

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LES DÉTERMINANTS DE LA SANTÉ EN CHINE :
UNE APPROCHE DE FRONTIÈRES STOCHASTIQUES

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉCONOMIQUE

PAR
DIANE COFFESSOU YAO

SEPTEMBRE 2009

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je voudrais remercier spécialement mon directeur de mémoire, le Professeur Pierre Ouellette du département des sciences économiques à l'Université du Québec à Montréal. Sa disponibilité, sa patience et surtout ses commentaires m'ont été indispensables tout au long de la réalisation de ce mémoire. Travailler sous sa direction a été un privilège extraordinaire et je ne trouverai jamais les mots exacts pour exprimer toute ma gratitude à son égard. Merci infiniment d'avoir cru en moi et de m'avoir assistée dans ce projet.

Je remercie également le Colonel Mirassou Jean-Claude de la Présidence de la République de Côte d'Ivoire pour tout le support financier qu'il a apporté à la réalisation de mes études universitaires. Sa patience, sa confiance et surtout ses conseils m'ont permise de concentrer mes efforts sur mon but principal. Merci pour tout ce que tu as accompli pour moi. Je te suis infiniment reconnaissante.

Je remercie les membres de ma famille d'avoir été patients et d'avoir fait confiance à mon désir de poursuivre mes études, notamment messieurs Yao Michel et Ian Gingras pour la persévérance et le soutien dont ils ont fait preuve durant mon projet de recherche. Finalement, je n'oublierai jamais de mentionner ma mère bien-aimée Yao Amoa Amélie qui n'est plus de notre monde à présent, mais qui a été ma principale source d'inspiration.

TABLE DES MATIÈRES

REMERCIEMENTS	ii
LISTE DES TABLEAUX	v
RÉSUMÉ	vi
INTRODUCTION	1
CHAPITRE I	
REVUE DE LITTÉRATURE	5
1.1 Première génération	6
1.2 Deuxième génération	8
1.3 Troisième génération	9
CHAPITRE II	
DÉPENSES EN SANTÉ ET SYSTÈME DE SANTÉ EN CHINE	14
CHAPITRE III	
DONNÉES	20
3.1 Indicateurs de la santé	22
3.2 Variables sociodémographiques	23
3.3 Variables économiques	24
3.4 Variables éducationnelles	25
3.5 Variables de ressources de santé	26
3.6 Variables de style de vie	27
3.7 Statistiques descriptives	27
3.8 Corrélation entre la variable dépendante et les variables explicatives	29
CHAPITRE IV	
ESTIMATIONS	31
4.1 Moindres carrés ordinaires (MCO)	31
4.2 Méthodes d'estimation des données de panel	32
4.2.1 Méthode de panel avec effets fixes	33
4.2.2 Méthode de panel avec effets aléatoires	34
4.2.3 Méthode de panel et racines unitaires	34
4.3 Modèle de frontières stochastiques avec des données de panel	36
4.3.1 Modèle de frontières stochastiques : données de panel	38

4.3.2	Modèle de frontières stochastiques: modèle invariant et modèle variable dans le temps-----	39
4.3.3	Modèle de frontières stochastiques avec variable d'output « négative » -----	41
CHAPITRE V		
RÉSULTATS -----		42
5.1	Résultats méthodologiques -----	42
5.1.1	Importance de la méthode-----	42
5.1.2	Importance de la définition des variables -----	46
5.2	Résultats sur les variables-----	48
5.3	Efficienc relative des déterminants de santé -----	51
CONCLUSION -----		55
RÉFÉRENCES -----		57

LISTE DES TABLEAUX

Tableaux du chapitre 3

Tableau 3.1 : Listes des régions étudiées en Chine

Tableau 3.2 : Statistiques descriptives

Tableau 3.3 : Définitions des variables

Tableau 3.4 : Tableau de corrélation

Tableaux du chapitre 5

Tableau 5.1 : Modèle en fonction de la méthode

Tableau 5.2 : Modèle en fonction de la définition de la variable d'éducation

Tableau 5.3 : Impact d'une variation de 1 % des inputs de santé sur le taux de mortalité

Tableau 5.4 : Efficience des inputs de santé par région en 2007

Tableau 5.5 : Efficience des inputs de santé par groupe de régions en 2007

RÉSUMÉ

Les pays développés ont les indicateurs de santé les plus favorables du monde, à savoir, des taux de mortalités infantiles très bas et des niveaux d'espérance de vie très élevés. Plusieurs études essayent de cerner la relation qui existe entre les ressources de santé mises à la disposition des populations et l'état de santé de ces mêmes populations. Il est possible de le faire en étudiant les dépenses de santé en relation avec l'état de santé. Toutefois, dans la plupart des études passées, des difficultés inhérentes à l'hétérogénéité des données ont fait en sorte qu'on obtient le plus souvent des résultats incertains. Il y a un manque de données de qualité, alors que le recours à des données homogènes est d'une très grande importance pour pouvoir mieux cerner l'impact des ressources en santé. Une des seules façons d'y arriver est d'utiliser des données provenant d'une seule source. Par exemple, le choix peut se porter sur des données provenant d'un seul pays, d'où l'intérêt d'utiliser des données régionales de la Chine. L'utilisation de ces données permet de tenir compte d'une part, du poids de la taille relative de chaque région dans l'analyse. Et d'autre part, elle permet d'éviter des problèmes liés à l'hétérogénéité des données. De ce fait, il n'y a pas de nécessité de conversion des unités monétaires, ni d'imprécision sur la notion de santé comme ce serait le cas dans une étude basée sur plusieurs pays. Les données chinoises sont fiables et de grande qualité et par conséquent peuvent être utilisées pour identifier les déterminants de la santé. Le fait que la Chine soit un pays émergent constitue un attrait en soi. Dans les pays développés, les investissements en santé sont extrêmement importants et de fait, le taux de rendement des dépenses en santé devient très faible (Crémieux *et al.*, 1999). L'identification des déterminants de la santé devient difficile car la relation est identifiable à une constante et donc peu significative. Comme les dépenses de santé et les niveaux de santé dans les pays en voie de développement et les pays en émergence sont plus faibles, les possibilités d'identification des déterminants de la santé deviennent plus fortes. C'est cette combinaison de données de qualité et homogènes ainsi que de taux de rendement positif qui explique le recours aux données chinoises.

Mots-clés : Frontières stochastiques, déterminants de la santé, taux de mortalité.

INTRODUCTION

La question de santé se situe au centre des grandes préoccupations de la majorité des gouvernements et des citoyens dans le monde. Au Canada, le problème du financement dans le système de santé a beaucoup retenu l'attention des gouvernements, tant fédéral que provinciaux. Plusieurs commissions (Kirby, Castonguay, Rochon, Clair, Ménard, Romanow)¹ ont été mises en place pour étudier le financement et l'organisation du domaine de santé et jusqu'à présent, les consensus sur les solutions ont été plutôt rares. L'absence de tels consensus découle de la grande difficulté qu'il y a à évaluer l'impact des décisions sur la santé des populations. En effet, l'origine fondamentale d'un tel problème se situe dans la difficulté à mesurer la santé et à trouver une façon appropriée de quantifier l'impact des politiques de santé.

Concept difficile à définir et à cerner, la santé est le résultat de plusieurs décisions et de variables génétiques et environnementales, habituellement hors de la portée des individus. Les gouvernements influencent le plus souvent plusieurs de ces variables. Les dépenses de santé sont l'une d'entre elles, mais pas la seule, loin de là, car plusieurs politiques telles que celles d'éducation ont des impacts sur la santé des populations, tout comme les politiques d'hygiène publique (eau, aqueducs, déchets, etc.), de même que celles de la redistribution du revenu.

Lorsqu'un gouvernement veut intervenir en santé, il doit être en mesure d'identifier quels sont les déterminants de la santé les plus importants et quelles sont les régions où les effets sont les plus importants. Est-ce que les politiques sont plus efficaces dans les régions pauvres que dans les régions riches ?

Ce type de questions concerne non seulement les pays riches comme le Canada, mais aussi, et peut-être encore plus, les pays pauvres ou en émergence. En effet, en raison de leur faible

¹ <http://www.hc-sc.gc.ca/hcs-sss/hhr-rhs/strateg/kirby-fra.php>
http://www.financementsante.gouv.qc.ca/fr/rapport/pdf/RapportFR_FinancementSante.pdf
<http://dsp-psd.tpsgc.gc.ca/Collection-R/LoPBdP/BP/prb0037-f.htm>
<http://publications.msss.gouv.qc.ca/acrobat/II/documentation/2005/Rapportmenard.pdf>
<http://dsp-psd.pwgsc.gc.ca/Collection/H117-1-2002F.pdf>

capacité financière, ces pays doivent s'assurer que leur argent est investi là où le taux de rendement est le plus élevé. C'est le cas de la Chine.

La République Populaire de Chine, le plus souvent appelée la « Chine », est le plus grand état de l'Asie orientale et le troisième plus grand pays au monde en surface après la Russie et le Canada. La Chine² est également le pays le plus peuplé au monde, avec environ 1,3 milliard d'habitants, répartis sur une superficie de 9 596 960 de km².

Aujourd'hui, la Chine fait partie des plus grandes surprises économiques dans le monde. La croissance économique a favorisé la réduction du nombre de Chinois vivant dans la pauvreté absolue. L'augmentation de l'espérance de vie de la population, de même que la diminution du taux de mortalité infantile en Chine sont certains aspects remarquables de cette croissance. En 1978, le PIB était de 364 milliards de yuans chinois et en 2006, il se chiffrait à 21 180 milliards de yuans. Le PIB per capita était de 381 yuans en 1978, il est passé à 16 084 yuans en 2006³. Comparativement aux pays riches, ces montants sont peu élevés, mais en termes de croissance, il s'agit d'une augmentation extraordinaire de 4 121 %.

Selon le Centre de recherche pour le développement international (CRDI), le système de santé chinois nécessite des changements, car l'augmentation des coûts et la détérioration du système de prestation de soins limitent l'accès aux services de santé. En effet, le ministère de la santé en Chine a affirmé qu'il y a 41 % des citoyens chinois qui n'ont pas accès à des soins médicaux lorsqu'ils sont malades⁴. Même si la Chine a entrepris des efforts avec la réforme de son système de santé depuis 1978, il n'en demeure pas moins qu'il y a quelques lacunes à combler. La Chine a connu ces dernières années une croissance économique extraordinaire, mais le développement social et humain n'a pas suivi cette expansion. Cependant, la Chine montre une image en matière de santé plus acceptable que celle de la majorité des pays pauvres qui vivent quant à eux, une situation alarmante. C'est d'ailleurs dans un tel contexte que les gouvernements chinois prônent l'amélioration de la santé via un développement économique.

² http://fr.wikipedia.org/wiki/R%C3%A9publique_populaire_de_Chine: Population pour l'année 2007

³ <http://www.stats.gov.cn/tjsj/ndsjsj/2007/indexeh.htm>

Chapitre 3 : national accounts: tableau : 3.1

⁴ <http://archive.idrc.ca/books/reports/2001/04-02f.html>

Étant donné que le revenu augmente, on peut penser que les dépenses de santé s'accroîtront dans le futur. Ainsi, la question de savoir comment les budgets du gouvernement chinois seront alloués devient un objet de recherche en soi. Les gouvernements doivent par conséquent être en mesure de se demander comment et où investir dans le système de santé. Quels sont les déterminants de la santé et quelle est l'importance d'investir dans le domaine de la santé ?

Au-delà des questions purement politiques de santé et de la question chinoise, ce mémoire renferme aussi des implications méthodologiques. Il fait une comparaison des modèles déterministes versus les modèles stochastiques. (Gerdtham *et al.*, 2000) ont regroupé les approches méthodologiques en trois générations d'études. Auparavant, la plupart des études réalisées dans le domaine de la santé étaient basées sur des méthodes économétriques de base telles que les moindres carrés ordinaires (MCO). Quelques années après, ces méthodes ont été remplacées par des méthodes économétriques plus raffinées telle que l'estimation des données de panel avec effets fixes ou aléatoires. À la suite des estimations en panel fixe ou aléatoire, survinrent les méthodes avec racines unitaires. Toutes ces études supposaient l'efficacité de la technologie de production de santé. Greene (2003) remet en question cette hypothèse et propose d'estimer une fonction de production sans hypothèse d'efficacité, en utilisant un modèle de frontières stochastiques. Les modèles de frontières stochastiques constituent donc une quatrième génération de modèles.

Dans ce mémoire, nous estimerons une fonction de production de santé pour des régions de la Chine afin de comparer l'efficacité de l'état de santé, ainsi que l'impact relatif des divers déterminants de santé. La démarche utilisée exige une étude des données longitudinales, donc une estimation de données de panel dans un modèle de frontières stochastiques.

À cette fin, ce document est organisé de la façon suivante. Le premier chapitre est une revue de littérature. Il discute des différentes méthodes utilisées, donc des différentes générations proposées par (Gerdtham *et al.*, 2000) en matière d'études réalisées dans le domaine de la santé, ainsi que des résultats qui en découlent. Le deuxième chapitre est un aperçu de l'histoire du système de santé en Chine. Il discute brièvement du fonctionnement du système de santé chinois et montre quelques statistiques sur les dépenses de santé.

Cette étude prend en considération deux groupes de facteurs inter-reliés qui sont les facteurs socioculturels et les facteurs économiques. Ainsi, le troisième chapitre est une description de la banque de données et des facteurs précédemment mentionnés utilisés pour la réalisation de ce document. Enfin, les chapitres quatre et cinq discutent respectivement de la méthode d'estimation et de l'interprétation des résultats trouvés.

CHAPITRE I

REVUE DE LITTÉRATURE

Il existe une vaste littérature économique dans le domaine de la santé essayant d'identifier les déterminants de la santé. À première vue, ces études sont plutôt disparates. Des différences importantes quant à la nature des données et des méthodes sont évidentes à la première lecture. Cependant, on peut les unifier en proposant le concept de fonction de production de la santé. Dans cette optique, l'état de santé (approximé de différentes façons par les auteurs) devient le résultat de décision et de consommation de la part des individus ou de groupes d'individus. Bien entendu, diverses variables de style de vie comme par exemple l'hygiène auront aussi un impact de même que des facteurs environnementaux. La fonction de production a pour but de relier ces facteurs à l'état de santé. On peut l'écrire de la façon suivante :

$$S \equiv f(x_1, \dots, x_n),$$

où S est l'état de santé (parfois un vecteur) et $x \equiv (x_1, \dots, x_n)$ est le vecteur des déterminants de la santé. Ce concept permet d'unifier les études sur les taux de rendement des ressources en santé.

À partir de cette représentation (utilisée implicitement par la plupart des auteurs) mais explicitée par (Crémieux *et al.*, 1999)⁵, il devient facile de situer les différentes contributions. En fait, les études se caractérisent à la fois par la nature des déterminants retenus (le vecteur x), par le nombre et le type de pays étudiés et par la nature de la structure des termes d'erreur utilisée pour l'estimation.

⁵ Crémieux, Ouellette et Pilon. *Health Economics*, 1999

Leu (1986) a utilisé la même méthode que Newhouse, mais avec un échantillon plus grand, composé de 19 pays de l'OCDE. Il a analysé l'état de santé via la relation qui existe entre les dépenses de santé et le revenu. De plus, il a considéré plusieurs variables autres que le revenu telles que les variables sociodémographiques dans le but d'éviter les problèmes liés au biais d'omission de variables significatives. La méthode d'estimation utilisée est celle des moindres carrés ordinaires, mais dans ce cas, il s'agit de coupes transversales. Ses résultats ne sont pas très différents de ceux de Newhouse, car la majeure partie des variations des dépenses en santé était expliquée par celles du revenu. Autrement dit, l'état de santé dépendait de l'évolution des dépenses effectuées en santé, expliquée grandement par les variations du revenu. Les résultats de Leu montraient néanmoins la pertinence d'autres variables dans les variations de l'état de santé comme les variables sociodémographiques.

Gerdtham et Jönsson (1992) ont tenté d'expliquer l'état de santé en réalisant une autre étude sur le lien entre les dépenses en santé et le revenu per capita. Ils ont utilisé un échantillon de 22 pays de l'OCDE. Ces auteurs ont trouvé qu'il existait une corrélation positive entre les dépenses en santé en parité du pouvoir d'achat et le revenu par habitant⁹. Cela leur permit de conclure qu'empiriquement, le ratio dépenses en santé et revenu augmente avec le niveau d'industrialisation des pays.

(Gerdtham *et al.*, 1992) réalisèrent des travaux presque similaires à la précédente étude sur les déterminants de l'état de santé en utilisant un échantillon de 19 pays de l'OCDE. Ils trouvèrent que le revenu par habitant expliquait de façon significative la variance des dépenses en santé et par conséquent l'état de santé, mais que l'urbanisation était un autre facteur explicatif pertinent. D'autres variables telles que institutionnelles et sociodémographiques avaient aussi un effet non négligeable sur la santé. Par conséquent, l'état de santé n'était plus uniquement déterminé par les dépenses de santé, mais aussi par d'autres facteurs sociaux.

Toutes les études mentionnées ci-dessus ont eu recours à des méthodes de moindres carrés ordinaires (MCO), même dans le cas de coupes transversales répétées, où des techniques de panel auraient dû être appliquées. D'où l'apparition d'une seconde vague d'étude ayant

⁹ Gerdtham. U.G. et Jönsson. B. (1991). Conversion Factor Instability in International Comparisons of Health Care Expenditure. *Journal of Health Economics*, Vol. 10, page 227 à 234.

recours aux données de panel. (Gerdtham *et al.*, 2000) regroupent ces études en une seconde génération.

1.2 Deuxième génération

Cette deuxième génération s'est aussi inspirée du modèle de Newhouse, toutefois en changeant la méthode des moindres carrés ordinaires pour une méthode d'estimation de données de panel, en raison surtout de la disponibilité des données pour plusieurs périodes. Autrement dit, les méthodes économétriques se sont raffinées avec cette seconde génération.

Hitiris et Posnett (1992) ont travaillé sur un échantillon composé de 20 pays de l'OCDE de 1960 à 1987. Ils disposaient donc d'un total de 520 observations. Ils ont répliqué les études précédentes (utilisant des MCO) avec des données de panel, c'est-à-dire plusieurs données sur plusieurs périodes. Ils ont trouvé d'un côté, qu'il existait une corrélation positive entre les dépenses de santé par habitant et le revenu per capita et d'un autre côté, que l'élasticité du revenu se situait autour de 1. En conséquence, les services de santé ne sont pas un bien de luxe, mais semble plutôt se rapprocher de la définition d'un bien nécessaire ou normal. Ils ont démontré l'importance d'autres facteurs comme par exemple le taux de mortalité.

Leur étude a été critiquée en raison de l'hétérogénéité des données. En effet, les données provenant de différents pays ne sont pas comparables entre elles et la notion de santé peut être floue d'un pays à un autre.

Crémieux, Ouellette et Pilon (1999) ont réalisé une étude sur le lien existant entre les dépenses de santé et l'état de santé au Canada. Ils ont utilisé des données nationales en vue de pallier les problèmes liés à l'homogénéité des données. De plus, à la différence des études précédentes, ils ont considéré dans leur analyse des variables sociodémographiques, des variables nutritionnelles et des variables de style de vie, dans le but d'éviter des biais d'omission de variables significatives. Ils ont trouvé qu'il y avait une corrélation positive entre les dépenses en santé et le niveau de santé des populations.

(Crémieux *et al.*, 2004) ont effectué la même étude, mais cette fois avec des données plus nombreuses des États-Unis. Aussi, ont-ils tenu compte de nouvelles variables telles que le nombre de médecins, le nombre de produits pharmaceutiques etc.

1.3 Troisième génération

(Gerdtham *et al.*, 2000) distinguent une troisième génération constituée de recherches surtout méthodologiques, sur la nature stochastique des données. En fait, on regarde si les données sont stationnaires. Si ce n'est pas le cas, il faut corriger les données afin de les rendre stationnaires. Une procédure de test se fait en vérifiant la présence de racines unitaires.

Roberts¹⁰ (1999) identifie 3 principales lacunes dans la majorité des recherches économétriques réalisées sur les dépenses de soins de santé qui sont : le manque d'analyse dynamique sur la relation entre les dépenses de santé, le revenu et d'autres variables potentielles, la non-considération de l'hétérogénéité des données et la négligence à l'égard des tests de sensibilité. Elle effectue une étude sur la dynamique des données hétérogènes pour des variables non-stationnaires, et précisément, sur la relation dynamique entre les dépenses de santé per capita, le revenu per capita, l'état de santé et d'autres variables potentielles telles que l'âge de la population et les dépenses publiques. Évidemment, son modèle prend en considération les trois principales lacunes. Son échantillon est composé de 20 pays de l'OCDE durant les années 1960 à 1993. Pour séparer l'effet prix et l'effet revenu, elle inclut dans son modèle une composante représentant le prix relatif des soins de santé (ratio indice de prix de santé et indice de prix du PIB) afin de voir si les pays riches dépensent plus en ressources de santé ou simplement, s'ils dépensent dans des facteurs où les prix sont les plus élevés. Roberts trouve au terme de ses travaux que le revenu est le premier facteur corrélé positivement avec les dépenses de santé. Les tests dynamiques semblent suggérer que l'élasticité-revenu de long-terme est environ égale à 1. L'âge de la population ne semble pas avoir d'effets à court terme, mais une proportion des dépenses publiques a un impact de long-terme sur la santé via les dépenses en santé.

¹⁰ J. Roberts (1999). Sensitivity of Elasticity Estimates for OECD Health Care Spending: Analysis of a Dynamic Heterogeneous Data Field. *Health Economics*. Vol. 8, page 459-472

Les études précédentes font ressortir un impact faible et souvent non significatif des dépenses de santé sur l'état de santé. Dans son mémoire, Berthé (2003) a montré que le lien entre les ressources de santé et l'état de santé pose un réel problème lorsqu'on utilise des données de certains pays industrialisés en raison du niveau de dépenses élevées effectuées dans leur système de santé. Ces pays développés investissent massivement en santé et le taux de rendement des dépenses de santé tend à devenir nul à la marge. Dans le cas des pays moins riches, l'effet est plutôt contraire. Berthé (2003) a montré que le taux de rendement des dépenses de santé dans ces pays en émergence demeure très élevé.

Les travaux antérieurs se sont tous inspirés des analyses de Newhouse et ont beaucoup contribué à l'évolution de la théorie économique en ce qui concerne la santé. Cependant, la plupart de ces études, avec l'exception notable de Berthé (2003), ne concernent que les pays de l'OCDE. Cela pose la question de la validité des conclusions (un effet négligeable des dépenses de santé sur l'état de santé des populations) pour les autres pays. Toutes ces études mentionnées ci-haut sont sujettes à d'immenses problèmes de qualité de données, de méthodes économétriques biaisées, d'hétérogénéité des données, de biais d'omission de variables significatives, d'absence de modèles théoriques etc. On peut donc conclure que les résultats dépendent de la qualité des données et des méthodes retenues.

Les études des trois générations retenues par Gerdtham ont comme caractéristiques de supposer que tous les pays (ou régions) se trouvent sur la frontière de production. Il s'agit d'une hypothèse d'efficacité. Techniquement, cela revient à ajouter un terme d'erreur à la fonction de production :

$$S = f(x_1, \dots, x_n) + u,$$

où u , distribué normalement et d'espérance nulle, ne représente que des erreurs de mesure. Selon que les données soient des coupes transversales, des séries temporelles et des panels, on aura recours à une méthode plutôt qu'une autre en tenant compte de la possibilité de racines unitaires.

Greene (2004a) propose de mettre de côté cette hypothèse en utilisant des méthodes de frontières stochastiques. Cette proposition peut constituer une quatrième génération. Selon

cette méthode, les observations ne se retrouvent plus nécessairement sur la frontière et il faut par conséquent distinguer les observations efficaces de celles qui ne le sont pas. On procède en reformulant le modèle comme suit :

$$S = f(x_1, \dots, x_n) + u - v,$$

où v représente la distance entre l'observation et la frontière efficiente. Par définition, il s'agit d'un terme unilatère ($v \geq 0$). La question de la distribution de ce terme et la méthode d'estimation seront abordées dans le chapitre 4.

Greene (2004b) a réalisé une étude sur le système des soins de santé des pays membres de l'organisation mondiale de la santé (OMS) en utilisant le modèle de frontières stochastiques. En effet, Greene (2004b) affirme que l'évaluation de l'efficience des politiques et des réformes dans le système de santé fait face à deux obstacles majeurs qui sont, la quantification des buts et objectifs de ces politiques et le choix des variables pouvant expliquer ses réformes. Il est parti du fait que la majorité des études réalisées avec le modèle de frontières stochastiques étaient limitées parce qu'elles n'arrivaient pas à distinguer l'hétérogénéité des pays de l'inefficience. Il a utilisé un échantillon composé de 191 pays sur une période de cinq ans (de 1993 à 1997), afin de pouvoir mesurer l'inefficience du système de soins de santé dans les pays membres de l'OMS. Il a estimé l'inefficience en appliquant le modèle de frontières stochastiques avec effets fixes, puis celui avec effets aléatoires. L'auteur a trouvé que, pour arriver à différencier l'hétérogénéité de l'efficience, il faut identifier et estimer l'inefficience technique en relâchant les hypothèses établies auparavant sur les résidus. Il a utilisé la densité de la population et l'éducation pour estimer les dépenses publiques en santé. Il a trouvé que lorsqu'il s'agit des pays membres de l'OCDE, les soins de santé devenaient un bien de luxe avec une élasticité supérieure à 1 (1,23) et un bien normal pour les pays émergents avec une élasticité autour de 1. Par le choix de ses variables explicatives, il montre aussi l'importance d'autres variables dans les services de soins de santé notamment la densité de la population ainsi que des variables liées à l'éducation.

Au terme de cette section, il faut retenir que la qualité des données recueillies dans le cas de la plupart des études réalisées en santé demeure un sujet de discussion. Il est capital de cibler

le type de ressources ayant le plus haut rendement en termes de santé. Depuis 15 ans, l'accent a été mis sur le développement et l'utilisation de méthodes appropriées, mais le problème résiduel concerne l'obtention de données fiables. Ces données proviennent d'organismes internationaux (OCDE, ONU, OMS, BAD)¹¹ qui recueillent des données de divers pays pour réaliser différentes études. Cependant, elles ne sont pas entièrement comparables parce que la définition des ressources change d'un pays à un autre et la notion de santé est floue (inclusion ou non de la maladie mentale, etc.). En résumé, il y a un manque de données de qualité, alors que le recours à des données homogènes est d'une très grande importance pour pouvoir mieux cerner l'impact des ressources en santé. Une des seules façons d'y arriver est d'utiliser des données provenant d'une seule source. C'est ainsi que se démarque notre étude en utilisant des données d'un seul pays en voie de développement. Cependant, la question n'est pas de savoir quel pays est étudié, s'il est riche ou pauvre, mais consiste avant tout à satisfaire la condition primordiale qui est l'obtention de données de confiance, dans le but de faire ressortir l'impact économique de l'investissement en santé sur l'état de santé dans un pays. En utilisant les données d'un seul pays, on maintient la définition de celles-ci sans avoir à convertir les unités monétaires. Dans le cas contraire, le choix du taux de conversion des unités monétaires aurait une grande influence dans l'étude. Dans le but de pallier les problèmes d'hétérogénéité et de manques de données, deux études ont été menées avec des données nationales : (Crémieux *et al.*, 1999) avec des données canadiennes et (Crémieux *et al.*, 2001) avec des données américaines.

Des données chinoises seront utilisées pour calibrer notre modèle. En effet, elles sont un grand avantage de ce mémoire dans la mesure où elles permettent de disposer de données homogènes et fiables. Ce travail se distingue donc des travaux précédents, car il tient compte des défauts des études passées. La démarche utilisée exige des données longitudinales, donc une méthode d'estimation avec des données de panel. Cependant, l'hypothèse d'efficience ne sera pas imposée et le modèle retenu dans ce mémoire est celui de Greene, donc un modèle d'estimation de frontières stochastiques avec des données de panel.

¹¹ http://www.oecd.org/home/0,3305,fr_2649_201185_1_1_1_1_1.00.html
<http://www.un.org/french/>
<http://www.who.in/fr/>
http://www.afdb.org/portal/page?_pageid=473,1&_dad=portal&_schema=PORTAL

Avant d'aborder des questions d'ordres économiques, il est fondamental de se pencher sur les aspects sociaux qui s'y rattachent. Quelles sont les caractéristiques sociales de la Chine ? Comment fonctionne le système de santé ?

CHAPITRE II

DÉPENSES EN SANTÉ ET SYSTÈME DE SANTÉ EN CHINE

Zhang Binghui affirme que la structure sociale en Chine a subi des changements et s'est diversifiée, surtout en raison de son ouverture vers l'extérieur¹². Ce dernier stipule que les différences sont fortement caractérisées par des changements socio-économiques et surtout professionnels à cause de la diversification des métiers. La famille traditionnelle, dit-il, s'est transformée en famille avec « enfants uniques » et celle-ci a accédé plus facilement à une certaine indépendance économique, plus empreinte à des changements et à la nouveauté. Ce phénomène a inévitablement mis en place deux entités distinctes : la famille dite urbaine et celle dite rurale¹³.

De nos jours, la Chine fait partie de l'une des plus grandes surprises au monde en matière de croissance économique. L'Organisation de coopération et de développement économique (OCDE) constate dans sa première étude économique réalisée sur la Chine que le rythme de la croissance économique de cette dernière s'établissant à 9 % en moyenne par année depuis vingt ans ne semble guère ralentir¹⁴. Cependant, le niveau de revenu est encore faible et il existe des inégalités dans plusieurs domaines sociaux économiques en Chine tel que l'éducation, la culture et surtout le système de santé publique¹⁵.

La croissance économique en Chine a permis de réduire le nombre de Chinois vivant dans la pauvreté absolue. L'espérance de vie et le taux de mortalité infantile ont respectivement augmenté et diminué au sein de la population. De 1949 à 2001, l'espérance de vie a augmenté d'environ 35 années et le taux de mortalité infantile a diminué de 200 pour mille naissances¹⁶.

¹² <http://netx.u-paris10.fr/actuelmarx/m4zhang.htm>

¹³ <http://netx.u-paris10.fr/actuelmarx/m4zhang.htm>

¹⁴ http://www.oecd.org/document/15/0,3343,fr_2649_201185_35363613_1_1_1_1.00.html

¹⁵ Étude économique de la Chine 2005. « OECD Economics surveys China » : Publiée le 16 septembre 2005 ;

<http://www.oecd.org/dataoecd/51/9/35308753.pdf>

¹⁶ *The Republic of China Yearbook*. 2004; Bibliothèque Centrale de l'Université du Québec à Montréal (UQAM) : Centrale Référence. YY CHINI A1 C583/(2004). page 5 à 49. page 6

Cependant, cette amélioration s'est surtout produite de 1960 à 1970, car à partir des années 1980, les gains sont devenus plus modestes. Il y a eu d'énormes investissements en santé, mais le système présente toujours quelques faiblesses. La défaillance des infrastructures médicales, le manque de capitaux, de main d'œuvre qualifiée et d'accès aux soins de santé chez les communautés rurales sont de multiples exemples de la faiblesse du système de santé en Chine¹⁷.

Les difficultés à contrôler les épidémies sont un exemple concret pour décrire une facette de la santé publique chinoise. À la fin de l'année 2002, une maladie s'est déclenchée dans la province de Guangdong sous le nom de Pneumonie atypique ou de syndrome respiratoire aigu sévère (SRAS). Quelques mois plus tard, celle-ci s'est déclarée dans toute la Chine et même dans d'autres métropoles du monde. Plusieurs politologues chinois ont attribué cette épidémie à la faiblesse du système de santé de leur pays.

Les dépenses de santé en Chine sont quasi exclusivement provinciales. En d'autres termes, les dépenses de l'ensemble des gouvernements provinciaux en services de soins de santé sont plus élevées comparativement à celles du gouvernement fédéral. Selon le site du Bureau national de la statistique de Chine (NBSC), en 2006, les dépenses totales en santé et en éducation se chiffraient à 743 milliards de yuans (environ 110 milliards de dollars canadiens)¹⁸. Les parts respectives des gouvernements locaux et provinciaux et du gouvernement central (ou fédéral) étaient de 671 et 72 milliards de yuans chinois¹⁹. Toujours en 2006, sur un total de 48 milliards de yuans de dépenses dans le domaine de la science et de 132 milliards de yuans de dépenses en soins de santé, les dépenses des gouvernements locaux et provinciaux étaient de 17 milliards en science et 130 milliards en soins de santé. Celles du gouvernement central étaient de 31 en science et de 2 milliards en soins de santé. En termes de pourcentage, les dépenses de santé du gouvernement fédéral représentaient environ 1 % et celles des gouvernements locaux et provinciaux étaient de 98 %.

¹⁷ <http://www.oecd.org/dataoecd/51/9/35308753.pdf>

¹⁸ 6.70 par dollar canadien : Taux de change du 31 décembre 2006 selon la Banque du Canada.
<http://www.banqueducanada.ca/fr/taux/exchfo-f.html>

¹⁹ <http://www.stats.gov.cn/tjsj/ndsjsj/2007/indexch.htm>

Chapitre 8 : finances du gouvernement: tableau 8.13 : dépenses en santé du gouvernement local et central

En 1978, les dépenses totales du gouvernement chinois toutes catégories confondues se chiffraient à 112 milliards de yuans. Les parts respectives du gouvernement fédéral et des gouvernements provinciaux et locaux étaient de 53 et 59 milliards de yuans. À cette époque, les dépenses du gouvernement fédéral représentaient 47 % des dépenses totales du pays et celles des gouvernements provinciaux environ 53 %. Au cours des années qui ont suivirent l'écart entre ces deux pourcentages a augmenté. Ainsi, en 2006, pour un total des dépenses publiques du gouvernement en Chine de 4 042 milliards de yuans, la part du gouvernement fédéral était de 999 milliards de yuans et celle des gouvernements locaux et provinciaux était de 3 043 milliards. Les pourcentages étaient de 25 pour le gouvernement central et 75 pour les gouvernements locaux et provinciaux²⁰. Comparativement aux normes occidentales ces montants sont très peu élevés, car pendant ces mêmes périodes, les dépenses en santé publiques en millions de dollars américains à parité de pouvoir d'achat du Canada, des États-Unis et de la France étaient respectivement de 35 885, 269 616 et 66 796 millions pour l'année 1990 et de 54 187, 563 033 et 118 358 millions pour l'année 2000²¹. L'évolution du pourcentage des dépenses de santé effectuées par les gouvernements provinciaux et l'écart de ces dernières par rapport aux dépenses du gouvernement fédéral chinois, permettent de confirmer l'idée selon laquelle les dépenses de soins de santé en Chine proviennent en grande partie des gouvernements des provinces. Entre 1978 et 2006, les dépenses du gouvernement (provincial et fédéral) ont augmenté dans l'ensemble. Cette augmentation dans le système chinois était causée par les investissements majeurs réalisés dans les infrastructures médicales. Par exemple, de 1990 à 2000, il y a eu une hausse de 21 % de lits dans les hôpitaux et les centres de santé et une augmentation de 15 % de professionnels en santé en Chine²². Cependant, les infrastructures médicales aujourd'hui sont devenues désuètes et il y a un manque d'équipements médicaux²³.

En 1950, 3 200 personnes par tranche de 100 000 habitants souffraient de maladies infectieuses telles que la tuberculose. À partir de 1990, il n'y avait que 292 personnes sur

²⁰ <http://www.stats.gov.cn/tjsj/ndsj/2006/indexeh.htm>; Chapitre 8 : finances du gouvernement; tableau 8.11 : dépenses du gouvernement local et central et ratio des dépenses

²¹ http://www.msss.gouv.qc.ca/statistiques/stats_sss/index.php?id=150.133.0,0,1,0&PHPSESSID=f48277d815a8226

²² *The Republic of China Yearbook*. 2004.

²³ *The Republic of China Yearbook*. 2004.

100 000 souffrant de cette maladie. Le pourcentage d'infectés a donc chuté, mais depuis cette année, la baisse s'est stabilisée et dans quelques régions, le taux a même tendance à augmenter. L'incidence des maladies infectieuses démontre une autre faiblesse que connaît le système de santé publique chinois durant ces dernières années.

Selon les autorités chinoises, un million de personnes sont infectés par le virus de l'immunodéficience humaine (VIH) et ce nombre est susceptible d'augmenter de 10 à 20 millions d'ici 2010 si la croissance de personnes atteintes n'est pas freinée. La situation des maladies endémiques est aussi un cas problématique à considérer dans l'avenir de la Chine.

Plusieurs études subdivisent la Chine en deux parties, certains la qualifient d'économie binaire, car il y a des différences notables entre les sociétés urbaines et les sociétés rurales. Il existe en Chine des villes et des cantons séparés par un système bien établis de permis de résidence. La principale source des dépenses en santé dans la république de Chine est le gouvernement provincial et non central. Ainsi, le niveau de bien-être produit par chaque région est déterminé par le niveau de revenu économique de cette dernière. Les dépenses en santé qui découlent du revenu économique réalisé vont différer d'une région à une autre. Par exemple, le niveau le plus élevé de dépenses en santé au cours de l'année 2006 provenait de la province du Guangdong avec 1 035 579 millions de yuans chinois et celui le plus bas, de la province de Ningxia avec 70 085 milles yuans²⁴.

Avant 1949, les villages de la Chine avaient un accès réduit à la médecine moderne. Après la création de la république populaire de Chine (RPC), les gouvernements ont insisté sur la promotion de la santé via le développement économique. Ce qui a nettement amélioré l'état de santé des villageois. Une parfaite illustration est la distribution de lits dans les hôpitaux. En 1965, 40 % de lits étaient distribués dans les villages, dix ans après, le pourcentage passait à 60²⁵.

Le système de santé rurale en Chine recevait jadis moins d'attention par rapport au système de santé urbain. En effet, la plupart des revenus du gouvernement provenaient de l'économie

²⁴ <http://www.stats.gov.cn/tjsj/ndsj/2006/indexeh.htm>; Chapitre 8 : finances du gouvernement: tableau 8.15 : dépenses du gouvernement par région

²⁵ *The Republic of China Yearbook*, 2004.

urbaine. Donc, les dépenses du gouvernement surtout dans les services publics s'adressaient d'abord aux résidents urbains. Le niveau des fonds transférés par le gouvernement dépendait du niveau économique de la province, de la ville et des cantons. Il y a en l'occurrence, une très faible participation du secteur rural. En 1998, les populations urbaines étaient de 379 millions d'habitants et chaque personne recevait à peu près en moyenne 130 yuans du gouvernement en services de santé. Quant aux populations rurales, elles étaient les plus nombreuses avec 866 millions d'habitants et chacun en moyenne recevait environ 11 yuans²⁶.

L'OCDE estime que la Chine doit augmenter ses dépenses dans les domaines de la santé et de l'enseignement et répartir ses fonds publics de manière plus efficace. En effet, les dépenses publiques consacrées à l'enseignement et surtout à la santé sont faibles et inefficaces pour satisfaire les besoins de la Chine en termes de développement. Les autorités locales, chargées du financement des programmes sociaux, de santé et d'enseignement en Chine, ne disposent pas de ressources suffisantes et sont le plus souvent restreintes par leur manque d'autonomie. Ainsi, selon l'OCDE : « Bien que les dépenses publiques officielles dans leur ensemble aient augmenté rapidement en Chine passant de 17,7 % du PIB en 1995 à 27,4 % du PIB en 2003, elles restent nettement en deçà de la moyenne de la zone OCDE, qui s'établit à 44,5 %. Cette accélération de la croissance des dépenses publiques chinoises a été particulièrement forte pour les investissements d'infrastructures et l'administration publique.²⁷ »

Il faut retenir que dans les années passées, la Chine a connu et connaît toujours une croissance économique surprenante, qui lui a permis de rehausser le niveau de vie de sa population, de diminuer le taux de mortalité infantile, d'augmenter l'espérance de vie de sa population, d'améliorer les infrastructures médicales et d'accroître ses dépenses en santé. Tel que mentionné précédemment, les autorités chinoises ont prôné l'amélioration de la santé, surtout dans les zones rurales, via le développement économique. Les questions de santé sont importantes dans la République populaire de Chine et doivent être l'objet de multiples attentions. Les responsables du système de santé en Chine doivent être en mesure de déterminer les ressources nécessaires à la santé de la population et se poser les questions suivantes. Quels sont les déterminants de la santé ? Comment se procurer les ressources

²⁶ *The Republic of China Yearbook*, 2004.

²⁷ http://www.oecd.org/document/13/0,3343,fr_2649_201185_36131966_1_1_1_1.00.html

nécessaires à la santé ? Quelle est l'importance et le taux de rendement de l'investissement en santé ? À quel niveau et où doit-on investir efficacement dans le système de santé ?

CHAPITRE III

DONNÉES

La principale source de collecte de données pour la réalisation de ce travail provient du site Internet du Bureau national de la statistique de Chine²⁸ (*National Bureau of Statistics of China*). Évidemment d'autres sources ont été utilisées en vue de compléter celle mentionnée précédemment. Il s'agit entre autres des fichiers : *China Health Yearbook 2005* et *2006*, *Health Yearbook 2005*, *China Health Data 2003 to 2007* et *General Yearbook_health Part 2001 to 2006*.

Les données sont essentiellement de nature régionale. L'étude se focalise sur 22 provinces, 5 régions autonomes et 4 municipalités, soit 31 régions au total sur une période de treize années c'est-à-dire de 1995 à 2007. L'ensemble des régions étudiées dans l'analyse, leur statut administratif et leur superficie sont présentés dans le tableau 3.1 ci-dessous.

Certaines régions ont dues être écartées de l'étude, entre autres les régions administratives spéciales de Hong-Kong et de Macao, ainsi que la province de Taïwan. L'absence de données pour certaines variables capitales sur plusieurs années est la principale cause de leur élimination. Tel que mentionné précédemment, pour éviter les problèmes liés à l'homogénéité, cette analyse ne tient compte que de données nationales. De ce fait, il n'est pas nécessaire de convertir les unités monétaires et d'uniformiser la notion de santé selon les régions, car elle est la même dans tout le pays. En sommes les données ont été recueillies pour des variables sociodémographiques, économiques, éducationnelles, de ressources de santé et de style de vie. Ces variables seront définies dans les prochaines sections.

Tableau 3.1 : Liste des régions étudiées en Chine

²⁸ <http://www.stats.gov.cn/english/statisticaldata/yearlydata/>

PROVINCE	CHEF-LIEU	SUPERFICIE (Km ²)
Municipalité de Beijing	Beijing	16 808
Municipalité du Tianjin	Tianjin	11 920
Province du Hebei	Shijiazhuang	187 700
Province du Shanxi	Taiyuan	156 800
Région autonome de Mongolie-Intérieure	Hohhot	1 183 000
Province du Liaoning	Shenyang	145 900
Province du Jilin	Changchun	187 400
Province du Heilongjiang	Harbin	460 000
Municipalité de Shanghai	Shanghai	6 340
Province du Jiangsu	Nanjing	102 600
Province du Zhejiang	Hangzhou	101 800
Province du Anhui	Hefei	139 400
Province du Fujian	Fuzhou	121 400
Province du Jiangxi	Nanchang	166 900
Province du Shandong	Jinan	156 700
Province du Henan	Zhengzhou	167 000
Province du Hubei	Wuhan	185 900
Province du Hunan	Changsha	211 800
Province du Guangdong	Guangzhou	177 900
Région autonome du Guangxi	Nanning	236 700
Province du Hainan	Haikou	33 920
Municipalité de Chongqing	Chongqing	82 300
Province du Sichuan	Chengdu	485 000
Province du Guizhou	Guiyang	176 100
Province du Yunnan	Kunming	394 100
Région autonome du Tibet	Lhasa	1 228 400
Province du Shaanxi	Xian	205 800
Province du Gansu	Lanzhou	454 000
Province du Qinghai	Xining	721 000
Région autonome du Ningxia	Yinchuan	66 000
Région autonome du Xinjiang	Urumqi	1 660 000

Sources : <http://www.info-Chine.net/provinces-chine.html>

<http://fr.wikipedia.org/wiki/Special:Recherche?search=&fulltext=Rechercher>

3.1 Indicateurs de la santé

L'objectif est d'estimer une fonction de production de santé pour chaque région en Chine. Cela revient à la relation suivante :

$$S_i = f(X_i),$$

où S (variable dépendante) est la fonction de production de santé qui dépend de plusieurs variables explicatives représentées par le vecteur (X) . Il n'est pas évident d'identifier une mesure par excellence du niveau de santé des individus. L'espérance de vie et le taux de mortalité (TM) sont souvent pris comme indicateurs de santé, mais il faut bien comprendre que ces indicateurs de quantité de vie sont imparfaits et mettent de côté les aspects qualitatifs.

La fonction de production de santé, approximée dans notre mémoire par le taux de mortalité, pour une région spécifique en tenant compte du temps s'écrit :

$$TM_{it} = f(X_{it}), \forall i = 1, \dots, N \text{ et } t = 1, \dots, T,$$

où t et i désignent respectivement l'année et la région. Le vecteur X_{it} indique l'ensemble des variables utilisées au temps t de la région i concernée.

Hitiris et Posnett (1992) utilisent le taux de mortalité dans leur étude en démontrant son impact dans l'explication de l'élasticité du revenu par rapport aux dépenses de santé. Newhouse (1977) utilise les mêmes indicateurs afin de démontrer l'importance du revenu per capita dans les dépenses de santé. Dans ce mémoire, le taux de mortalité des hommes et de femmes est utilisé comme indicateur de santé et donc représente la variable dépendante. Statistique Chine définit le taux de mortalité utilisé dans l'étude comme étant le nombre de décès moyen de la population au cours d'une période de temps exprimé par dix mille habitants. Une fois encore, pour des raisons de disponibilité de données, le taux de mortalité a été choisi en lieu et place de l'espérance de vie. En plus, les causes de mortalité dans un pays sont majeures, les habitudes de vie, l'hygiène, les valeurs sociales, les décisions politiques, la nature, etc., sont autant de facteurs pouvant être cités à cet effet.

À ce stade, la question est de déterminer le vecteur de variables explicatives (les déterminants de la santé). Dans la suite, nous passerons en revue les principales catégories de variables explicatives.

3.2 Variables sociodémographiques

Il n'est pas évident d'aborder des questions économiques d'un pays sans prendre en considération le côté sociale et même culturel qui s'y rattache. Le fonctionnement de la structure sociale étant différente d'une région à l'autre, il est donc très important d'en tenir compte dans ce genre de travail. Par exemple, le fait que la population de Beijing soit plus jeune ou vieille que celle de Tianjin est un facteur très important à savoir dans un projet en santé. Il en est de même pour la densité de la population (ratio du total de la population exprimée en 10 000 habitants et de la superficie en kilomètre carré). Une plus grande concentration de population pour une même superficie peut engendrer des impacts significatifs sur le bien-être et l'état de santé des individus. Par exemple, plus de pollution, plus de fléaux socioéconomiques, moins d'infrastructures médicales etc.

(Gerdtham *et al.*, 1992) ont utilisé des variables sociodémographiques dans la construction de leur modèle. Ainsi, ils ont trouvé que des variables institutionnelles et sociales avaient un effet non négligeable sur l'état de santé. Leu (1986) également utilise des variables de la même catégorie dans l'analyse de ses données. Il a trouvé que le revenu avait un impact sur les dépenses de santé, mais qu'il existait aussi d'autres variables de nature sociodémographique telle que l'urbanisation qui influaient sur l'état de santé.

Les variables sociodémographiques utilisées dans notre analyse sont la densité de la population et le ratio de dépendance des personnes âgées. Roberts (1999) utilise l'âge de la population dans son analyse de données comme variable sociodémographique. Généralement, une population relativement vieille requiert plus de services de soins de santé qu'une population relativement jeune. Par exemple, une population vieille nécessitera l'implantation de plus d'établissements pour personnes âgées, plus de personnels pour l'entretien de ceux-ci, donc plus de salaires à payer pour un gouvernement et plus de cotisations en pension de retraite pour la population active, donc plus de dépenses publiques. Selon Statistique Chine,

le ratio de dépendance des personnes âgées est un indicateur qui reflète l'impact social du vieillissement de la population dans l'économie. Quant à la densité de la population, elle correspond à la population à la fin de l'année par superficie en km².

Cette variable a été utilisée par (Crémieux *et al.*, 1999) au cours de leurs travaux. Berthé (2003) tient compte de l'âge de la population dans son mémoire. Selon lui, la structure d'âge d'un pays est très importante car il existe une tendance dans la pyramide des âges corrélée avec le niveau du développement des pays. La question du développement des pays renvoie à l'importance des variables économiques.

3.3 Variables économiques

Dans *The Republic of China Yearbook*,²⁹ il est écrit que la principale source des dépenses en santé dans les régions chinoises provient des gouvernements provinciaux. Autrement dit, les dépenses en santé dépendent du revenu réalisé dans chaque région. Lorsque les revenus des gouvernements sont élevés, les dépenses en santé sont aussi élevées. Au Québec, selon le ministère de la santé, en pourcentage des dépenses de programme, les dépenses totales effectuées dans le domaine de la santé sont passées de 31 % en 1995-1996 à 38 % en 2006-2007³⁰. L'état de santé des Québécois s'est nettement amélioré durant cette période, le taux de mortalité infantile a baissé et le niveau d'espérance de vie des femmes et des hommes a augmenté. Pour 1 000 naissances dans la région de Québec, on a dénombré un taux de mortalité infantile de 4,5, alors qu'au Niger, on compte un taux de 123 mortalités infantiles pour 1 000 naissances.

Newhouse (1977) utilise des variables économiques dans son étude. En effet, il a trouvé que les dépenses de santé variaient en fonction du PIB par habitant, faisant donc de la santé un bien de luxe. Le revenu à lui seul expliquait plus de 90 % des variations dans les dépenses en santé.

²⁹ *The Republic of China Yearbook*, 2004.

³⁰ http://www.msss.gouv.qc.ca/statistiques/stats_sss/index.php?id=135,96,0,0,1

Les pays les plus riches au monde affichent des niveaux de santé les plus satisfaisants. Les États-Unis peuvent être cités comme exemple. En effet, c'est le pays qui dépense le plus en santé et le niveau de santé de sa population est nettement supérieur à celui des pays en voie de développement. Ainsi, plus il y a de revenu, plus il y a de services de soins de santé accessibles à l'ensemble de la population et plus l'état de santé s'améliore. Les variables économiques sont par conséquent importantes pour expliquer l'état de santé. Tel que mentionnée en début d'analyse, pour améliorer l'état de santé, les gouvernements doivent investir dans le domaine et s'assurer que cet investissement soit efficace. Pour cela, des variables tels que le revenu par habitant ou les dépenses du gouvernement ont une importance capitale dans la prise des décisions.

La variable économique prise en compte dans ce mémoire est le produit intérieur brut (PIB) par habitant corrigé de l'inflation. Hitiris et Posnett (1992) ont utilisé le revenu par habitant et Roberts (1999) a utilisé les dépenses publiques comme variables explicatives dans leurs analyses de données. Le PIB par habitant corrigé de l'inflation a été divisé par l'indice de prix du produit intérieur brut proposé par le Bureau national de la statistique de Chine (NBSC)³¹.

3.4 Variables éducationnelles

L'éducation est une base fondamentale à l'aboutissement de tout projet. Par exemple, sensibiliser les communautés sur les bienfaits de la médecine moderne, nécessite d'abord l'éducation des populations en vue de faciliter la compréhension. Pourquoi est-il plus important pour une femme d'accoucher dans un hôpital que chez le sorcier du village ? Comment savoir lire et appliquer la médication recommandée par un médecin ? Un individu analphabète aura du mal à comprendre parfaitement le système de santé et ses fonctionnements. Ce sont tant de notions importantes faisant appel à une base éducationnelle.

³¹ National Bureau of Statistics of China : <http://www.stats.gov.cn/enGLISH/>

(Crémieux *et al.*, 1999) utilisent des données sur l'éducation, entre autres, le niveau d'éducation et le nombre de diplômés per capita dans le cadre de leurs travaux sur les dépenses en soins de santé dans les provinces canadiennes.

Plusieurs variables liées à l'éducation seront prises en compte dans cette étude pour toutes les raisons qui précèdent. Ce sont le pourcentage de la population analphabète, le pourcentage de la population non scolarisée (ratio du total de la population non-scolarisée sur le total de la population), le pourcentage de la population éduquée (ratio de la population âgée de 15 ans et plus ayant un niveau d'éducation (primaire, secondaire, universitaire etc.) divisée par le total de la population) et le pourcentage des femmes analphabètes. Le choix s'est porté sur les femmes parce que dans la majorité des pays, l'éducation des enfants est associée directement à la mère. Dans les pays en voie de développement, tandis que les hommes vaquent à leurs occupations professionnelles, les femmes s'affèrent à l'éducation des plus petits. Même dans les pays riches, la profession d'éducatrice en garderie est occupée en grande majorité par des femmes. Selon une étude publiée par le développement des ressources humaines Canada³², le milieu professionnel de garderie est occupé par près des deux tiers de la population féminine.

Sur le site de l'UNESCO, la définition pour le taux d'analphabétisme et la population analphabète correspond à « la proportion et le nombre de personnes dans la population qui sont incapables de lire et d'écrire un exposé simple et bref, de faits en rapport avec leur vie quotidienne. » Pour toutes les raisons qui précèdent, l'éducation est par conséquent une variable clé prise en compte dans l'analyse des données de ce mémoire. Comme toutes ces variables sont par définition imparfaites, nous avons opté pour une série d'estimation où chacune de ces variables sera utilisée en solo afin de tester la robustesse des résultats.

3.5 Variables de ressources de santé

La variable de ressources de santé utilisée dans cette analyse est le nombre total d'institutions de santé par 10 000 habitants. Le Bureau national de la statistique de Chine décrit le total des institutions de santé comme étant l'ensemble des hôpitaux, des cliniques externes, des centres

³² <http://dsp-psd.pwgsc.gc.ca/Collection/RH4-12-2001F.pdf>

de santé locaux, des centres de santé communautaires, des centres de prévention et de contrôle des maladies, des organisations médicales et des centres de maternités et de la petite enfance. (Crémieux *et al.*, 2004) utilisent le même type de variables à savoir le nombre de médecins et le nombre de produits pharmaceutiques, pour estimer le lien entre les dépenses de santé et l'état de santé aux États-Unis.

Il est très important pour les gouvernements de connaître le domaine dans lequel il serait plus efficace d'investir. Faut-il investir plus dans le nombre de médecins, ou dans le nombre d'institutions de santé ? Faut-il favoriser plus la sensibilisation sur l'hygiène des populations en faisant plus des campagnes publicitaires en vue d'améliorer l'état de santé ? La notion de médecine est liée directement à l'état de santé des individus, d'où l'importance d'en tenir compte au cours de cette recherche via le nombre d'hôpitaux, de centres de contrôle et de prévention des maladies, de cliniques externes etc.

3.6 Variables de style de vie

Plusieurs rapports de l'organisation mondiale de la santé interpellent les pays sur les conditions d'hygiène et des habitudes de vie. En effet, la majorité des pays pauvres ne disposent pas de systèmes sanitaires adéquats et de systèmes d'épuration des eaux performants. Des études révèlent que les conditions dans lesquelles les femmes au Niger accouchent sont affreuses, de sorte que le taux de mortalité infantile est très élevé dans le pays. Le lien entre l'hygiène, les habitudes de vie et la santé est bien connu, ce qui fait de cette variable un facteur important à tenir compte dans le modèle. (Crémieux *et al.*, 1999) ont marqué une différence en ajoutant dans leurs travaux ces variables liées aux habitudes de vie. La variable de style de vie prise en considération dans l'analyse est le volume d'eau potable offert à la population par 10 000 habitants.

3.7 Statistiques descriptives

Le tableau ci-dessous présente les statistiques descriptives de l'ensemble des variables utilisées dans le modèle et les modèles connexes. Une liste de définition des différentes

appellations des variables du modèle est présentée pour faciliter la compréhension de ces dernières.

Le tableau 3.2 est une description des variables dont l'appellation a été modifiée pour faciliter la manipulation dans le logiciel *Stata* (version 10) et le tableau 3.3 présente les résultats de la statistique descriptive des variables utilisées dans le mémoire.

Tableau 3.2 : Définitions des variables

Liste de définition des variables	
Variabes	Définitions
<i>tx_mort</i>	Taux de mortalité de la population par 10 000 habitants (hommes et femmes)
<i>pour_anphl</i>	Pourcentage de la population analphabète
<i>pour_fem_anphl</i>	Pourcentage des femmes analphabètes
<i>pour_nscol</i>	Pourcentage de la population non-scolarisée
<i>pour_pop_educ</i>	Pourcentage de la population éduquée
<i>pib_corr_infl</i>	Produit intérieur brut par habitant corrigé de l'inflation
<i>rt_dep_vieu</i>	Ratio de dépendance des personnes âgées
<i>densite</i>	Densité de la population par 10 000 habitants
<i>tot_inst_mil</i>	Total des centres de santé par 10 000 habitants
<i>vol_eau_10mil</i>	Volume d'eau potable offert par 10 000 habitants

Tableau 3.3 : Statistiques descriptives

Variabes	Moyenne	Écart-type	Minimum	Maximum	Observations
<i>tx_mort</i>	1,82	0,11	1,50	2,17	401
<i>pour_anphl</i>	2,45	0,58	1,21	4,19	340
<i>pour_fem_anphl</i>	2,82	0,55	1,61	4,31	340
<i>pour_nscol</i>	0,79	0,056	0,65	0,95	340
<i>pour_pop_educ</i>	3,27	0,20	1,82	3,52	309
<i>pib_corr_infl</i>	3,14	1,12	-0,74	5,60	401
<i>rt_dep_vieu</i>	2,38	0,23	1,66	3,09	339
<i>densite</i>	-4,0057	1,47	-8,54	-1,23	401
<i>tot_inst_mil</i>	1,12	0,082	0,93	1,64	395
<i>vol_eau_10mil</i>	1,45	0,10	1,28	1,74	371

La statistique descriptive ci-dessus décrit l'ensemble des variables utilisées dans l'analyse. Elle montre la moyenne, l'écart-type, le minimum et le maximum et le nombre d'observations de chaque variable prise en compte dans le mémoire. On retrouve dans le tableau plusieurs variables ajoutées dans l'analyse afin de mesurer le biais d'omission des variables et la robustesse des résultats. Ce sujet sera discuté plus amplement dans la section réservée à l'interprétation des résultats. Le tableau de définition est la liste des variables utilisées dans l'étude et leurs descriptions. Par exemple, tx_mort représente le taux de mortalité des hommes et des femmes par 10 000 habitants. Le pourcentage de la population (hommes et femmes) analphabète est représenté par l'expression pour_anphl. L'analyse pourrait continuer ainsi pour les autres variables mentionnées dans les tableaux. La section qui suit développe sur la corrélation entre le taux de mortalité et les variables explicatives.

3.8 Corrélation entre la variable dépendante et les variables explicatives

En première analyse, nous avons calculé le tableau des corrélations entre le taux de mortalité et les variables explicatives retenues. Tel qu'attendu, le taux de mortalité est corrélé positivement avec le pourcentage de la population analphabète, le pourcentage de personnes non-scolarisées, le pourcentage de femmes analphabètes, la densité de la population et le ratio de dépendance des personnes âgées. Une augmentation de chacune de ces variables entraîne une augmentation du taux de mortalité. À l'inverse, le pourcentage de la population éduquée, le produit intérieur brut par habitant corrigé de l'inflation, le total des institutions de santé par 10 000 habitants et le volume d'eau potable offert à la population par 10 000 habitants sont corrélés négativement avec le taux de mortalité. Une augmentation de chacune de ces variables entraîne une diminution du taux de mortalité. Le tableau 3.4 ci-dessous présente les résultats de la corrélation entre la variable dépendante représentée par le taux de mortalité et les variables explicatives.

Tableau 3.4 : Tableau de corrélation

	tx_mort
<i>pour_anphl</i>	0,45
<i>pour_fem_anphl</i>	0,46
<i>pour_nscol</i>	0,49
<i>pour_pop_educ</i>	-0,26
<i>pib_corr_infl</i>	-0,085
<i>densite</i>	0,0012
<i>rt_dep_vieu</i>	0,14
<i>tot_inst_mil</i>	-0,15
<i>vol_eau_off_10mil</i>	-0,41

CHAPITRE IV

ESTIMATIONS

L'un des objectifs de ce travail est d'estimer la frontière de production de santé via un modèle de frontières stochastiques. Le modèle choisi pour cet exercice a été proposé par Greene (2004b). De ce fait, ce mémoire se classerait dans la quatrième génération telle que vu dans le premier chapitre. Avant la présentation du modèle, analysons pourquoi le choix s'est porté sur le modèle de frontières stochastiques et non sur les autres modèles économétriques, entre autres les moindres carrés ordinaires (MCO) et les moindres carrés avec des données de panel. Les différentes estimations ont été réalisées à partir du logiciel *Stata* (version 10).

4.1 Moindres carrés ordinaires (MCO)

Le modèle de régression le plus élémentaire est celui des moindres carrés ordinaires. Nous présentons brièvement ce modèle qui nous servira de référence pour développer les autres méthodes proposées pour estimer une fonction de production. Dans le cas de la fonction de production présentée plus tôt, la méthode des MCO se présente comme suit (nous faisons apparaître la constante pour des raisons de simplification de présentation des méthodes subséquentes) :

$$S_i = \beta_0 + f(X_i, \beta_1) + \mu_i \quad \forall i = 1, \dots, N.$$

S_i , l'état de santé, est la variable dépendante, X_i et $\beta \equiv (\beta_0, \beta_1)$ sont respectivement le vecteur des déterminants de l'état de santé et le vecteur des paramètres à estimer et μ_i est le terme d'erreur. Le terme d'erreur suit une distribution normale de moyenne nulle et sa variance est la même pour toutes les observations. Il n'existe pas de relation entre le terme d'erreur et les variables explicatives:

$$\mu \sim N\left(0, \sigma_\mu^2\right), E(\mu_i \mu_j) = 0 \text{ et } E(\mu_i | x_i) = 0.$$

La méthode des moindres carrés ordinaires revient à minimiser la somme des carrés des termes d'erreurs par choix des paramètres. Formellement, cela revient à :

$$\min_{\beta} \sum_{i=1}^N [S_i - \beta_0 - f(X_i, \beta_1)].$$

La solution à ce problème sera un vecteur $\hat{\beta}$ sans biais et à variance minimale.

La nature des données utilisées dans ce mémoire ne permet pas le choix de la méthode des moindres carrés ordinaires. En effet, les données sont en petit nombre pour une trentaine de régions et quelques années. Le recours à des méthodes de panel est incontournable

4.2 Méthodes d'estimation des données de panel

Les données de panel peuvent être définies comme des données observées sur un ensemble d'individus et de périodes. Autrement dit, c'est un ensemble de données compris dans un échantillon observé sur plusieurs années. Par exemple, des données sur l'état de santé des pays de l'OCDE de 1990 à 2007. Une façon simpliste d'utiliser les données de panel est d'estimer le modèle de régression suivant :

$$S_{it} = \beta_0 + f(X_{it}, \beta_1) + \mu_{it},$$

où i et t représentent respectivement la région et la période d'observation de la variable dépendante et des variables explicatives.

L'utilisation des données de panel est avantageuse dans la mesure où il y a plus d'observations à exploiter. Contrairement au modèle de régression simple à une seule variable, les données de panel permettent de tenir compte de l'hétérogénéité des observations et de saisir les effets à court et à long terme de l'échantillon.

Plusieurs méthodes d'estimation peuvent être utilisées avec des données de panel. On pourrait comme ci-haut utiliser l'estimation par la méthode des moindres carrés ordinaires en mettant en commun (*pooling*) les variables sans égard à leur provenance. Cependant, cela reviendrait à standardiser les comportements des agents, des pays et des régions dans l'échantillon, de les rendre homogènes sans tenir compte de l'hétérogénéité de la nature des

données. En d'autres termes, cela supposerait que les paramètres estimés sont les mêmes pour toutes les régions et les années. Une autre possibilité serait d'estimer les équations de façon individuelle par la méthode des moindres carrés ordinaires. Une fois encore, cette méthode fait défaut en raison du trop grand nombre de paramètres à estimer étant donné le petit nombre d'observations par région. Pour tenir compte de l'hétérogénéité, la plupart des économètres ont recouru à des modèles d'estimation avec effets fixes et avec effets aléatoires qui posent l'hypothèse d'exogénéité stricte des variables par rapport aux termes d'erreurs ou résidus.

4.2.1 Méthode de panel avec effets fixes

Le modèle d'estimation avec effets fixes ou modèle de la covariance, consiste à modéliser l'hétérogénéité à l'aide d'un effet individuel. On ajoute au modèle précédent, une variable fictive qui ne change pas d'une période à une autre. Dans ce cas, le modèle devient :

$$S_{it} = \beta_0 + f(X_{it}, \beta_1) + z_i + \mu_{it}.$$

z_i spécifie des effets constants, non aléatoires qui modifient simplement la valeur de la constante β_0 dans l'équation précédente selon la valeur de i . L'élément fixe peut être corrélé avec les variables explicatives. Une autre façon de présenter le modèle avec effets fixes permet de mieux comprendre les phénomènes en action. On peut réécrire l'équation précédente comme suit :

$$\begin{aligned} S_{it} &= (\beta_0 + z_i) + f(X_{it}, \beta_1) + \mu_{it} \\ &= \beta_{0i} + f(X_{it}, \beta_1) + \mu_{it} \end{aligned}$$

Ainsi, on obtient un modèle où la constante est propre à chaque région. À partir de cette fonction de production, on peut maintenant définir la matrice de covariance de diverses façons et obtenir divers estimateurs en découlant. Un cas extrême consiste à supposer que toutes les erreurs proviennent d'une même distribution. Dans ce cas, il suffit d'estimer avec la méthode des MCO en ajoutant une variable binaire par région. D'autres possibilités sont bien entendu envisageables, comme supposer que les distributions sont propres à chaque région et que seul le paramètre $\hat{\beta}_1$ est commun entre les régions. Le désir d'atténuer les

hypothèses sur la distribution est contraint par le nombre de degrés de liberté parfois trop faible quand le nombre de régions et/ou de périodes est petit.

4.2.2 Méthode de panel avec effets aléatoires

Dans le cas du modèle avec effets aléatoires ou modèle à erreurs composées, l'effet n'est plus un élément fixe, mais plutôt un élément aléatoire, non-observé. L'hypothèse d'homogénéité totale est donc rejetée. L'équation s'écrit :

$$S_{it} = \beta_0 + f(X_{it}, \beta_1) + \mu_{it} + w_{it} .$$

Cette fois, l'élément aléatoire est représenté par w_{it} qui n'est plus un effet constant et n'est pas corrélé avec les variables explicatives. Le modèle suppose que w_{it} et μ_{it} sont véritablement aléatoires. Les hypothèses de base dans ce modèle supposent que w_{it} et μ_{it} sont centrés, c'est-à-dire, sont d'espérance nulle, homoscédastiques et d'écart-type égale à σ_w et σ_{μ} . Ils sont indépendants et non corrélés l'un avec l'autre.

Lorsque la variance de w et μ est différente pour chaque observation et lorsqu'il y a une corrélation non nulle entre les deux termes, on peut estimer ce modèle par la méthode des moindres carrés généralisées (MGC). Une autre façon d'estimer ce modèle est la méthode des variables instrumentales qui consiste à introduire une variable corrélée avec les variables explicatives mais pas avec le terme d'erreur. Il faut mentionner que la dernière méthode a l'inconvénient d'être difficile à réaliser en raison du choix de la variable instrumentale qui n'est pas facile à identifier.

4.2.3 Méthode de panel et racines unitaires

Tel que mentionné en début de chapitre, les études qui ont suivi celles de la deuxième génération se sont majoritairement intéressées à la nature stochastique des données de panel. Ainsi, un certain nombre d'études ont testé la présence de racines unitaires dans les panels (voir Roberts, 1999). L'une des façons de procéder est le processus autorégressif (AR). Greene (1993) définit le processus autorégressif d'ordre 1, comme étant tout processus stationnaire X_t qui vérifie la relation suivante :

$$X_{it} = \varphi X_{it-1} + \mu_{it},$$

où μ_{it} représente les résidus et est par hypothèse un bruit blanc³³. Si ce processus est remplacé dans l'expression de l'écriture matricielle de l'équation ci-dessus, on a :

$$\begin{aligned} S_{it} - \varphi S_{it-1} &= \beta_{0i} + f(X_{it}, \beta_1) - \varphi(\beta_{0i} + f(X_{it-1}, \beta_1)) + \mu_{it} \\ &= \beta_{0i}(1 - \varphi) + f(X_{it}, \beta_1) - \varphi f(X_{it-1}, \beta_1) + \mu_{it} \end{aligned}$$

Lorsque $\varphi = 1$, alors il y a présence de racine unitaire et le processus est dit non stationnaire. Dans ce cas, il faut le rendre stationnaire en faisant des régressions de première différence. Le test de la présence des racines unitaires implique le test de $\varphi = 1$. Bien entendu, on peut adapter ce modèle au cas des méthodes de panel avec ou sans effets fixes ou aléatoires (voir Im *et al.*, 2003).

Le choix d'une méthode par rapport à une autre peut dépendre de plusieurs considérations comme la nature de l'échantillon, le nombre d'unités statistiques, etc. Le plus souvent, le modèle avec effets aléatoires sera plus prisé car considéré comme plus efficient surtout si la corrélation entre les variables explicatives et l'élément ajouté (z_i) est nulle. Le modèle avec effets fixes et celui avec effets aléatoires permettent de tenir compte de l'hétérogénéité des données, toutefois, les hypothèses sur la nature spécifique des effets ne sont pas les mêmes d'un modèle à un autre. Dans un premier, les effets peuvent être corrélés avec les variables explicatives et dans un autre, ils ne le sont pas. À cet effet, le test de Hausman³⁴ permet de tester laquelle de ces deux hypothèses est appropriée et par conséquent, permet de choisir laquelle des méthodes serait la plus adéquate selon la nature des données.

Par-dessus l'analyse qui précède, il ne faut pas perdre le but principal de ce mémoire qui est d'expliquer les déterminants de l'état de santé dans les régions chinoises via le modèle de frontières stochastiques. La méthode des moindres carrés ordinaires en panel avec effets fixes et aléatoires permet de tenir compte de l'hétérogénéité des données. Toutefois, elle peut s'avérer compliquée dans le cas où il y a non-disponibilité de données soit pour l'ensemble

³³ Dans un processus d'ordre k , μ est un bruit blanc si : $E(\mu) = 0$ et $E(\mu_t, \mu_{t-k}) = \sigma^2$ lorsque $k=0$ et

$E(\mu_t, \mu_{t-k}) = 0$ lorsque $k \neq 0$ (Greene, 1993).

³⁴ Notes de cours : Analyse statistiques des données de panels: Centre interuniversitaire québécois de statistiques sociales (CIQSS) (2008)

des agents, soit pour l'ensemble des années ou pour les deux ensembles. Si la notion de fonction de production est appliquée dans ces modèles, ceux-ci partent de l'hypothèse qu'il y a efficacité de la fonction de production. Mais qu'en est-il lorsqu'il y a inefficacité de la fonction de production ?

4.3 Modèle de frontières stochastiques avec des données de panel

La santé de chaque agent (ici les régions) va dépendre de la performance avec laquelle ce dernier arrive à combiner les facteurs déterminant la santé. Cependant, ce processus peut dépendre de la capacité des décideurs publics à utiliser de manière efficace les ressources à leur disposition. La performance à produire la santé peut être définie comme la plus ou moins grande proximité du maximum atteignable tel que définie par la frontière de production. On en arrive à définir la performance comme une distance entre la situation réelle ou observée et la situation optimale telle que décrite par la frontière de production. Cette distance est notée v_i et on obtient ainsi (Greene, 2004b) :

$$S_{it} = \beta_0 + f(X_{it}, \beta_1) + \mu_{it} - v_{it}.$$

Cet ensemble peut aussi être appelé l'erreur composée, car il comprend deux facteurs. D'une part, il y a μ_{it} qui est une perturbation aléatoire résultant d'événements favorables ou défavorables externes, tels que le climat, la nature ou même la performance des équipements, au cas où le producteur de santé en utilise dans son processus de fabrication. Dans la littérature économétrique, c'est un terme avec une distribution de type symétrique. D'autre part, il y a v_{it} , le terme d'inefficacité qui peut être interprété comme la déviation de la performance observée par rapport à la performance optimale sur la frontière S_i . C'est un terme avec une distribution strictement non négative.

μ_{it} est distribué normalement et indépendamment de v_{it} . L'inefficacité v_{it} est supérieure ou égale à zéro et sa distribution est donnée par la borne comprise entre zéro et la valeur maximale de la fonction de production. Étant donné que l'output ne peut être négatif, alors on observera une production comprise entre zéro et la valeur de la fonction de production sur la frontière. Plusieurs hypothèses de distribution peuvent être appliquées dans le modèle de

frontières stochastiques. Dans son analyse, Stevenson (1980) a utilisé une distribution conjointe de type normale-exponentielle où u est distribué normalement et v de façon exponentielle, tandis que Greene (1990) a appliqué une autre distribution conjointe de type normale-gamma. D'autres distributions telles que la normale tronquée, la log-normale, la chi carrée, etc. peuvent aussi être appliquées pour représenter l'inefficience. Pour des petites déviations, Greene (2004b) suggère de spécifier le modèle en logarithme.

Chaque déviation est le résultat de facteurs échappant au contrôle du producteur de santé. Cela peut être sa volonté de travail ou son effort. Dans un modèle d'offre de la firme, on parlera de la volonté et de l'effort de l'employeur et de ses employés dans l'accomplissement de leur tâche. La frontière elle-même peut varier selon le type de producteur de santé. Elle peut aussi varier à travers le temps pour un même producteur de santé. Kim et Schmidt (2000) suggèrent une semi-paramétrisation de l'inefficience en ajoutant des effets fixes.

Dans le modèle de panel avec effets fixes, un élément individuel et constant est ajouté afin de modéliser l'hétérogénéité des données. Dans celui avec effets aléatoires, c'est un élément aléatoire qui est introduit dans le modèle pour tenir compte de l'hétérogénéité. Ces deux modèles supposent qu'il y a efficience de la frontière de production. Autrement dit, toutes les observations se retrouvent sur la frontière de production et sont donc efficaces. Les méthodes de panel avec effets fixes et avec effets aléatoires permettent de tenir compte de l'hétérogénéité, dépendamment des hypothèses formulées, tout comme le modèle de frontière et de mesurer la nature stochastique des données. Cependant, elles ne permettent pas de mesurer l'inefficience car celle-ci est écartée par l'hypothèse d'efficience de la fonction de production. À l'opposé, le modèle de frontières stochastiques ne suppose pas que toutes les observations sont efficaces, donc ne se retrouvent pas sur la frontière de production. Il permet non seulement de tenir compte de l'hétérogénéité des données, mais de mesurer la distance entre les observations inefficaces et celles efficaces, c'est-à-dire, mesurer l'inefficience des observations. Les modèles de panel avec effets fixes et aléatoires peuvent être un outil complémentaire dans le modèle de frontières stochastiques. En effet, plusieurs hypothèses peuvent être relâchées et il devient possible d'adapter le modèle de frontières stochastiques à des effets fixes et aléatoires en panel.

4.3.1 Modèle de frontières stochastiques : données de panel

Greene (2004b) stipule que deux restrictions sont formulées dans le modèle de frontière. Il n'y a pas d'hétérogénéité dans le temps et l'inefficience elle-même ne change pas dans le temps. Tout comme dans les MCO en panel avec effets fixes, un terme constant (élément fixe z_i) peut être ajouté dans le modèle de frontières stochastiques qui devient un modèle de frontières stochastiques avec effets fixes. Il ne varie pas dans le temps et est corrélé avec les variables explicatives du modèle. Celui-ci est estimé par maximum de vraisemblance :

$$S_{it} = \alpha + X_{it}'\beta + z_i + \mu_{it} - v_{it}.$$

Selon Greene, ce modèle contient des limites. En effet, l'hypothèse selon laquelle l'inefficience ne varie pas dans le temps peut être raisonnable seulement pour un court laps de temps. Le fait d'introduire une variable fictive constante dans toutes les observations, exclut la nature hétérogène des données. Par conséquent, Greene propose d'introduire l'hétérogénéité des données à travers la distribution de l'inefficience, ce que nous verrons dans la section suivante.

La nature régionale des données appliquées dans notre analyse nous permet d'éviter une fois encore le problème d'hétérogénéité des données. Cependant, il y a un biais du paramètre qui augmente dans ce genre de modèle non linéaire avec effets fixes lorsque le nombre de périodes est relativement petit. À ce sujet, l'auteur propose l'estimation du modèle de frontières stochastiques avec effets aléatoires.

L'utilisation d'effets aléatoires ne modifie pas l'écriture du modèle, mais seulement le traitement statistique du terme idiosyncratique. Comme cela a déjà été vu précédemment, il est inutile de répéter cette discussion dans ce paragraphe. Notons cependant que dans le modèle d'estimation de frontières stochastiques avec effets aléatoires, la variable fictive constante ne change pas dans le temps et n'est pas corrélée avec les autres variables incluses dans le modèle. L'inefficience composée varie librement à travers le temps. Greene (2004b) stipule que l'hétérogénéité est introduite dans ce modèle de deux façons. D'une part, en supposant que l'hétérogénéité affecte la frontière et, d'autre part, en l'introduisant à travers la distribution du terme d'inefficience. Par conséquent, la variation de l'inefficience est obtenue

en modifiant les hypothèses de restriction sur l'erreur idiosyncratique et permet de tenir compte de la variation du temps. Le modèle devient :

$$S_{it} = \alpha + X'_{it} \beta + w_i + \mu_{it} - v_{it}.$$

4.3.2 Modèle de frontières stochastiques: modèle invariant et modèle variable dans le temps

Le modèle de frontières stochastiques en panel avec effets aléatoires pourrait être retenu pour plusieurs raisons. D'abord la définition de frontières stochastiques permet de supposer qu'il existe effectivement une fonction de production avec certaines observations efficaces. Mais qu'il existe aussi d'autres observations qui ne sont pas efficaces et donc qu'il y a inefficience de la production. Par conséquent, il faut distinguer les observations efficaces de celles qui ne le sont pas. Les effets aléatoires dans le modèle de frontières stochastiques avec des données de panel permettent de tenir compte de l'effet du changement du temps capté par la constante qui y est introduite. De ce fait, on parle de l'estimation du modèle de frontières stochastiques en panel avec effets aléatoires. La nature en panel du modèle d'estimation dépend de la nature des données du mémoire, qui sont des données observées pour plusieurs observations (les régions) et sur plusieurs périodes (de 1995 à 2007). La constante peut être par exemple les politiques de système de santé en vigueur sur tout le territoire chinois et le sexe des populations (homme ou femme) qui a la même définition dans toutes les régions en Chine.

La distinction entre le modèle avec effets fixes et celui avec effets aléatoires dépend du traitement appliqué aux variables fictives z et w qui y sont ajoutées. Greene (2004a) affirme que l'estimation avec les effets fixes admet une distribution libre qui nécessite seulement des spécifications sur la moyenne conditionnelle. Cela permet de généraliser toutes les observations et de masquer l'identité de l'inefficience estimée. Les effets aléatoires permettent un paramétrage individuel et précis sur le terme d'inefficience. Toutefois, ces deux modèles spécifient deux défauts qui sont l'invariabilité et l'absence d'hétérogénéité dans le temps. Ces défauts sont pris en considération dans le modèle de frontières stochastiques.

La différence entre le modèle invariant et celui qui autorise l'inefficience à varier dans le temps dépend de la façon de traiter le terme d'inefficience. Dans le modèle où l'inefficience est une constante, on a :

$$v_{it} = v_i .$$

Dans le modèle qui autorise l'inefficience à varier dans le temps, Greene (2004b) propose le modèle de Battese et Coelli (1995) dans lequel le terme d'inefficience est modélisé comme une variable aléatoire ayant une distribution de type normale-tronquée, multiplié par un terme spécifique fonction du temps. On obtient ainsi (Greene, 2004b) :

$$v_{it} = \exp \left\{ -\eta (t - T_i) \right\} v_i ,$$

où η représente un terme spécifique fonction du temps qui multiplie l'inefficience v_i .

Dans le modèle de frontières stochastiques, lorsque le terme spécifique fonction du temps qui multiplie l'inefficience η (modèle de frontières stochastiques variable dans le temps) est non significatif dans les résultats de l'estimation, il suffit de considérer seulement le modèle de frontières stochastiques invariant. Si le terme v_i est non significatif, alors le modèle de panel sans terme d'inefficience sera de mise.

Dans les estimations de l'analyse, le taux de mortalité (*tx_mort*) détermine la variable dépendante et les variables explicatives sont caractérisées par le taux de la population éduquée par 10 000 habitants, la densité de la population, le ratio de dépendance des personnes âgées, le produit intérieur brut par habitant corrigé de l'inflation, le total des institutions de santé par 10 000 habitants et le volume d'eau potable offert à la population par tranche de 10 000 habitants. D'autres variables sont ajoutées facultativement dans le modèle afin de mesurer leurs effets quelconques sur le taux de mortalité. Ce sont le taux de la population non scolarisée, le pourcentage de femmes analphabète et le pourcentage de la population analphabète. Ceci est dans le but de renforcer la validité des résultats obtenus. Au final, nous estimerons la relation suivante :

$$tx_mort = f(pour_pop_educ, pib_corr_infl, densite, rt_dep_vieu, tot_inst_mil, vol_eau_off_10mil).$$

Généralement, le modèle de frontières stochastiques est spécifié sous forme de logarithme (Greene, 2004a et 2004b). Aussi, pour obtenir des résultats à partir du modèle de frontières stochastiques dans *Stata*, les données doivent être définies en forme logarithmique. Ainsi l'output représenté par le taux de mortalité (tx_mort), correspond à $\ln(tx_mort)$ et chaque variable explicative correspond à $\ln(x_{it})$. Nous estimons par conséquent un modèle log-linéaire.

4.3.3 Modèle de frontières stochastiques avec variable d'output « négative »

Le taux de mortalité a été choisi dans notre analyse comme étant une mesure d'output. Cependant la véritable mesure d'output est le taux de survie obtenue comme suit :

$$\text{taux de survie} = 1 - \text{taux de mortalité} .$$

Ces deux mesures sont équivalentes théoriquement, mais en termes statistiques, il y a deux différences lorsqu'elles sont utilisées. La première différence est qu'à l'exception de la constante, les coefficients ne changent pas sauf pour le signe selon qu'on utilise le taux de mortalité ou le taux de survie. La deuxième différence est que lorsqu'on utilise le taux de mortalité plutôt que le taux de survie comme mesure d'output, comme c'est le cas dans notre analyse, il y a changement dans le signe du terme d'erreur unilatère et cela devient fondamental lorsque l'on utilise *Stata*. Dans un tel contexte, il faut soit estimer le modèle comme étant une fonction de coût et cela revient à activer simplement dans le logiciel *Stata* la commande *cost* à la fin de la régression (commande *xtfrontier*) ou il faut redéfinir la variable d'output comme étant le taux de survie (1-taux de mortalité). Dans notre analyse, nous avons estimé le modèle de frontières stochastiques décrit précédemment comme étant une fonction de coût.

CHAPITRE V

RÉSULTATS

L'interprétation des résultats se divise en deux parties. Les résultats méthodologiques discutent de l'importance du choix de la méthode d'estimation et de la définition des variables. Les résultats sur les variables discutent des faits saillants qui ressortent des coefficients du modèle, interprètent les effets marginaux et l'efficacité relative des inputs du modèle.

5.1 Résultats méthodologiques

Plusieurs tableaux sont présentés afin d'exposer les résultats des différentes estimations. Ces résultats sont fonction de la méthode économétrique utilisée et de la définition. Chacun de ces facteurs est analysé dans les sections ci-dessous.

5.1.1 Importance de la méthode

Le tableau 5.1 présente le modèle retenu pour l'étude des données analysées avec les méthodes d'estimations vues précédemment. Ces méthodes font ressortir des résultats importants. Le premier résultat est que les coefficients estimés sont sensibles à la méthode utilisée. Dans les moindres carrés ordinaires (*pooling*), la variable sociodémographique (ratio de dépendance des personnes âgées) est significative à 99 % alors qu'elle ne l'est plus quand on introduit des effets fixes. À l'inverse, le coefficient de la variable représentant le taux de densité de la population n'est pas significatif dans la méthode de moindres carrés ordinaires (*pooling*) et redevient significatif dans celle avec effets fixes. De plus, le signe du coefficient de la variable n'est pas le même d'une méthode à une autre. Les coefficients sont positifs dans une méthode et négatifs dans une autre. L'ensemble des variables explique

faiblement la variable dépendante avec un R^2 de 0,42 dans la méthode de MCO (*pooling*), de 0,33 dans celle avec effets fixes et de 0,31 dans la méthode avec effets aléatoires.

Il est possible de tester si les modèles de frontières stochastiques contribuent à l'explication des phénomènes. Le coefficient ν_i , le coefficient constant de l'inefficience (modèle invariant) est significatif, ce qui nous amène à accepter la modélisation du terme d'erreur unilatère. Pour sa part, η , le coefficient de l'inefficience relié au temps (modèle variable dans le temps) n'est pas significatif, ce qui nous indique que l'inefficience est constante dans le temps. En conséquence, les modèles de moindres carrés (*pooling*, effets fixes et effets aléatoires) sont mis de côté au profit du modèle de frontières stochastiques. Puisque le coefficient η est non significatif, le modèle de frontières stochastiques invariant dans le temps a été retenu comme modèle d'estimation de référence. Tous les coefficients de ce modèle sont significatifs.

Comme il a été dit, lorsqu'on analyse le tableau ci-dessous en commençant par les méthodes des moindres carrés ordinaires, on constate qu'au fur et à mesure qu'on relâche les hypothèses, les coefficients deviennent tous significatifs et se stabilisent. Ces résultats rejoignent ceux de Greene (2004b). D'où le résultat : *moins on fait d'hypothèses, plus les coefficients deviennent significatifs et plus il y a de coefficients significatifs*. De plus, les coefficients se stabilisent au fur et à mesure que l'on relâche les hypothèses. Par exemple, la densité de la population par 10 000 habitants est corrélée positivement avec le taux de mortalité et est significative avec la méthode des moindres carrés ordinaires (*pooling*). La variable devient négativement non-significative avec la méthode des moindres carrés ordinaires avec effets fixes tel qu'attendu et à nouveau, positivement significative avec celle avec effets aléatoires. Le sens de la corrélation reste le même dans les deux modèles de frontières stochastiques. Certaines variables ressortent significatives quelque soit le modèle d'estimation comme c'est le cas avec le coefficient du pourcentage de la population analphabète qui est significatif dans toutes les méthodes d'estimation. Aussi, la méthode des moindres carrés ordinaires avec effets aléatoires présentent des coefficients significatifs et presque similaires à ceux de la méthode de frontières stochastiques, mais une fois encore, cette méthode est limitée par l'hypothèse d'efficience vue dans le chapitre précédent et explique faiblement le modèle avec un R^2 de 0,31.

Par ailleurs, les résultats ne dépendent pas seulement de la méthode utilisée, mais aussi de la définition des variables. Cela est l'objet de la section suivante.

Tableau 5.1: Résultats en fonction de la méthode d'estimation

Variables explicatives (en log)	taux de mortalité : 308 observations									
	MCO Pooling $R^2=0,42$		MCO avec effets fixes $R^2=0,33$		MCO avec effets aléatoires $R^2=0,31$		Frontières stochastiques : modèle invariant avec effets fixes		Frontières stochastiques: modèle variable dans le temps avec effets aléatoires	
	coefficients	<i>t-stat</i>	coefficients	<i>t-stat</i>	coefficients	<i>t-stat</i>	coefficients	<i>t-stat</i>	coefficients	<i>t-stat</i>
<i>pour_pop_educ</i>	-0,18	-4,37	-0,13	-2,66	-0,15	-3,47	-0,15	-3,39	-0,12	-2,73
<i>pib_corr_infl</i>	-0,044	-4,61	-0,053	-3,67	-0,064	-6,15	-0,070	-6,22	-0,016	-1,05
<i>densite</i>	0,012	1,83	-0,18	-1,97	0,031	3,30	0,031	3,09	0,01	0,10
<i>rt_dep_vieu</i>	0,22	6,74	0,064	1,45	0,11	2,97	0,11	2,91	0,17	4,44
<i>tot_inst_mil</i>	-0,24	-2,65	-0,18	-2,47	-0,19	-2,61	-0,18	-2,63	0,16	-2,73
<i>vol_eau_off_10mill</i>	-0,60	-9,44	-0,44	-2,34	-0,58	-5,31	-0,56	-4,91	-0,52	-4,79
<i>constante</i>	3,20	17,89	2,35	5,16	3,43	15,72	3,19	12,2 3	1,99	1,44
v_i (constante de l'inefficience)	nsp	nsp	nsp	nsp	nsp	nsp	0,22	2,40	0,75	0,60
η (coefficient de l'inefficience relié au temps)	nsp	nsp	nsp	nsp	nsp	nsp	nsp	nsp	0,011	0,66

5.1.2 Importance de la définition des variables

Cette section découle du fait que notre approche est essentiellement *macro*. Nous étudions avant tout des relations entre des variables qui représentent chacune des aspects qui dans les faits sont plus subtils. La question de la représentativité de la variable par rapport au phénomène envisagé devient rapidement objet de débat. Pour cette raison, il est important de s'assurer que nos résultats sont robustes et qu'ils ne dépendent pas d'un choix anecdotique des variables. La capacité d'effectuer ce type d'analyse de sensibilité dépend de la possibilité de mettre la main sur diverses variables représentant toutes le même phénomène. Cela est habituellement hors de portée, mais dans notre cas, nous avons quelques possibilités pour représenter l'éducation (voir chapitres des données). Le tableau 5.2 ci-dessous présente la stabilité dans les résultats du modèle d'estimation lorsqu'on change la définition de la variable d'éducation dans la méthode d'estimation de frontières stochastiques. Par exemple, les modèles 1, 2 et 3 du tableau 5.2 sont les mêmes que le modèle présenté dans le tableau 5.1 ci-haut, sauf que la définition de la variable d'éducation change. Le pourcentage de la population avec niveau d'éducation est utilisé dans le modèle de base (tableau 5.1), le pourcentage de la population non scolarisée est utilisé dans le modèle 1, le pourcentage de la population analphabète est utilisé dans le modèle 2 et celui des femmes analphabètes est appliqué dans le modèle 3. Ce qui ressort de cette démarche est la constance des résultats. Les variables utilisées dans le modèle de frontières stochastiques sont toutes significatives. Par exemple, lorsqu'on change la définition de la variable d'éducation, c'est-à-dire lorsqu'on passe du pourcentage de la population avec un niveau d'éducation à celui de la population non-scolarisée, le coefficient de la variable économique (PIB corrigé l'inflation par 10 000 habitants) demeure toujours significatif, ainsi que celui de la nouvelle variable d'éducation utilisée. Cette constance ne se reflète pas seulement pour ces deux variables mais pour toutes les autres variables du modèle.

Tableau 5.2 : Modèle en fonction de la définition de la variable d'éducation

Catégories de variables (en log)	taux de mortalité: 308 observations	
	Coefficients	t-stat
	Modèle 1	
pour_nscol	0,43	3,16
<i>pib_corr_infl</i>	-0,070	-6,97
<i>densite</i>	0,029	2,94
<i>rt_dep_vieu</i>	0,10	2,74
<i>tot_inst_mil</i>	-0,16	-2,30
<i>vol_eau_off_10mil</i>	-0,55	-4,89
<i>constante</i>	2,33	8,18
	Modèle 2	
pour_anphl	0,068	3,72
<i>pib_corr_infl</i>	-0,049	-3,75
<i>densite</i>	0,026	2,71
<i>rt_dep_vieu</i>	0,093	2,61
<i>tot_inst_mil</i>	-0,16	-2,24
<i>vol_eau_off_10mil</i>	-0,54	-4,84
<i>constante</i>	2,42	9,11
	Modèle 3	
pour_fem_anphl	0,066	3,52
<i>pib_corr_infl</i>	-0,051	-3,95
<i>densite</i>	0,024	2,51
<i>rt_dep_vieu</i>	0,092	2,58
<i>tot_inst_mil</i>	-0,16	-2,26
<i>vol_eau_off_10mil</i>	-0,55	-4,93
<i>constante</i>	2,42	9,02

Il faut retenir des ces sections précédentes que les résultats diffèrent selon la méthode appliquée. Les méthodes des moindres carrés ordinaires (*pooling*), avec effets fixes et avec effets aléatoires ne permettent pas aux variables d'expliquer le taux de mortalité en raison de la faible valeur du R^2 de ces dernières discutées ci-haut.

Le modèle de frontières stochastiques invariant avec effets fixes assure la stabilité et la significativité des coefficients tout en étant moins contraignantes que les méthodes de

moindres carrés. Les résultats sont de plus en plus significatifs lorsqu'il y a moins d'hypothèses. Les coefficients sont stables lorsque la définition des variables change, ce qui nous rassure sur la validité des coefficients estimés. Rappelons que le modèle de frontières stochastiques avec effets aléatoires qui admet une variation du temps a encore moins d'hypothèses que celui invariant et aurait pu être retenu pour l'estimation. Cependant, dans le cas de notre étude, le coefficient rattaché au temps qui différencie le modèle variable dans le temps de celui invariant est non significatif (tableau 5.1). Par conséquent, seul le modèle de frontières stochastiques invariant suffit pour l'estimation de la frontière.

L'analyse qui précède nous permet de conclure sur le choix de la méthode d'estimation retenue dans le mémoire. Le modèle de frontières stochastiques en panel avec effets fixes qui n'autorise pas une variation de l'inefficience dans le temps est la méthode la plus stable et la plus générale. Pour cette raison, c'est celle pour laquelle nous optons en vue d'interpréter les résultats sur chaque variable. Le groupe de variables explicatives retenu est celui du tableau 5.1.

5.2 Résultats sur les variables

Cette section s'intéresse à l'interprétation des résultats à partir des coefficients et des effets marginaux (tableau 5.1). Nous aborderons aussi les questions de l'élasticité et de l'efficience relative des déterminants de santé ainsi que de la mesure d'inefficience.

L'état de santé dans les régions de la Chine est déterminé par l'éducation, les caractéristiques sociales, l'économie, le style de vie (l'hygiène) et les ressources en santé disponibles.

La variable d'éducation est significative et est reliée négativement au taux de mortalité. Le signe négatif s'explique par le fait que la variable est par définition « positive » : il s'agit du pourcentage de la population avec un niveau d'éducation. À l'inverse, le ratio de dépendance des personnes âgées est significatif et son coefficient est corrélé positivement avec le taux de mortalité. Lorsqu'on applique le test de sensibilité aux coefficients, les valeurs restent les mêmes ou ne sont que légèrement modifiées. Par exemple dans le tableau 5.2 lorsque la définition de la variable d'éducation change (modèle 1, 2 et 3), les coefficients des autres variables du modèle demeurent toujours significatifs et ne varient que légèrement. Nous

pouvons affirmer que quelque soit la définition de la variable appliquée dans le modèle, les coefficients demeurent stables dans l'ensemble. En général, les coefficients sont significatifs aux seuils de 1 % et 0,1 %.

Par ailleurs, étant donné la valeur des coefficients, les variables n'ont pas la même importance sur le taux de mortalité. Les résultats montrent que certaines variables ont plus d'effets sur la variable dépendante, comparativement à d'autres dont l'effet est moindre. Les variables de style de vie, de ressources de santé et d'éducation ont les coefficients les plus élevés et sont les trois principaux indicateurs de l'état de santé du modèle. Ces résultats ne sont pas contradictoires avec ceux de Newhouse (1977). En fait, ils les complètent en montrant qu'il y a d'autres facteurs que le revenu (le PIB par habitant) aussi importants pour déterminer l'état de santé. Nous citons entre autres le style de vie et l'hygiène des populations, représentés dans notre analyse par le volume d'eau potable offert à la population, l'ensemble des institutions de santé et l'éducation de la population. Ces dernières ont un impact capital sur le taux de mortalité. Leurs coefficients ressortent plus fortement dans le modèle d'estimation. Les ressources de santé, le style de vie et l'éducation représentés dans le modèle diminuent le taux de mortalité de la population dans les régions de Chine. Les coefficients des variables économiques et sociodémographiques font de celles-ci des déterminants aussi importants de l'état de santé. Ces résultats sont les mêmes que ceux trouvés par Leu (1986) qui affirment que le revenu est un facteur clé dans l'explication de l'état de santé, mais les variables sociodémographiques ont aussi des effets considérables.

Le tableau 5.1 ci-dessus permet d'examiner également l'élasticité des variables explicatives sur le taux de mortalité. Puisque les données ont été transformées en forme logarithmique, les valeurs de l'élasticité des variables sont les mêmes que les résultats de l'estimation et pour toutes les régions. Chaque ligne du tableau représente la sensibilité du taux de mortalité par rapport à chacune des variables. En d'autres termes, il s'agit du lien entre la variable dépendante et les variables explicatives.

Les variations dans le taux de mortalité sont grandement expliquées par celles des variables de style de vie et de ressources de santé. Le signe de chaque valeur décrit la relation entre les variations des variables explicatives et celles du taux de mortalité. Par exemple, une variation négative des ressources de santé entraîne une variation positive du taux de mortalité. Les

variations du PIB par habitant aussi petites soient-elles, sont très importantes dans les variations de l'état de santé. Dans l'analyse, l'élasticité de la variable économique n'est pas proche de 1. Par conséquent, l'état de santé dans les régions chinoises n'est pas un bien de luxe, mais plutôt un bien nécessaire. Ces résultats sont similaires à ceux trouvés par Hitiris et Posnett (1992). De même, Greene (2004b) a trouvé dans son analyse que les variations de l'éducation et de la densité de la population étaient aussi des facteurs importants dans celles de l'état de santé des pays de l'OMS.

Le tableau 5.3 présente la sensibilité du taux de mortalité suite à une variation de 1 % de chacune des variables explicatives. La deuxième colonne du tableau 5.3 représente la moyenne des variables explicatives du modèle d'estimation. La troisième colonne correspond à une variation de 1 % dans les moyennes de ces variables et la dernière colonne représente la variation correspondante dans le taux de mortalité. Lorsque les inputs de santé varient de 1 %, le taux de mortalité est plus affecté par la variation des inputs d'éducation, car leur impact est plus grand.

Ce qui ressort de notre travail, c'est que les variables les plus importantes pour améliorer la santé ne sont pas médicales mais bien sociales. Au premier rang, loin devant les autres déterminants, on retrouve l'éducation. Puis viennent le revenu, l'accès à l'eau (un indicateur de salubrité du milieu) et le ratio de dépendance de personnes âgées. Finalement, même si elles sont significatives, la densité de population et le nombre d'institutions de santé n'ont qu'un impact marginal.

Cela n'est pas surprenant. Les conditions de vie de tous les jours ont un effet plus important que les services de santé qui ne servent que lors des événements affectant la santé de façon ponctuelle. Améliorer les conditions d'hygiène (ici, le volume d'eau offert à la population) et la capacité des individus à prendre soin d'eux-mêmes (le revenu et l'éducation) ont des effets structurants plus importants à long terme que le nombre de médecins ou d'hôpitaux qui sont utilisés principalement en situation de crise ou en fin de vie.

Tableau 5.3 : Impact d'une variation de 1 % des inputs de santé sur le taux de mortalité

Variabes	Moyenne des variables	Variation de 1 % de la variable explicative	Variation du taux de mortalité correspondant
<i>pour_pop_educ</i>	26,62	0,27	-0,25
<i>pib_corr_infl</i>	37,76	0,38	-0,16
<i>vol_eau_off_10mil</i>	4,28	0,043	-0,15
<i>rt_dep_vieu</i>	11,10	0,11	0,075
<i>tot_inst_mill</i>	3,08	0,031	-0,034
<i>densite</i>	0,037	0,00038	0,000070

N.B. La moyenne du taux de mortalité est 6,18

5.3 Efficience relative des déterminants de santé

Le tableau 5.4 montre les résultats sur l'efficience des variables en 2007 pour chacune des régions classées selon l'ordre croissant du revenu par habitant. Le tableau 5.5 présente l'efficience des inputs de santé par groupe de régions pour la même année. Il s'agit de comparer les régions tout en analysant l'efficience relative des inputs de santé.

Mesurer l'efficience technique revient à évaluer la capacité de chaque région à transformer les inputs de santé en output (l'état de santé), ce qui peut être traduit comme la performance à transformer les facteurs de santé en état de santé.

À la suite de l'estimation du modèle de frontières stochastiques, on récupère en premier le taux d'inefficience via la relation suivante (Battese et Coelli, 1992) :³⁵

$$efficience_{it} = E \left\{ \exp(v_{it}) \middle| \mu_{it} \right\}.$$

Ensuite, $\exp(tx_mort)$ calcule l'output observé en niveau (rappelons que tx_mort est le log du taux de mortalité) et le taux d'inefficience permet de calculer l'output efficient selon la relation $\left\{ \exp(tx_mort_{it}) / efficience_{it} \right\}$.

³⁵ Avec la commande *predict efficiency, te* (pour *technical efficiency*) dans *Stata*.

Dans notre analyse, nous rappelons que l'output est représenté par le taux de mortalité qui est un output « négatif ». Les résultats sont différents d'une région à une autre. Cependant, les taux de mortalité, efficient et observé, sont plus élevés dans les régions à faible revenu, ce qui est conforme à notre résultat reliant positivement le taux de mortalité au revenu. Cependant, et cela est un résultat qui peut paraître surprenant, *il n'y a pas de relation systématique entre le niveau d'inefficience et le niveau de revenu.*

Les écarts de performance observés dans les résultats entre les régions peuvent être attribuables à une mauvaise allocation des ressources dans l'investissement des déterminants de santé, à des investissements inadéquats dans les ressources de santé, à des prises de décisions inadéquates de la part des dirigeants politiques, etc. Cela peut se résorber par le biais d'une réallocation des facteurs de santé adéquate.

En résumé, les résultats du tableau ci-dessous signifient d'une part que l'inefficience reste constante dans le temps. Une fois de plus, les résultats sont les mêmes que les conclusions formulées dans le modèle de frontière stochastique de Greene (2004b) : *Il n'y a pas beaucoup d'hétérogénéité dans le temps et l'inefficience elle-même ne varie pas dans le temps.* D'autre part, le taux de mortalité observé est plus élevé dans les régions pauvres. L'écart dû à l'inefficience est plus élevé dans la région la plus pauvre et dans celles les plus riches. Dans la majorité des régions où le revenu par habitant est supérieur à 10 000 et inférieur à 20 000, l'écart dû à l'inefficience est faible.

Tableau 5.4 : Efficience des inputs de santé par région en 2007

Régions	Revenu moyen par habitant	Taux de mortalité observé	Taux de mortalité efficient	Écart dû à l'inefficience
Shanghai	66 367	6,03	4,75	1,28
Beijing	58 204	4,92	4,27	0,65
Tianjin	46 122	5,86	4,69	1,17
Zhejiang	37 411	5,57	4,83	0,74
Jiangsu	33 928	7,07	5,60	1,47
Guangdong	33 151	4,66	4,35	0,31
Shandong	27 807	6,11	5,22	0,89
Liaoning	25 908	5,36	4,37	0,99
Fujian	25 729	5,90	5,37	0,53
Mongolie Intérieure	25 393	5,73	4,87	0,86
Hebei	19 877	6,78	5,64	1,14
Heilongjiang	19 383	5,39	4,79	0,60
Jilin	18 478	5,05	4,49	0,56
Xinjiang	16 999	5,01	4,36	0,65
Shanxi	16 945	5,97	5,11	0,86
Henan	16 206	6,32	5,51	0,81
Hubei	16 012	5,96	5,13	0,83
Hainan	14 660	5,71	5,21	0,50
Chongqing	14 649	6,30	5,51	0,79
Shaanxi	14 607	6,16	5,40	0,76
Hunan	14 555	6,71	5,41	1,30
Ningxia	14 492	5,04	4,89	0,15
Qinghai	14 257	6,13	5,01	1,12
Jiangxi	12 893	5,99	5,20	0,79
Sichuan	12 633	6,29	5,37	0,92
Tibet	12 555	5,10	4,26	0,84
Guangxi	12 109	5,99	5,16	0,83
Anhui	12 045	6,40	6,02	0,38
Yunnan	10 346	6,22	5,18	1,04
Gansu	10 540	6,65	5,63	1,02
Guizhou	6 915	6,60	5,68	0,92

Tableau 5.5 : Efficience des inputs de santé par groupe de régions en 2007

Régions	Revenu moyen par habitant	Taux de mortalité observé	Taux de mortalité efficient	Écart dû à l'inefficience
Provinces dont le revenu par habitant est supérieur à 30 000 : (6 régions)	38 557	5,68	4,83	0,85
Provinces dont le revenu par habitant est supérieur à 20 000 et inférieur à 30 000 : (4 régions)	26 711	5,88	5,05	0,83
Provinces dont le revenu par habitant est supérieur à 10 000 et inférieur à 20 000 : (20 régions)	14 865	5,96	5,21	0,75
Provinces dont le revenu par habitant est inférieur à 10 000 : (1 région)	6 915	6,60	5,68	0,92

CONCLUSION

L'objectif principal de cette étude était d'identifier des déterminants de la santé dans plusieurs régions chinoises et le niveau d'inefficience dans l'allocation des ressources en santé. À l'aide de plusieurs données recueillies sur plusieurs années, pour chacune des régions, de nombreux outils ont été utilisés dans l'accomplissement de ce travail. Les études antérieures réalisées dans le domaine de la santé ont été classées en trois grandes générations par (Gerdtham *et al.*, 2000) et ceci en fonction de la méthode économétrique utilisée, des hypothèses émises et des variables prises en considération. L'hypothèse d'inefficience de la frontière stochastique soumise par Greene (2004a) va constituer une quatrième génération et c'est son modèle économétrique qui a été retenu pour l'analyse présente. Les données ont été recueillies à l'aide de plusieurs fichiers et sites statistiques pour trente et une régions en Chine. Plusieurs autres régions ont dû être écartées en raison de l'absence de données. Par conséquent, l'étude a retenu des variables sociodémographiques, économiques, éducationnelles, de ressources de santé et de style de vie pour analyser l'état de santé des populations via le taux de mortalité, sur un intervalle de douze années, c'est-à-dire, de 1995 à 2007.

Les résultats qui ressortent de cette étude peuvent être fonction soit de la méthode économétrique appliquée, soit de la définition des variables. Alors, au fur et à mesure qu'on relâche les hypothèses, les coefficients du modèle deviennent significatifs et robustes. Il y a une relation inverse entre le taux de mortalité et le pourcentage de la population éduquée, le PIB par habitant corrigé de l'inflation, le nombre total d'institution de santé par dix mille habitants et le volume d'eau potable offert par dix mille habitants. Une augmentation dans chacune de ces variables entraîne une diminution du taux de mortalité des populations. À l'inverse, le pourcentage de la population analphabète, le pourcentage de la population non scolarisée, le pourcentage des femmes analphabètes, la densité de la population, et le ratio de dépendance des personnes âgées, ont une relation positive avec le taux de mortalité de la population. Par exemple, une hausse du pourcentage de la population analphabète entraîne

également une hausse du taux de mortalité. Les variables liées aux habitudes de vie (volume d'eau potable offert), à l'éducation (analphabétisme et niveau d'éducation), aux ressources de santé (institutions de santé), à l'économie (produit intérieur brut par habitant) et aux facteurs sociodémographiques (densité) sont les indicateurs les plus importants pour déterminer l'état de santé dans le modèle.

Les variables de revenu, d'hygiène et d'éducation sont les plus sensibles du modèle. Cependant, le sens de la variation des inputs reste le même dans toutes les régions et conserve la même relation que celle des coefficients de l'estimation du modèle envers le taux de mortalité. L'état de santé dans les régions de Chine n'est pas un bien de luxe, mais plutôt un bien nécessaire.

L'inefficience relative des inputs de santé est constante dans le temps mais ne semble pas montrer de relation évidente avec le niveau de revenu. La différence dans les résultats sur l'efficience, permet de spécifier l'objectif que doivent viser les autorités du système de santé dans chacune des régions.

Cette étude est une contribution à l'amélioration du système de santé non seulement pour les autorités régionales en Chine, mais aussi pour tout pays désireux de mieux cibler leurs investissements en santé.

Il faudra investir prioritairement dans l'éducation, les ressources en santé et dans l'amélioration de l'hygiène des populations, sans toutefois négliger les facteurs médicaux des régions.

RÉFÉRENCES

- Battese G. E. et Coelli T. J., A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, *Empirical Economics*, Volume 20, pp. 325-332, 1995.
- Battese G. E. et Coelli T. J., Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India, *The Journal of Productivity Analysis*, Volume 3, pp. 153-169, 1992.
- Berthé A., *Lien entre le revenu, les dépenses de santé et l'état de santé des populations*, Mémoire de maîtrise en science économique, Département des sciences économiques, Université du Québec à Montréal, 2003.
- Crémieux P.-Y., Ouellette P. et Pilon C., Health Care Spending as Determinants of Health Outcomes, *Health Economics*, Volume 8, pp. 627-639, 1999.
- Crémieux P.-Y., Ouellette P., Meilleur M.-C., Leong S., Greenberg P., et Binrbaum H., *Investing in Health : The Social and Economic Benefits of Health Care Innovation, Research in Human Capital and Development*, dans I. Farquhar, K Summers et A Sorkin (Eds.), Elsevier Science, Volume 14, 2001.
- Crémieux P.-Y., Meilleur M.-C., Ouellette P., Petit P. et Zelder M., Public and Private Pharmaceutical Spending as Determinants of Health Outcomes in Canada, *Health Economics*, Volume 14, pp. 107-116, 2004.
- Gerdtham U.G., Sögaard J., Anderson F. et Jönsson B., Econometric Analysis of Health Expenditure: A Cross-Section Study of the OECD Countries, *Journal of Health Economics* Volume 11, pp. 63-84, 1992.
- Gerdtham U.G., Sögaard J., Jönsson B. et Andersson, A pooled Crosses Section Analysis of the Health Expenditure of the OECD Countries, dans P. Zweifel et H. Frech (Eds.), *Health Economics*, Dordrech: Kluwer Academic Publishers, 1992.

- Gerdtham U.G. et Jönsson B., International Comparisons of Health Expenditure: Theory, Data, and Econometric Analysis, dans Culyer A. et Newhouse J., (Eds.), *Handbook of Health Economics*, Elsevier Science Publishers, 2000.
- Greene W., A Gamma Distributed Stochastic Frontier Model, *Journal of Econometrics*, Volume 46, 1, pp. 141-164, 1990.
- Greene W., The Econometric Approach to Efficiency Analysis, The Measurement of Productive Efficiency: Techniques and Applications, dans Fried HO, Lovell CAK, Schmidt, SS (Eds.), Oxford University Press, New York, pp. 68–119, 1993.
- Greene W., *Econometric Analysis, fifth edition*, Upper Saddle River NJ: Prentice Hall, 2003.
- Greene W., Fixed and Random Effects in Stochastic Frontier Models, *Journal of Productivity Analysis*, Volume 23, pp. 7–32, 2004a.
- Greene W., Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems, *Health Economics*, Volume 13, pp. 959–980, 2004b.
- Hitiris T. et Posnett J., The Determinants and Effects of Health Expenditure in Developed Countries, *Journal of Health Economics*, Volume 11, pp. 173-181, 1992.
- Im K.S., Pesaran M.H. et Shin Y., Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, Volume 115, pp. 53-74, 2003.
- Kim Y. et Schmidt P., A Review and Empirical Comparison of Bayesian and Classical Approaches to Inference on Efficiency Levels in Stochastic Frontier Models with Panel Data, *Journal of Productivity Analysis*, Volume 14, 2, pp. 91-118, 2000.
- Leu R. E., The Public-Private Mix and International Health Care Costs.” dans A.J. Culyer et B. Jönsson (Eds.), *Public and Private Health Services*, Oxford: Basil Blackwell, Oxford, R. E, 1986.
- Newhouse J., Medical Care Expenditure: A Cross-National Survey, *Journal of Human Resources*, Volume 12, pp. 115-125, 1977.
- Roberts J., Sensitivity of Elasticity Estimates For OECD Health Care Spending: Analysis of a Dynamic Heterogeneous Data Field, *Health Economics*, Volume 8, pp. 459-472, 1999.

Stevenson R.E., Likelihood Functions for Generalized Stochastic Frontier Estimation,
Journal of Econometrics, Volume 13, pp. 57-66, 1980.