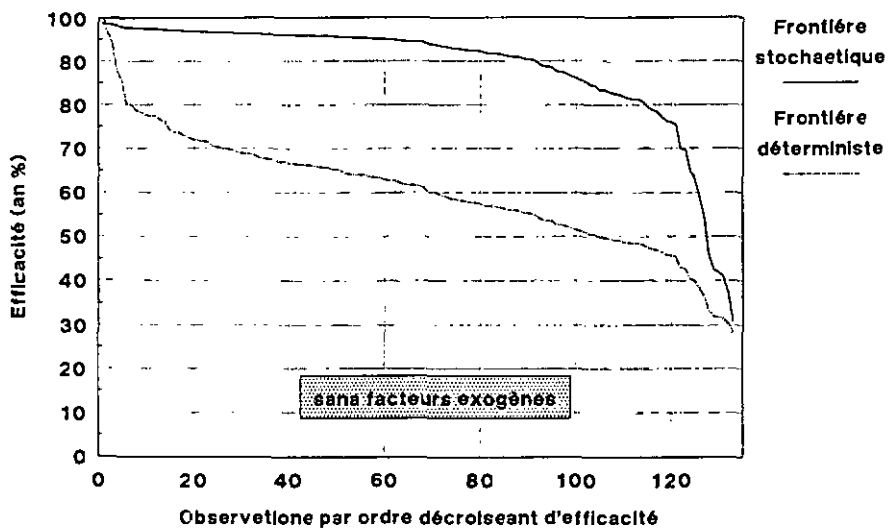


Figure 14.12 Efficacité productive selon le type de frontière – sans facteurs exogènes

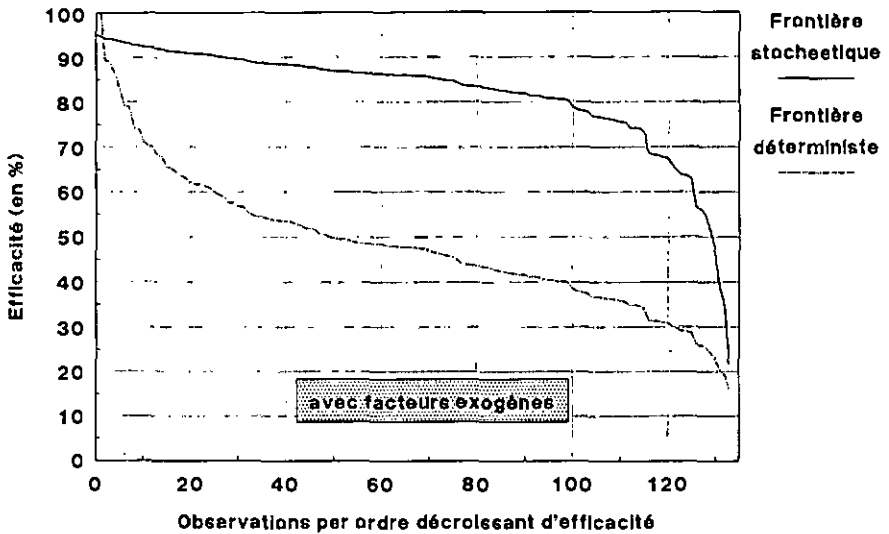


Figure 14.13 Efficacité productive selon le type de frontière – avec facteurs exogènes

14.5. Élasticités de substitution et rendements d'échelle

Les estimations des élasticités de substitution partielles entre les facteurs de production (σ_{ij}) évaluées pour les valeurs moyennes des observations sont toutes significativement différentes de zéro (tableau 14.6). Nous remarquons que le facteur travail est un substitut aux deux autres facteurs et qu'il y a complémentarité entre le capital et l'énergie⁶. Par conséquent, avec la structure de production utilisée par les entreprises de transports urbains, nous pouvons dire que le choix des moyens de transport détermine la quantité d'énergie utilisée.

⁶ Lorsque $\sigma_{ij} > 0$, les inputs i et j sont substituables et si $\sigma_{ij} < 0$, ils sont complémentaires, voir McFadden, 1978, p. 48.

Tableau 14.6 Elasticités de substitution entre les facteurs de production et élasticités-prix de la demande (erreurs-type entre parenthèses)

Facteurs	Travail	Capital	Energie
Travail	-0.13425 (0.00251)	0.64893 (0.00297)	0.37265 (0.02464)
Capital		-3.66077 ((0.02313)	-0.11487 (0.05130)
Energie			-6.45884 (1.47956)
Elasticité-prix de la demande	-0.10914 (0.00176)	-0.28654 (0.01670)	-0.52243 (0.01119)

L'élasticité-prix de la demande d'énergie est la plus importante. Cette particularité est quelque peu déconcertante puisque, ce facteur étant celui qui représente la plus faible part du coût total, son prix ne devrait pas être un élément fondamental dans les prises de décision des entreprises. Toutefois, les arguments de protection de l'environnement et d'économie d'énergie sont probablement suffisamment puissants pour incite les gestionnaires à s'intéresser plus attentivement à la consommation d'énergie de leurs véhicules.

A l'opposé, la demande de travail est la moins sensible aux variations de prix. Bien que cet input représente environ 80% du coût total des entreprises de notre échantillon, deux raisons permettent d'expliquer la rigidité de la demande de travail par rapport à son prix. En premier lieu, rappelons que la majorité des entreprises est fortement lié au secteur public (contributions financières pour couvrir les déficits, participation des collectivités publiques locales dans le capital social, aménagement d'infrastructure par les communes, etc). Ces relations étroites ont contribué à rendre les politiques salariales des entreprises de transports urbains aussi rigides que celles des administrations publiques: grille des salaires, résistance des gestionnaires aux licenciements, etc). En second lieu, l'activité même de ce type de firmes les contraint à l'immobilité géographique, d'où l'obligation de s'approvisionner sur le marché local de l'emploi. Cet inconvénient associé au fait que les salaires sont souvent plus élevés dans les grandes villes – où se situent précisément les transports urbains – est certainement la cause principale de l'inélasticité-prix de la demande de travail.

Nous avons vu précédemment (chapitre 7) qu'il y avait monopole naturel lorsque la fonction de coût d'une industrie était sous-additive. Pour savoir si la présence

d'une seule firme est réellement souhaitable dans le cas des transports publics urbains, nous avons examiné la fonction de coût global de notre échantillon afin de déterminer si elle était caractérisée par des rendements d'échelle croissants, mesure moins rigoureuse que celle de la sous-additivité mais plus facile à appliquer. Pour un output moyen (403.3 millions de places-km), notre analyse fait apparaître des rendements d'échelle de 0.747. Sur tout l'intervalle de production observé, les rendements d'échelle sont toujours positifs et varient extrêmement peu puisqu'ils passent de 0.718 pour l'output le plus faible (5 millions de places-km) à 0.766 pour l'entreprise la plus importante (4542.2 millions de places-km). Malheureusement, ce résultat doit être considéré avec beaucoup de prudence car le paramètre β_{YY} est non significativement différent de zéro et négatif. L'augmentation proportionnelle des rendements d'échelle avec l'output provient du fait que ce sont des entreprises en réseau dont la limite de capacité du processus de production n'est jamais atteinte quelle que soit leur taille (voir figure 14.4). Ce résultat semble démontrer l'importance croissante des coûts fixes par rapport à l'étendue du réseau de transport.

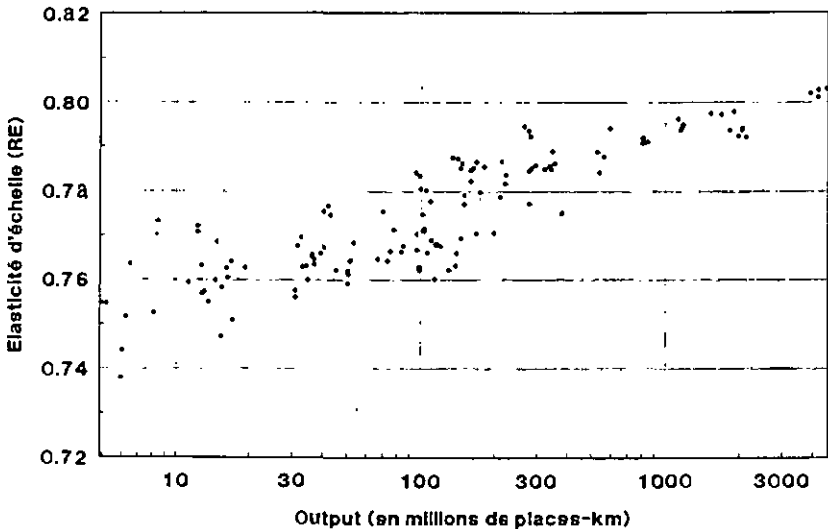


Figure 14.14 Rendements d'échelle

Une autre façon d'illustrer ce phénomène d'économies d'échelle est de présenter la courbe de coût moyen (figure 14.15) d'une entreprise représentative de notre

échantillon. Cette courbe a été obtenue en évaluant la fonction de coût moyen sur l'intervalle de production observé en maintenant constant le prix des facteurs à leur niveau moyen⁷. Nous pouvons constater que le coût moyen est fortement décroissant sur tout l'intervalle de production observé.

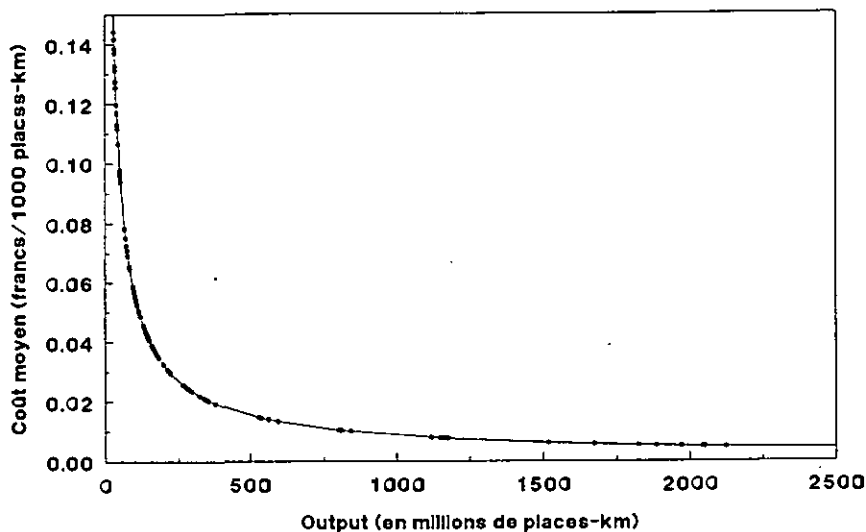


Figure 14.15 Courbe de coût moyen de l'échantillon

⁷ La pente de la courbe du coût moyen est suffisante pour montrer l'existence de rendement d'échelle croissants puisque $RE=1-(\text{coût marginal/coût moyen})$. Par conséquent, un coût moyen décroissant démontrera la présence d'économies d'échelle

15. Performance des entreprises de transports urbains en fonction du mode de fourniture

Les transports publics urbains examinés dans notre échantillon sont fournis selon les modes de fourniture suivants: la gestion directe (15 observations), la sous-traitance (6 observations) et la concession (119 observations). Nous avons par ailleurs divisé ce dernier mode en concession à une entreprise publique (44 observations), à une entreprise d'économie mixte (56 observations) et à une entreprise privée (19 observations). Ces différents modes de fourniture ont ensuite été répartis en deux secteurs; l'un à dominance publique (55 observations), l'autre privée (78 observations). Il faut encore noter, d'une part, que la sous-traitance et la concession à une entreprise privée ne concerne que les compagnies n'utilisant que des autobus et, d'autre part, que la majorité des grandes entreprises de transports urbains sont des entreprises publiques bénéficiant d'une concession (dans 35 cas sur 55).

15.1 Comparaison des productivités partielles des facteurs de production

Pour chacun des trois inputs, une comparaison entre secteurs public et privé a été effectuée. Toutefois, nous n'avons pas jugé utile d'examiner la productivité de chaque mode de fourniture en raison du manque de précision de cette mesure de la performance.

En ce qui concerne la productivité partielle du travail aucune différence significative n'a été constatée entre les deux secteurs. Par contre, les productivités partielles du capital et de l'énergie font apparaître une assez grande différence entre secteurs public et privé (tableau 15.1). Il ressort en effet de l'analyse que le premier est plus performant que le second à un niveau de signification de 99%. Il s'agit toutefois de rester prudent quant à toute interprétation hâtive d'un tel résultat. Pour le facteur énergie en particulier, la plupart des entreprises exploitant des véhicules fonctionnant à l'énergie électrique sont classées dans le secteur public or le rendement d'un moteur électrique est supérieur à celui d'un moteur à combustion.

Tableau 15.1 Productivité partielle des inputs pour les secteurs public et privé

	Productivité partielle du travail		Productivité partielle du capital		Productivité partielle de l'énergie	
	Secteur public	Secteur privé	Secteur public	Secteur privé	Secteur public	Secteur privé
Moyenne	1749.3	1872.7	73.9	47.1	32.0	24.8
Ecart-type	427.7	791.5	23.2	17.9	6.9	4.0
Minimum	491.2	852.3	29.9	10.5	15.4	12.1
Maximum	2951.9	3287.3	114.0	90.8	40.5	36.9

15.2 Comparaison de l'efficacité productive selon le mode de fourniture

Nous avons comparé la performance relative des différents modes de fourniture des transports publics urbains en nous basant sur la mesure de l'efficacité productive obtenue à partir des frontières de coût déterministe et stochastique. Dans cette partie, seules les fonctions de coût intégrant les variables exogènes au processus de production ont été retenues, afin de mettre toutes les observations sur un même pied d'égalité.

Dans une première étape, l'analyse de la performance est effectuée entre les secteurs public et privé. Dans une seconde, la performance de chaque mode de fourniture est considérée.

La comparaison de l'efficacité productive entre les observations provenant du secteur public et du secteur privé n'a révélé aucune différence significative. Quelle que soit la méthode retenue – déterministe ou stochastique –, nous avons constaté que les performances moyennes de chacun des deux secteurs sont extrêmement proches l'une de l'autre. Il semble donc que, dans le domaine des transports publics urbains, la propriété privée des moyens de production ne permette pas de s'approcher davantage d'une situation économiquement optimale que la propriété publique.

Tableau 15.2 Efficacité productive moyenne des secteurs public et privé

	A partir d'une frontière stochastique		A partir d'une frontière déterministe	
	Secteur public	Secteur privé	Secteur public	Secteur privé
Moyenne	90.2	87.5	60.3	60.4
Ecart-type	12.2	14.4	10.3	14.8
Minimum	30.9	37.3	28.3	30.3
Maximum	98.2	98.8	87.9	100.0

Grâce à une analyse plus fine, la seconde étape révèle l'existence de différences d'efficacité entre les modes de fourniture. En premier lieu, nous avons remarqué que l'efficacité moyenne de la sous-traitance est la plus élevée et que celle de la concession à une entreprise d'économie mixte est la plus faible, ceci quel que soit le type de frontière retenu (voir tableau 15.3). Cependant, les tests de comparaison de l'efficacité moyenne ne sont pas significatifs lorsque celle-ci est obtenue à partir d'une frontière stochastique alors qu'avec une frontière déterministe, les tests s'avèrent significatifs à 95% et plus.

En comparant les modes de fourniture deux à deux, quelques résultats intéressants se dessinent pourtant. Relevons en particulier que l'efficacité productive moyenne de la concession publique estimée à partir d'une frontière stochastique est significativement supérieure à celle de la concession mixte.¹

Une autre démarche consiste à comparer entre elles les entreprises exploitant plusieurs moyens de transports, c'est-à-dire les transports publics urbains des grandes agglomérations suisses. Nous avons constaté ici encore que les entreprises d'économie mixte bénéficiant d'une concession étaient significativement moins performantes que l'ensemble des autres firmes et ceci quel que soit le type de frontière utilisé. A l'opposé, la même analyse montre que la gestion directe est plus efficace que les autres modes de fourniture.

¹ A relever que le faible nombre d'observations (6) en régime de sous-traitance ne permet pas d'obtenir des résultats concluants.

Tableau 15.3 Efficacité productive des modes de fourniture

	A partir d'une frontière stochastique			
	Moyenne	Ecart-type	Minimum	Maximum
Gestion directe	87.6	21.1	30.9	96.8
Sous-traitance	97.2	0.8	96.0	98.5
Concession publique	91.6	6.4	76.0	98.2
Concession mixte	86.4	13.5	42.0	98.8
Concession privée	89.3	17.5	37.3	98.6
	A partir d'une frontière déterministe			
	Moyenne	Ecart-type	Minimum	Maximum
Gestion directe	60.6	12.7	28.3	70.5
Sous-traitance	76.4	9.9	66.4	94.3
Concession publique	60.9	9.3	45.7	87.9
Concession mixte	57.8	13.5	31.8	100.0
Concession privée	66.2	17.0	30.3	96.8

A la lumière de ces résultats, nous constatons que le mode de fourniture a une influence directe sur l'efficacité productive des entreprises de transports publics urbains. La concession à une entreprise d'économie mixte semble être la solution la plus défavorable. Cette situation est tout de même surprenante car ce mode de fourniture laisse, en théorie du moins, une relativement grande liberté d'action aux gestionnaires de l'entreprise. Les causes d'inefficacité sont probablement diverses et une analyse individuelle plus poussée serait nécessaire à leur identification.

Tableau 15.4 Efficacité productive moyenne des entreprises exploitant plusieurs moyens de transport

	Nombre d'observations	Frontière stochastique		Frontière déterministe	
		Moyenne	Ecart-type	Moyenne	Ecart-type
Gestion directe	8	95.6	6.0	65.5	6.9
Concession publique	35	90.4	6.6	58.4	8.0
Concession mixte	16	86.2	3.6	52.0	2.8

Il convient cependant de se poser quelques questions au sujet des relations entre les gestionnaires de l'entreprise et les autorités de contrôle. Tout d'abord, il s'agit de déterminer avec certitude si les coûts ont été correctement révélés. Il peut y avoir en effet des incitations à sous-estimer ou à surestimer les coûts totaux. Dans le premier cas, nous pouvons très bien supposer qu'en situation de gestion directe, l'office responsable des transports n'intègre pas la totalité des coûts qui lui incombent. Par exemple, certaines tâches effectuées par l'administration générale de la commune ne seraient pas entièrement imputées à cet office qui ferait alors apparaître une efficacité plus élevée qu'en réalité. Nous pensons en particulier aux frais financiers pris en charge par l'administration générale et aux aménagements réalisés pour le compte des transports urbains par les services des ponts et chaussées. Dans le second cas, il est également facile d'imaginer que l'entreprise en concession surestime ses coûts dans le but, soit d'obtenir une contribution de la collectivité publique plus importante, soit de justifier une augmentation des prix ou encore une réduction de la qualité des prestations.

Se pose donc le problème du contrôle du monopole par les autorités publiques. Le système de la sous-traitance est peut-être une solution à développer même lorsque les réseaux sont de grande taille. En effet, si les entreprises de transports par autobus en régime de sous-traitance se sont montrées les plus efficaces ce n'est pas par hasard. L'organe de contrôle de la commune peut faire jouer la concurrence pour l'attribution du contrat car les véhicules peuvent servir aussi bien comme moyen de transport urbain que comme moyen de transport privé (absence de coûts irrécupérables). Par conséquent, l'entrée et la sortie du marché sont sans coût. Il est dès lors possible d'imaginer un mécanisme semblable dans le cas de réseaux plus rigides tels ceux utilisant des tramways et des trolleybus. Si la commune possède les installations fixes (rails, lignes électriques) et le

matériel roulant, elle peut sous-traiter leur exploitation à une entreprise privée qui n'aurait à pas à se préoccuper de problèmes dus aux coûts irrécupérables (*sunk cost*). De cette façon, la collectivité publique peut aisément faire jouer la concurrence et laisser planer la menace d'une rupture de contrat dans le cas où le sous-traitant ne répondrait plus aux conditions déterminées en début de période.

Bien que cette solution permette aux autorités de contrôle d'exercer une plus grande pression sur les coûts d'exploitation de l'entreprise, elle ne permet pas cependant de résoudre totalement le problème de la révélation exacte des coûts.

16. Conclusions

Au cours de cette étude, nous avons examiné les différentes possibilités de fournir un service en réseau. Nous avons ensuite cherché à déterminer parmi les alternatives possibles, laquelle était la mieux adaptée à ce type de prestation en prenant comme critère de choix l'efficacité productive, c'est-à-dire un indicateur permettant d'isoler les observations qui produisent au coût le plus bas.

L'identification du mode de fourniture le plus efficace exige tout d'abord une formulation théorique des effets économiques de chacun de ces modes sur le comportement des acteurs. Jusqu'ici, les économistes se sont surtout penchés sur les problèmes relatifs au marché et, dans une moindre mesure, à ceux liés à l'intervention publique. Ils ont pratiquement ignoré le domaine se situant dans le domaine un peu flou qui mêle simultanément activité privée et publique. Seule la concession a fait l'objet d'une attention plus marquée de la part des économistes américains en raison de la réglementation particulière de leur pays.

Toutefois, aucune approche théorique n'a pu démontrer la supériorité d'un mode de fourniture sur les autres. Chacun a ses atouts mais aussi ses faiblesses tant du point de vue de l'optimalité économique que du coût et de l'efficacité des instruments de réglementation qu'ils nécessitent. Ce sont avant tout les caractéristiques du service et la structure du marché qui vont déterminer le ou les modes de prestation les mieux adaptés. Le rapport entre les moyens mis en oeuvre et les résultats effectifs sur le comportement des agents est donc une variable à ne jamais négliger lors de toute prise de décision. Le moindre coût de l'intervention doit, en effet, rester un des principaux objectifs des pouvoirs publics étant donné qu'un équilibre économiquement optimal ne peut, en pratique, jamais être atteint.

Quel que soit le mode d'intervention utilisé par la collectivité publique, de nombreux problèmes dus à l'existence d'inefficacité dans le processus de production persistent. Il s'agit alors d'examiner l'ampleur de cette inefficacité afin de prendre les mesures qui conviennent le mieux à chaque situation. L'analyse de la performance, qui fait l'objet de la partie empirique, devrait non seulement permettre de faire un choix parmi les divers modes de fourniture mais aussi de définir les actions à entreprendre.

Notre principal objectif dans l'élaboration d'une mesure de l'efficacité fut de construire un indicateur qui soit le plus neutre possible et qui permette de tenir

compte, autant que possible, des nombreux éléments d'un processus de production. Les récents développements de la théorie de la dualité conviennent particulièrement bien à une analyse d'entreprises en réseau réglementées puisque l'output peut être considéré comme une variable exogène aux décisions des gestionnaires tout en traitant les proportions d'inputs choisies comme une variable endogène. L'utilisation d'une frontière stochastique représente également un élément essentiel de cette recherche car elle réduit considérablement l'importance d'une éventuelle observation aberrante. Dans tout échantillon, des erreurs de mesure subsistent et peuvent biaiser considérablement un résultat d'autant plus que celui-ci se base sur une seule observation comme c'est le cas avec les méthodes déterministes. Enfin, l'introduction de variables de qualité et d'environnement dans la fonction de coût rend chaque firme indépendante du milieu dans lequel elle évolue et évite qu'elle soit pénalisée par une normalisation de son cadre économique et géographique. Ceci d'autant plus que les entreprises observées évoluent dans des contextes forts différents (grandes agglomérations, zone urbain ou rural, etc) et offrent des services de qualité très inégale. A notre connaissance, l'emploi de variables d'environnement dans une étude de la performance n'a été réalisée jusqu'ici que par Deprins et Simar (1989) ainsi que par Burgat et Jeanrenaud (1990). Ces auteurs ne se sont toutefois pas intéressés aux avantages offerts par l'utilisation d'un système d'équations.

Les résultats obtenus soulignent d'ailleurs l'influence des facteurs d'environnement et de qualité puisque l'efficacité productive s'accroît d'environ 8% lorsque ceux-ci sont pris en compte. De même, la différence entre méthodes déterministe et stochastique donne une idée de l'importance des observations extrêmes dans la détermination des résultats puisque nous avons constaté une efficacité moyenne de presque 50% supérieure lorsque la seconde méthode est appliquée. Nous devons cependant insister sur leur cohérence. En effet, les deux approches parviennent aux mêmes conclusions en ce qui concerne les entreprises les moins efficaces et celles qui se sont avérées les plus performantes. Cela signifie que les firmes occupent un classement pratiquement identique quelle que soit la méthode retenue.

Lorsque les facteurs exogènes sont introduits dans un modèle de frontière stochastique, le degré d'inefficacité moyen de notre échantillon est alors relativement faible (environ 10%). Bien que ce résultat semble indiquer une bonne prestation d'ensemble des transports publics urbains en Suisse, il ne doit pas faire oublier que si les 133 entreprises de notre échantillon agissaient de

manière totalement efficace, il leur serait possible d'épargner environ 370 millions de francs. Ceci devient d'autant plus intéressant lorsque l'on sait que les autorités publiques versent à ces compagnies une contribution représentant 31% de leurs coûts totaux. Ainsi, l'amélioration de la performance permettrait de réduire cette part de plus de 36%. Par conséquent, même si les entreprises de transports urbains semblent être relativement bien gérées, de nombreux efforts de rationalisation devraient être réalisés, ne serait-ce que pour devenir moins dépendantes des décisions des autorités politiques locales.

L'absence de différences d'efficacité entre secteurs public et privé n'est en soit pas une surprise. D'une part, les transports urbains sont largement tributaires des contributions financières des collectivités publiques, elles sont donc soumises à des règles et à des contraintes identiques quel que soit le mode de propriété. D'autre part, rappelons que les coûts totaux divulgués par les gestionnaires peuvent être biaisés par des surestimations ou des sous-estimations souvent liées à l'absence de comptabilité analytique. Nous ajouterons qu'en général le secteur public a accès à des capitaux meilleur marché que le secteur privé.

Par ailleurs, nous aimerions insister sur le fait que nous n'avons pas cherché à mesurer l'efficacité, c'est-à-dire le résultat d'une politique des transports, mais uniquement l'efficacité productive. Cette distinction est extrêmement importante car une entreprise peut très bien être gérée efficacement mais offrir un service ne correspondant pas du tout à la demande effective de transports urbains. Dans ce cas, la performance de la firme devrait être reconsidérée ainsi que la conception même des transports urbains. Il est évident que pour atteindre un état économiquement optimal, l'objectif d'efficacité productive ne suffit pas et tout écart entre l'offre et la demande devrait être surveillé de près afin de pouvoir être corrigé. A ce propos, le principal problème est celui du dimensionnement du réseau de transport et de sa capacité. Nous savons que tout réseau est excédentaire puisqu'il doit répondre à d'importantes fluctuations quotidiennes de la demande, la question est donc de savoir quelle est la meilleure manière d'adapter sa capacité à ces variations.

L'analyse de la performance des entreprises de transports urbains nous a permis de relever le comportement particulièrement efficace des entreprises en régime de sous-traitance. Nous pensons à ce propos que cette alternative devrait être examiné plus attentivement par les autorités publiques soucieuses d'efficacité économique. En effet, en cette période de restrictions budgétaires, la sous-traitance d'un réseau de transport devrait laisser aux collectivités publiques

locales l'entière responsabilité de l'établissement des prix et, par conséquent de ses propres contributions à la firme sous-traitante, tout en lui épargnant les problèmes de gestion de l'entreprise. De cette manière, la commune aurait la possibilité de se concentrer sur la gestion globale des transports sur son territoire tout en gardant la responsabilité de sa politique.

REFERENCES

- Aigner, D.J., Lovell, C.A.K. et Schmidt, P., "Formulation and Estimation of Stochastic Frontier Production Function Models", *Journal of Econometrics*, 6, 1977, p. 21-37.
- Allen, R., *Mathematical Analysis for Economists*, MacMillan, Londres, 1938.
- Arrow, K.J., *Essays in the Theory of Risk-Bearing*, North-Holland, Amsterdam, 1970.
- Arrow, K.J., "The Organisation of Economic Activity: Issues Pertinent to the Choice of Market versus Non-Market Allocation", in R.H. Haveman et J. Margolis (éds), *Public Expenditure and Policy Analysis*, Markham, Chicago, 1971.
- Atkinson, S.E. et Halvorsen, R., "A Test of Relative and Absolute Price Efficiency in Regulated Utilities", *Review of Economics and Statistics*, 62, 1980, p. 81-88.
- Averch, H. et Johnson, L.L., "Behaviour of the Firm under Regulatory Constraint", *American Economic Review*, 52, 1962, p. 1053-1059.
- Bailey, E., *Economic Theory of Regulatory Constraint*, Lexington Books, Lexington, Mass., 1973.
- Banker, R.D. et Maindiratta, A., *Maximum Likelihood Estimation of Monotonic and Concave Production Frontiers*, School of Urban and Public Affairs, Carnegie-Mellon University, 1985.
- Barten, A.P., "Maximum Likelihood Estimation of a Complete System of Demand Equations", *European Economic Review*, 1, 1969, p. 7-73.
- Baumol, W.J. et Klevorick, A.K., "Input Choices and Rate-of-Return Regulation: An Overview of the Discussion", *Bell Journal of Economics*, 1, 1970, p. 162-190.
- Baumol, W.J., Panzar, J.C. et Willig, R.D., *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Jovanovich, New York, 1982.
- Berg, S.V. et Tschirhart, J., *Natural Monopoly Regulation: Principles and Practice*, Cambridge University Press, Cambridge, 1988.

- Berndt, E.R. et Christensen, C.R., "The Translog Function and the Substitution of Equipment, Structures and Labor in U.S. Manufacturing 1929-68", *Journal of Econometrics*, 1, 1973, p. 81-114.
- Berndt, E.R. et Wood, D.O., "Technology, Prices and the Derivated Demand for Energy", *Review of Economics and Statistics*, 57, 1975, p. 259-268.
- Boadway, R.W., *Public Sector Economics*, Little, Brown and Company, Boston, 1979.
- Borcherding, T.E., Pommerehne, W.W. et Schneider, F., "Comparing the Efficiency of Private and Public Production: The Evidence from Five Countries", *Zeitschrift für Nationalökonomie (suppl.)*, 2, 1982, p. 127-156.
- Bös, D., *Public Enterprise Economics*, North-Holland, Amsterdam, 1986.
- Braeutigam, R.R., "Commentary on Natural Monopoly Measures and Regulatory Policy", in K. Nowotny, D.B. Smith et H.M. Trebing (éds), *Public Policy Regulation*, Kluwer Academic Publishers, Boston, 1989, p. 217-222.
- Bruggink, T.H., "Public versus Regulated Private Enterprise in the Municipal Water Industry: A Comparison of Operating Costs", *Quarterly Review of Economics and Business*, 22, 1982, p. 11-125.
- Buchanan, J.M. et Stubblebine, W.C., "Externality", *Economica*, 29, 1968, p. 371-384.
- Burgat, P. et Jeanrenaud, C., *Mesure de l'efficacité productive et de l'efficacité - coût: cas des déchets ménagers en Suisse*, Working paper no 9002, Institut de recherches économiques et régionales, Neuchâtel, 1990.
- Buschor, E., *Die öffentliche Finanzwirtschaft zwischen Automatismen und Mittelwerknappung*, Haupt, Berne, 1983.
- Caves, D.W. et Christensen, L.R., "The Relative Efficiency of Public and Private Firms in a Competitive Environment: The Case of Canadian Railroads", *Journal of Political Economy*, vol. 88, 1980, pp. 958-976.
- Charnes, A., Cooper, W.W. et Rhodes, E., "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2, 1978, p. 429-444.
- Charnes, A., Cooper, W.W. et Rhodes, E., "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through", *Management Science*, 27, 1981, p. 688-697.

- Christensen, L.R. et Greene, W.H., "Economies of Scale in U.S. Electric Power Generation", *Journal of political Economy*, 84, 1976, p. 655-676.
- Cowing, T.G. et Stevenson, R.E., "Introduction: Productivity Measurement and Regulated Industries", in T.G. Cowing et R.E. Stevenson (éds), *Productivity Measurement in Regulated Industries*, Academic Press, New York, 1981.
- Crain, W.M. et Zardkoohi, A., "A Test of the Property Rights Theory of the Firm: Water Utilities in the United States", *Journal of Law and Economics*, 21, 1978, p. 395-408.
- Crain, W.M. et Zardkoohi, A., "Public Sector Expansion: Stagnant Technology of Property Rights", *Southern Economic Journal*, 46, 1980.
- Cullis, J.G. et Jones, P.R., *Microeconomics and the Public Economy: A Defence of Leviathan*, Basil Blackwell, New York, 1987.
- Curien, N. et Gensollen, M., "De la théorie des structures industrielles à l'économie des réseaux de télécommunications", *Revue économique*, 38, 1987, p. 521-578.
- Delis, C., "Productivity and Economies of Scale in the Athens-Piraeus Area Electric Buses", *Annales de l'économie publique, sociale et coopérative*, 1988, p. 493-508.
- Demsetz, H., "Why Regulate Utilities?", *Journal of Law and Economics*, 11, 1968, p. 55-66.
- Densison, E.F., *Accounting for United States Economic Growth, 1929 to 1969*, The Brookings Institution, Washington D.C., 1974.
- Deprins, D. et Simar, L., "Estimating Technical Inefficiencies with Correction for Environmental Conditions, with an Application to Railways Companies", *Smash*, no 8901, Bruxelles, 1989.
- DiLorenzo, T.U. et Robinson, R., "Managerial Objectives Subject to Political Market Constraints: Electric Utilities in the U.S.", *Quarterly Review of Economics and Business*, 22, 1982, p. 111-125.
- Encoua, D. et Moreaux, M., "Concurrence et monopole naturel: une approche par la théorie des jeux", *Annales d'économie et de statistiques*, 8, 1987, p.89-116.
- Evans, D.S. et Heckman, J.J., "A Test for Subadditivity of the Cost Function with an Application to the Bell System", *American Economic Review*, 74, 1984, p. 615-623.

- Färe, R., Grosskopf, S. et Logan, J., "The Relative Performance of Publicly-Owned and Privately-Owned Electric Utilities", *Journal of Public Economics*, 26, 1985, p. 86-106.
- Färe, R. et Hunsaker, W., "Notions of Efficiency and their Reference Sets", *Management Science*, 32, 1986, p. 237-243.
- Farrell, M.J., "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society A*, 120, 1957, p. 253-281.
- Farrell, M.J. et Fieldhouse, M., "Estimating Efficient Production Functions under Increasing Returns to Scale", *Journal of the Royal Statistical Society A*, 125, 1962, p. 252-267.
- Feldstein, M., *The American Economy in Transition*, University of Chicago Press, Chicago, 1980.
- Fisk, D., Kiesling, H. et Mueller, T., "Private Provision of Public Services: An Overview", in J.H. Carr (éd.), *Crisis and Constraint in Municipal Finance*, Center for Urban Policy Research, New Brunswick, 1984.
- Fitts, H.C. et Lindveit, E.W., *Alternative Service Delivery Approaches Used by Local Government, A resource Paper*, Governmental Capacity Sharity Program, Office of Policy Development and Research (U.S. Department of Housing and Urban Development), HUD-0003629, Washington D.C., 1984.
- Foreman-Peck, J. et Waterson, M., "The Comparative Efficiency of Public and Private Enterprise in Britain: Electricity Generation between the World Wars", *Economic Journal* (suppl.), 95, 1985, p. 83-95.
- Forsund, F.R. et Hjalmarsson, L., "On the Measurement of Productive Efficiency", *Swedish Journal of Economics*, 76, 1974, p. 141-154.
- Forsund, F.R., Lovell, C.A.K. et Schmidt, P., "A Survey of Frontier Production Functions and of their Relationship to Efficiency Measurement", *Journal of Econometrics*, 13, 1980, p. 5-25.
- Forsyth, P.J., "Airlines and Airports: Privatisation, Competition and Regulation", *Fiscal Studies*, 5, 1984, p. 61-75.
- Gabrielson, A., *On Estimating Efficient Production Functions*, Working Paper no A-85, Chr. Michelsen Institute, Bergen, 1975.
- Gathon, H.J., "Indicators of Partial Productivity and Technical Efficiency in the European Urban Transit Sector", *Annales d'économie publique, sociale et coopérative*, 60, 1989, p. 43-59.

- Greene, W.H., "Maximum Likelihood Estimation of Econometric Frontier Functions", *Journal of Econometrics*, 13, 1980, p. 27-56.
- Greene, W.H., *Econometric Analysis*, MacMillan, New York, 1990.
- Hanoch, G., "The Elasticity of Scale and the Shape of Average Costs", *The American Economic Review*, 65, 1975, p. 492-497.
- Hatry, H.P., *A Review of Private Approaches for Delivery of Public Services*, The Urban Institute Press, Washington D.C., 1983.
- Hatry, H.P. et Valente, C.F., "Alternative Service Delivery Approaches Involving Increased Use of the Private Sector", *The Municipal Yearbook*, International City Management Association, Washington D.C., 1983, p. 199-217
- Hicks, J.R., *The Theory of Wages*, MacMillan, Londres, 1963.
- Jondrow, J., Lovell, C.A.K., Materow, I.S. et Schmidt, P., "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model", *Journal of Econometrics*, 19, 1982, p. 233-238.
- Judge, G.G. et al., *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, New York, 1988.
- Kendrick, J.W., *Postwar productivity Trends in the United States, 1948-1969*, National Bureau of Economic Research, New York, 1973.
- Kumbhakar, S.C., "Production Frontiers and Panel Data: An Application to U.S. Class 1 Railroads", *Journal of Business and Economic Statistics*, 5, 1987, p. 249-255.
- Kumbhakar, S.C., "Economic Performance of US Class 1 Railroads: A Stochastic Frontier Approach", *Applied Economics*, 21, 1989, p. 1433-1446.
- Leibenstein, H., "Allocative Efficiency versus X-Inefficiency", *American Economic Review*, 56, 1966, p. 392-415.
- Lewis, T.R. "Reputation and Contractual Performance in Long-Term Projects", *Rand Journal of Economics*, 17, 1986, p. 141-157.
- Manzini, A., *Efficacité publique et privée: analyse théorique et vérification empirique dans le cas du transport aérien*, Thèse de doctorat, Université de Genève, Genève, 1990.

- McFadden, D., "Cost, Revenue, and Profit Functions", in M.Fuss et D. McFadden (éds), *Production Economics: A Dual Approach to Theory and Applications*, vol. 1, North-Holland, Amsterdam, 1978, p. 3-109.
- Maddala, G.S., *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, 1983.
- Meeusen, W. et van den Broeck, J., "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", *International Economic Review*, 18, 1977, p. 435-444.
- Meyer, R.A., "Publicly Owned versus Privately Owned Utilities: A Policy Choice", *Review of Economics and Statistics*, 57, 1975, p. 391-399.
- Millward, R. et Ward, R., *The Comparative Performance of Public and Private Gas Enterprises in the Late Nineteenth Century*, mimeo, University of Salford, 1987.
- Mood, A.M., Graybill, F.A. et Boes, D.C., *Introduction to the Theory of Statistics*, McGraw-Hill, New York, 1974.
- Musgrave, R.A. et Musgrave, P., *The Theory of Public Finance*, McGraw-Hill, New York, 1980.
- Neuberg, L.G., "Two Issues in the Municipal Ownership of Electric Power Distribution Systems", *Bell Journal of Economics*, 8, 1977, 303-323.
- Nyskanen, W.J., *Bureaucracy and Representative Government*, Adline Atherton, Chicago, 1977.
- Pescatrice, D.R. et Trapini, J.M., "The Performance and Objectives of Public and Private Utilities Operating in the United States", *Journal of Public Economics*, 13, 1980, p. 259-276.
- Posner, R.A., "Theories of Economic Regulation", *Bell Journal of Economics*, 5, 1974, p. 335-358.
- Pryke, R., "The Comparative Performance of Public and Private Enterprise", *Fiscal Studies*, 3, 1982, p.68-81.
- Quigley, J.M. et Kemper, P., *The Economics of Refuse Collection*, Ballinger, Cambridge, Mass., 1976.
- Richmond, J., "Estimating the Efficiency of Production", *International Economic Review*, 15, 1974, p. 515-521.

- Savas, E.S. et Stevens, B., *Evaluating the Organisation of Service Delivery: Solid Waste Collection and Disposal*, Columbia University, New York, 1976.
- Savas, E.S., *Privatizing the Public Sector: How to Shrink Government*, Chatham House, Chatham, N.J., 1982.
- Savas, E.S., *Privatization, The Key to Better Government*, Chatham House, Chatham, N.J., 1987.
- Schmidt, P., "Frontier Productions Functions", *Econometric Reviews*, 4, 1986, p. 289-328.
- Schwab, N. et Christe, L., "Réduire le coût des services publics", in C. Jeanrenaud et W. Moesen (éds), *Gérer l'austérité budgétaire*, Economica, Paris, 1990.
- Sharkey, W.W., *The Theory of Natural Monopoly*, Cambridge University Press, Cambridge, 1982.
- Shephard, W.W., *Cost and Production Functions*, Princeton, University Press, Princeton, 1953.
- Sherman, R., *The Regulation of Monopoly*, Cambridge University Press, Cambridge, 1989.
- Silberberg, E., *The Structure of Economics: A Mathematical Analysis*, McGraw-Hill, New York, 1990.
- Spence, A.M., "Monopoly, Quality and Regulation", *Bell Journal of Economics*, 6, 1975, p. 417-429.
- Stevenson, R.E., "Likelihood Functions for Generalized Stochastic Frontier Estimation", *Journal of Econometrics*, 13, 1980, p. 57-66.
- Thiry, B. et Tulkens, H., "Productivité, efficacité et progrès technique: notions et mesures dans l'analyse économique", in *Rapport préparatoire de la Commission 5 "Efficacité et management"*, 8e Congrès des économistes belges de langues française, CIFOP, Charleroi, 1988, p. 17-51.
- Uzawa, H., "Production Function with Constant Elasticities of Substitution", *Review of Economic Studies*, 29, 1962, p. 291-299.
- Varian, H.R., *Microeconomic Analysis*, W.W. Norton and Company, New York, 1984.

- Vickers, J et Yarrow, G., *Privatization and the Natural Monopolies*, Public Policy Centre, Londres, 1985.
- Vickers, J et Yarrow, G., *Privatization - An Economic Analysis*, The MIT Press, Cambridge, Mass., 1988.
- Viton, P.A., "A Translog Cost Function for Urban Bus Transit", *Journal of Industrial Economics*, 29, 1981, p. 287-304.
- Waterson, M., *Regulation of the Firm and Natural Monopoly*, Basil Blackwell, Oxford, 1988.
- Williamson, O.E., "Franchise Bidding for Natural Monopolies - In General and with Respect to CATV", *Bell Journal of Economics*, 7, 1976, p. 73-104.
- Wolf, C., *Markets or Governments: Choosing Between Imperfect Alternatives*, The Rand Corporation, Santa Monica, 1986.
- Yarrow, G., "Privatization in Theory and Practice", *Economic Policy*, 2, 1986, pp. 323-377.
- Yunker, J.A., "Economic Performance of Public and Private Enterprise: The Case of U.S. Electric Utilities", *Journal of Economics and Business*, 28, 1975, pp. 60-67.
- Zellner, A., "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of the American Statistical Association*, 57, 1962, p. 348-368.
- Zellner, A. et Revankar, N.S., "Generalized Production Functions", *Review of Economic Studies*, 39, 1969, p. 241-250.