Université Clermont Auvergne, Clermont-Ferrand
École Doctorale des Sciences Économiques, Juridiques, Politiques et de Gestion
Centre d'Études et de Recherches sur le Développement International (CERDI)

POTENTIAL FOR IMPROVEMENT OF EFFICIENCY IN HEALTH SYSTEMS: THREE EMPIRICAL STUDIES

Thèse Nouveau Régime
Présentée et soutenue publiquement le 26 septembre 2017
Pour l'obtention du titre de Docteur *ès* Sciences Économiques

Par

Laurène PETITFOUR

Sous la direction de
Mme Martine AUDIBERT et de Pr. Jacky MATHONNAT

Membres du jury

Jean-Claude BERTHÉLÉMY	Professeur, Université Paris I Sorbonne	Rapporteur
Hervé LELEU	Directeur de recherche, Lille Économie Management,	Rapporteur
Carine MILCENT	Chargée de recherche, PSE-ENS	Suffragante
Damien DE WALQUE	Senior Economist, Banque Mondiale	Suffragant
Jacky MATHONNAT	Professeur Émérite, Université Clermont Auvergne	Directeur
Martine AUDIBERT	Directrice de recherche, Université Clermont Auvergne	Directrice

Université Clermont Auvergne, Clermont-Ferrand
École Doctorale des Sciences Économiques, Juridiques, Politiques et de Gestion
Centre d'Études et de Recherches sur le Développement International (CERDI)

POTENTIAL FOR IMPROVEMENT OF EFFICIENCY IN HEALTH SYSTEMS: THREE EMPIRICAL STUDIES

Thèse Nouveau Régime
Présentée et soutenue publiquement le 26 septembre 2017
Pour l'obtention du titre de Docteur *ès* Sciences Économiques

Par

Laurène PETITFOUR

Sous la direction de
Mme Martine AUDIBERT et de Pr. Jacky MATHONNAT

Membres du jury

Jean-Claude BERTHÉLÉMY	Professeur, Université Paris I Sorbonne	Rapporteur
Hervé LELEU	Directeur de recherche, Lille Économie Management,	Rapporteur
Carine MILCENT	Chargée de recherche, PSE-ENS	Suffragante
Damien DE WALQUE	Senior Economist, Banque Mondiale	Suffragant
Jacky MATHONNAT	Professeur Émérite, Université Clermont Auvergne	Directeur
Martine AUDIBERT	Directrice de recherche, Université Clermont Auvergne	Directrice

Remerciements

5 années de thèse, 4 bureaux occupés, 3 logiciels utilisés, 2 pays découverts, 1 thèse.

On fait le bilan, calmement, et il peut sembler maigre réduit à des chiffres. Pourtant...

Trois chapitres de cette thèse sont issus d'études réalisées avec mes directeurs de thèse. Ils cachent des heures de discussions parfois animées, et des versions successives de questionnaires et de rapports, des visites d'hôpitaux dans la campagne chinoise, des traversées enneigées d'Oulan-Bator. Un grand merci à Mme Audibert et M. Mathonnat pour la confiance qu'ils m'ont accordée et pour leur investissement dans les projets que nous avons menés ensemble. Ils ont rendu ces années de thèse extrêmement riches et stimulantes, loin de toute routine.
Je remercie aussi vivement Jean-Claude Berthélémy, Hervé Leleu, Carine Milent et Damien de Walque d'avoir accepté de lire cette thèse. Vos commentaires permettront d'améliorer les travaux qui la composent.
Deux enquêtes de terrain sont au cœur de cette thèse. Elles n'auraient pas eu lieu sans les équipes de l'Université médicale et du Bureau de la santé à Weifang (Chine), et de l'Ecole de Santé Publique d'Oulan-Bator (Mongolie). Je les remercie chaleureusement pour leur accueil, le temps passé à nous guider et à collecter les données, et pour leur investissement. Je n'oublie pas Xiezhe, qui a su endosser sans faillir la casquette d'interprète en plus de celle de doctorant pendant nos séjours à Weifang. Je te souhaite le meilleur pour la suite.

Cette thèse fait partie de six années passées au Cerdi, au milieu d'une foule sentimentale avec soif d'idéal que je salue: les doctorants, les enseignants-chercheurs, les étudiants, et Chantal, Agnès, Johan, Marie et Martine pour organiser tout cela, et rendre les couloirs vivants malgré leur moquette entre gris clair et gris foncé. Certains connaissent déjà bien cette thèse. Une pensée et la patrie reconnaissante à Hervé, Aurore, Jérôme et Marlène qui ont subi moult de mes questions et qui m'ont relue... Sur les pages typographiées... Coquilles, rages et crustacés...

Certains de mes camarades ont eu plus de chance que les autres et ont partagé leur bureau avec moi pendant un temps. A Pierre (sucré pimenté), Sekou (désolée j'ai épuisé mon stock de blagues sur ton nom), Christelle, Clément, Maria et Antoine, mes salutations. Je vous souhaite juste... Une minute de silence. J'embrasse aussi

tous les autres cerdiens qui ne sont pas timides, tu sais: Vincent (et Tom Cruise), Jérôme (and the Jets), Claire, Jules, Hughes, Martha, Clément, Victor S., Hippolyte, Ababacar, Pierre, et Charlotte et Seb, ces magnifiques Jambons de triathlon. Et bien sûr, merci à Chloë car elle l'a elle l'a, ce je n'sais quoi.
Mes hommages également à la bande de Pich': Antoine, Victor, Sophie, Flo, Gwen, Estelle, Alex. Il n'y aura jamais un copain de trop dans l'équipe à Pierrot.
Ce lustre doctorant aura été un lustre de vie clermontoise. J'ai touché le fond d'la piscine, et y ai trouvé des gens formidables. A Romain, antisocial qui garde son sang-froid, Baptiste, équipier rêvé de Swimrun, Elsa, Aude, Johan, Morane, Rodolphe, Morgane, Pauline et tous ces nageurs en bonnet bleu, beaucoup d'affection chlorée. Les Piel's, je vous écris aussi une petite bafouille, pour pas que vous vous fassiez de mourons. D'autres sont restés au sec mais m'ont rendue Clermont très attachante: Alex dont la petite entreprise ne connaitra pas la crise, Dorota, Marion dont le coup de folie n'est pas fini, Ju, et Angélique, pendant que les champs brûlent et bien après.

Des amis d'avant ont continué de me parler quand je suis partie à Clermont, là-bas, où tout est neuf et tout est sauvage. J'en remercie grandement Rémi, Dr Levionnois, Cyrille, Lola, Cyrielle, Lisa, Antoine, Robin, Célia, Charles et Christophe.

Enfin, je souhaite remercier ma famille d'être toujours là. Passez leur soutien à la machine, faites-le bouillir, et on voit que les couleurs d'origine vont toujours revenir. Une grosse pensée à mes parents, à mes grands-parents, et à mon petit frère qui veut des bottes de sept lieues.

Résumé

Dans l'optique du troisième Objectif de Développement Durable ("Santé et Bien-Etre pour tous"), il est nécessaire d'augmenter les ressources consacrées à la santé dans les pays à faible revenus, mais aussi de s'assurer que ces ressources sont allouées de façon optimale. Pour cela, les mesures d'efficience sont un outil d'analyse adapté pour évaluer la performance des systèmes au niveau macroéconomique ou des établissement de santé au niveau microéconomique, afin d'obtenir "plus de santé pour son argent" (Organisation Mondiale de la Santé, 2010). Au travers de ses quatre chapitres, cette thèse s'inscrit dans la littérature empirique de l'évaluation de l'efficience des systèmes de santé.

Le premier chapitre est une revue méthodologique des mesures non-paramétriques d'efficience, utilisées dans les trois chapitres empiriques qui suivent. Le second chapitre estime l'efficience d'un échantillon de 120 pays à revenu faible ou intermédiaire de 1997 à 2014. On considère que les systèmes de santé produisent des résultats en termes de santé (de la survie maternelle et infantile) grâce à des dépenses de santé. Les résultats montrent que, pour un état de santé identique, les pays de l'échantillon pourraient dépenser 20% de ressources en moins en moyenne, et que l'inefficience augmente avec le niveau de développement. Les deux derniers chapitres sont des études de cas. Le troisième porte sur un échantillon d'hôpitaux municipaux à Weifang, dans la province chinoise du Shandong. Il met en lumière, grâce à des données d'enquête, le potentiel d'amélioration en termes de performance, et le rôle de certains facteurs sur l'inefficience des hôpitaux: la demande de soins, et la part de subventions dans leur revenu. Le quatrième chapitre traite de l'efficience des établissements de soins de santé primaires à Oulan-Bator, en Mongolie. Avec les mêmes ressources, ils pourraient produire 30% de soins supplémentaires en moyenne. Le bassin de desserte est positivement associé au niveau d'efficience, mais la faible rémunération du personnel, ainsi qu'un équilibre sous-optimal entre personnel médical et non-médical semblent freiner l'efficience des établissements de santé.

Mots-clefs: Financement de la santé ; Efficience ; Soin de santé primaire; Analyse en frontière partielle; Réforme pharmaceutique chinoise.

Summary

In the perspective of the third Sustainable Development Goal ("Good Health and Well-being"), it is necessary to increase financial resources for health in low income countries, but also to ensure that those resources are optimally allocated. To this purpose, efficiency measures appear as a useful tool to assess the performance of healh systems at the macroeconomic level, or of health facilities as the microeconomic level to get "more health for the money" (WHO,2010). Through its four chapters, this thesis provides some empirical evidence to the assessment of the efficiency of health system.

The first chapter is a methodological review of nonparametric efficiency measures, used in the three empirical studies that follow. The second chapter assesses the efficiency of a sample of 120 low and middle income countries over the 1997/2014 period. Production function is defined as health expenditures producing health outcomes (maternal and juvenile survival). It concludes that, for the same health outcomes, countries could spend more than 20% for the same health outcomes, and that inefficiency increases with the level of development of coutries.
The last two chapters are case studies. The third one focuses on Township Health Centers in Weifang, Shandong province, China, relying on survey data. It highlights the potential for performance improvement and the role of demand side determinants and of the share of subsidies in incomes to explain efficiency scores. The fourth chapter deals with the efficiency of primary healthcare facilities in Ulan-Bator, Mongolia. It concludes that efficiency could be spurred by about 30%. Demand side factors are positively associated to efficiency, but low levels of staff remuneration, as well as a suboptimal balance between medical and non medical staff seem to hinder activity and efficiency of health facilities.

Keywords: Health financing ; Efficiency ; Primary healthcare; Robust frontier analysis; Chinese pharmaceutical reform.

Contents

List of Tables

List of Figures

List of acronyms

CERDI	Centre d'Etudes et de Recherches sur le Développement International
WHO	World Health Organization
SDG	Sustainable Development Goal
SFA	Stochastic Frontier Analysis
GDP	Gross Domestic Product
THC	Township Health Center
NCMS	New Cooperative Medical Scheme
NEML	National Essential Medicine List
FHC	Family Health Center
DMU	Decision Making Unit
DEA	Data Envelopment Analysis
FDH	Free Disposal Hull
WB	World Bank
DHS	Demographic and Health Survey
PPP	Purchasing Power Parity
CMS	Cooperative Medical Scheme
VHS	Village Health Station
CHS	Commune Health Center
NRCMS	New Rural Cooperative Medical Scheme
PHF	Primary Healthcare Facility
NEDP	National Essential Drug Policy
NHFPC	National Health and Family Planning Commission
NDRC	National Development and Reform Commission
BOR	Bed Occupancy Ratio
ALOS	Average Length Of Stay
TFP	Total Factor Productivity
ECG	ElectroCardioGram
EHR	Electronic Health Records
PCA	Principal Component Analysis
GAI	Global Activity Index
USSR	Union of Soviet Socialist Republics
MoH	Ministry of Health
FDI	Foreign Direct Investments

Chapter 1

Introduction

1.1 A health system: multiple components to articulate towards defined objectives

1.1.1 A health system: an interaction of multiple and various actors...

The definitions of what a health system encompasses are multiple and have been progressively enriched. The first and strictest definition considers as synonyms the health system and the healthcare system. In this perspective, it includes all the activities and people whose primary purpose are to promote, restore and/or maintain health (Reinhardt and Cheng, 2000). The concerned activities have different natures: personal medical health services are health care delivered individually; non-personal health services are delivered collectively (smoking or alcohol consumption prevention programs against for example); intersectoral actions are not directly health services but aim at improving health: sanitation infrastructures, legislation around the road traffic fro instance.

Even with this narrow definition of a health system, it involves several actors: health care providers at their widest sense (public and/or private medical staff of course, working in public and/or private entities, but also family or neighborhood caregivers), political authorities (Ministry of Health or equivalent), and financing institutions (private insurance companies or social security funds). Figure 1.1 sums up the different actions included in the health system.

1.1.2 ...At the service of a common objective

For a long time, the health system was considered to have one main objective: improvement of the health status of the population it serves. It was thus considered in the light of this primary goal. The Alma-Ata declaration in 1978 is a good example of this conception: it gives to developing countries the objective of a universal access to health by 2000, at a reasonable cost for users and societies, but with no detail about the financing model to implement to make such healthcare systems

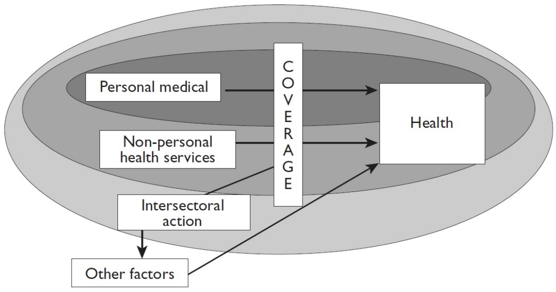


Figure 1.1: Boundaries of the health system
Source : WHO (2007)

sustainable. They were mostly financed by general government expenditures, and then jeopardized when those latter collapsed in the 1980's due to the debt crisis. As a consequence, quality of healthcare provided in those countries deteriorated.
In 1987, the Bamako initiative introduced the idea of user and communitarian participation, in order to ease health cost recovering for providers, and to improve the quality of provided care. Ridde (2004) highlights the fact that equity was neglected in the implementation of this reform. Cost of healthcare increased for households, and so did inequalities in access to health care and health indicators, both within developing countries, and between developing and developed countries: Figure 1.2 highlights the inequalities in life expectancy at birth. Figure 1.3 shows that in the poorest countries, households have to pay themselves for health services through out-of-pocket payments, which represent about a half of total health expenditures[1]. This proportion is much lower in upper middle income countries and high income countries.

In this context, it was necessary to rethink the health system concept, and to define what it encompasses. This was done by Frenk and Murray (1999) and Reinhardt and Cheng (2000). The main change was the definition of the mission of a health system, which was enlarged. Three objectives are described as "intrinsically valuable", i.e. they are at least partially independent from the others (it is partially possible to increase their attainment without changing one of the other objectives), and it is always desirable to improve their attainment (WHO). Other objectives are defined as instrumental, they serve the realization of the intrinsic objectives.

[1]Out-of-pockets are defined as payments borne by patients and the moment of the health service delivery. They are opposed to prepayments, realized through taxes or insurance premiums

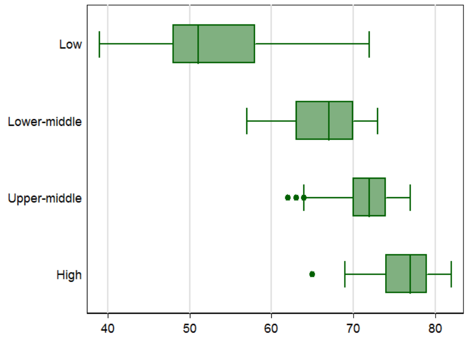


Figure 1.2: Distribution of life expectancy at birth by World Bank income group in 2000

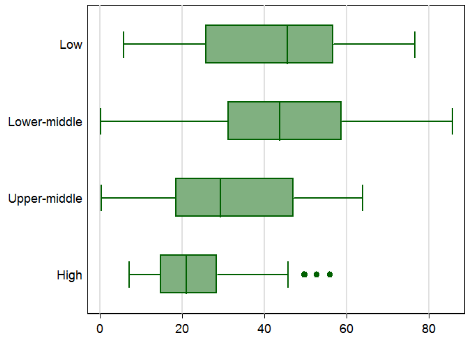


Figure 1.3: Distribution of the share of out-of-pocket payments in total health expenditures by World Bank income group in 2000

1.1.3 The enriching of health systems objectives

First, the "ultime" objective, improvement of health is decomposed into two parts: the first is an improvement in average, and the second a reduction of health inequalities (importance of the *distribution*). It was already implicit in the Alma-Ata declaration (through the notion of *universal* access to healthcare), but is now an explicit goal.

A second valuable objective of health systems is the responsiveness to the population needs and expectations. Again, it is decomposed between the average improvement and the reduction of inequalities.

Finally, health systems aim at a fair financing of the provided services to protect people against the risk of illness. The importance of the reduction inequalities in terms of health status and satisfaction of expectations, associated to the idea of a *fair* financing, suggests that equity has become a priority in the health system objectives.

To reach those objectives, the different actors of the health system fulfill four main functions: *delivery* of personal and non-personal health services, *investment* in health resources (human resources and capital, infrastructure and equipment), *financing* of the service delivery by raising and allocating resources, and *oversight* (or stewardship) of the different actors and resources involved in the system to meet its objectives. One actor may be involved in one or several of those functions. For instance, health facilities may be only responsible for delivering health services, or they can be responsible for service delivery and human resources creation through training. Public authorities, by organizing the relations between the different actors, are involved in the four functions. The logical chains from functions to objectives are depicted in Figure 1.4.

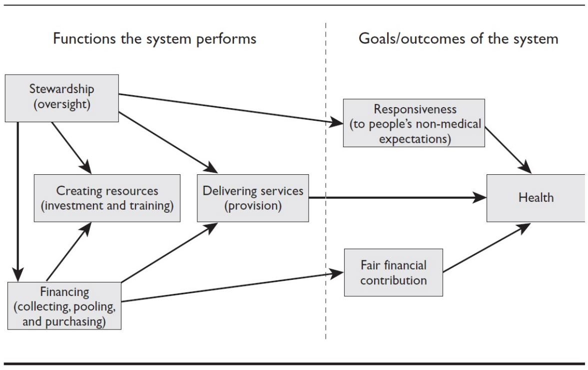


Figure 1.4: Articulation of the health system functions and goals
Source : WHO (2007)

When such functions and objectives are defined, then it is possible to assess the performance of a global health system and of its different actors, according to the missions they are each supposed to fulfill.
The first step of health system assessment is attainment: respectively to its objectives and functions, how does it perform? Kruk and Freedman (2008) provides a literature review of indicators used by studies assessing health systems and actors of the health system in the realization of the effectiveness and equity objectives. The evaluation of those realizations is to be evaluated in light of the outcomes indicators (e.g. health status information for effectiveness), or of output indicators (production of health services for effectiveness). The choice between the two depends upon the focus level of the study. For instance, considering health facilities, the improvement of health does depend upon their activity and performance, but not entirely. Therefore, it is more accurate to assess only their outputs.

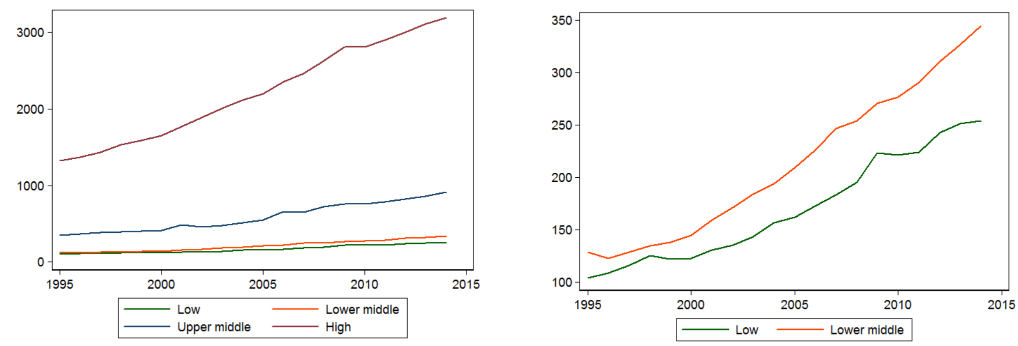
1.1.4 Universal Health Coverage and the Sustainable Development Goalth

WHO (2010) gives a framework for health reforms and recommendations: the Universal Health Coverage. It is defined as the goal that "people have access to services and do not suffer financial hardship paying for them" by the Member States of the World Health Organization in 2005, and encompasses all of the five previous objectives. In 2015, Universal Health Coverage is included in the Sustainable Development Goals (SDGs). It is the cornerstone of SDG 3 ("Good Health and Well-Being"), and is directly involved in SDG 1 ("No poverty"), 2 ("End hunger, achieve food security and improved nutrition") 10 ("Reduce inequalities within and among countries"). As such, its importance has been reinforced as a priority in the policy agenda.

1.2 How to improve health status and equity ?

1.2.1 The necessity to increase health expenditures

In order to improve the attainment of the five intrinsic objectives, a first instrumental objective is to increase the quantity of resources dedicated to health. This is even more crucial in developing countries, where the level of health expenditures per capita is extremely low, especially compared to high income countries (see Figure 1.5). In 2014, four countries still spent less than 50 annual dollars per capita in health: The Central African Republic, The Democratic Republic of the Congo, Eritrea and Madagascar, and 19 countries spent less than 100 annual dollars per capita in health (NHA).



All the groups   Low and Lower-middle income

Figure 1.5: Evolution of the average annual total health expenditures per capita by World Bank income groups, 2010 PPP dollars

More specifically, an objective is to increase public health expenditures as a lever for action. Private expenditures rely almost exclusively on out-of-pockets payments in developing countries. Moving towards equity in health financing implies the decrease in out-of-pocket payments. Indeed, they represent a financial burden for individuals and can impoverish them if they have to face catastrophic health expenditures (Wagstaff and Doorslaer, 2003; Xu et al., 2003) [2]. Consequently, in the objective of

[2]Catastrophic health expenditures are defined as a share of a household budget dedicated to healthcare that exceeds a certain threshold, and makes the household fall into poverty. The measure is calculated as the share of the available income allocated to healthcare (the threshold is commonly fixed around 10%, or by the share of the *ability to pay* dedicated to healthcare. In this case, the

fair financing of the health system, public health expenditures are the priority.

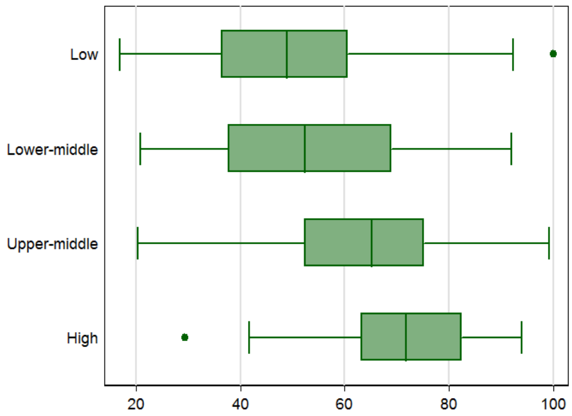


Figure 1.6: Percentage of government health expenditures in total health expenditures in 2014

Several levers for action are available to public authorities to increase public resources and dedicate them to health (WHO 2010):

- To give a higher priority to the health sector in the allocation of public resources. The declaration of Abuja in 2001 gave the objective of 15% of public expenditures allocated to health. This objective is far from being reached. Since 2005, the trend is even declining (see Figure 1.7). In 2001, only two African low income countries satisfied the criteria (Madagascar and Tanzania). In 2011, there were only six (Ethiopia, Liberia, Madagascar, Malawi, Swaziland and Uganda).

- To increase health expenditures through fiscal space (Heller, 2005; Mathonnat, 2010; Tandon and Cashin, 2010). Fiscal is defined as the *budgetary potential to gather additional resources for a given purpose* (here, health expenditures) *without harming the financial sustainability of the government*. This can come from (i) favorable economic conditions (economic growth) that create wealth and provide taxes, (ii) from a reallocation of government resources, (iii) from an increase in collected resources. In developing countries, where the informal sector is important, improving the low tax compliance may be a way to raise new resources, allocated to health, without decreasing other expenditures. Using the fiscal system is also advocated by WHO (2010) to seek for the Universal Health Coverage goal.

ability to pay is the available income minus the subsistence expenditures, it is associated to non-food expenditures. In this case, the threshold of 40% is commonly used. See Wagstaff and Doorslaer (2003) and Xu et al. (2003)

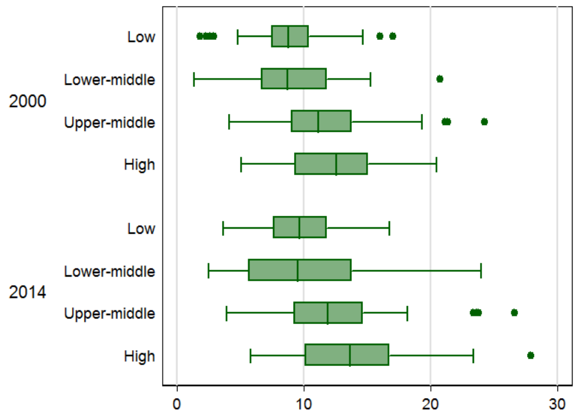


Figure 1.7: Evolution of the proportion of government health expenditures in government expenditures

Source : WHO (2007)

1.2.2 The amount of expenditures is not sufficient: the need for a greater efficiency of health systems

Nevertheless, the increase in health expenditures does not mechanically improve the health status, because health systems have been proved to use their resources inefficiently, leading to major wasted outlays. In this perspective, an improvement in the efficiency of health systems can be a major path to gains in health and equity, for the same quantity of inputs.

The interest of efficiency measures can be seen in two close but different ways. First, it is a tool to compare performance of health systems or their components, relatively to their resources and the constraints they face. Indeed, as pointed in WHO (2010), comparing outcomes is not sufficient to assess performance. For a same health status, a country using fewer resources than another should be ranked higher. Efficiency measures take into account the quantity of resources injected, and, to some extent, the constraints they face[3]. In this matter, they are a useful tool to compare the performances of health systems.

Second, efficiency is an instrumental goal of the health system, which aims at improving the attainment of the intrinsic objectives. An efficient system, i.e. a system which allocates its resources optimally produces the maximal level of health services expectable given those resources. This is favorable to global health status and to equity. If healthcare is too expensive, some people will not seek health care for financial reasons, therefore both health and equity deteriorate. If healthcare is of

[3]The presentation of those models is the core of Chapter 2

poor quality, people will not seek it either, or they will go to more expensive facilities. Again, both health and equity deteriorate.

It is thus crucial to move from the macro-economic diagnosis of the identification of the sources of this efficiency. This issue is developed by the WHO in its 2010 Health Report, focused on the development of the Universal Health Coverage. Reaching this Universal Health Coverage requires optimal use of current resources.
Aiming at getting "More Health for the money", WHO (2010) identifies ten major sources of inefficiency:

1. underuse of generics and excessive prices for medicines,
2. use of substandard and counterfeit medicines,
3. inappropriate and ineffective use of medicines,
4. overuse of healthcare equipment and procedures,
5. inappropriate staff mix, unmotivated workers,
6. inappropriate hospital admissions and length of stay,
7. inappropriate hospital size,
8. medical errors and suboptimal quality of care,
9. health system leakages (waste, corruption, fraud),
10. inefficient mix, inappropriate level of strategy.

From this statement, empirical work has to be developed through case studies to determine which sectors in which countries are affected by one or several of these source of inefficiency, in order to identify reforms to solve this issue. Efficiency and performance improvement of the health workforces are the first objectives given by WHO (2016), which underlines the accuracy of this issue.
To address those flaws of resources, a relevant organization of the relationship between the different healthcare providers is desirable. Several of those relationships are agency relations, and induce some agency costs from moral hazard and informaion asymmetry Those costs are sources of inefficiency of the health system[4]. Liu

[4]Agency costs appear between two economic actors when the first one (the "principal") hires the second one (the "agent") to produce a good or service on its behalf. The well-being of the principal positively depends upon the effort of the agent, which partly determines the output, and negatively depends on the wage he pays to its agent . On the contrary, the well-being of the agent depends positively on its revenue and negatively from his effort. The agency costs comes from the

and Mills (2007) analyze all the 11 health actors interactions that can be associated to agency relations. In this context, the stewardship function of health authorities is essential. It has to take into consideration every agency relation in order to anticipate strategic behaviors and to avoid them while defining provider payment mechanisms.

This thesis provides empirical evidences about two examples of the importance of dealing with information asymmetry and moral hazard to spur the efficiency of health institutions. The first study is of the Chinese pharmaceutical supply chain until 2009, widely reformed in 2009 (see section 4.1.2.1). The second study is on the contract binding primary healthcare facilities in Ulan-Bator, Mongolia to their missions, which are private entities contracted by the health authorities to provide health care but the ministry has little control over the effort of the medical staff (see Section 5.2.2). Another chapter deals with the performance of global health systems.

1.3 Structure of the thesis

This thesis consists in a methodological chapter and three empirical studies. Those three empirical works are part of the literature that use efficiency measures to assess the performance of health systems or facilities, and analyze those performances through different scopes.

Chapter 2 is a methodological review of the nonparametric efficiency models, used in the empirical Chapters 3, 4 and 5. Nonparametric efficiency models are part of *frontier* models, coming from the firm theory. It is focused on the production function of any good or service considered as the transformation of a quantity of inputs into a quantity of outputs. In this context, the production frontier is best possible performance, i.e. the maximal output for the quantity of inputs, or the minimal input quantity for the level of outputs. It is reached when inputs and outputs are allocated optimally. Efficiency measures are the estimation of the distance of an observation to its frontier, to its best possible performance. It relies on the estimation of the production frontier, and then on the computation of the distance from a point to this frontier.

Two majors kinds of models are implemented: Stochastic Frontier Models (SFA) and nonparametric deterministic models. The first category develops an economet-

divergence of interests, and from the fact that the principal does not observe the effort of the agent. This latter benefits from an information asymmetry. To maximize its well-being, the principal has to find payment mechanisms that make the agent reveal its effort. See Jensen and Meckling (1976) for the first definition of the agency relation.

ric estimation for the frontier, while the second category solves a linear program to compare the different observations in a sample. In this thesis, nonparametric models were chosen, the empirical reasons for each study's method are explained in Chapters 3, 4 and 5[5]. Therefore, nonparametric methods are presented, with the main stakes that drive their evolution. Some flaws have been coped with, and some issues remained at least partially unsolved, mainly the introduction of environmental variables.

Chapter 3 is a macroeconomic study on the efficiency of a sample of 120 low and middle income countries according to the World Bank. Relying on three year averages from 1997 to 2014, it aims at assessing the global performance of those countries, and to estimate the waste in resources. In order to take into account the economic capacities of the each country, the Gross Domestic Product is introduced in the model. The study concludes to an increasing technical efficiency over the period, driven by an improvement of the health status of many countries.

Chapter 4 is a case study, relying on survey data consisting of a sample of 30 Township Health Centers (THCs) situated in the rural prefecture of Weifang, Shandong province, China. During the study period (2006/12), a major pharmaceutical reform was lead, which aimed at improving the performance of pharmaceutical system. As result of two decades of unhappy combinations of health reforms, in the 2000 decade the Chinese health system was extremely costly, with low levels of health outcomes. It particularly suffered from overprescription and overpricing of medicines, poor medicine quality and excessive use of high-technology medical, induced by perverse incentives of healthcare providers. The reimbursement of a part of health expenditures through the 2003 New Cooperative Medical Scheme, an insurance system, accelerated the rise of costs. Through the implementation of a National Essential Medicine List (NEML) in 2009, and the rebuilding of the pharmaceutical supply chain, the government aimed at improving the efficiency of the system. This Chapter encompasses directly or indirectly each of the ten sources of inefficiency presented in section 1.2.2.
The study uses a two-stage procedure. The first stage is the assessment of the efficiency level of the 30 THCs (primary healthcare facilities) of the sample through a nonparametric frontier analysis. Contrary to what was expected, no efficiency increase is noticed across the period. In average, sampled THCs could produce 30% more with the same level of inputs.

[5] The advantages and drawbacks of each of the two methods are presented in Chapter 2. Broadly, if the health production is unknown, it is more relevant to use nonparametric frontier estimators.

The second stage is the estimation of the determinants of those efficiency scores. Adequation between demand and supply is essential to explain performance, as the most efficient THCs are those in which there are the most people per bed. The share of subsidies in the THC income is negatively associated to THC performance, highlighting a problem of incentives for healthcare providers.

Chapter 5 is another case study, using survey data from primary health care facilities of Ulan-Bator, Mongolia (so-called *Family Health Centers* (FHCs)), from 2011 to 2012. Mongolia is a former Soviet republic, with a tradition of free basic healthcare. The liberalization of the system led to the health sector being dependent on macroeconomic conditions through the amount of available resources. In the 1990's, the State kept on financing the health system, but macroeconomic difficulties emerged in the 2000's, with a relative decrease in public health expenditures, offset by private expenditures. In the context where raising new resources is difficult, efficiency gains are a major path to ensure universal health coverage to the population. This Chapter mainly encompasses sources 5, 8 and 10 of inefficiency in the frame of section 1.2.2.

The study uses a two-stage procedure, to first assess the efficiency scores through a nonparametric frontier approach and explain the scores in a second stage. Efficiency is quite low in the sample, and decreases over the study period. Demand, through the catchment area of each FHC, is positively correlated to efficiency, as well as to supply side variable concerning the composition and remuneration of the staff. The conclusion is that it is necessary to find some accurate incentives mechanisms for the staff, and to understand the knowledge and perceptions of the population relative to FHCs to spur activity of primary healthcare facilities.

Bibliography

Frenk, J. and Murray, C. J. (1999). A WHO framework for health system performance assessment. Technical report, World Health Organization.

Heller, P. (2005). Understanding fiscal space. IMF Policy Discussion Papers 05/4, International Monetary Fund.

Jensen, M. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.

Kruk, M. E. and Freedman, L. P. (2008). Assessing health system performance in developing countries: A review of the literature. *Health Policy*, 85(3):263 – 276.

Liu, X. and Mills, A. (2007). *Public Ends, Private Means: Strategic Purchasing of Health Services*, chapter Agency theory and its Applications in Health Care, pages 151–178. Preker, Alexander, Washington DC, World Bank Edition.

Mathonnat, J. (2010). Disponibilités de ressources financières pour la santé dans les pays d'Afrique subsaharienne. Notes et documents, Agence Française de Développement.

Reinhardt, U. and Cheng, T. (2000). The World Health Report 2000 – Health systems: improving performance. Bulletin of the World Health Organization 78(8): 1064, World Health Organization.

Ridde, V. (2004). L'initiative de Bamako 15 ans apres - un agenda inachevé. HNP discussion paper, World Bank.

Tandon, A. and Cashin, C. (2010). Assessing public expenditure on health from a fiscal space perspective. Working paper, World Bank.

Wagstaff, A. and Doorslaer, E. v. (2003). Catastrophic and impoverishment in paying for health care: with application to Vietnam 1993–1998. *Health Economics*, 12(11):921–933.

WHO (2007). Everybody business : Strengthening health systems to improve health outcomes : WHO's framework for action. Technical report, World Health Organization, Geneva.

WHO (2010). Health systems financing: the path to universal coverage. World Health Report, World Health Organization.

WHO (2016). Global strategy on human resources for health: Workforce 2030. Technical report, World Health Organization.

Xu, K., Evans, D. B., Kawabata, K., Zeramdini, R., Klavus, J., and Murray, C. J. (2003). Household catastrophic health expenditure: a multicountry analysis. *The Lancet*, 362(9378):111 – 117.

Chapter 2

Nonparametric efficiency analyses and environmental variables: the available tools

2.1 Introduction

Since the beginning of the nineties, efficiency has become a crucial stake in Health Economics, both in academic and institutional works (WB, 1996, 1997; Hollingsworth, 2008; Giuffrida and Gravelle, 2001). The concept has been applied both at the microeconomic (the facility: health center, hospital for instance) and macroeconomic level (looking at national health systems). In a context of financial constraint, improvement of health care systems efficiency levels is a major path to reach universal access to healthcare, and two of the three first Sustainable Developement Goals: "no poverty" (considering the risk of impoverishment linked to health expenditures) and "good health and well-being".
The emergence of this literature was fed by the development of new methodological tools that enlarge the possibilities of statistical analysis. In this perspective, this chapter proposes a methodological review of efficiency analyses. The first section focuses on the assessment of efficiency by nonparametric methods. The second one focuses on the introduction of external and environmental variables into those efficiency models.

The concept of technical efficiency Efficiency, in a general meaning, deals with the managerial strategy of a firm, a so-called *Decision Making Unit* (DMU hereafter) concerning the transformation of a certain quantity of resources (or inputs) into a certain quantity of outputs through the production process. DMUs may be many kinds of organization: firms, but also, in this thesis, health facilities. They can also be defined at a regional or national level: a health, or education system, a fiscal entity for instance.
The economic efficiency, that reflects the global managerial performance of DMU

includes two components:

- *technical efficiency.* It focuses on the production function, i.e. on the transformation of a quantity of inputs into a certain quantity of outputs. In an input orientation, an efficient DMU uses the smallest amount of inputs to produce a certain level of output. In a output orientation, it produces the greatest amout of output given their level of inputs. The prices are not taken into account, only quantities.

- *allocative efficiency.* It enlarges the concept of technical efficiency efficiency by taking into consideration the ratio of the prices. In an input orientation, it looks at the optimal combination of inputs (i.e. the combination that minimizes production costs) to produce a given level of output, considering the set of input prices. In an output orientation, it looks at the optimal combination of outputs (i.e. the combination that maximizes the income) for a given level of inputs, considering the set of output prices.

In the empirical analyses of this thesis, it was not possible to have complete and detailed information concerning the prices of each resource and each medical output. Therefore, the focus was on the quantity of delivered health services and resources, aiming at maximizing health facilities activity. The case studies will only estimate technical efficiency, thus this chapter only deals with technical efficiency.

Parametric and nonparametric analyses Empirical models proposed in the literature are divided in two main types: parametric and nonparametric. The most common parametric model is the Stochastic Frontier Analysis (SFA), introduced by the pioneering works of Aigner et al. (1977) and Meeusen and van den Broeck (1977). Parametric models are econometric approaches. They rely on a hypothesis about the functional form between the quantity of inputs and the quantity of outputs.
For instance, let the production function of a firm be composed of two inputs, labor L and capital K and of one output Y. f is the functional form of the relation between a set of inputs (L,K) and a quantity of output Y, i.e. that satisfies $Y = f(L, K)$.
Assuming that f is a Cobb-Douglas function the production function is written as follows:

$$Y = f(L, K) = kL^{\alpha}K^{\beta}; k > 0; 0 \leq \alpha \leq 1; 0 \leq \beta \leq 1$$

The empirical model explains the quantity of output by the quantity of inputs and a residual. If the chosen functional form is a Cobb-Douglas function, the model is the following one:

$$lnY_i = \alpha + \beta_1 lnL_i + \beta_2 lnK_i + \varepsilon_i$$

In the case of a SFA model, the residual has a particular property. It is built so that $\varepsilon_i = v_i - u_i$ with $cov(v_i, u_i) = 0$. v_i is a stochastic component that captures random events which are not under the control of the DMU: "climatic conditions((Aigner et al., 1977)), random equipement failure (Greene, 1993), errors in identifying or measuring explanatory variables (...) or just pure chance." (Jacobs et al., 2006). It is normally distributed. u_i is a non-negative term that represents the distance of the individual i to the production frontier (the maximum output given the quantity of inputs), the maximum feasible. Several distributions are possible (half-normal, truncated normal...). Parametric models therefore rely on the empirical knowledge of the form of the production function. In some fields of application, production functions are well known. In some others, including health economics, the literature does provide any theoretical or empirical knowledge about the functional form of the production function.
The choices of the functional form of the production function and of the distribution of the inefficiency term are not neutral and can impact the results. In this way, it represents a weakness of parametric models.
For those reasons, the case studies of this thesis use only nonparametric models. This chapter aims at presenting the main methodological stakes of nonparametric efficiency models in their empirical implementation. The first part of this chapter deals with the different models developed by the literature. The second part tackles the introduction of environmental variables in nonparametric models, a crucial part of efficiency assessment.

2.2 Nonparametric efficiency models: a first assessment

2.2.1 The reference models

Nonparametric models were first developed by Farrell (1957), through the introduction of the concept of technical efficiency. They were not empirically used until Charnes et al. (1978), introduced Data Envelopement Analysis (DEA) only in the case of constant returns to scale. Banker et al. (1984) extended the model to the variable returns to scale case. The main objective is to study the behavior of DMUs inside a sample and to benchmark them through a scoring process. those methods are only *data-driven* since there is no assumption on the functional form of the production function (nonparametric characteristic). The whole gap between the production frontier (the best possible performance) and a particular DMU is attributed to inefficiency, i.e. to managerial factors. It implies that nonparametric models described in this chapter are *deterministic* (contrary to the *Stochastic* Frontier Analysis).

2.2.1.1 Data Envelopment Analysis

The Data Envelopment Analysisb (DEA thereafter) is the most used nonparametric model in the empirical literature. It consists in resolving a linear optimization program to get the efficiency score for each DMU. The first version (Charnes et al., 1978) only considered constant returns to scale. The next sections rely on Daraio and Simar (2007) and Simar and Wilson (2008) for the presentation of the formalized models.
Let us consider a set of inputs X and a set of outputs Y. The production frontier is defined as follows :

$$\widehat{\Psi}_{DEA} = \{(x, y) \in \Re_{+}^{p+q} | y \leq \sum_{i=1}^{n} \gamma_i y_i ; x \geq \sum_{i=1}^{n} \gamma_i x_i\}$$

There two possible computations of efficiency scores, according to the chosen orientation: the **input orientation** aims at minimizing the inputs for a fixed level of output, while the **output orientation** aims at maximizing the outputs for a fixed level of inputs. The two model are detailed below.

The input orientation The input efficiency score θ_{DEA} of a DMU producing at (x_0, y_0) will be obtained by resolving the following program:

$$\widehat{\theta_{DEA}}(x_0, y_0) = min\{\theta | y_0 \leq \sum_{i=1}^{n} \gamma_i Y_i ; \theta x_0 \leq \sum_{i=0}^{n} \gamma_i X_i ; \theta > 0\}$$

$\widehat{\theta_{DEA}}(x_0, y_0)$ represents the radial distance between (x_0,y_0) and the input target, the input level that the DMU should use, for the same level of output, to be on the production frontier. The interpretation is the following: a DMU exhibiting a score of 0.8 could decrease of 20% its input quantities, while reaching the same level of output.

The output orientation The output efficiency score θ_{DEA} of a DMU producing at (x_0, y_0) will be obtained by resolving the following program:

$$\widehat{\theta_{DEA}}(x_0, y_0) = max\{\lambda | \lambda y_0 \leq \sum_{i=1}^{n} \gamma_i Y_i ; x_0 \geq \sum_{i=1}^{n} \gamma_i X_i ; \lambda > 0\}$$

$\widehat{\theta_{DEA}}(x_0, y_0)$ represents the radial distance between (x_0,y_0) and the output target, the output level that the DMU should reach, for the same level of input, to be on the production frontier. The interpretation is the following: a DMU exhibiting a score of 1.2 could increase of 20% its quantity of output, with the same level of input.
Conditions were later added by Banker et al. (1984) to introduce variable, non increasing or non decreasing returns to scale. They are the following ones:

- for variable returns to scale, $\sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n$
- for non increasing returns to scale, $\sum_{i=1}^{n} \gamma_i \leq 1; \gamma_i \geq 0, i = 1, ..., n$
- for non decreasing returns to scale, $\sum_{i=1}^{n} \gamma_i \geq 1; \gamma_i \geq 0, i = 1, ..., n$

Those conditions allow to decompose inefficiency between technical inefficiency and scale inefficiency, i.e. the fact of not operating at an optimal scale of production

2.2.1.2 Free Disposal Hull

The DEA relies on the convexity assumption of the data hull, of the envelop. On the contrary, the Free Disposal Hull (FDH) only relies on the free disposability assumption of inputs and outputs. If a DMU operates at (x, y), then another operating at (x', y') with $(x' \geq x)$ and $(y' \leq y)$ will also be part of the production set. In other words, it is always possible to produce less with more inputs, to waste resources.
The expression of the production frontier becomes the following one (with the sample $\chi =: \{(X_i, Y_i), i = 1, ..., n)\}$):

$$\widehat{\Psi}_{FDH} = \{(x, y) \in \Re_{+}^{p+q} | y \leq Y_i ; x \geq X_i, (X_i, Y_i) \in \chi\}$$

In this case, there is no possibility to assume any nature of the returns to scale, as the production frontier takes the form of stairs linking every DMU that uses the smallest quantity of inputs, among those which produces at least the same quantity of output in the input orientation (that produces the greatest quantity of outputs among those which use at most the same quantity of inputs). The frontier is the closest possible to the observed points of the sample.
Figure 2.1 gives a graphical representation of DEA and FDH frontiers, in the case of a production function with only one input and one output. The major trend is the same with both of the two techniques, but with some differences. A point, for

instance, is efficient under the FDH estimator, but not with the DEA estimator. This difference is due to the drawing of the production frontier. The DEA makes a linear combination of DMUs, raising the frontier in comparison to the FDH case. The choice between DEA and FDH depends on the convexity assumption about the production set. On the one hand if the true set is convex, both DEA and FDH are consistent, but FDH has a slower rate of convergence. On the other hand if the true set is not convex, the DEA estimator is not consistent (Daraio and Simar, 2007). Empirically, most studies assume a convex set and use a DEA model (Hollingsworth, 2008).

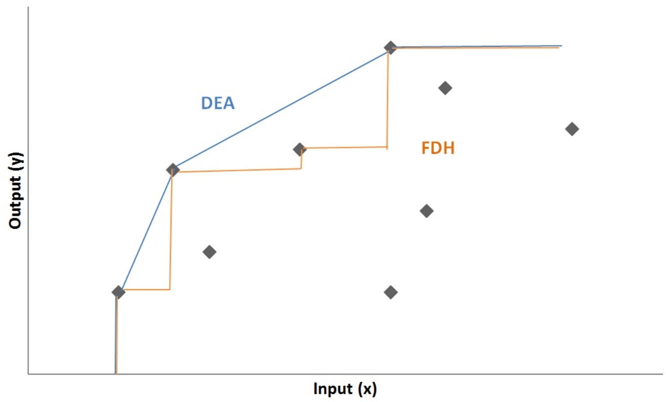


Figure 2.1: DEA and FDH frontiers

2.2.2 Limits of nonparametric models and the robust models

From their data-driven dimension, nonparametric methods suffer from several weaknesses.

2.2.2.1 The weaknesses of nonparametric deterministic measures

The sensitivity to the definition of the production function Jacobs et al. (2006) underlines the importance of the chosen inputs and outputs on the efficiency scores. A way to avoid this potential bias is to test for alternative expressions of the production function, and to see to which extent the scores are affected by thos changes.

The curse of dimensionality Deterministic nonparametric estimators have very slow rates of convergence, especially in compared the stochastic parametric estimators. They require much more observations to obtain relevant estimates of the efficiency score, particularly when the number of inputs or outputs increases (Simar and

Wilson, 2008). To tackle this issue, additional models were proposed to bootstrap both DEA and FDH efficiency scores (Simar and Wilson, 1998). Partial frontier models, developed in section 2.2.2.2, also exhibit some interesting properties in this matter.

The sensitivity to outliers and extreme values As the production frontier is defined by the best performers in the sample, the production hull is necessarily sensitive to the presence of outliers. Outliers are observations that may be subject to a measurement error, or that do not respond to the same production function as the others. They exhibit much better performances than the others. Thus, they cannot be analyzed with the same Data Generating Process, otherwise, in a deterministic model, they would biased all the scores. It would distord the production frontier, and the scores of all their peers would be underestimated.
Several techniques have been developed by the literature to detect those outlying values. Those techniques can either be graphical (with the limit that graphical analysis can only include two dimensions)) or of a jackknife approach (Wilson, 1993, 1995). Zhu (1996) and Zhu (2001) proposes a super-efficiency approaches, for DEA estimates. Finally, after the introduction of partial frontier models by Cazals et al. (2002), Simar (2003) uses the different parameters of the model to detect outliers.
The difficulty of the detection of outlying DMUs comes from the fact that they exhibit multiple dimensions, making the issue more complex.

The introduction of environmental variables The last and major drawback of deterministic nonparametric models is the difficulty to introduce environmental variables in the efficiency analysis. In parametric models, those variables are in the regression, so their role in DMUs performances can be easily identified.
On the contrary, in nonparametric models, the introduction of environmental variables is still a open issue, especially in empirical terms. Section 2.3 discusses the issue into more details.

2.2.2.2 Robust frontier models

Starting from Cazals et al. (2002), several nonparametric models were proposed in the literature, named "robust" or "partial" frontier analysis. Their main characteristic is that the data hull does not envelop all the DMUs of the samples; some of them are left outside the hull. In this way, they cannot distord the frontier, hence the robustness of the frontier and of the scores. They show better statistical properties than the DEA or FDH estimators.

Those models can be read in the light of the probabilistic reformulation of traditional nonparametric models proposed by Cazals et al. (2002) and extended to the multivariate case by Daraio and Simar (2005). The following section presents the probabilistic expression of nonparametric efficiency models, and then more precisely the two main families of robust frontier analyses: order-m and order-α.

Probabilistic expression of nonparametric efficiency models Cazals et al. (2002) proposes a new formulation of the efficiency set of DMUs, and of efficiency scores, using of a probability function. Here, following Daraio and Simar (2007) and Simar and Wilson (2008), the multivariate case of Daraio and Simar (2005) will be developed. It relies on the following probability function:

$$H_{XY}(x, y) = Prob(X \leq x, Y \geq y)$$

It can be interpreted as the probability, for a DMU operating at (x, y), to be dominated by another, i.e. the probability that another DMU produces at least as much output with at most the same quantity of input.
$H_{XY}(x, y)$ can be decomposed as follows:

$$H_{XY}(x, y) = Prob(X \leq x|Y \geq y)Prob(Y \geq y) = F_{X|Y}(x|y)S_Y(y)$$
$$H_{XY}(x, y) = Prob(Y \geq y|X \leq x)Prob(X \leq x) = S_{Y|X}(y|x)F_X(x)$$

where $S_Y(y)$ and $S_{Y|X}(y|x)$ are survivor and conditional survivor functions of Y, and $F_X(x)$ and $F_{X|Y}(x|y)$ are distribution and conditional distribution functions of X. Efficiency scores can be computed from those expression, in both of two possible orientations.
In the input orientation, with $S_Y(y) \geq 0$

$$\tilde{\theta}(x, y) = inf\{\theta|F_{X|Y}(\theta x|y) > 0\} = inf\{\theta|H_{XY}(\theta x, y) > 0\}$$

In the output orientation, with $F_Y(y) \geq 0$

$$\tilde{\lambda}(x, y) = sup\{\lambda|S_{Y|X}(\lambda y|x) > 0\} = sup\{\lambda|H_{XY}(x, \lambda y) > 0\}$$

Simar and Wilson (2008) explicits the meaning of the efficiency score: in a input (output) orientation $\tilde{\theta}(x, y)$ ($\tilde{\lambda}(x, y)$) is the proportion in which a DMU operating at (x, y) can decrease (increase) the quantity of inputs (outputs), output level (input level) fixed, to reach a null probability of being dominated. This expression gives an estimation of the scores identical to the one of the FDH estimator. Indeed, in the input orientation the frontier is the minimum of the support of the conditional distribution function $F_{X|Y}(x|y)$, i.e. the minimal quantity of inputs used by DMUs producing at least y output. This definition is strict and makes the estimator sensitive to the presence of any extreme value. The point of partiel frontier estimators is to soften this condition, to get a robust frontier.

Order-m frontiers The order-m model is introduced by Cazals et al. (2002). Let us consider m independent and identically distributed random variables among the distribution of X, and producing at least the same quantity of output.

The production frontier will be defined as the *expected* minimal achievable quantity of input among those m observations.

The full frontier production $\phi(y)$ can be written as:

$$\phi(y) = \theta(x, y)x = inf\{|F_{X|Y}(x|y) > 0\}$$

With the order-m model, the new frontier is the *expected* lower boundary of the minimum of m random variables $X^1, ..., X^m$ from the distribution function of X and satisfying $Y \geq y$. The definition of the frontier becomes:

$$\phi_m(y) = E(min(X^1, ..., X^m)|Y \geq y) = \int_0^{\infty} (1 - F_{X|Y}(x|y))^m dx$$

with

$$\lim_{m \to \infty} \phi_m(y) = \phi(y)$$

This relation is still valid for the empirical expression of the estimator, $\phi_m(y)$, built with m i.i.d. random variables given by the empirical distribution function of X given that $Y \geq y$, $\widehat{F}_{X|Y}(x|y))^m$.

For one level of output y, and m i.i.d. random variables X_i, $i = 1, ..., m$ from the conditional n-variate distribution function $F_X(\cdot|y)$, the production set is:

$$\Psi_m(y) = \{(x', y') \in \mathbb{R}_+^{N+M} | x' \geq X_i, y' \geq y, i = 1, ..., m\}$$

The definition comes back to the free disposal hull, for m firms producing at least y. The efficiency input score for the given y and the production set $\Psi_m(y)$ is then defined by:

$$\tilde{\theta}_m(x,y) = inf\{\theta | (\theta x, y) \in \Psi_m(y)\}$$

This score is obtain, for a certain set of m values of X_i, the corresponding input score $\tilde{\theta}_m(x,y)$ is computed as such:

$$\tilde{\theta}_m(x,y) = \min_{i=1,\dots,m} \{ \max_{j=1,\dots,p} (\frac{X_i^j}{x^j}) \}$$

From a computational point of view, the final scores are obtained following the following four-step procedure (Daraio and Simar, 2007):

1. For a fixed value of y, a sample of m DMUs is drawn from the sample, among those which satisfy $Y_i \geq y$.
2. Compute $\tilde{\theta}_m^d(x,y)$.
3. Repeat 1 and 2 D times, with a large value of D. [1]
4. Compute the final score as the average of all the preliminary ones.

Order-α frontiers Order-α frontiers are introduced by Aragon et al. (2005) and extended to the multivariate case by Daouia and Simar (2007). Aragon et al. (2005) proposes a nonparametric estimator of the efficient frontier, allowing for some DMUs to be above the frontier ("partial frontier"). But, contrary to the order-m model, it is the proportion α of DMUs remaining above the frontier that will be fixed (and not the number m of DMUs in the production set). In other words, in a input orientation a DMU operating at (x,y) is compared to the input level not exceeded by $(1-\alpha) * 100\%$ of DMUs producing at least y of output.
Supposing that $S_Y(y) > 0$ and for $0 > \alpha \leq 1$, the efficiency frontier is:

$$\phi_\alpha(x,y) = inf\{x | F_{X|Y}(x|y) > 1-\alpha\}$$

So the input efficiency score of the DMU operating at (x,y) will be:

$$\theta_\alpha(x,y) = inf\{\theta | F_{X|Y}(\theta x|y) > 1-\alpha\}$$

One advantage of the order-α method is the economic interpretation of the efficiency score. A DMU exhibiting a score of 0.8, with $\alpha = 90\%$ can reduce by 20 % its quantity of inputs to use the same quantity as 90% of the firms producing a least

[1] Daraio and Simar (2007) proposes D=200 as a reasonable value.

the same level of output. This means that the closer α gets to 1, the closer the order-α estimator gets to the FDH estimator, to a full frontier model.
The estimator of the score is the following:

$$\widehat{\theta}_{\alpha,n}(x,y) = inf\{\theta|\widehat{F}_{X|Y,n}(\theta x|y) > 1-\alpha\}$$

Graphically, the frontiers computed by the different nonparametric estimators are depicted in Figure 2.2.

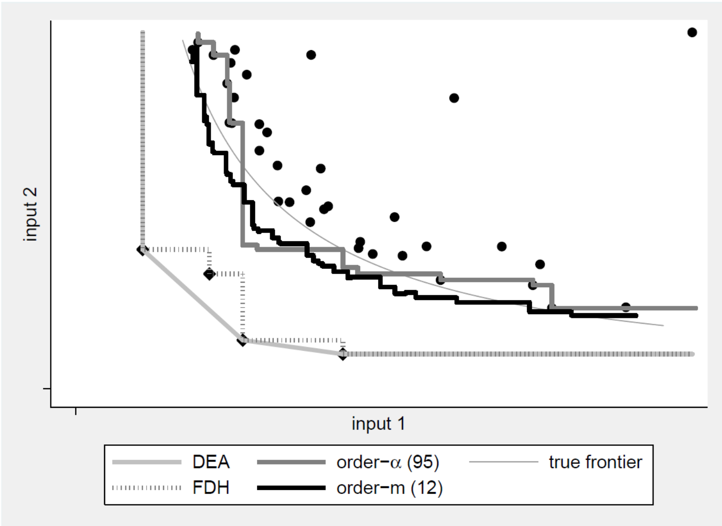


Figure 2.2: Non parametric estimates of production frontier, for a production function of two inputs

Notes: Tauchman, 2012.

Super-efficient DMUs and detection of outliers A crucial point of partial frontier analyses is that they allow for DMUs to be above the production frontier, which means that efficiency scores can be higher than 1, for so-called *super-efficient* units. This property can be used to detect potential outliers. Simar (2003) proposes a method using the order-m analyses, to discriminate super-efficient and outliers. First, not all DMUs with scores higher than 1 are potential outliers. A threshold τ must be defined, and a DMU will be considered as potential outlier if its score is higher than $1+\tau$. Then, the method relies on several elements:

- the "leave-one-out score", i.e. the score of a certain DMU when this same DMU is kept out of the reference set, and this for several values of m, and in both input and output orientation. DMUs whose score remains very high even when m increase are probably outliers. Then, the proportion of DMUs exhibiting a "leave-one-out score" higher than $1+\tau$ is expressed depending upon m, for every value of τ.

- the number of peers for each DMU, i.e. the number of DMU producing at least the same quantity of output in a input orientation

If a DMU exhibits a leave-one-out score higher than $1+\tau$ for several values of τ and both in input and output, then it is likely to be an outlier. Looking at the proportion of high scores acoording to m, any disruption in the curv certainly shows the presence of outliers in the sample.

Which value for m ? Which value for α ? Conversly to the DEA and the FDH estimators, the partial frontier models require to choose some value for the "trimming" parameters, m and α. As α has an easy economic interpretation, the value of α depends upon how "severe" or absolute the frontier is expected to be. It depends on the proportion of DMUs left above the production frontier. On the contrary, the m parameter does not have any direct economic meaning, so the choice of its value may seem more arbitrary. Daraio and Simar (2007) and Simar and Wilson (2008) propose to rely on the proportion of DMUs above the frontier (with a score higher than 1) to choose an value of m. This proportion decreases as m increases. It is possible to select the smallest m value that gives the target proportion of super-efficient DMUs. In the absence of a precise target value, the selected value of m can be a threshold value from which the proportion does not change. Simar and Wilson (2008) underline that most of the time, the value of m does not affect much the proportion.

Order-m or order-α ? Order-m and order-α models give two close but different answers to the same issue. They both have been proved to be bias-robust (Daouia and Ruiz-Gazen, 2006). Then, which one of the two should be preferred? This trade-off is tackled by Daouia and Gijbels (2011), with no absolute answer. Order-m frontiers are less robust to extreme values than order-α while estimating the full frontier (when $\alpha \to 1$ or $m \to \infty$). If the order-α breaks down, then it becomes more sensitive to outliers than the order-m frontier. Also, the order-m estimator exhibits better statistical properties. To compare the two analyses run with both of the two methods, they propose to chose a value $\alpha = (\frac{1}{2})^{\frac{1}{m}}$ for a chosen m value. In the empirical works of this thesis, both of the two methods have been tested and compared. They yield very similar results.

Empirical implementations For those various advantages, robust frontier methods have been implemented in the literature, but not to a very large extent. The following papers can be quoted: Xia et al. (2014) on the energy regulation in China, Roudaut and Vanhems (2012) on ivoirian firms, and Pilyavsky and Staat (2008) on community hospitals in Ukraine.
The assessment of efficiency in the empirical literature mainly relies on "reference" models. As pointed out by Hollingsworth (2008), the DEA method has been used in 80% of the efficiency analyses in the health sector between 1978 and mid-2006. Varabyova and Schreyögg (2013) sum up several of its advantages. Health care is the second sector in terms of number of DEA applications, behind the banking sector, up to 2013 (Liu et al., 2013).

2.3 The introduction of environmental variables in efficiency analyses: first attempts

Efficiency analyses assess performances of DMUs, supposing that the whole distance between a DMU and its projection on the production frontier is managerial inefficiency. But many exogenous factors can impact a performance, from climate to economic circumstances. Taking those factors into account is crucial in terms of policy implications. Introducing some additional variables in a nonparametric efficiency analysis is a major stake in the literature. Contrary to SFA models, where they can be added to the model, environmental or managerial variable do not fit easily into deterministic models. This section aims at summarizing the main possibilities, and their empirical feasability.

2.3.1 One-stage models

The first introduction of environmental variables is proposed by Banker and Morey (1986b). It proposes to distinguish, in a classic DEA model, the inputs and outputs under the control of the manager (discretionary) from the others. The formal model is close to the DEA, with the difference that non-discretionary inputs cannot be minimized, or outputs maximized. This method requires to know the effect of the environmental variable : a variable supposed having a positive impact on efficiency will be introduced as an input. If a DMU exhibits a great quantity of this variable, then it will be more likely to reach a good performance. So, if it exhibits similar discretionary inputs and outputs to another DMU, but a greater value of non-discretionary input, it will get a lower efficiency score. In the case where the environmental variable is not continuous, Banker and Morey (1986a) propose a similar

method to deal with categorical environmental variable.
This kind of method has some limits. First, as shown in Coelli et al. (2005), it requires to know, *ex ante* the effect of the environmental variable on efficiency, which is not always the case. An alternative method is to introduce the environmental variable under the form of a neutral one. In this case, a DMU is only compared with DMUs exhibiting the same characteristics. This allows not to make hypothesis on the impact of the environmental variable, but it reduces to a great extent the number of peers to which a DMU can be compared.
Second, one-stage models suggest a relation of substituability between discretionary and non-discretionary variables, that might not be obvious empirically. They also assume at least the free disposability of external variables, and, in the case of DEA model, the convexity of their relation with the output.
Finally, the number of environmental variables to be introduced in the model is very limited. First, the number of inputs and outputs in a production function has to be low. Second, it is hard to read if there are several variables.

2.3.2 Two-stage models

Two-stage models are the most used methods in the empirical literature to explain efficiency scores. They mostly consist in a traditional DEA model to get efficiency scores. Then, in a second stage, those scores are the dependent variables of an econometric regression, where environmental variables are explicative factors. Different estimators are used at this stage. Most of the literature uses Tobit model to estimate the impact of environmental factors, due to the distribution of efficiency scores: comprised between 0 and 1, and with a large proportion of 1. But some papers stand for linear models (Hoff, 2007), claiming that Tobit is a mis-specification, and does not give direct effects of the potential determinants. They considered that efficiency scores are not *censored* in an economic meaning, and thus that the Tobit model is not appropriate (McDonald, 2009).
Two-stage models are also criticized in a more general way for several reasons (Simar and Wilson, 2013):

- the estimated efficiency scores are biased estimators of the true one;
- the estimated efficiency scores are serially correlated;
- as the quantity of inputs an outputs are correlated to the environmental variables, the residual of the second stage is correlated to the explicative variables

They propose a single and a double bootstrap procedures to cope with those limits (Simar and Wilson, 2007) and improve the statistical properties of the second-stage.

Another answer is given by Ramalho et al. (2010), with McDonald (2009). McDonald (2009) distinguishes two ways of considering efficiency scores in the literature.

- The first one, the so-called *conventionalist* one, looks at the estimated scores as estimates of the true scores. It is advocated by Simar and Wilson (2013), for instance.
- The second school, the *instrumentalist* one, considers scores as "simply descriptive measures of the relative performance of the units in the sample". In this case, many of the issues linked to the second stage disappear.

Ramalho et al. (2010) belong to this second group, and propose an estimator based on the fraction regression model, first proposed by Papke and Wooldridge (1996). Their estimator is used in the empirical analyses of this thesis.

2.3.3 Three-stage model

One of the drawback of two-part models is that efficiency score are rough, they are not purged from the effect of environmental variables. It is possible to know the impact of each variable, but the score does not directly reflect the managerial performance of the DMU. Yet, this information is crucial in terms of policy recommandation, or benchmarking of firms.
In this perspective, Fried et al. (2002) proposes a three-stage procedure that allows to get some purged scores that reflect the managerial performances of DMUs. Three potential determinants of DMUs are considered:

- the managerial performance (the efficiency);
- the environmental components that impacts the global performance;
- the good or bad luck, or omitted variables

The objective is to disantangle the three effects. The procedure is the following one:

1. A "classic" DEA model is run, without any environmental component. "Rough" scores are computed, as well as total slacks (radial and non-radial). Input slacks are considered whether the input orientation is used, output slacks otherwise.
2. Total slacks are regressed on environmental variables using a SFA estimator. This allows to distinguish the effect of environmental variables from the managerial inefficiency and the statistical noise in the performance of DMUs.

From the results of the second stage, adjusted inputs or outputs (according to the chosen orientation) are computed. The adjustment puts every DMUs in the best possible situation in terms of environment and of luck, using the coefficients of Stage 2.

3. A new DEA model is run, using the adjusted inputs or outputs. The new scores are then purged of the impact of environment and luck and only reflect the managerial performance.

Definitive scores are always higher than the rough scores, as DMUs go from their actual situation to the best environment and luck. In their empirical application, the mean score rises from 0.52 to 0.905 from Stage 1 to Stage 3, and the number of efficient DMUs went from 86 to 222 out of 395. This means that luck and environment explains a very large part of DMUs' inefficiency. It also shares the limits of parametric approaches, as the second stage requires assumptions about the functional form of the "inefficiency" term.
Considered in a "conventionalist" way, this method suffers from the same flaws as the typical two-stage approaches, about the estimated nature of the scores that need to be taken into account in each possible second stage, and the correlation issue between regressors and the residual term of the second stage.
The last but crucial limit of two and three stage-models is the assumption of *separability* they make between the inputs and outputs on one part, and the environmental variables on the other part (Simar and Wilson, 2011). In other words, environmental variables do not affect the production frontier itself (the best possible performance), it only affects the distance from one DMU to its projection on the frontier, i.e. its score. This is the strongest weakness of those methods in theoretical terms, because this assumption seems quite unrealistic. It is explicitely made in the model of Simar and Wilson (2007), but it is implicit for the relevance of two-stage approaches.
If this assumption is not satisfied, then Simar and Wilson (2011) advocate for the use of contional efficiency models, extended from Daraio and Simar (2005).

2.3.4 Conditional models

Conditional order-m Conditional efficiency models are part of the order-m efficiency models, proposed by Cazals et al. (2002) and extended to the multivariate case by Daraio and Simar (2005). The only required assumption is the free disposability assumption (see section 2.2.2.2).
For one level of output y, Z the vector of environmental variables and m i.i.d. random variables X_i, $i = 1, ..., m$ from the conditional n-variate distribution function $F_{X|Y,Z}(x|y, z)$, the production set is:

$$\tilde{\Psi}^z_m(y) = \{(x', y') \in \mathbb{R}^{N+M}_+ | x' \geq X_i, y' \geq y, i = 1, ..., m\}$$

The production set then depends on the value of z since $Z = z$ is fixed. The efficiency input score for the given y and production set $\Psi^z_m(y)$ is then defined by:

$$\theta_m(x, y|z) = inf\{\theta | F_{X|Y}(\theta x|y, z) > 0\}$$

with $F_{X|Y} = Prob(X \leq x|Y \geq y, Z = z)$. These score is obtained, for a certain set of m values of X_i, the corresponding input score $\tilde{\theta}_m(x, y)$ is computed as such:

$$\tilde{\theta}_m(x, y|z) = \min_{\{i|yi \geq y, |zi - z| \leq h\}} (\max_{j=1,...,N} (\frac{x^j_i}{x^j}))$$

where h is a bandwith that has to be chosen carefully to "optimize" in a certain sense the estimation of the density of Z" (Daouia and Simar, 2007). The four-step procedure (Daraio and Simar, 2007) used to compute the definitive scores is the same as in the classic order-m model.

Conditional order-α It is also possible to include environmental variables in an order-α model. The conditional input efficiency measure for Z=z is defined, for a given value of y, by:

$$\theta_\alpha(x, y|z) = inf\{\theta | F_{X|Y}(\theta x|y, z) > 1 - \alpha\}$$

The interpretation is the same as in the simple order-α model: $\theta_\alpha(x, y|z)$ is the proportionate input reduction that a DMU operating at (x, y) should implement to reach the input efficient frontier of $\alpha * 100\%$, conditionally to z.

Those conditional methods are more robust and require less hypothesis on the relationship between environmental variables and components of the production function. They have been recently used in several papers including Minviel and Witte (2017); Guerrini et al. (2016); Matousek and Tzeremes (2016). To the best of our knowledge, it has not been used in health literature.

As they take the environmental variables into account in the estimation of the production frontier itself, conditional models require large dataset, to get a sufficient variability in the several dimensions of the model (inputs, outputs, environmental variables). The larger the number of variables in the model, the larger the dataset needs to be, which can be an empirical difficulty.

2.4 Conclusions

The literature about non-parametric efficiency assessment enlarged during the last decades, providing some new statistical tools for empirical analysis. Those tools aim at adressing some unsolved issues in the reference models, DEA and FDH: their sensitivity to the dataset, and the introduction of environmental variables mainly.
In this thesis, three empirical works are realized, using the most accurate estimator possible according to each case.
Chapter 3 considers national health system of a sample of 120 developing countries as DMUs and assess the efficiency of their health expenditures in the production of "good health", proxied by maternal and juvenile survival. To take into account the available resources, and the level of development of the different, the GDP per capita is introduced as a non-discretionary input. To control for the presence of potential outliers, and because of the size of the sample, robust frontier methods are implemented.
Chapter 4 estimates the efficiency of a sample of 30 Township Health Centers (THC) in Weifang, China, and studies its determinants. A two-stage method is then adopted, because of the number of potential determinants of efficiency. Robust frontier methods are used in the first stage again, to deal with outliers and size of the sample.
Finally, Chapter 5 focuses on the efficiency of Family Health Centers (FHC), the first level health facilities in Ulan-Bator, Mongolia. The objective is to understand the determinants of efficiency so a two-stage method is adopted, with robust frontier estimator in the first stage.

Bibliography

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Aragon, Y., Daouia, A., and Thomas-Agnan, C. (2005). Nonparametric frontier estimation: A conditional quantile-based approach. *Econometric Theory*, 21(2):358–389.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9):1078–1092.

Banker, R. D. and Morey, R. C. (1986a). Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research*, 34(4):513–521.

Banker, R. D. and Morey, R. C. (1986b). The use of categorical variables in Data Envelopment Analysis. *Management Science*, 32(12):1613–1627.

Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1):1–25.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of Decision Making Units. *European Journal of Operational Research*, 2(6):429–444.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., and Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. Springer US.

Daouia, A. and Gijbels, I. (2011). Robustness and inference in nonparametric partial frontier modeling. *Journal of Econometrics*, 161(2):147–165.

Daouia, A. and Ruiz-Gazen, A. (2006). Robust nonparametric frontier estimators: Qualitative robustness and influence function. *Statistica Sinica*, 16.

Daouia, A. and Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140(2):375–400.

Daraio, C. and Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis*, 24(1):93–121.

Daraio, C. and Simar, L. (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Springer.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3):253–290.

Fried, H. O., Lovell, C. A. K., Schmidt, S. S., and Yaisawarng, S. (2002). Accounting for environmental effects and statistical noise in Data Envelopment Analysis. *Journal of Productivity Analysis*, 17(1):157–174.

Giuffrida, A. and Gravelle, H. (2001). Measuring performance in primary care: econometric analysis and DEA. *Applied Economics*, 33(2):163–175.

Greene, W. H. (1993). *The measurement of productive efficiency*, chapter The Econometric Approach to Efficiency Analysis, pages 68–119. Oxford University Press, Oxford.

Guerrini, A., Romano, G., Mancuso, F., and Carosi, L. (2016). Identifying the performance drivers of wastewater treatment plants through conditional order-m efficiency analysis. *Utilities Policy*, 42:20 – 31.

Hoff, A. (2007). Second stage DEA: Comparison of approaches for modelling the DEA score. *European Journal of Operational Research*, 181(1):425–435.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, 17(10):1107–1128.

Jacobs, R., Smith, P. C., and Street, A. (2006). *Measuring Efficiency in Health Care: Analytic Techniques and Health Policy*. Cambridge University Press.

Matousek, R. and Tzeremes, N. G. (2016). CEO compensation and bank efficiency: An application of conditional nonparametric frontiers. *European Journal of Operational Research*, 251(1):264 – 273.

McDonald, J. (2009). Using least squares and tobit in second stage DEA efficiency analyses. *European Journal of Operational Research*, 197(2):792 – 798.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2):435–44.

Minviel, J. J. and Witte, K. D. (2017). The influence of public subsidies on farm technical efficiency: A robust conditional nonparametric approach. *European Journal of Operational Research*, 259(3):1112 – 1120.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632.

Pilyavsky, A. and Staat, M. (2008). Efficiency and productivity change in Ukrainian health care. *Journal of Productivity Analysis*, 29(2):143–154.

Ramalho, E., Ramalho, J., and Henriques, P. (2010). Fractional regression models for second stage DEA efficiency analyses. *Journal of Productivity Analysis*, 34(3):239–255.

Roudaut, N. and Vanhems, A. (2012). Explaining firms efficiency in the Ivorian manufacturing sector: a robust nonparametric approach. *Journal of Productivity Analysis*, 37(2):155–169.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.

Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1):49–61.

Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1):31–64.

Simar, L. and Wilson, P. W. (2008). *The Measurement of Productive Efficiency and Productivity Change*, chapter Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. Oxford Scholarship Online.

Simar, L. and Wilson, P. W. (2011). Two-stage DEA: caveat emptor. *Journal of Productivity Analysis*, 36(2):205–218.

Simar, L. and Wilson, P. W. (2013). Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. *Foundations and Trends(R) in Econometrics*, 5(3–4):183–337.

WB (1996). China - issues and options in health financing. Economic Report 15278, World Bank, Washington, DC.

WB (1997). Financing health care: Issues and options for china. China 2020 series, The World Bank, Washington, D.C.

Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business & Economic Statistics*, 11(3):319–323.

Wilson, P. W. (1995). Detecting influential observations in Data Envelopment Analysis. *Journal of Productivity Analysis*, 6(1):27–45.

Xia, X. H., Chen, Y. B., Tasawar, H., Alsaedi, A., and Q., C. G. (2014). Energy regulation in China: Objective selection, potential assessment and responsibility sharing by partial frontier analysis. *Energy Policy*, 66:2014.

Zhu, J. (1996). Robustness of the efficient DMUs in Data Envelopment Analysis. *European Journal of Operational Research*, 90(3):451 – 460.

Zhu, J. (2001). Super-efficiency and DEA sensitivity analysis. *European Journal of Operational Research*, 129(2):443 – 455. A Global View of Industrial Logistics.

Chapter 3

The efficiency of health systems in middle and developing countries

[1]

3.1 Introduction

As said in the introduction of this thesis, the final objective of a health system is to improve the health condition of a population. To this purpose, financial resources are injected in the system, by various agents (households, governments, international institutions, private companies) and under different forms (funding of social security funds or private insurance prepayments, spendings in health infrastructures and staff for instance).

This chapter assesses the global health system of a sample of 120 middle and low income countries (according to the World Bank classification) in order to evaluate the performance of those countries in terms of healh expenditures. In order words, which are the countries that use most efficiently the resources they inject in their health system? Or, for a certain health status, which are the countries that spend the least resources? In a context of financial constraint, especially in developing countries, this issue is crucial.

Nevertheless, this assessment needs to take into account the propoer characteristics of each country, because they are likely to affect the performance of their health system. Those characteristics are various. They include the level of economic human and developement (wealth, level of education), or demographic aspects for instance. In this study, GDP per capita is included in the production function, to control for the level of development in the efficiency assessment.

The production function is made of private and public health expenditures as inputs, GDP per capita as non-discretionary input, and juvenile (under-five) and maternal

[1] This chapter is based on a reseach project on health fiscal space, asked by the WHO and the FDA (Mathonnat J, Petitfour L. et Y. Tapsoba, *Espace budgétaire et dépenses publiques pour la santé - Mesures, évolutions et déterminants*, december 2016, 107 p). It uses and complements the section about efficiency measure of health expenditures, developing the methodological aspects presented in Chapter 2 of this thesis. I remained solely responsible of this analysis presented here.

mortality as outputs. Six production functions are estimated: one for each three-year subperiod from 1997 to 2014, on a sample of 120 low and middle income countries.

The results highlight some large disparities in health systems performances, but with an increasing trend: the average technical efficiency went from 0.632 in 1997/99 to 0.770 in 2012/14. Another result is the higher technical efficiency level of low income countries (mostly sub-saharian countries) compared to the rest of the sample, due to their very low level of health expenditures.

Section 3.2 presents the production function of our model, Section 3.3 the data used for the analysis. Section 3.4 details the empirical analysis, and Section 3.5 gives the results of the study, and Section 3.6 concludes it.

3.2 The production function

3.2.1 Inputs

The literature around the efficiency of health systems often includes sevral kinds of inputs: the number of physicians and inpatient beds per 1000 inhabitants (Retzlaff-Roberts et al., 2004; Ravangard et al., 2014). Yet, physicians and inpatients beds are directy financed by health expenditures. Therefore, there is no relation of substituability between those variable, which is a problem in the perspective of a non-paramatric efficiency analysis. They were thus excluded as inputs from this analysis, to keep only the two types of health expenditures.

Here, the inputs of the production function were chosen to be an exhaustive expression of the resources injected in the health systems. Two inputs were introduced: public health expenditures, and private health expenditures. They are the widest expression possible of the resources injected in the health care system.

Public health expenditures include government (central and local) spendings, external resources and Social Security Funds financing (source: WHO).

Private health expenditures include out-of pocket payments, charitable donations, private insurance spendings and direct service payments by private donations (source: WHO).

It was chosen to consider both private and public health expenditures even though, in this efficiency analysis, the Decision Making Units are governments, i.e. public entities. This reason is justified by three main reasons. First, households spend a significant part of their health expenditures in public facilities. Second, in many countries, the staff of the public facilities has some private activities. Third, public funds finance to some extent (according to the countries) private health facilities. For those reasons, it is not sufficient to consider only public, and not private health expenditures to assess the performances of health systems, as the links are tigh between the two of them. Moreover, private health expenditures can represent a very large part of total health expenditures in developing countries. Introducing them as separate inputs in the efficiency analysis gives a more precise represention of the production function of the firm.

3.2.2 Outputs

The output of a health system is the health status itself, so the chosen variables are indicators of the health status of the population. Three indicators are commonly used in the literature: life expectancy at birth, maternal mortality, and under five mortality.
Life expectancy was used by Alexander et al. (2003); Ravangard et al. (2014); Cheng and Zervopoulos (2014). It is a good proxy of the global health status and condition, but it does not react very quickly to a policy change, for two kinds of reasons.

- The first reason is that life expectancy is a very global and long run indicator, so it takes time for a precise health policy to have a important impact on it. In this study, the different subperiods last only three years, which is certainly too short to see any change. Therefore, if there is a big increase of health expenditures in t, it will not be seen in the life expectancy before *t+1*. In t, a decrease will be observed.

- This slow reaction of life expectancy is also due to the it is estimated nature. It is computed with health indicators related to a past health production process. The life expectancy at birth in t depends upon health policies implemented in *t-1*, so the reaction to a policy will not be immediate

As one of the objectives of the study was to compare the peformances across subperiods, this lagged reaction of life expectancy is a major concern, so it was not included in the production function.

The other indicators that are often used are mortality indicators: mainly maternal mortality (Cheng and Zervopoulos, 2014; Dukhan, 2010) and infantile or juvenile mortality (Cheng and Zervopoulos, 2014; Dukhan, 2010; Retzlaff-Roberts et al., 2004). They have the advantage of being very reactive to a push in health inputs, so both of them were selected as outputs of the model.

An issue emerged from this choice. Mortality ratios have to be minimized, which is not compatible with the characteristics of an output in an effiency model. They have to be transformed to become proper outputs. The main two solution are the inversion of the variable, and, when it is a ratio, its complementary expression. For instance, a maternal mortality rate of x deaths per 100 000 births, becomes a survival rate of (100 000 - x) surviving women per 100 000 births. This last expression is preferred for two reasons. First, it has an "concrete" interpretation, it represents a real quantity, which more relevant to introduce in a production function. Second, the distorsion of the quantity is more important with a inverted variable.
Let's consider the case of a mortality rate, with four different countries. Countries A and B exhibit very low mortality rates, with a difference of 1 between the two. Countries C and D both exhibit high mortality rates, but also with a difference of 1 between the two. The difference will remain 1 with a transformation in a survival rate. On the contrary, with an inversion, the difference between A and B will become greater than between C and D. In economic terms, this is quite counter-intuitive, because it means that an decrease in mortality for countries with poor health conditions will be considered as less valuable than for countries with better health status.

Thus, the two outputs of the health production function of this study are the maternal and the juvenile (under five) survival rate.

3.2.3 Environmental variables

Health systems face very different constraints and situations, that needs to be taken into account in the assessment of their performances. Indeed, there are various determinants to the health status, and multiple channels through which policies can improve (or not) health condition. Dupas and Miguel (2017) underline the role of environmental and infrastructural variables on health. For instance, sanitation infrastructures have a significant on morbidity, and their magnitude mainly depends upon the economic conditions. Several other variables can be considered as determinants of health system performances[2].

3.2.3.1 Education

Education is a key element of the effectiveness of health policies, through the channels of health behaviors, and also of the qualification of the health care providers.

[2]This section only deals with the choice of the variables. Methodological concerns are dealt with in section 3.4.1.1

- Education allows a better access to information, and thus, to knowledge about "healthy behaviors". Thomas et al. (1991) rely on Brazilian DHS data. They highlight the role of the mother education on child height in Brazil and identify that the access to information (through the radio, the media) is the main transmission channel (and not the income). They also find a substitution effect between the education of the mother and the availability of health services. On the contrary, the education of the mother and the availability of sanitation infrastructure are found to be complementary in their effect on the child height.

- Education makes people more adaptable in their behaviors, it improves their ability to learn. Rosenzweig (2004) concludes that investments in health resources have to be implemented jointly with opportunities to learn so that they induce actual changes in health behaviors

It is introduced as an input of the health system production function in several papers. For instance, Retzlaff-Roberts et al. (2004) introduces the school expectancy.

3.2.3.2 Demography

Some demographic variables might constraint the performances of health care systems, through one main channel: the unit cost. Reaching the same objective in terms of health status does not require the same means in a big city and in a rural area. Two variables can be considered in this perspective: the *density of population* and the *urbanization rate*. Two main factors can be highlighted, whose impacts are contrary:

- The accessibility of health care facilities. In remote areas, maintaining health facilities is more expensive than in the cities. In cities, the catchement area is often larger (in number of patients) and allows some economies of scale, and lower unit cost for the same health services. To this difference of unit cost must be added the transportation costs borne by households, much higher in rural areas. In this case, it takes more resources to reach a certain health objective.

- The structure of the health care pyramid. In rural areas, health care facilities are mainly primary facilities. Patients are referred to higher level facilities if needed, for severe pathologies. On the contrary, in rural areas, thanks to the geographical proximity, people can easily bypass the healthcare pyramid and directly go to second or third level facilities, where unit costs are higher. As a consequence, the healthcare system is more expensive in cities., and technological resources, are concentrated in the cities. The unit cost is therefore higher in densely populated areas.

A main preoccupation with those two variables is that there is no certainty about the way they affect the heath production function. This means that it is not possible to introduce it directly in the production function. An input necessarily has to be minimized, and an output to be maximized.

3.2.3.3 Economic conditions

The last environmental input introduced in efficiency analyses is the economic contraint faced by a governement, often approximated by the GDP per capita. Indeed, in the short run the GDP per capita is not under the control of the government. Yet, it directly affects the health system performances through different channels:

- It constraints both public (through the amount of collected taxes, and available public resources) and private (through the effect on the purchasing power) health expenditures.

- It affects indireclty health status through the other environmental variables, especially education and infrastructures.

For instance, Retzlaff-Roberts et al. (2004) introduces the school expectancy, the gini coefficient and the tobacco use as inputs of the health production function.

3.3 The data

3.3.1 The source of the data and the selection of the sample

A large part of the literature around efficiency analyses on health systems focus on developed countries (Afonso and Aubyn, 2005, 2006; Bhat, 2005). Some papers use samples composed of both developing and developed countries (Grosskopf et al., 2006; Gupta and Verhoeven, 2001; Kirigia et al., 2007). Here, the focus is placed on low and middle income countries, which also suffers from waste of resources in their health systems, and face harder financial contraints, and challenges in terms of health status than developed coutries.
The data comes from the World Health Organization. All the observations from 1997 to 2014 were considered. The initial sample included all the countries classified by the World Bank as Low Income, Lowere Middle Income or Upper Middle Income

(135 countries). Then, some countries were excluded, because the information was not available for every variable and every period.

Three year averages were computed from 1997 to 2014: 1997/99, 2000/02, 2003/05, 2006/08, 2009/11 and 2012/14. Indeed, there may be some conjonctural variations in health expenditures, not necessarily linked to the global performance of the health system (mostly, variations in external resources allocated to health that are above the control of governements). Moreover, for some isolated years data was missing. With three year averages, conjonctural variations are smoothed compared to annual data.

Finally, the definitive sample is constituted of 120 countries to be perfectly balanced. The initial list and final one are presented in Tables 3.1, 3.2 and 3.8. The sample in constituted of 26 low income countries (mostly sub-saharian countries, except for Cambodia and Nepal), 48 lower middle income countries, and 46 upper middle income countries.

Table 3.1: Countries of the sample

Country	**WB income group**	**Country**	**WB income group**
Albania	Upper middle	Egypt. Arab Rep.	Lower middle
Algeria	Upper middle	El Salvador	Lower middle
Angola	Upper middle	Eritrea	Low
Armenia	Lower middle	Ethiopia	Low
Azerbaijan	Upper middle	Fiji	Upper middle
Bangladesh	Lower middle	Gabon	Upper middle
Belarus	Upper middle	Gambia. The	Low
Belize	Upper middle	Georgia	Lower middle
Benin	Low	Ghana	Lower middle
Bhutan	Lower middle	Grenada	Upper middle
Bolivia	Lower middle	Guatemala	Lower middle
Bosnia and Herzegovina	Upper middle	Guinea	Low
Botswana	Upper middle	Guinea-Bissau	Low
Brazil	Upper middle	Guyana	Lower middle
Bulgaria	Upper middle	Haiti	Low
Burkina Faso	Low	Honduras	Lower middle
Burundi	Low	India	Lower middle
Cabo Verde	Lower middle	Indonesia	Lower middle
Cambodia	Low	Iran. Islamic Rep.	Upper middle
Cameroon	Lower middle	Jamaica	Upper middle
Central African Republic	Low	Jordan	Upper middle
Chad	Low	Kazakhstan	Upper middle
China	Upper middle	Kenya	Lower middle
Colombia	Upper middle	Kiribati	Lower middle
Comoros	Low	Kyrgyz Republic	Lower middle
Congo. Dem. Rep.	Low	Lao PDR	Lower middle
Congo. Rep.	Lower middle	Lebanon	Upper middle
Costa Rica	Upper middle	Lesotho	Lower middle
Cote d'Ivoire	Lower middle	Liberia	Low
Cuba	Upper middle	Libya	Upper middle
Djibouti	Lower middle	Macedonia. FYR	Upper middle
Dominican Republic	Upper middle	Madagascar	Low
Ecuador	Upper middle	Malawi	Low

Table 3.2: Countries of the final sample, 2

Country	WB income group	Country	WB income group
Malaysia	Upper middle	Sao Tome and Principe	Lower middle
Maldives	Upper middle	Senegal	Lower middle
Mali	Low	Sierra Leone	Low
Mauritania	Lower middle	Solomon Islands	Lower middle
Mauritius	Upper middle	South Africa	Upper middle
Mexico	Upper middle	Sri Lanka	Lower middle
Micronesia. Fed. Sts.	Lower middle	St. Vincent and the Grenadines	Upper middle
Moldova	Lower middle	Sudan	Lower middle
Mongolia	Upper middle	Suriname	Upper middle
Montenegro	Upper middle	Swaziland	Lower middle
Morocco	Lower middle	Syrian Arab Republic	Lower middle
Mozambique	Low	Tajikistan	Lower middle
Myanmar	Lower middle	Tanzania	Low
Namibia	Upper middle	Thailand	Upper middle
Nepal	Low	Togo	Low
Nicaragua	Lower middle	Tonga	Upper middle
Niger	Low	Tunisia	Upper middle
Nigeria	Lower middle	Turkey	Upper middle
Pakistan	Lower middle	Turkmenistan	Upper middle
Panama	Upper middle	Uganda	Low
Papua New Guinea	Lower middle	Ukraine	Lower middle
Paraguay	Upper middle	Uzbekistan	Lower middle
Peru	Upper middle	Vanuatu	Lower middle
Philippines	Lower middle	Viet Nam	Lower middle
Romania	Lower middle	Yemen. Rep.	Lower middle
Rwanda	Low	Zambia	Lower middle
Samoa	Lower middle		

3.3.2 Health status

3.3.2.1 Main evolutions

The evolution of maternal and under five mortality in the sample are depicted in Figures 3.1 and 3.2. They both decreased constantly from 1997/99 to 2012/14, but their distribution remained the same across the subperiods.

The mortality decreased by 35% in average and 31% in median from 1997/99 to 2012/14 (see Table 3.10). In 1997/99, in average 375.9 women over 100 000 died giving birth, in 2012/14 it was 243.8 (198.7 and 136.3 in median).

The distribution of the variable over the sample remained the same across the subperiods: the concentration is the strongest among the 50% lowest maternal mortality rates, and the inter-quartile ratio is stable around 8. Half of the sample is under 200 deaths per 100 000 births, see Table 3.9). The remaining quarter are much more disparate, especially for the highest one, up to a maximal outlying value[3].

Under five mortality also decreased in average and median across the period (see Figure 3.2 and Table 3.10), to a larger extent than maternal mortality (45% in average and median). In average, 83.3 children over 1000 died before their fifth birthday in 1997/99 (32 in median), in 2012/14 this figure fell to 45.9 (33.9 in median).

Contrary to the distribution of maternal mortality, the distribution of under five mortality is concentrated around the median value for every subperiod (i.e. between the 25^{th} and the 75^{th} percentile). As a piece of evidence, the inter-quartile ratio is comprised between 3 and 4 according to subperiods, which is smaller than the same ratio for maternal mortality (ratios computed from Table 3.9).

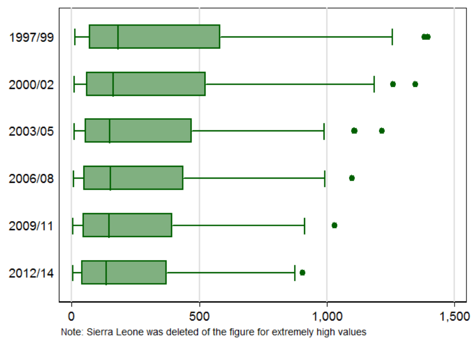


Figure 3.1: Distribution of maternal mortality per 100 000 births

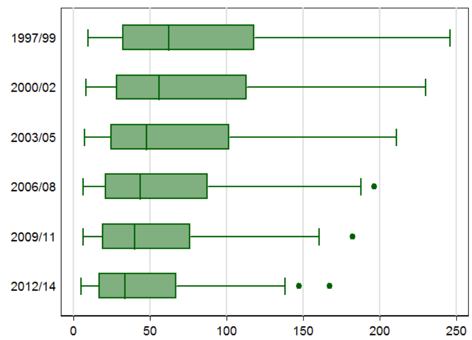
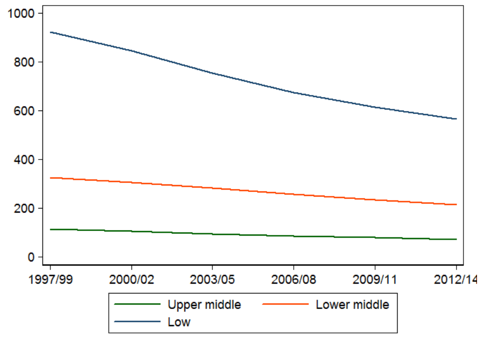


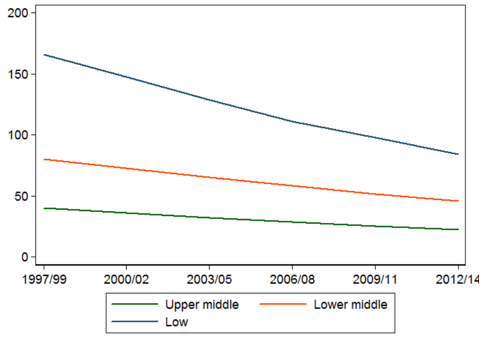
Figure 3.2: Distribution of under-five mortality per 1000 births

The analysis of the evolution of juvenile and maternal mortalities highlights the strong improvement of health conditions in low income countries (Figure 3.3a and 3.3b). It led to a tightening of the gap between this group and the middle income countries of the sample. Maternal mortality fell from 924 to 565 in the first group, from 327 to 216 in lower-middle income countries, and from 40 to 22 in upper-middle countries.

[3]Sierra Leone exhibits extremely high maternal mortality rates for each subperiods, thus it was not included in Figure 3.1 to keep it clear. It decreased constantly across the study period, from 2813 to 1460 deaths per 1000 births, which is still much higher than the rest of the sample. In terms of under five mortality, Sierra Leone is part of the highest values of the sample, but it does not apart from the other countries.



a. Maternal mortality (per 100 000 births)

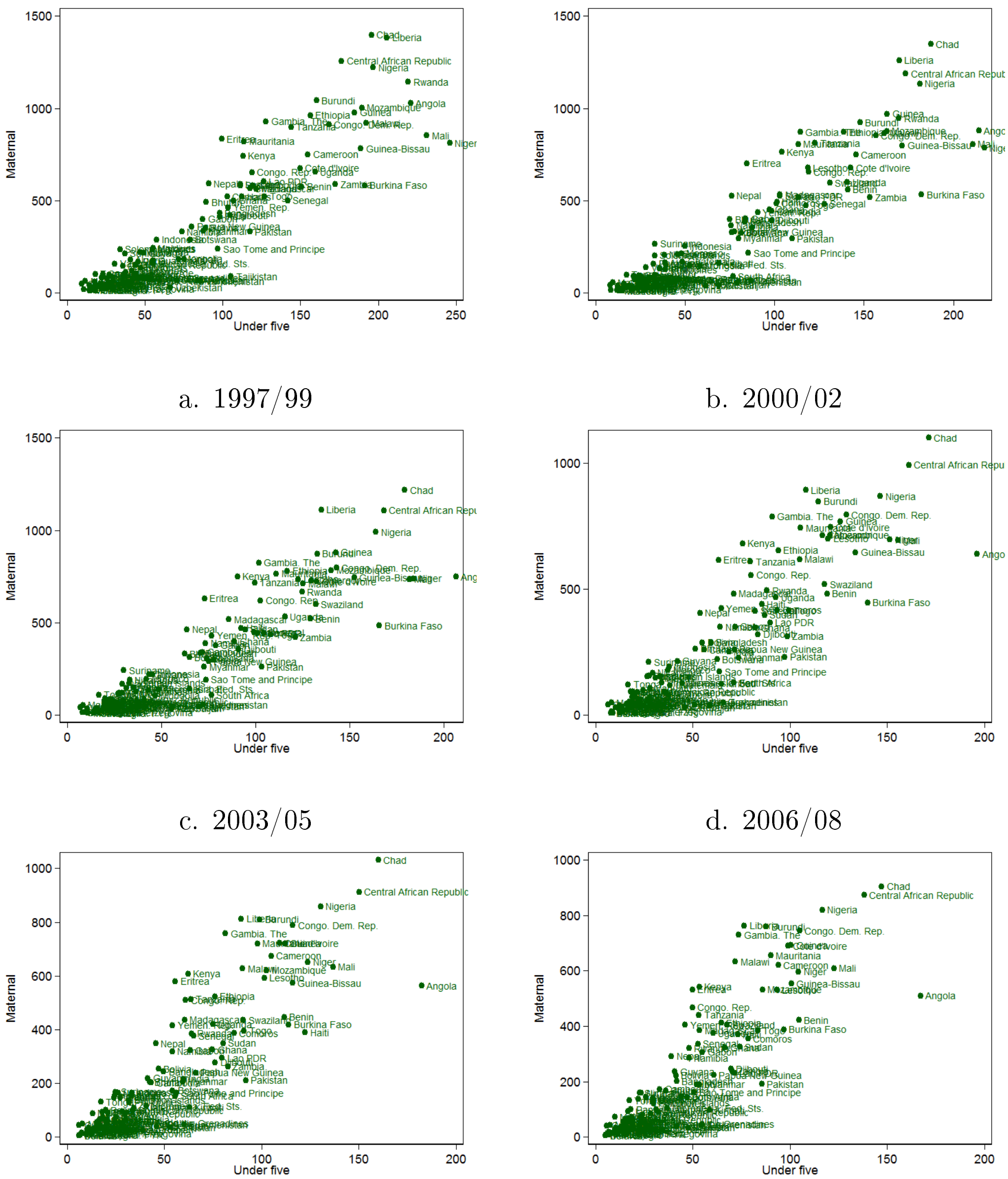


b. Under 5 mortality per 1000 births

Figure 3.3: Evolution of mortality variables per WB income groups (in average)

3.3.2.2 Strong correlation between the two indicators

There is a strong positive correlation between maternal and under-five mortality, for each subperiod (see Figure 3.4). Most of the sample is concentrated in the left-bottom part of the figures (low maternal mortality, low under-five mortality). No country exhibits a low maternal mortality and a high under-five mortality, or the inverse situation, confirming the strong correlation between the two indicators.
A group of countries can be identified with a poorer health status, it is situated in the right-upper part of the figures (high maternal mortality, high under-five mortality): Chad, Central African Republic, Angola. Those countries are all situated in sub-saharian Africa, and for most of they suffered from conflicts during the period.



a. 1997/99
b. 2000/02
c. 2003/05
d. 2006/08
e. 2009/11
f. 2012/14

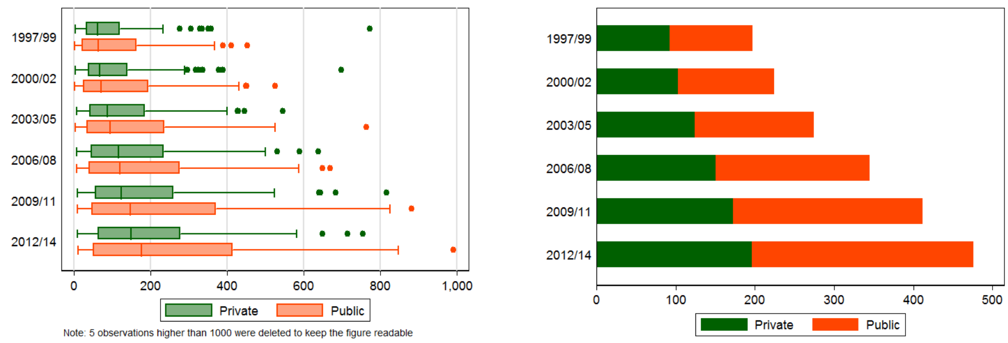
Figure 3.4: Maternal and under five mortality, per period (deaths per 1000 births for maternal mortality, deaths per 100 000 births for under five mortality)

3.3.3 Health expenditures

3.3.3.1 Main evolutions

Total health expenditures doubled from 1997 to 2014 (see Figure 3.5). They sored from 196 to 481 annual 2010 PPP dollars per capita (Table 3.9). This increase

is mainly due to public expenditures. They represent the greater part of health expenditures in the majority of the sample (around 60%), and experienced a stronger increase than private expenditures between the first and the last subperiod (increase of 181 % against 106 in average, and 218 against 137% in median, see Table 3.10). In average, governements of the sample spend 292 annual 2010 PPP dollars per capita in health sector in 2012/2014, while private agents spend 190 annual 2010 PPP dollars per capita.
The two indicators exhibit almost the same distribution for the three first subperiods: more than a half of the subsample spends less than 200 (100 until 2005) dollars per year in health, for private and public expenditures. From 2006 to 2014, public health exenditures sore in some countries (the 75^{th} percentile went from 162.5 in 1997/99 to 331.2 in 2006/08, to 451.0 in 2012/14, see Table 3.9 in Annex). The dispersion increased for public health expenditures[4].



a. Distribution by period b. Composition by period

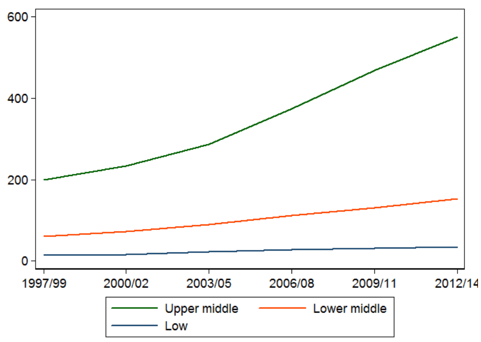
Figure 3.5: Evolution of health expenditures, annual 2010 PPP $

It is also interesting to decompose these evolutions according to the World Bank income classification. Figure 3.6 shows the evolution of health expenditures separately for the three groups. Public expenditures were almost multiplied by 3 in upper middle income countries, from 200 to 600 annual PPP dollars. In lower-middle income countries, it went from 60 to 152 dollars in average, and in low income countries it went from 14 to 35 dollars. This increase seems marginal in absolute terms, particularly compared to the other groups, but in relative terms public health expenditures more than doubled across the period.
The same trend is observed for private health expenditures, there some important differences between the three groups. They rose from 28 to 57 annual PPP dollars

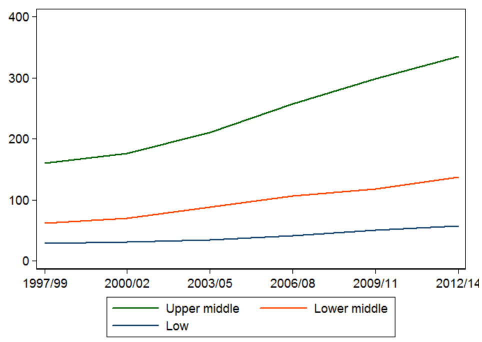
[4] Five values of the public health expenditures are extremely high and were removed from Figure 3.5. They correspond to the following observations: Cuba 200/08, 2009/11, 2012/14 ; Maldives 2012/14; Panama 2012/14.

for low income countries, from 61 to 138 in lower-middle income countries, and from 160 to 335 for upper-middle countries in average. Within the study period, the gap widened between the countries of the sample.

Though the gap widened between the different groups, even in low income countries health expenditures sore in relative terms. Figure 3.7 highlights their evolution for low income countries only, underlining the increasing trend.



a. Public health expenditures



b. Private health expenditures

Figure 3.6: Evolution of health expenditures per capita per WB income group, 2010 PPP dollars

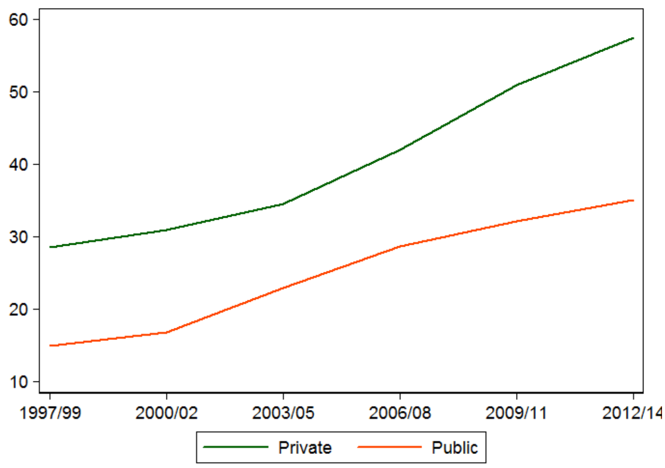
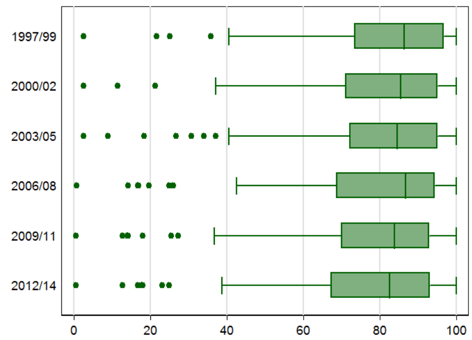


Figure 3.7: Evolution of health expenditures in low income countries

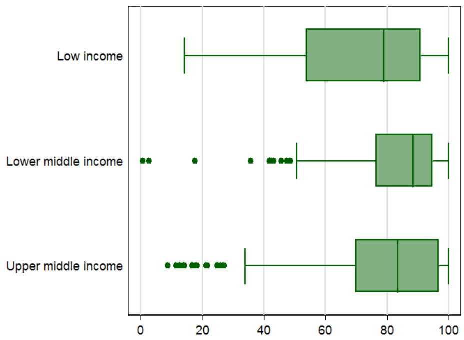
3.3.3.2 Composition of health expenditures

Private expenditures are mostly composed of out-of pocket payments in the sample (the median in over 80%). There in no real evolution of this trend across subperiods

(Figure 3.8a), and no real difference according to the level of development (Figure 3.8b).



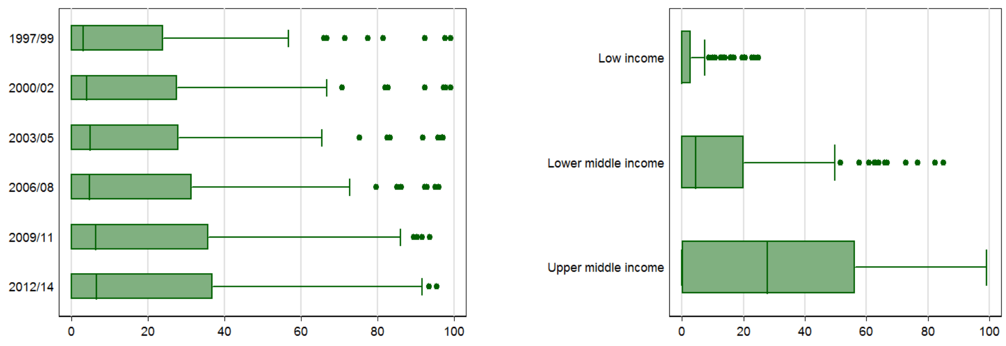
a. By period



b. By WB group

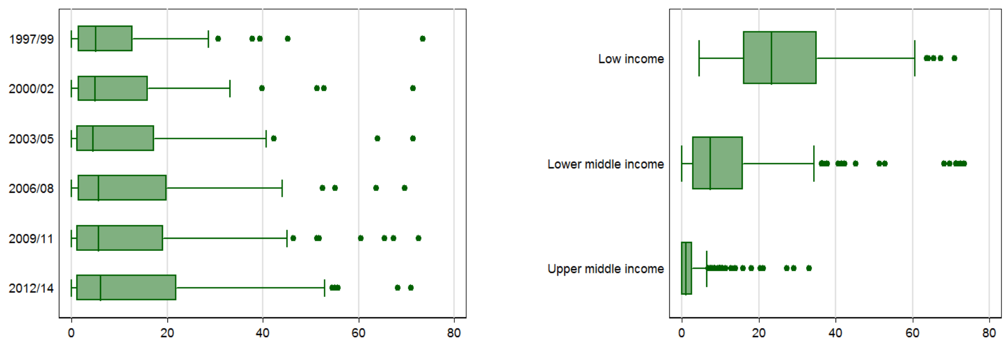
Figure 3.8: Distribution of the proportion of out-of-pocket expenditures in private health expenditures

Public expenditures are partly composed of Social Security financing. Their proportion in Public health expenditures increased over the period from 16.6% in 1997/99 to 21.5% in 2012/14 in the sample (see Figure 3.9a). Contrary to the proportion of out-of-pocket expenditures, their importance in strongly linked to the level of development of the country. Figure 3.9b highlights the difference of distribution of the proportion of Social Security financing in public health expenditures. It is higher for lower-middle, and much higher for upper-middle income countries than for low income countries (in average, respectively 3, 13 and 33 % for low, lower-middle and upper-middle income countries).
Public expenditures are also made of external resources for health (including Official Development Assistance). Their share is stable across the period, but is strongly related to the level of development. The median is 25% for low income countries (with values going to more than 60%), while it is less than 10% for middle income countries. For upper-middle countries especially, the 75^{th} percentile is smaller than 5%.



a. By period b. By WB group

Figure 3.9: Distribution of the share of Social Security Funds health expenditures in public health expenditures



a. By period b. By WB group

Figure 3.10: Distribution of the share of external resources for health in total health expenditures

3.3.3.3 Complementarity or substituability between private and public health expenditures ?

There is a clear common trend between private and public expenditures, that can be seen graphically for each period in Figure 3.11. The Pearson correlation coefficients are significant at the 95^{th} percentile, and positive for every period (Table 3.3). The correlation is stronger for the first subperiods than for the last.
Some countries spend exceptionally high amounts of money on health:

- Lebanon exhibits very high private health expenditures, around 600 $ per capita for each subperiod. This value is much higher than any other country of the sample until 2000/02. Public expenditures are stable across the period, around 400 $ per capita, which is in the highest values of the sample for the first three subperiods.

- The Cuban government spends a lot of money on health. From 2000, it represents the largest public health expenditures in the sample. The gap with the second country widened across the subperiods.

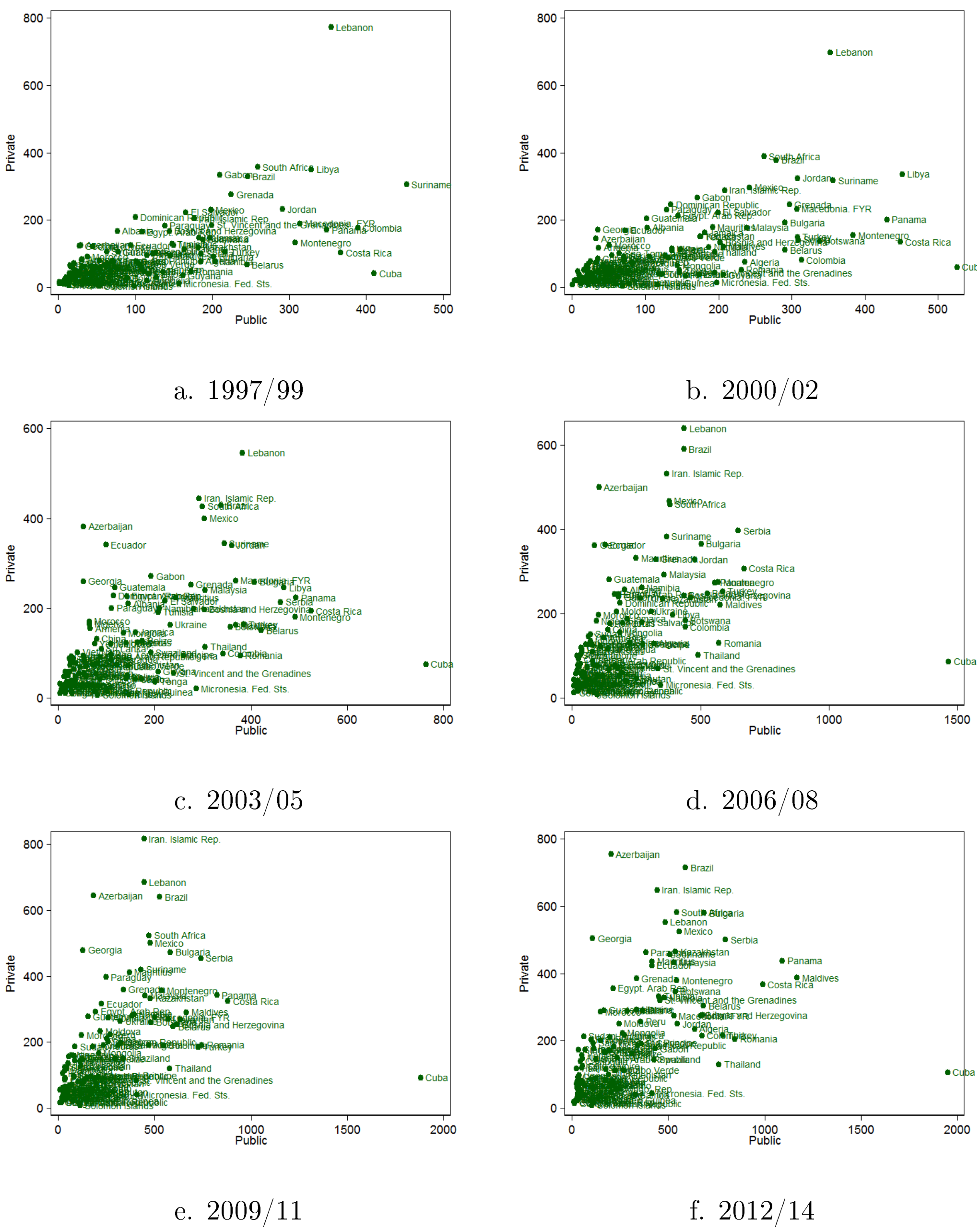
- Other countries increased their private health expenditures in a very large extent compared to the rest of the sample: Azerbaijan, Brazil, Iran, Mexico, South Africa.

Table 3.3: Pearson rank correlation between private and public health expenditures

Subperiod	**97/99**	**00/02**	**03/05**	**06/08**	**09/11**	**12/14**
Pearson corr.	0.648	0.6088	0.5371	0.4926	0.4758	0.5144

The proportion of private health expenditures ranges between 0.4 and 0.5, with a declining trend between subperiods (0.51 in 1997/99, 0.46 in 2012/14). Half of the sample exhibits a proportion of private health expenditures roughly comprised between 0.4 and 0.6 (see Figure 3.12a) for every subperiods. Very low values are to be noticed, while very high values are decreasing across the period the minimal value remains stable, while the maximal values declines).
Looking at the composition of health expenditures according to World Bank income classification, low income countries health expenditures are mostly borne by private agents (three quarter of the group exhibit a proportion of private expenditures greater than 50%, see Figure 3.12b). On the contrary, the median proportion of private health expenditures among total health expenditures is 0.4 for upper-middle income countries. This confirms the idea that the poorest countries, households have to face themselves the financial risks of illness (Dupas, 2011).

Public and private health expenditures are rather complementary in the financing of the health system, from their strong correlation. This correlation is possibly due to the economic conditions of each country, hence the necessity to take those latter into account in the performance assessment.

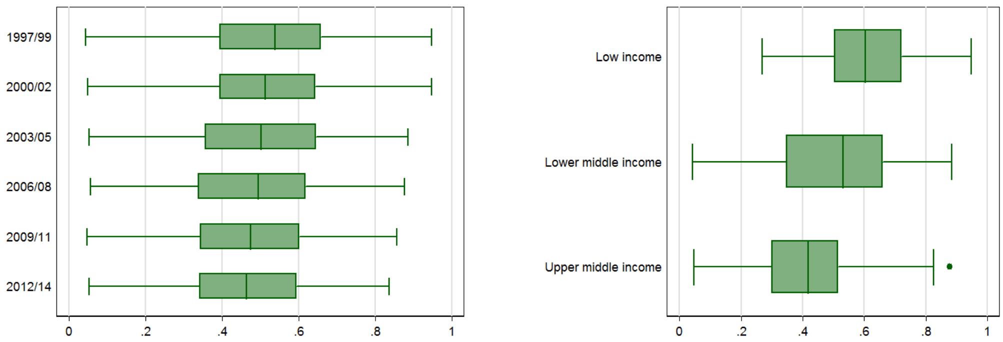


a. 1997/99 b. 2000/02

c. 2003/05 d. 2006/08

e. 2009/11 f. 2012/14

Figure 3.11: Health Expenditures, 2010 PPP $



a. By subperiod b. By WB income group

Figure 3.12: Distribution of the proportion of private health expenditures in total health expenditures

3.3.4 GDP per capita

The average GDP per capita increased from 1997/99 to 2012/14 in the sample from 4627.4 to 7244.6 annual PPP dollars (increase of 57%). The variation rate is even greater for the median (73%) that rose from 3218.0 to 5568.1 annual PPP dollars. The distribution of the variable is given in Figure 3.13[5].
Half the sample is concentrated between 0 and 5000 for every subperiod. But the disparity is very large for the upper half of the sample: the gap between the minimum and the 25t̂h percentile is always smaller than the one between the median and the 75t̂h, and this for the six subperiods. Moreover, the gap between the 75t̂h and the maximum is always bigger than the one between the minimum and the 75t̂h percentile.

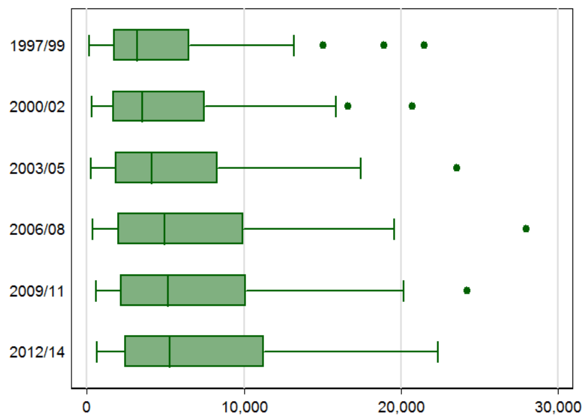


Figure 3.13: Distribution of the GDP per capita across subperiods, annual 2010 PPP $

3.4 Measurement of efficiency

3.4.1 The empirical model

3.4.1.1 How to take the environmental variable into account ?

Environmental variables are always introduced as inputs in the literature about health systems efficiency, implicitely assuming that they are part of the production function.
The objective of this study is to compute efficiency scores that take into account the impact of the GDP per capita. A first solution is to estimate several production

[5]A few values are very high, they correspond to the following observations: Lybia for all periods but 2012/14, Kazakhstan in 2012/14, Malaysia in 2009/11 and 2012/14.

frontiers, so that countries would only be compared to similar peers. The problem is that the GDP per capita is a continuous variable, so the choice of a threshold is completely arbitrary, and that id would impossible to compare the scores obtained in the two subsamples. This solution was rejected.
The most common in the literature is to introduce the environmental variable in the production process itself, by considering them as inputs. Nevertheless, it is necessary to be careful with the implications of this technique. Indeed, in a production function, inputs have to be minimized, and outputs have to be maximized. But environmental variable are not, by definition, under the control of Decision Making Units. So according to the orientation of the model, some issues have to be delt with.
For example, in an output orientation, the idea is to maximize the quantity of outputs with a given level of inputs. In this case, there is no computation problem. The quantity of output produced by country A will be compared to the quantity of output produced by countries spending at least the same amount of money as A in health, and exhibiting at least the same GDP per capita. This makes sense.
But, in the case of an input orientation, the idea is to minimize the quantity of inputs used by DMUs producing at most the same quantity of output. Thus, if two countries exhibit similar levels of output, similar public and private health expenditures, but one has a smaller GDP per capita, its score will be higher than the other one. Yet, the GDP per capita is here considered as exogenous data, so it cannot be minimized. It is introduced as a non-discretionary input.

3.4.1.2 The parameters and assumptions

The choice of the estimator As the functional form of the health production function is unknown, nonparametric estimator was chosen. Because of the size of the sample, and of the potential presence of outliers, partial frontier estimators were used to compute efficiency scores.

The choice of the orientation An input orientation was chosen in order to compare countries with similar health status and economic capacities, and to assess their performance in terms of health expenditures. The score reflects the waste in health resources.

The number of estimated production frontiers In almost 20 years, the environment and the health production functions changed a lot. Thus, it is not reasonable to consider all the observations relatively to the same production frontier. On the contray, there is no reason to think that the health production function changed drastically within three years. One production frontier per subperiod was thus estimated.

The choice of the m parameter In partial frontier model, the value of the m, seen as a "trimming parameter", is crucial. Simar and Wilson (2013) gives a method to choose the optimal value : "The final value of m can be chosen in terms of the desired level of robustness", i.e., of proportion of DMUs above the production frontier. The upper limit is the number of observations.
Here, all possible values from 10 to 110 were tested, with a range of 10. As expected, the proportion of DMUs above the frontier (with a score higher than one) decreases as m increases, and so does the mean efficiency in the sample (see Figure 3.14), for the last subperiod). From m=40, there are less than 10% of super-efficient DMUs in the sample, and from m=50, this proportion is very low and the mean efficiency does not vary (see Figure 3.14). The value of 40 was chosen for the rest of the final analysis , as from m equal to 40 the scores remained identical whatever the chosen value of m.

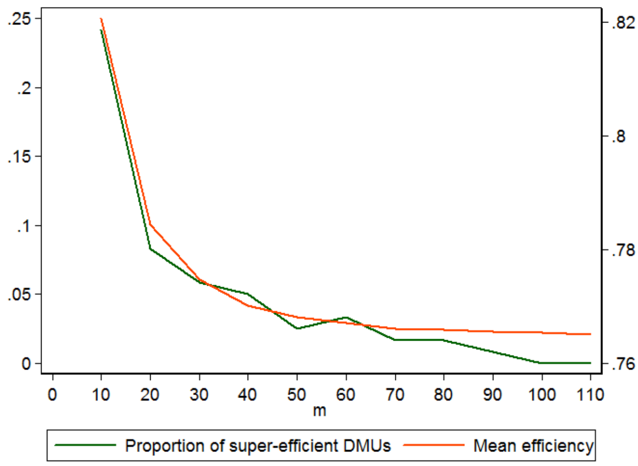


Figure 3.14: Proportion of super-efficient DMUs and mean efficiency according to the value of m

The choice of the *alpha* parameter Running the order-*alpha* analysis, the value of the parameter has to be chosen. The higher the *alpha*, the closer the method is to the Free Disposal Hull. In this study, the proportion of super-efficient DMUs is very low, and even equal to 0 from $\alpha = 5\%$ (see Table 3.4 [6]).

[6] see Section 3.4.3 for the description of the different models

Value of alpha	87	89	91	93	95	97	99
Model 1	0.067	0.042	0.017	0.017	0	0	0
Model 2	0.042	0.025	0	0	0	0	0
Model 3	0.05	0.042	0.025	0.008	0	0	0
Model 4	0.017	0	0	0	0	0	0

Table 3.4: Proportion of super-efficient DMUs according to the model and the value of *alpha*

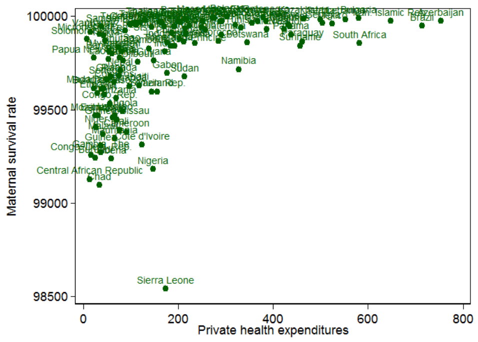
3.4.2 Detection of outliers

Partial frontier analyses give some robustness to the estimation regarding the presence of outliers. Indeed, outliers will exhibit a score higher than 1 so that they are easily detectable and do not distord the production frontier (and bias the scores of the other DMUs).
A first graphical analysis gives an idea of potentially atypical DMUs regarding the relation between the quantity of inputs and outputs. The limit of this analysis is that it can only handle two dimensions at a time. However, it gives an idea of the sample homogeneity. The objective is to detect the potential super-efficient DMUs, that will distord the production frontier because they do not respond to the same production function.
The only atypical countries that can be highlighted in Figure 3.15 are not potential outliers in the efficiency analysis. Indeed, they seem to be partucularly inefficient (high level of input, ou low level of output) so they are not likely to distord the production frontier. Moreover, the robust frontier analyses, both with the order-m and order-*alpha* estimators give very few, or no super-efficient DMUs (countries with a score higher than 1, see Figure 3.14 and Table 3.4). The trends are exactly the same for the six subperiods, leading to the conclusion that the study does not suffer from the presence of any outlier.

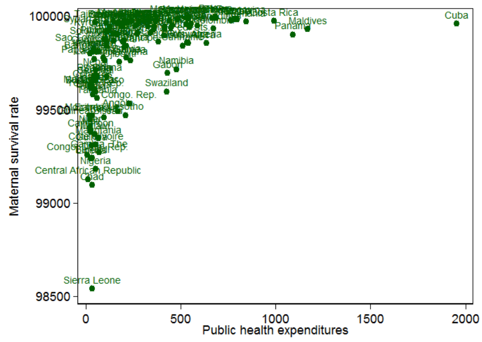
3.4.3 Robustness of the model

As stated in Chapter 2, as they are data-driven, non-parametric efficiency models are sensitive to the definition of the production function (see paragraph 2.2.2.1). To test the validity of the efficiency scores, several alternative scores, computed wih alternative inputs or outputs, were compared to the initial one[7]. All the tested

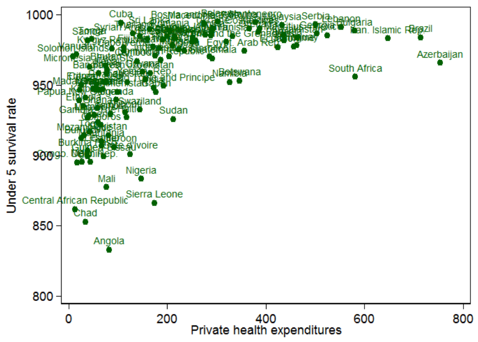
[7]An additional alternate model was tested, including the prevalence of HIV in the population as an output, following Dukhan (2010). The reason to introduce it in that for countries heavily affected, the costs may be a financial burden that other countries don not have to face. It can increase their level of health expenditures and explain a low performance, therefore they are to be compared to countries facing the same situation. In our sample, nine countries suffer from higher HIV prevalence then the other countries (prevalence higher than 10%): Botswana, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, and The Gambia). The introduction



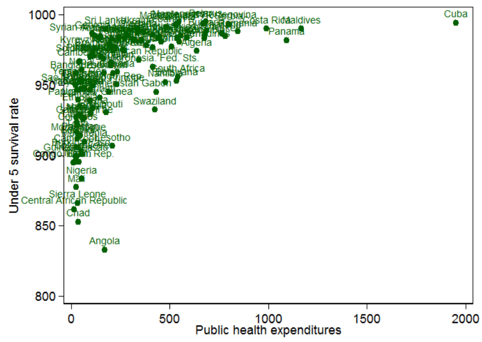
a. Maternal survival rate/Private expenditures



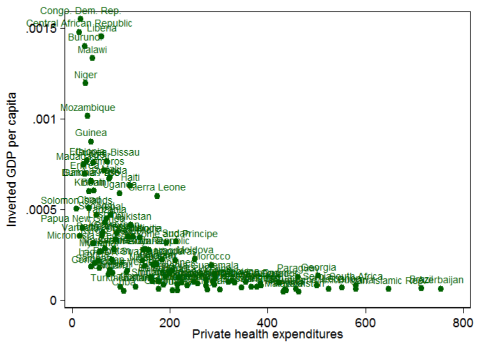
b. Maternal survival rate/Public expenditures



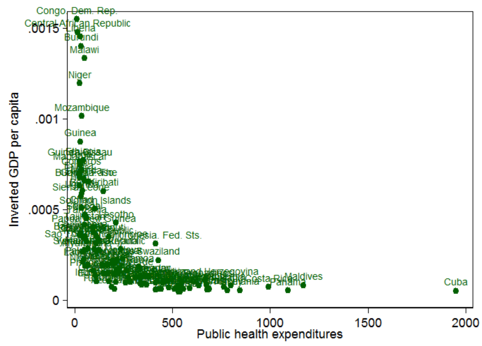
c. Under-five survival rate/Private expenditures



d. Under-five survival rate/Public expenditures



e. Inverted GDP per capita/Private expenditures



f. Inverted GDP per capita/Public expenditures

Figure 3.15: Graphical detection of outliers

models are summed up in Table 3.5.

of HIV prevalence in the production function did not affect their score by more than 0.02. The correlation coefficient of scores obtained with model 1 was 0.94, and the Spearman rank correlation was 0.89. The results remained the same.

Moreover, there are two candidate methods, order-m and order-*alpha* frontier analysis. The scores were thus computed, and compared with both of the two methods.

	Model 1	Model 2	Model 3	Model 4
Inputs				
Public health expenditures	$\surd$	$\surd$	$\surd$	
Private health expenditures	$\surd$	$\surd$	$\surd$	
Total health expenditures				$\surd$
GDP per capita	$\surd$	$\surd$	$\surd$	$\surd$
Outputs				
Maternal survival	$\surd$		$\surd$	$\surd$
Juvenile survival	$\surd$	$\surd$		$\surd$

Table 3.5: Tested models for the production function

In this perspective, efficiency scores were computed for four production functions, with order-m and order-*alpha* estimations. They were compared both through a Pearson correlation matrix (Table 3.11), and through a matrix of Spearman rank correlation (Table 3.12). Correlations are here given for the last subperiod only, but lead to the same conclusions for the other ones.

The analysis of the Pearson correlations confirms the validity of our results, as the coefficients are very high. Indeed, between scores computed with various production functions but a same method, they range between 0.657 and 0.893 for order-m estimations, and 0.651 and 0.895 for order-*alpha* estimations. The lowest correlation is obtained between Models 2 and 3, when only juvenile and maternal mortality are successively taken into account. This suggests that the relevance of including both of the two indicators in the production function. The correlation is the highest between Models 1 and 4, when the inputs are modified (public and private separately, of total health expenditures).

In terms of rank correlations, the same conclusion can be drawn. When the method is fixed and only the production function changes, the Spearman rank correlation ranges from 0.658 to 0.870 for order-m scores, and 0.674 to 0.903 for order-*alpha* scores. Similarly to the Pearson coefficients, the correlation is the weakest between Models were only juvenile and maternal mortality are introduced as outputs. The highest correlation is also found between Models 1 and 4.

It is also to be noticed that the scores estimated with the same production function but with a different method are always highly correlated. The Pearson correlation coefficient are all 0.999, and the Spearman rank correlations ranges between 0.982 (Model 1) and 0.998 (Model 2). Therefore, the choice of one method over the other does not impact the scores at all.

Order-m scores were used for the rest of the analysis, considering the better statistical properties of this estimator (see paragraph 2.2.2.2).

3.5 Efficiency scores

3.5.1 Evolution of efficiency between 1997 and 2014

Technical efficiency increased progressively across the subperiods (Table 3.6), from 0.632 to 0.770 in average. This increase is mainly explained by the evolution of the lowest scores of the sample. In 1997/99 and 2000/02, the 25^{th} percentile were 0.316 and 0.388, which is extremely low. It means that a quarter of the sample (30 countries) could, for the same level of maternal and juvenile mortality, and for the same GDP per capita, decrease their health expenditures by 60% at least. On the opposite side of the distribution, 40 countries are on the production frontier (Table 3.7). There is no country in the sample that uses less financial resources for the same level of maternal and juvenile mortality and for the same GDP per capita. The disparity in terms of performances is very strong in the sample.
The gap tightened in the last subperiods: the median increased to 0.987 in 2012/14. Almost half of the 120 countries are on the production frontier in the last subperiod. The 25^{th} percentile soared to 0.518, meaning that there is a smaller number of very inefficient countries. The average efficiency score is went to 0.770, indicating that in average, in 2012/14, middle and low income countries wasted almost a third of their health expenditures, considered their health status and their economic situations.

period	mean	N	min	max	p25	median	p75	sd
1997/99	0.632	120	0.133	1	0.316	0.582	1	0.321
2000/02	0.683	120	0.116	1	0.388	0.694	1	0.301
2003/05	0.725	120	0.119	1	0.472	0.815	1	0.291
2006/08	0.744	120	0.124	1	0.488	0.927	1	0.288
2009/11	0.753	120	0.160	1	0.485	0.946	1	0.286
2012/14	0.770	120	0.169	1	0.518	0.987	1	0.274

Table 3.6: Mean descriptives statistics about efficiency scores

Table 3.7: Mean descriptives statistics about efficiency scores, inefficient countries only

period	mean	N	min	max	p25	median	p75	sd
1997/99	0.447	80	0.133	0.976	0.281	0.401	0.582	**0.228**
2000/02	0.499	76	0.116	0.994	0.342	0.456	0.675	0.224
2003/05	0.528	70	0.119	0.992	0.376	0.499	0.690	0.227
2006/08	0.521	64	0.124	0.997	0.348	0.490	0.667	0.218
2009/11	0.551	66	0.160	0.999	0.365	0.512	0.754	0.241
2012/14	0.555	62	0.169	0.999	0.361	0.526	0.732	0.221

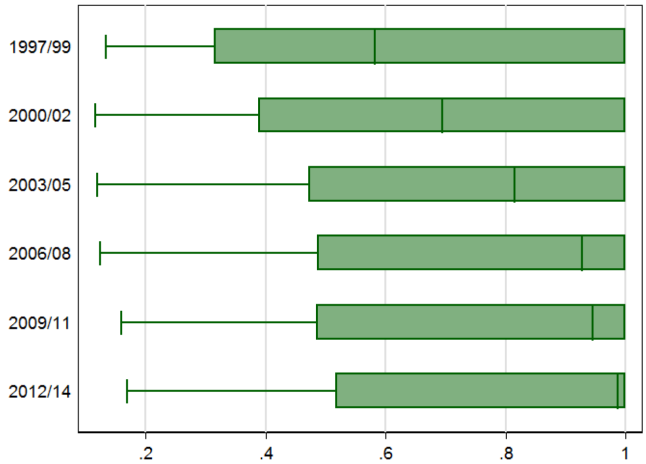


Figure 3.16: Evolution of the distribution of efficiency scores

3.5.2 Different country profiles

It is important to check whether there are some significant differences of performances between the three World Bank income groups. Tables 3.15 give some information about this issue. There does not seem to be any difference on the first subperiods, but in the last ones the gap has widened (0.898 for low income countries, 0.769 for lower-middle incomes and 0.698 for upper-middle incomes). This might be related to the unequal increase in health expenditures, that makes some countries very inefficient compared to other ones.

The difference of mean between low income countries and the rest of the sample is statiscally significant at a 95^{th} level of confidence, and for every subperiod but the first one. The difference between upper-middle income countries and the rest of the sample is significant for the last three subperiods, at a 95^{th} level of confidence also.

Those results are not very surprising, given the very important gaps in terms of health expenditures, compared to the differences in mortality.

3.5.3 Potential limits to the model

Several explanations can be given to this fact. The first one is that the difference of performance decreased among the countries of the sample across the study period. But there is also a risk of lack of comparability between the different countries. Indeed, in an input orientation, countries are compared to other countries exhibiting at least the same output, i.e., in this analysis, at most the same juvenile and maternal mortalities, and the same GDP per capita. Yet, according to the distibution of the two mortality variables, there are few countries exhibiting very low mortality rates, so there are few peers to compare them.

Another limit of the model may rely in those mortality variables. Indeed, after a certain threshold of economic development, there are not many differences between countries with very different level of developement in terms of maternal and juvenile mortality. Countries experience an Inequalities in health status may be better expressed by the prevalence of cardiovascular disease or diabetes for example. It is very likely that mortality variables are not sufficient to capture health inequalities for countries that reached a certain level of development and well-being!;

3.6 Conclusion

As a conclusion, from 1997 to 2014 the gap in terms of health system performances tightened in average, between the least and the most efficient countries. But this global evolution hides various evolutions according to countries level of development. Low income countries lowered their mortality ratios to an important extent, while their health expenditures remained very low in absolute terms. Their efficiency scores therefore sore across the period. On the contrary, the mortality ratios of upper-middle income countries very slightly decreased as they already were low for most of them, while their health expenditures, both private and public, skyrocketed. The evolution of their efficiency scores is thus less impressive.

From those conclusions, several points can be drawn. First, there is a need for a better definition of a health production function to get relevant efficiency assessment. The first assumption of any efficiency assessment is that the definition function includes all the resources and the productions of all the Decisions Making Units. Health systems certainly aim at producing some "health", or avoiding the production of bad health. But the measurement of this health is very complicated, as shown in section 3.2.2. But, more important, the transformation of health resources into "health" is subjected to many external elements that entirely under the control of governments, or public authorities. Many studies tried to deal with this issue, here

we include the GDP per capita in the production function as a proxy of the level of developement.
But the relation between the external variables and the performances are rarely well known, so the assessment of the performance does have some limits.
A response to this issue is the focus not on the production of "health", which is the ultimate objective of a health system, but on the delivery of health services, as an intermediate output. In this case, the Decision Making Unit is not a country anymore, but a health facility, whose production function is easier to establish. This kind of analysis is the core of Chapters 4 and 5, focusing on survey data from Towhship Health Centers in Weifang, China and on Family Health Centers in Ulan-Bator, Mongolia.

Bibliography

Afonso, A. and Aubyn, M. S. (2005). Non-parametric Approaches to Education and Health Expenditure Efficiency in OECD Countries. *Journal of Applied Economics*, 8(2):227–246.

Afonso, A. and Aubyn, M. S. (2006). Relative Efficiency of Health Provision: a DEA Approach with Non-discretionary Inputs. Working Papers Department of Economics 2006/33, ISEG - School of Economics and Management, Department of Economics, University of Lisbon.

Alexander, C. A., Busch, G., and Stringer, K. (2003). Implementing and interpreting a Data Envelopment Analysis model to assess the efficiency of health systems in developing countries. *IMA Journal of Management Mathematics*, 14(1):49.

Bhat, V. N. (2005). Institutional arrangements and efficiency of health care delivery systems. *The European Journal of Health Economics*, 6(3):215–222.

Cheng, G. and Zervopoulos, P. D. (2014). Estimating the technical efficiency of health care systems: A cross-country comparison using the directional distance function. *European Journal of Operational Research*, 238(3):899 – 910.

Dukhan, Y. (2010). *Améliorer l'efficience des systèmes de santé et la protection financière contre le risque maladie dans les pays en développement*. PhD thesis, CERDI, Université d'Auvergne.

Dupas, P. (2011). Health behavior in developing countries. *Annual Review of Economics*, 3(1):425–449.

Dupas, P. and Miguel, E. (2017). Chapter 1 - Impacts and determinants of health levels in low-income countries. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Economic Field Experiments*, volume 2 of *Handbook of Economic Field Experiments*, pages 3 – 93. North-Holland.

Grosskopf, S., Self, S., and Zaim, O. (2006). Estimating the efficiency of the system of healthcare financing in achieving better health. *Applied Economics*, 38(13):1477–1488.

Gupta, S. and Verhoeven, M. (2001). The efficiency of government expenditure: experiences from africa. *Journal of Policy Modeling*, 23(4):433 – 467.

Jacobs, R., Smith, P. C., and Street, A. (2006). *Measuring Efficiency in Health Care: Analytic Techniques and Health Policy.* Cambridge University Press.

Kirigia, J. M., Asbu, E. Z., Greene, W., and Emrouznejad, A. (2007). Technical efficiency, efficiency change, technical progress and productivity growth in the national health systems of continental African countries. *Eastern Africa Social Science Research Review*, 23(2):19–40.

Ravangard, R., Hatam, N., Teimourizad, A., and Jafari, A. (2014). Factors affecting the technical efficiency of health systems: A case study of Economic Cooperation Organization (ECO) countries. *International Journal of Health Policy Management*, 3(2):63–69.

Retzlaff-Roberts, D., Chang, C. F., and Rubin, R. M. (2004). Technical efficiency in the use of health care resources: a comparison of OECD countries. *Health Policy*, 69:55–72.

Rosenzweig, M. R. (2004). Why Are There Returns to Schooling? *The American Economic Review*, 85(2):153–158.

Simar, L. and Wilson, P. W. (2013). Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. *Foundations and Trends(R) in Econometrics*, 5(3–4):183–337.

Thomas, D., Strauss, J., and Henriques, M.-H. (1991). How Does Mother's Education Affect Child Height? *Journal of Human Resources*, 26(2):183–211.

3.7 Appendix

Table 3.8: Countries of the initial sample but excluded in the final one

Country	**WB income group**
American Samoa	Upper middle
Dominica	Upper middle
Korea. Dem. Rep.	Low
Kosovo	Lower middle
Marshall Islands	Upper middle
Palau	Upper middle
Somalia	Low
South Sudan	Low
St. Lucia	Upper middle
Tuvalu	Upper middle
West Bank and Gaza	Lower middle
Iraq	Upper middle
Afghanistan	Low
Timor-Leste	Lower middle
Zimbabwe	Low

Table 3.9: General statistics about main indicators

	Mean	Minimum	Maximum	p25	p50	p75	Std. Dev.
Private health expenditures							
1997/99	92.321	4.049	772.401	32.734	60.967	120.035	99.708
2006/08	146.100	4.293	638.799	43.082	101.361	222.954	132.195
2012/14	189.884	4.800	754.118	59.856	144.935	277.151	165.267
Public health expenditures							
1997/99	104.245	0.973	452.359	22.198	63.742	162.506	102.673
2006/08	224.411	6.025	2350.042	40.484	134.185	331.234	289.122
2012/14	292.962	10.710	1951.648	52.771	202.592	451.004	299.719
Maternal mortality rate							
1997/99	375.942	12.333	2813.333	72.833	198.667	582.667	418.354
2006/08	293.137	8.000	1803.333	50.667	154.333	467.000	311.640
2012/14	243.775	4.000	1460.000	46.000	136.333	385.000	263.173
Under 5 mortality rate							
1997/99	83.298	9.400	246.033	32.000	62.033	118.250	61.804
2006/08	58.878	6.433	196.300	21.100	39.700	88.233	45.262
2012/14	45.893	5.000	167.167	17.300	33.867	67.600	35.631
GDP per capita							
1997/99	4627.401	150.592	21499.850	1726.555	3217.991	6514.668	3902.079
2006/08	6474.620	341.345	27951.970	2100.965	4975.489	9976.538	5126.777
2012/14	7274.579	646.035	22355.590	2593.007	5568.144	10997.060	5490.948
Urbanization rate							
1997/99	42.061	7.828	85.705	26.883	41.500	56.223	18.521
2006/08	46.114	9.866	87.900	29.907	45.917	62.104	19.426
2012/14	48.405	11.474	87.548	32.753	48.428	63.893	19.656

Expenditures and GDP per capita in 2010 PPP $, maternal mortality per 100 000 births, under 5 mortality per 1000 births, Urbanization in percentage

Table 3.10: Variation of the main indicators

	Mean			Median			Std Dev			Rel std dev.		
Variable	**97/99**	**Var(%)**	**12/14**	**97/99**	**Var (%)**	**12/14**	**97/99**	**Var(%)**	**12/14**	**97/99**	**Var(%)**	**12/14**
Priv. health exp.	92.32		189.88	60.97		144.94	99.71		165.27	1.08		0.87
		105.68			137.73			65.75			-19.41	
Pub. health exp.	104.24		292.96	63.74		202.59	102.67		299.72	0.98		1.02
		181.03			217.83			191.92			3.87	
Maternal mort.	375.94		243.78	198.67		136.33	418.35		263.17	1.11		1.08
		-35.16			-31.38			-37.09			-2.99	
Under 5 mort.	83.30		45.89	62.03		33.87	61.80		35.63	0.74		0.78
		-44.91			-45.41			-42.35			4.64	
GDP per cap.	4627.40		7274.58	3217.99		5568.14	3902.08		5490.95	0.84		0.75
		57.21			73.03			40.72			-10.49	

Expenditures and GDP per capita in 2010 PPP \$, maternal mortality per 100 000 births, under 5 mortality per 1000 births, Urbanization in percentage

Table 3.11: Correlation matrix between the scores of the different models for the 2012/14 subperiod

	M1	M2	M3	M4	A1	A2	A3	A4
M1	1.000							
M2	0.778	1.000						
M3	0.867	0.657	1.000					
M4	0.893	0.760	0.798	1.000				
A1	0.999	0.779	0.864	0.893	1.000			
A2	0.771	0.999	0.648	0.757	0.773	1.000		
A3	0.869	0.659	0.999	0.802	0.868	0.651	1.000	
A4	0.894	0.763	0.798	0.999	0.895	0.762	0.803	1.000

Note: M is for order-m (m=40), and A for order-*alpha* (*alpha* = 99%).
The figures refer to those in Table 3.5. All the correlations
are significantly different from 0 with a 95% degree of confidence.

Table 3.12: Spearman rank correlation matrix between the scores of the different models for the 2012/2014 subperiod

	M1	M2	M3	M4	A1	A2	A3	A4
M1	1.000							
M2	0.727	1.000						
M3	0.870	0.658	1.000					
M4	0.862	0.758	0.779	1.000				
A1	0.982	0.790	0.869	0.899	1.000			
A2	0.719	0.998	0.652	0.759	0.786	1.000		
A3	0.870	0.679	0.995	0.802	0.877	0.674	1.000	
A4	0.858	0.767	0.780	0.997	0.903	0.769	0.804	1.000

Note: M is for order-m (m=40), and A for order-*alpha* (*alpha* = 99%).
The figures refer to those in Table 3.5. All the p-values
for the bilateral independence tests are 0.000.

Table 3.13: Mean descriptives statistics about efficiency scores, low income countries only

period	mean	N	min	max	p25	median	p75	sd
1997/99	0.621	26	0.222	1	0.333	0.582	1	0.293
2000/02	0.793	26	0.116	1	0.656	0.947	1	0.273
2003/05	0.887	26	0.124	1	0.884	1.000	1	0.224
2006/08	0.894	26	0.336	1	0.935	1.000	1	0.205
2009/11	0.891	26	0.285	1	0.821	1.000	1	0.200
2012/14	0.898	26	0.361	1	0.845	1.000	1	0.194

Table 3.14: Mean descriptives statistics about efficiency scores, lower middle income countries only

period	mean	N	min	max	p25	median	p75	sd
1997/99	0.645	48	0.133	1	0.275	0.766	1	0.346
2000/02	0.671	48	0.187	1	0.377	0.646	1	0.296
2003/05	0.688	48	0.119	1	0.454	0.624	1	0.279
2006/08	0.730	48	0.124	1	0.498	0.749	1	0.271
2009/11	0.753	48	0.201	1	0.521	0.900	1	0.272
2012/14	0.769	48	0.169	1	0.507	0.848	1	0.260

Table 3.15: Mean descriptives statistics about efficiency scores, upper middle income countries only

period	mean	N	min	max	p25	median	p75	sd
1997/99	0.624	46.000	0.134	1	0.348	0.482	1	0.316
2000/02	0.633	46.000	0.138	1	0.366	0.587	1	0.310
2003/05	0.671	46.000	0.149	1	0.432	0.711	1	0.308
2006/08	0.675	46.000	0.181	1	0.369	0.706	1	0.318
2009/11	0.674	46.000	0.160	1	0.355	0.753	1	0.317
2012/14	0.698	46.000	0.196	1	0.351	0.744	1	0.304

CHAPTER 4

Activity and efficiency of Township Health Centers in Weifang, China

[1]

This chapter focuses on the efficiency of a sample of 30 Township Health Centers in Weifang, China, relying on survey data from 2006 to 2012. During this period, China implemented a large pharmaceutical reform, which is the main institutionel frame of this case study. Section 4.1 presents the major evolutions of the Chinese health system since the comunist era. Section 4.2 is a literature review around health care facilities efficiency, and their determinants. Section 4.3 details the content of the collected database, and gives information about the activity of the THC of the sample. Section 4.4 deals with the first stage of the efficiency analysis, i.e. the computation of the scores, and Section 4.5 presents the second stage, the estimation of the determinants of efficiency.

[1] This chapter comes from a research program on the effects of the essential drug policy on a sample of Township Health Centers of the prefecture of Weifang, Shandong province, under the joint supervision of Pr. J. Mathonnat (CERDI) and Pr. A. Ma, Head of Weifang Medical University. It benefited from funds from the National Research Agency through the program "*Investissement d'Avenir*" (ANR-10-LABEX-14-10-01) by the CERDI and from financial support from the FERDI (*Fondation pour les Études et les Recherches sur le Développement International*), as well as from the Bureau of Social Affairs of the French Embassy in Beijing. This program was conducted in collaboration with a team from the Health Bureau of Weifang and from searchers of the Weifang Medical University, and Pr. N. Chen, Deputy Director General, Department Policies and Legislation, National Health and Family Planning Commission in Beijing. A first version of this study was at the core of a research report in French and Chinese, whom I am a co-author (Audibert M, Huangfu X., Mathonnat J, Petitfour L., *Politique Nationale de Médicaments Essentiels et activités, financement et efficience d'un échantillon d'hôpitaux municipaux de la préfecture de Weifang*, FERDI, december 2015, 85 p). I was involved in its presentation and discussion with the Health Bureau from Weifang and with the National Health and Family Planning Commission in Beijing. The study also lead to a Working Paper of the FERDI (Petitfour L, Huangfu X, Audibert M, Mathonnat J., *Efficiency of township hospitals in China in the context of the drug policy reform: Progress should not get bogged in midstream - A case study from a survey in Weifang prefecture*, Working Papers Serie, 2017). An article, for which I am the corresponding author has been resubmitted to *International Journal of Health Economics and Management*. I remained solely responsible of this analysis presented here.

4.1 Background

Section 4.1 provides information about the historical context of the Chinese Health sytem, from the comunist era to the 2009 pharmaceutical reform.

4.1.1 The dislocation of the communautary health system

4.1.1.1 The Chinese health system under the communist regim (1960-1978)

The Chinese Health system exhibits numerous specificities, partly inherited from the communist era. Indeed, primary health care was a corner stone of the Maoist health system. In urban areas, households were covered by the Free Medical Service program –financed by governments- if civil servants, and by the Labor Medical Service program –financed by employers -if industrial employees. This last system was first proposed to state-owned enterprises workers, and then enlarged to collective-owned enterprises workers. In both of the two systems, households benefitted from free primary health services (Dong, 2009).
In rural areas, the health insurance took the form of the Cooperative Medical Scheme (CMS), and was integrated, at the commune level (the working unit), to the collective agricultural system. It relied on its pyramidal structure, as a so-called three tier structure (Liu and Hsiao, 1995). The first tier was the Village Health Station (VHS) at the brigade level, composed of barefoot doctors, members of the commune who had received a short training to provide preventive and basic curative healthcare (Bloom and Xingyuan, 1997). Barefoot doctors worked part-time as health providers, part-time as peasants, and they were paid for their two activities by the commune. If needed, they referred the patients to the Commune Health Center (second tier) where they would be healed by physical assistants. Then, for serious illnesses, patients were referred to the County Hospital (third tier), staffed with fully qualified physicians (Zhu et al., 1989). Commune Health Centers also had the responsibility of managing and training the barefoot doctors of their area.

This pyramidal system ensured the accessibility of health facilities (basic and sophisticated), and a control of health costs at the same time. Unit costs were higher at the County Hospital, so it xas more expensive for the CMS finances. As barefoot doctors were part of the commune, they did not have interest to refer unnecessarily their patients to County Hospitals. Thus only the most serious cases were referred. There were three main sources of financing of the CMS: individual premiums, commune welfare funds (agricultural incomes were collectivized and a part was dedicated to the CMS financing), and subsidies from upper governments. Communes could choose the organization of their CMS (provider payment, package benefit). In most of them it provided peasants with free basic healthcare, preventive services and other services, and reimbursement of drugs at the brigade level or in the reference system (Zhu et al., 1989).

The CMS system was politically and financially supported by the government, especially during the Cultural Revolution, so the percentage of communes with CMS rose in the 1960 decade from 32% in 1960 to 46 in 1962, and 80 in 1968. At the end of the system, in 1976, 90% of Chinese rural communes had a CMS. In terms of health status, it means that 90% of the Chinese rural population had access to immunization and basic healthcare, which was the cornerstone of a global improvement of Chinese health status.

This improvement appears in the evolution of life expectancy at birth (figure 4.1). In 1960, a Chinese expected to live 43 years at his birth, almost the same as an Indian citizen for instance, while the Indian GDP per capita is 50% higher than the Chinese one (192 international $ in China against 307 in India). In 1970, the Chinese life expectancy soared to 58 years, and reached 65 years in 1978, while the Indian life expectancy was only 53 years. The infant mortality also fell over this period. In 1949 it was estimated to be 200 per thousand live births, in 1982 it was 20.3 (Young, 1988). This improvement was realized despite of the degradation of the economic conditions during the Cultural Revolution. The GDP per capita remained very low in the 1960 decade and even decreased in the first part of the decade (Figure 4.2).

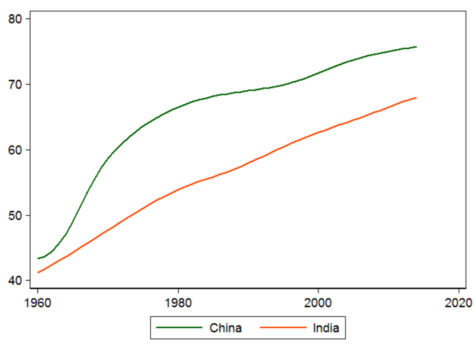


Figure 4.1: Evolution of life expectancy at birth in China and India

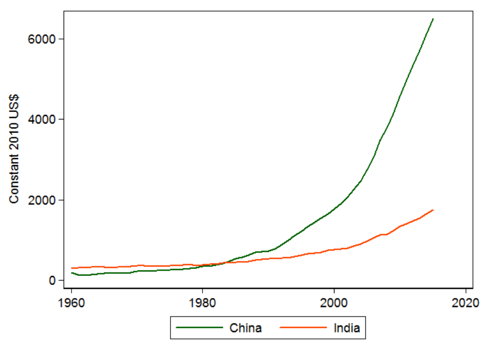


Figure 4.2: Evolution of GDP per capita in China and India

4.1.1.2 The liberalisation of the Chinese Health system (1980's)

The reforms and their direct impacts After the death of Mao in 1978, the important reforms of the system enhanced the end of the CMS, due to "a combination of financial, political and management problems" (Xueshan et al., 1995).

First, the system of communes as working units was replaced by the "household production responsibility system". In the communes (now called townships), each household was attributed a piece of land to cultivate, and sold the production to the government. The financial losses or benefits were individually assumed, so the CMS lost its major financing source, the commune welfare funds. This also means that barefoot doctors could not be compensated anymore for the time they work as health workers. They became full-time peasants (as agriculture became more income-generating) or full time doctors (through an upgrade of their skills) (Liu et al., 1995).

Second, political reasons also to the collapse of the CMS as they were associated to the Cultural Revolution era, which was completely rejected after Mao's death. The new regimen did not stand for the CMS system despite the improvements it allowed in access to health care and the Chinese rural health status. As a result, in 1986 only 4.8% of the villages had a CMS (while 87.8% still had a Village Health Stations).

The collapse of the CMS had some negative consequences on the access to healthcare in rural areas. As welfare funds disappeared, health institutions shifted to a fee-for-services system, with large out-of-pockets payments (Xueshan et al., 1995). Therefore the financial burden went from collective to individual.

The management model of rural health facilities also changed. The CMS system relied on the three-tier structure, a compelling referral system that aimed at limiting health expenses while ensuring an almost universal access to healthcare (see Section 4.1.1.1). But in the 1980 decade the referral pyramid was broken, as a large part of VHS and Township Health Centers (THCs, former Commune Health Centers) were sold (VHS to former barefoot doctors that upgraded their qualifications for example) or submitted to contract with private providers. In 1990, less than a half of VHSs were owned by the village (Xueshan et al., 1995).

As a result, they became independent institutions, with great financial autonomy. For the VHS, it means that their financial outcomes depended greatly upon the economic conditions of the areas as they rely mostly on fee-for-services payments. In poor areas, many VHSs disappeared for financial reasons. In richer areas, barefoot doctors continue working in the stations, but with fewer control from township or county authority in the health costs and follow-up of patients. In most villages, preventive activities were neglected because they were not profitable. On the contrary, they focused on curative services and drug selling, on which they could take a margin. In some areas rich enough to raise some funds for this purpose, a substitute to CMS was created. It generally implemented some incentives mechanisms to increase preventive follow-up (with a capitation amount dedicated to prevention for instance, or a reimbursement of antenatal visits) and prevent over prescription using co-payments(Young, 1988). This widened the gap between rich and poor areas. In their survey in 1989, Henderson et al. (1995) find a health insurance coverage of almost 50% in the most developed province of their sample, Jiangsu, and of 11.7% in the least developed, Guizhou.

As for THCs, as they were newly financially independent, they had to substitute former subsidies with new sources of revenue, i.e. activity incomes. They had to compete for patients with the County Hospitals (Liu et al., 1994), since there were no financial incentives for patients to visit them anymore. As households preferred to pay more but for better services, they directly went to County Hospitals, where the staff was more qualified and the medical equipment more sophisticated. The improvement of economic conditions, thanks to the rise of the mean GDP per capita, fed this new demand for high quality medical care, and the global inflation of healthcare expenses. The pressure of the demand in County Hospitals and urban facilities became really strong, whereas attracting patients was difficult for many THCs.

The nature of THC activities also changes as they had to spur their incomes. The price of the preventive and basic curative services was fixed by the government, to keep access to basic healthcare affordable for households. But the consequence was that they were not profitable for THCs, which shifted their activity towards more sophisticated acts, on which they could earn money (Liu et al., 2000) As they were also authorized a 15% mark-up on drug sale, they also started overprescribing drugs of patients.

Therefore, the gap widened between rich areas where THCs could afford to invest in equipment and propose some high quality services, and those in poorer areas which suffered from the lack of revenue. They could not invest, or even maintain their existing equipment, so the quality of their care deteriorated (Bloom and Xingyuan, 1997). Between 1980 and 1990, the number of THCs in China decreased from 55413 to 47749 (-14%). The decrease is more important for collectively owned THCs (-17%) than for state-owned THCs (-8%) (Xueshan et al., 1995). The case study of Tang, 1997 in the county of Donglan (Southwestern China), illustrates all those difficulties, and how THCs failed to accomplish the missions they were attributed by the Ministry of Health (MoH), especially in preventive and training activities.

Wider consequences on Chinese health system One major evolution of the Chinese health system is the cost escalation of health expenditures in the 1980 and 1990 decade that made health insurance organizations implode. **?** divide the factors into four categories: "increase in the number of covered people; inflation factor (...); the ageing of the covered population; residual factors". Inflation includes the price of the inputs, and of the fee itself. Disentangling the effects of each of those factors on the cost increase, they find that from 1985 to 1989, 74 % of the increase of social health insurance costs are due to inflation and residual factors.
Those "residuals factors" are the "increase of demand, use of new technology and drugs, and improvement in general quality of services". They can be considered from the supply and the demand side of health care. On the supply side, recently autonomous health facilities replaced governments subsidies with activity incomes and drug selling benefits, and tended to overprescribe expensive drugs and services (such as injections or high-tech exams). They transferred the financial load from the public authorities to health insurance for insured households, or to uninsured households.

On the demand side, patients asked for better quality of service, which implied an increase of expenses in inputs from the health facility. Moreover, the phenomenon of induced demand, regarding the (excessive) prescription of high-tech services, also caused some more expenses, that may have been unnecessary. Patients expected their doctor to give them antibiotic injection for instance, and would have had the impression of low quality care otherwise (Liu et al., 2000).

Increasing risk of impoverishment due to health expenses Chinese households progressively had to take in charge the financial burden of health costs escalation through out-of-pockets payments.

In the 1980's, the inequalities were very important between the rural and the urban areas. In the latter, most people benefitted from work-based insurance, so that they did not bear alone the costs of their health cares. On the contrary, in rural areas access to health care could be much more complicated when local facilities had closed or provided a poor quality of services, with few preventive activities. They could go to the VHC or County Hospital, but had to pay themselves for the services and the drugs, which was a potential source of impoverishment. In 1986, in rural areas the infant mortality rate was 27.3 per thousand live births, while 14.3 in urban areas (Young, 1988).

Nevertheless, the gap between urban and rural tightened in the 1990 decade, as a consequence of economic reforms in urban areas that jeopardized the insurance coverage of many Chinese employees. Du (2009) links the insurance coverage to the open door policy (progressive opening to international competition in the 1980 and 1990 decades), to the reforms of state-owned enterprises and the agricultural reform of the 1980 decade. Faced with competition and the sharp decrease of subsidies (Hu et al., 1999), many state-owned enterprises exhibited huge financial losses, and had to cut in their workforce and lower their funding of fringe benefits (for example, diminish the coverage or the benefit package).

Du (2009) highlights varieties in the health insurance coverage rate: in 1991 it is 11% in rural group, and 65% in urban ones. However, a strong decrease in the urban health coverage (44% in 2000) suggests that the situation regarding health expenses was difficult for both urban and rural groups at the end of the 1990 decade. Inequalities are mainly linked to working status: only 8% of insured unemployed households, and 6% for employees of foreign and/or private companies. This last figure can be explained by an overestimation due to the fact that some people lose their job and insurance but are still administratively registered as employees.

The figures of Du (2009) include full and partial coverage of health expenses, while Hu et al. (1999) makes the distinction between the two. From survey data on enterprise and individual workers in 1992, they found that more than a half of their sample benefitted from a partial coverage only. No significant difference is found between the workers with no coverage at all and those with partial coverage on the incidence of out-of-pocket expenditures, even though a significant difference between no coverage and a partial coverage is found on the mean amount of payments. Of course, this study doesn't include unemployed people, who mostly have no health insurance at all.

Because of those multiple factors, out-of-pockets payments became the first source of health expenditures in China. They represented 20% of total health spending in 1978, and progressively soared to 60% in 2009).

4.1.2 Health reforms in the 2000's

4.1.2.1 The NRCMS and its limits

The implementation of the NRCMS At the end of the 1990 decade, while the GDP per capita had been multiplied by more than four since 1980, health expenses were still a financial burden for many households. In the case of a serious illness or accident, the expenditures could be catastrophic and lead the household into poverty. Facing a growing discontent, the government launched a large health insurance scheme, the Basic Social Medical Insurance (BSMI) system. It includes a scheme for urban employees introduced in 1998, another for urban unemployed and a last for rural residents, the New Rural Cooperative Medical Scheme, implemented in 2003 (Dong, 2009). Though inspired by the CMS (it is managed at the county level, with counties managing their benefit package), it differs from it in several ways; for instance, it is voluntary-based, and not mandatory. It is also the first time that central and local governments take part in the financing of the rural health care (Dong, 2009; You and Kobayashi, 2009), through premiums paid to NCMS pools in Western and central China. Its implementation was spectacular. From 2003 to 2010, the enrollment rate rose from less than 10% up to more than 90% (WB, 2010; Yip et al., 2012). The copayments were higher in high level facilities, in order to rebuild the pyramidal reference system. An increase of THCs activity was expected, due to new patients: households that couldn't afford health care before the reform, and households that directly went to County Hospitals when sick. It was actually observed (Wagstaff and Yu, 2007; Pélissier et al., 2012). Milcent and Wu (2015) also find a positive impact of the NRCMS on the subjective health status, using a longitudinal household survey, that includes two channels: an "insurance effect" of the coverage (between insured and uninsured households, in counties where NRCMS is available), and a "general equilibrium effect" affecting every resident of a NRCMS county. But the NRCMS did not have its intended impact on the impoverishment due to health reasons. Yang and Wu (2015) emphasizes the increase of outpatient cost per case between 2004 (pre-enrollment in their study area) and 2009.

Limits of NRCMS in reducing impoverishment The NRCMS was not sufficient, in the way it was implemented, to reduce substantially the impoverishment induced by health expenses. There are several reasons to this unexpected result.

First, when insured, people increased their health expenditures, including out-of pockets payments (Wagstaff and Yu, 2007). Those who didn't seek health care at all now did, and those who already sought for care increased their expenses to get better quality of care (Lindelow, 2005). This is a positive sign in terms of universal access to health care but also needs to be addressed by a relevant incentive policy to control for health unit costs.

Second, the NRCMS was focused on avoiding catastrophic expenditures and preventing households from falling or sinking deeper into poverty because of health

shocks. Therefore, inpatient cares (hospitalizations, medical acts and drugs) are the most reimbursed, while outpatients expenditures are mainly left to out-of-pockets payments. Yet, at the beginning of 21rst century, most of Chinese provinces had achieved their epidemiological transition. The ageing of the population, and the change of lifestyles due to urbanization and the improvement of economic conditions (tobacco use, access to cheap rich food, lack of physical activity) for most of the population induced major changes in the main causes of mortality and morbidity that changed a lot within the last decades. Yang et al. (2008) found that while communicable diseases and maternal and perinatal conditions accounted for almost 30% of deaths in China in 1973, the figure decreased to less than 10% in 2005. On the contrary, cerebo-cardiovascular disease and cancers explained about 25% of deaths in 1973 and more than 50% in 2005. Chronic diseases (hypertension, diabetes) soared in the Chinese population, but their diagnosis and treatment are highly imperfect (Wang et al., 2005; Lei et al., 2012).
Sun et al. (2009), through a household survey in Ningxia and Shandong found that the enrollment in the NRCMS was not protective enough against catastrophic expenses linked to chronic diseases. In each of the two provinces, around 15% of NCMS members spent more than 40% of their annual non-food expenditures in treatment for chronic diseases. This irrelevance of the benefit package regarding the epidemiological profile of NCMS members is also shown by Yip and Hsiao, 2009. They compare the performance, in terms of impoverishment, of a typical benefit package of the NRCMS (high copayments on outpatient care and drugs, high deductibles and ceilings for inpatient care) with another they imagined, focused on primary and outpatient care (no deductible for drugs, coverage of primary care) with a similar premium. This last system proved to be more effective to reduce impoverishment due to medical reasons, because very high expenses are rare, while daily expenditures are a financial burden for households.

Third, the NRCMS did not tackle the financing scheme of health providers that spurred overpricing and overprescribing for many years. Since their liberalization, health facilities sought for activity incomes and drug sales benefit to reach a positive financial balance at the end of the year. If there was a financial surplus, it allowed to give some bonuses to health practitioners. Those phenomena put together created incentives for health practitioners to overprescribe high-technology tests and expensive drugs rather than cheaper ones, in order to increase the THC income, through user fees. Drug benefits became the main source of hospitals financing (Yip et al., 2010). Until 2009, health facilities were allowed to take a 15% mark-up on the drug purchasing price, and used to actually "take an average margin between 30% and 40%" (Wang Dongsheng, vice-director of the Social Development Division of National Development and Reform Commission in 2006).

Drug overpricing was also spurred by the national drug supply chain and its excessive number of wholesalers, intermediaries between the drug producers and health facilities, and by the absence of bidding (Yu et al., 2010). Since the liberalization, the Chinese pharmaceutical market was made of manufactures (up to 5000 in 2008), mostly small producers of generic drugs (Sun et al., 2008). They sold their production to a 'third-tier' who made the contact to another wholesaler, or to health facilities. Every actor of the supply chain taking a benefit, this structure participated in the very high prices of drugs in China (compared to the international prices). Moreover, the regulatory focus had been put on the price rather than the quality of drugs, so that many producers deliver poor quality drugs to contract their costs.

Those various mechanisms that lead to the generalization of overpricing and overprescribing are analyzed through a system dynamics model by Li et al. (2014b), by decomposing every actor of the Chinese pharmaceutical chain and their interests. They lead to a global inefficient system, very expensive for poor services. The authors stand for a cutoff of the benefit chain of drug suppliers between profits and prescribing behaviors. Empirical works, relying on audit studies (Currie et al., 2011, 2014), conclude that supply side factors (i.e. perverse financial incentives of health providers) are the main reason for antibiotic overprescription (far ahead the demand side factors, i.e. patient expectations)

4.1.2.2 The Health Care Reform Plan

The main objectives of the reform In 2009 a global plan concerning the health system is announced, with the strong idea that it should be government-led. It aims at universal access to basic health care, and equity in this matter. Huge financial resources are injected in the system at this purpose, related to several major aspects (Yip and Hsiao, 2009; Yip et al., 2012).

A first objective is to reform public hospitals. Second, the plan advocates primary health care and prevention, placed at the cornerstone of the new system. With new resources for infrastructure and staff, the goal is to reinstate the referral pyramid, with VHS and VHCs at gateways to the healthcare structure. This also means new responsibilities for these Primary Healthcare Facilities (PHF) in terms of vaccinations, antenatal care, follow-up of young and elder people, and of patients with chronic diseases. Diagnosis of those pathologies is also part of their missions.

Important resources are also dedicated to the health insurance institutions. New

targets are fixed in terms of per capita subsidy to both rural and urban insurance schemes. This aims at ensuring universal coverage, and allowing NCMS County Bureaus to widen their benefit package, particularly towards outpatient care.
Finally, the 2009 plan is the starting point to the National Essential Drug Policy (NEDP).
To cope with the excessive price of drugs and disconnect hospitals income from drug sales, the Chinese government implements the NEDP. The objective is to improve the drug supply system and ensure both equity in the access to basic care medicines and safety of drug utilization. A National Essential Medicine List (NEML) is released at the same time, updated in 2012. It includes three medicinal categories: chemical and biological drugs (317 drugs), traditional Chinese patent medicines (203 drugs) and traditional Chinese cut crude herbs (NHFPC, 2013). To meet regional specific needs, local governments are allowed to establish an additional list of essential drugs. In Shandong province, where this study area is located, the additional list (2010 version) consists of 216 drugs. All PHF must now prescribe exclusively essential drugs. As for other healthcare facilities, the utilization of essential drugs should be a priority, and the rate of essential drug utilization must reach the threshold defined by health authorities (NHFPC, 2009).
Since October 2008, the government has also gradually implemented a zero-markup policy for the sales of essential drugs (NDRC, 2008). The selling price in PHF is adjusted to the purchasing price, including delivery costs, and health facilities are not allowed to fix a higher selling price anymore. Moreover, the governement cuts the direct link between PHF and pharmaceutical supplyers, by the creation of a public bidding system at the provincial level. This policy, between 2009 and 2010, completely redefined the structure of PHF financial balance (Fang et al., 2013). Indeed, the loss of drug benefits is a huge hole in PHF revenue and could have widely disturbed their daily activity.

Compensation schemes for the loss of drug sales income To compensate losses due to the zero-markup policy, and ensure the stability of NEDP, different modes of financial compensation have been created (Yuan and Tang, 2012; Zhuo and Zou, 2012). Each county can either choose one, or mix several modes of compensation among the following.

Exclusive government compensation

The PHF with a financial deficit between incomes and expenditures will have their deficits financed exclusively by county government. This compensation mode provides a guarantee for implementation of NEDP, but also a lot of financial pressure for local governments, which could impedes the financial viability of the NEDP. If subsidies are allocated unconditionally, this could lead to a strong decrease of efficiency, as health care providers know that the government will finance them without regards to their performances.

Incentive system

The financial losses are not directly compensated, instead the PHF encourage their medical staff to develop profitable activities such as surgery, or color Doppler ultrasound for instance.

Multiple compensations

They mainly come from government subsidies but also from health insurance funds, or other complementary compensations like the increase of the cost of some treatments. For example, some health facilities increased the cost of tests and injections, other health facilities created a general consultation fee, which includes the fee for consultation and the fee for injection, or a fee for drug prescription.

The separation of revenue and expenditure system

Both the revenues and the expenditures of PHF are totally managed by the county government. In other words, all the incomes of PHF are paid to the county government, and their expenses are integrated into the government budget. The government handles the wages and the cost of the zero-markup policy. This compensation mode breaks the link between revenue and expenditure, so that profit-seeking behaviors could be avoided.

By lowering the unit price of healthcare for households and realigning the healthcare providers incentives, this reform aims at increasing THCs level efficiency through an increase in the demand for healthcare.

4.2 Literature review and the potential effects of the zero mark-up policy on township hospitals efficiency

4.2.1 Expected effects of the reform

Effects on demand of healthcare Because of the high cost of care, the healthcare demand at the TH level is considered as relatively price elastic to the amount of the residual cost borne by the households. The mandatory use of essential drugs, as well as the policy of zero mark-up, should lead to a decrease in the unit cost of care and a reduction in catastrophic costs, all things being equal. For these reasons, an increase in healthcare demand is expected, and was found in various studies (Li et al., 2013; Xiao et al., 2013). This can result from a demand that was previously not satisfied for financial reasons (renunciation to care, self-medication) or from a

transfer of demand from Village Health Stations, or county hospitals, to Township Health Centers. An increase may not be observed if patients, considering the decrease in unit costs, seek more sophisticated care (in county hospitals) than they would have otherwise. This effect was highlighted in the Gansu province following the development of NCMS (Wagstaff and Yu, 2007).
Considering that Township Hospitals are far from being saturated, an increase in activity could lead to an increase in their efficiency. Pélissier et al. (2012) find a decreasing bed occupancy ratio (BOR) from 2000 to 2004 (from 40% to 35%), then increasing to 60% in 2008. In such a context, a surge of activity should not mean new inputs, and an increase in efficiency is expected as the physical and human resources are newly utilized.
Yet, Audibert et al. (2013) demonstrates a positive and significant impact of the NRCMS on Township Hospital activity in Weifang, but a negative impact on its efficiency. It is therefore essential, in terms of public policy, to distinguish the two analyses.

Effects on efficiency In the precise context of the pharmaceutical reform, three main scenarios can be considered:

- Demand for care increases, all other things being equal leading to efficiency progress.
- Demand for care increases, but its expected positive effect on efficiency is offset by an increase in inputs. To anticipate increased attendance at THCs, the quantity of inputs (especially staff) rise, despite their low productivity. If the demand does not increase sufficiently, the effect on the efficiency is potentially negative. This seems likely in Weifang, based on our discussion with the local authorities.
- Demand remains unchanged, as does the level of personnel and equipment. A decline or stagnation in efficiency is expected for several reasons:
  - Compensation for loss of income from drugs is partial (see paragraph 4.1.2.2). Demand remains unchanged because THCs develop coping strategies that maintain high unit costs. For example they can increase medical activities (such as lab tests and drug injections) prices beacause they are not supervised by the reform. The unit cost of care borne by households does not decrease, therefore the demand remains the same.
  - Cost of care decreases. If there is partial compensation, and without THCs coping strategies, objective and/or perceived quality of care are

likely to decline. The negative effect on demand neutralizes or outweighs the positive effect coming from the decrease in the cost of care borne by households.

Nevertheless, the success of the reform relies upon several conditions in its implementation. First, there is a need for a comprehensive policy, that takes every aspect of the issue into consideration. The removal of perverse financial incentives is not sufficient. Drug providers have to be involved in the reform so that the drug quality would not be altered, and that there would not be any shortage in drug supply. The NEML also has to meet the needs of health care providers (Chen et al., 2014; Xiao et al., 2013). Indeed, if the competition is too hard in the bidding process, pharmaceutical firms may lower the quality of their drugs, so quality controls are necessary. It is also essential to ensure a permanent supply of health facilities in essential drugs. Otherwise, two scenarios are possible:

- health facilities activity is constrained because of the lack of medicines to provide, so every possible positive effects on efficiency is diminished;
- health providers prescribe drugs out of the essential medicine list, so patients have to buy them at a higher price in private pharmacies. The financial burden of drug consumption remains heavy and is still a barrier to healthcare.

4.2.2 Literature on health facilities efficiency

For decades, efficiency studies using Data Envelopment Analysis focused on developed countries, mainly US and European countries, as summed up in Hollingsworth (2008) and O'Neill et al. (2008). Pélissier (2012), studying the efficiency of a sample of 24 THCs in Weifang prefecture (Shandong, China), details three previous studies: Liu and Mills (2005), Ng (2011) and Hu et al. (2012). Since then, several studies have been published using efficiency analysis of health efficiency in China including Yang and Zeng (2014), Li et al. (2014a), Cheng et al. (2015a), Cheng et al. (2016). They provided some more information on the methodological and analytic point of view.

Hu et al. (2012) and Audibert et al. (2013) focus on the impact of NRCMS on health facilities, through a two-stage procedure (DEA + Tobit). The first uses data from 30 province hospitals across China, on the 2002/08 period, while the second studies a sample of 24 THCs in the Weifang prefecture, Shandong province, on the 2000/08 period. Audibert et al. (2013) uses the NRCMS coverage rate to estimate the effect of the reform, while Hu et al. (2012) uses both a dummy equal to 1 if the province implements the reform in a given year, and the coverage rate in the province. The NRCMS reform is found to have a positive effect on efficiency in regional-level hospitals (Hu et al), but a negative one in THCs (Audibert et al., 2013). This difference can be explained by the difference of nature of those two kinds of health facilities: if reimbursed with no incentives to go to PHFs, patients tend to go to high-level facilities to get better quality of care. The increase of activity is therefore stronger in high-level facilities than in PHFs, while resources increased at least in the same proportions in PHFs.

Li et al. (2014a) concludes to the same increasing trend in efficiency as Hu et al. (2012) in 12 third-grade public hospitals in Beijing from 2006 to 2009. They were able, through the Malmquist method, to decompose the Total Factor Productivity (TFP) increase into technological change (evolution of the production frontier) and pure efficiency change (evolution of the distance to the frontier). The first element drives the TFP increase, while pure efficiency change is even decreasing for some hospitals, suggesting that the gap has widened between hospitals that can provide high-tech and quality care and others.

Cheng et al. (2015a), also use a two-stage procedure to estimate, decompose and explain the efficiency of 114 County Hospitals in the province of Henan from 2010 to 2012. They find the same decomposition of TFP increase: a leading role of technological change as Li et al. (2014a), and a low average increase of pure efficiency gain (decreasing for several hospitals). They introduced productivity indicators in the second stage of the study (Average Length Of Stay (ALOS), Bed Occupancy Ratio (BOR) for instance) that are found to have a significant impact on efficiency.

Hu et al. (2012) and Cheng et al. (2015a) use inpatients days to measure activity, while the other quoted studies use the number of inpatients. The drawback of using inpatient days is that if a hospital increases its average length of stay to maximize its revenue, it will increase its level of efficiency. Yet, it will cost more to the patient and the insurance system for a same output. This is even more relevant given the hypothesis of a homogenous case-mix: in average, the ALOS should be more or less the same across facilities. In this study the number of admitted inpatients was thus chosen.

Table 4.1 sums up the main characteristics of the efficiency studies quoted in Section 4.2.2.

4.2.3 Potential determinants of efficiency

The literature related to this issue highlights two main types of factors: internal and external ones.

Table 4.1: Global review of recent efficiency analysis on Chinese health facilities

Article	Study area	Production function	Methodology	Main results
Audibert et al. (2013)	24 THCs in Weifang prefecture, Shandong 2000/2008	Inputs: preventive staff for the preventive model, curative staff, number of beds, Equipment Index. Outputs: number of vaccinations for the preventive model, number of outpatients, inpatients, home visits, lab tests and medical imaging for the curative model	DEA + Tobit	Decreasing technical efficiency positive impact of net income per capita, and of NRCMS coverage on technical efficiency
Yang and Zeng (2014)	70 General Hospitals in Shenzhen 2006/2010	Inputs: number of beds, doctors, nurses administrative staff, other staff. Outputs: number of outpatients and inpatients + ALOS and mortality rate as undesirable outputs. Control for the ownership and the size in the second stage.	Three stage DEA + Malmquist	Negative evolution of technical efficiency and quality of care. Environnemental factors in Stage 2 explains most of the output slacks: few possibilities for managers to increase efficiency. No evidence of a efficiency-quality trade-off.
Li et al. (2014)	12 third-grade Public Hospitals in Beijing, 2006/2009	Inputs: staff, number of beds. Outputs: number of outpatient cases and emergency visits, of discharged patients.	DEA + Malmquist	Productivity growth of most hospitals, but decreasing over the years. Technological change leads TPF increase, decrease of technical efficiency change.
Cheng et al. (2015)	114 County Hospitals in Henan province, 2010/2012	Input orientation. Inputs: number of physicians and nurses, number of open beds. Number of outpatients and emergency visits, number of inpatient days	DEA + Tobit	Increase of technical efficiency (0.75, 0.78 and 0.81 successively) Increase of TFP, but thanks to technological change mainly. Negative impact of large numbers of beds and of ALOS on Technical Efficiency, positive impact of productivity indicators.
Cheng et al. (2016)	48 THCs in Xiaogan prefecture, Hubei province, 2008/2014	Inputs: number of medical staff, of other technicians, of non-medical staff, number of beds. Outputs: number of outpatients an emergency visits, number ot inpatients, number of Electronic Health Records (EHR) under management, number of chronic diseases patients under management.	Bootstrapped DEA + Malmquist	Average technical efficiency of 0.51. Negative technological change and increase in technical efficiency change, for a decrease of TPF.

Internal factors Internal variables deal with the way THCs are managed and financed. Here two managerial variables are tested: the wage and bonus expenditures, per employee, and the proportion of licensed staff (i.e., proportion of doctors among the medical staff).

The financing structure of hospital income is also taken into account through its composition, between subsidies and activity revenue. In the situation of soft budget constraint (Kornai, 2009), by a mechanism of moral hazard, the dependence of hospitals on public subsidies has a negative impact on their efficiency. However, if subsidies are allocated according to THCs performances, their effect can be positive on efficiency and attributed to an incentive mechanism.

A methodological concern is the potential endogeneity of the effect of subsidies on technical efficiency, through reverse causality. In theory (according to the policy guidelines), the amount allocated to a THC is partly related to its performance, so the causality would rather be from efficiency to the amount of subsidy, with a positive sign. It can also be considered that in a soft budget constraint situation, if a THC is inefficient it will more likely be bailed out at the end of the year to offset its deficit and reach a financial balance. In this case, the amount of subsidies is still reliant on THC performances, but with a negative sign.

To check for the endogeneity of subsidies, the determinants of subsidies in the study are analyzed through a panel data model. The results highlight an absence of reverse causality between subsidies and efficiency scores. Neither the efficiency score, nor its lagged value significantly explain the amount of received subsidies. The endogeneity of subsidies is thus rejected. Indeed, the amount of the subsidies allocated to each THC is largely driven by the population of the township (the size effect), by the county GDP per capita (which has a positive effect on the financial capacities of the NCMS bureau), and other unobservable factors.

The importance of subsidies is captured through several variables: the proportion of subsidies in THC total income (used as a proxy for THC dependence on public financing), and the proportion of subsidies in THC expenditures. Other proxies are successively introduced to the model: amount of subsidies per capita in the catchment area, per bed, and per medical staff.

External factors External determinants of efficiency encompass all the aspects of a hospital's environment including competition with other healthcare facilities and the importance of potential demand in the catchment area. During a previous period (2000-2008), those variables were found to be the most determinant factors of TH efficiency (Audibert et al., 2013). Here, according to the available data, several variables were introduced:

- the population covered by the NCMS in the township;
- the density of population;
- the net income per capita;
- the number of Village Health Stations under the responsibility of the THC. This last variable can have two opposite effects: there can be a phenomenon of competition (a negative effect), or stimulation by the referring of patients (positive effect).

To check whether there was a shift in the determinants of efficiency between the beginning and the end of the period, the potential determinants were interacted with dummies corresponding to sub-periods. Determinants were also introduced at square to test non-linear relations.

4.3 Evolution of activity in Weifang prefecture Township Health Centers

Section 4.3 focuses on the THCs of a case study in Weifang, Shandong province. It presents the collected data and the main trends drifting the evolution of the activity and productivity of the THCs of the sample, and of their corresponding NCMS bureaus.

4.3.1 The database: survey and study area

This study relies on annual survey data from the rural part of the Weifang prefecture, a relatively rich coastal province, over the 2006-2012 period. Therefore, this study includes the implementation of the 2009 pharmaceutical reform. The Weifang prefecture includes 12 administrative divisions: four *urban* districts in the Weifang urban area (Weicheng, Kuiwen, Fangzi et Hanting), six *rural* county-level cities (Qingzhou, Zhucheng, Shouguang, Anqiu, Gaomi and Changy) and two *rural* counties (Changle and Linqu).[2] Each county or county-level city is divided into townships, and each township is associated one Township Health Center.

Because the urban healthcare issues are different (and so are the urban healthcare facilities), we used only the eight rural entities to select the THCs of the sample. A previous study had be conveyed on the same area in 24 THCs of the six county-level

[2]County-level cities are third-level administrative divisions. They are often composed of a city and its rural surroudings. They were created in the 1980's to reunite the institutions the city itself and of the former county. On the contrary, in a county, there is no major city. But both are administratively classified as rural areas.

County	Population
Anqiu	97.42857
Changle	60.61857
Changyi	60.94857
Gaomi	86.64571
Linqu	86.51143
Qingzhou	90.87714
Shouguang	104.28
Zhucheng	107.5771

Table 4.2: Population in Weifang rural counties, thousands of people (2006/12 mean)

cities (Audibert et al., 2013). It was therefore decided to keep those 24 THCs to have some information on a longer run for descriptive statistics. Six THCs were added, in the counties, to have a complete vision of rural administrative divisions in Weifang.

For Linqu and Changle, a list of all the THCs was dressed, with the GDP per capita associated to each township. We ranked them and selected six THCs for the sample through a systematic sampling method. A first version of the questionnaires was established in winter 2013, and discussed with local colleagues in a first visit at Weifang, to be sure of the availability of the data. Once the final version had been approved, "survey teams" were constituted, including Weifang Health Bureau staff and colleagues from Weifang Medical University. The questionnaire was presented to them, and the data was collected in the THCs at spring 2013. Sources of data include books and registers from THCs and the statistical and finance offices of townships and counties. Figure 4.3 gives the localization of each THC of the sample in the Weifang prefecture. Data analysis was realized in France after the survey, and the first results were discussed with the Weifang Health Bureau and the Weifang Medical University during another visit in Weifang in 2014, in order to prepare the final report of the study.

Every township in the sample is administratively classified as rural, therefore exhibits a large majority of rural population (around 90% throughout the period, Table 4.4). The populations and densities covered by each hospital vary, creating large differences in terms of potential demand for healthcare. The average net income per capita is homogenous among the 30 townships and almost doubled from 2006 to 2012 (5510 annual constant yuans in 2006, 10800 in 2012, Table 4.3).

The broad characteristics of the population remained stable across the period (Table 2). There is no major demographic evolution between 2006 and 2012. The stability of demographic indicators associated with the precise missions of THCs which focus on curative treatment of non-severe cases justifies a homogeneous case-mix across

Table 4.3: Global statistics about the 8 counties of the sample

	2006			2009			2012		
	Mean	Std dev	Rel std dev	Mean	Std dev	Rel std dev	Mean	Std dev	Rel std dev
Rate of under 6 pop (%)	7.83	3.83	48.89%	8	4.09	51.20%	8.32	4	48.06%
Rate of above 65 pop (%)	8.97	1.31	14.61%	9.81	1.37	13.91%	11.23	2.29	20.42%
Rural net income (yuans)	5326	574	10.78%	7507	626	8.34%	11735	824	7.02%
Infant mortality (per 1000)	5.19	3.08	59.29%	4.32	1.62	37.57%	3.3	1.13	34.23%
Juvenile mortality (per 1000)	5.7	3.21	56.23%	4.87	1.84	37.85%	4	1.16	28.91%
Inpatient mortality (per 1000)	4.91	2.63	53.61%	4.68	2.74	58.59%	3.74	1.92	51.27%

Source: Data from County registers; Rel. std dev= Std%mean

Table 4.4: Characteristics of the 30 THCs of the sample

	2006			2009			2012		
	Mean	Std dev.	Rel std dev.	Mean	Std dev.	Rel std dev.	Mean	Std dev.	Rel std dev.
Population of the Township	58599	29615	50.54%	72950	31666	43.41%	71359	33000	46.25%
TH catchment area	46233	17979	38.89%	55143	24453	44.34%	56317	25312	44.95%
Number of VHS under the TH supervision	28	13	48.41%	34	16	48.89%	36	19	50.98%
Density of the covered population (Inhab. per km^2)	483	187	38.79%	503	224	44.57%	495	232	46.78%
Density of the township population (Inhab. per km^2)	545	408	74.93%	563	451	80.05%	553	482	87.12%
Average net income	5511	1170	21.23%	7626	992	13.01%	10800	982	9.09%

Source: Data from Township registers. Rel std dev.= Std%mean

(*) Nominal values have been deflated by the General Retail Price Index in Shandong province, 2006=100 (from China Data Online)

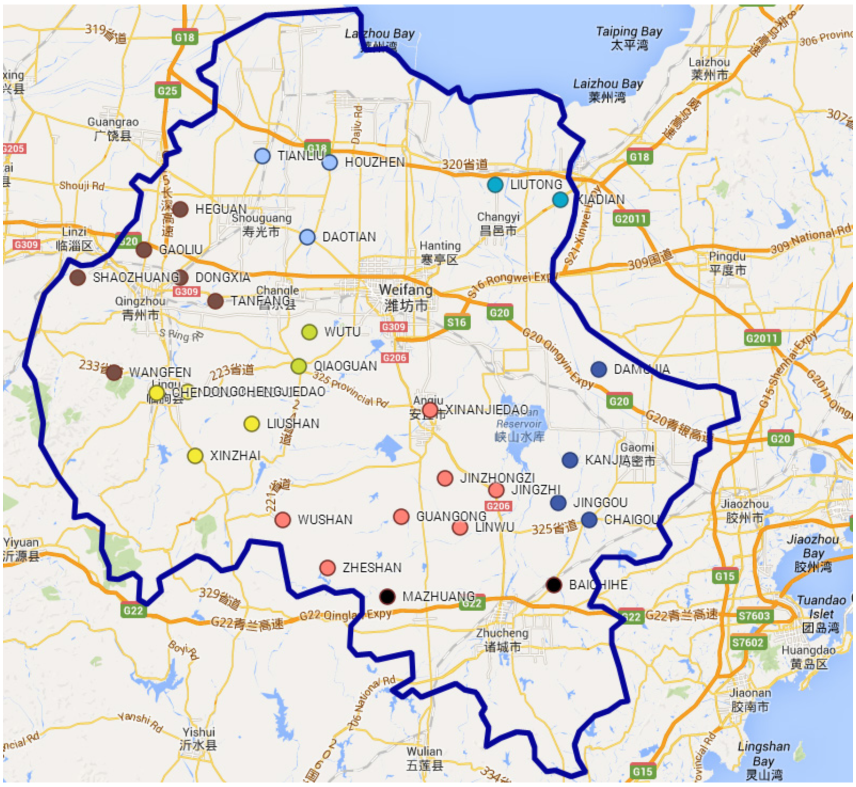


Figure 4.3: Localization of THCs of the sample

the THCs of the sample, and across the period. The choice of collecting THCs actitivity is linked to this case-mix issue. As it provides Primary Health Care, if the catching areas have some identical characteristics, the case-mix is likely to be the same across THCs. This is crucial for an afficiency analysis.

4.3.2 Important evolution of activity

4.3.2.1 The physical and human resources, and their evolution across the period

There were important investments in human and physical resources over the studied period. As a result, the quantity of staff, and the size of hospitals, represented by the number of available beds, increased in the sample. Regarding the number of beds, two categories of THCs can be distinguished in Figure 4.4: a group whose size did not change between 2006 and 2012 (on the diagonal axis), and another whose size increased a lot (Jinggou, Dongchengjiedao, Xinzhai for instance).
The evolution and staff is much less homogenous than the one of the number of beds (see Figures 4.5a and 4.5b). Several THCs saw their staff decreased over the period (those under the diagonal line). This can be linked to several factors. First,

it can be linked to a decrease of the *manning staff*, i.e. the quantity of employees allocated and paid by the local governement, based the covered population and the number of beds (both strongly correlated). Second, it can mean that THCs cannot find enough employees to fill complete manning staff (this problem was put forward by the heads of THC we met during the qualitative survey). They face difficulties to attract qualified medical staff to medical areas.

Figures 4.5a and 4.6 together give some elements of answer. For example, in Chengguanjiedao the decrease in manning staff mainly explains the decreasing staff (moreover, this decrease seems to be mainly linked to non-medical staff). Indeed, the global staff went from 150 to 110, following the decrease of manning staff (from 100 to 60 approximately). In Chaigou, on the contrary, the manning staff does not explain the decrease of medical staff (the global staff decreased from 100 to 80, while the manning staff remained at 110). The THC may have some difficulties to hire new staff.

If an increase in the actual staff is observed, it can also have several explanations. The government may have increased the manning staff of a precise THC, or the THC chose to hire some extra staff on its own budget, considering that the manning staff is not sufficient. In Tanfang for example, the manning staff soared in six years, while the increase of the actual staff was much smoother. It is possible that the THC substituted some manning staff to the extra staff he had to pay itself.

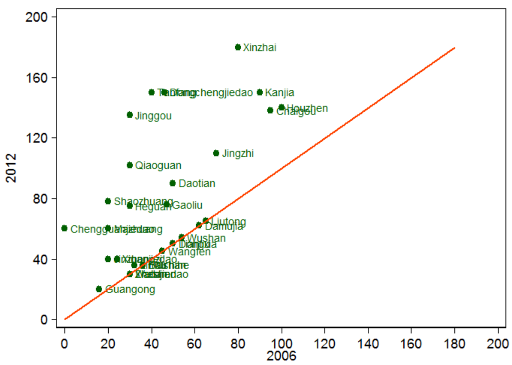
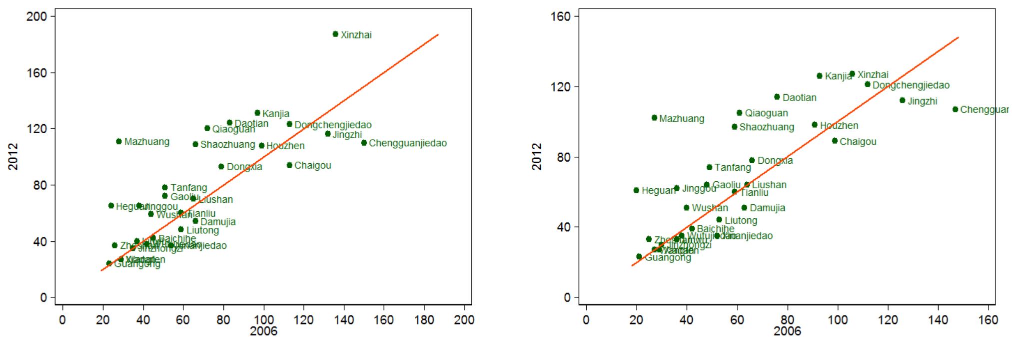


Figure 4.4: Evolution of the number of available beds between 2006 and 2012

As for equipment, six kinds of machines are considered:

- Two for medical imaging: radiology, echograph
- Four for functional explorations: computed tomography, electrocardiograph, endoscope, anesthetic machine, ECG monitoring.

Some of them are almost equally disdributed in the THCs of the sample, others are



a. Global staff b. Medical staff

Figure 4.5: Evolution of staff between 2006 and 2012

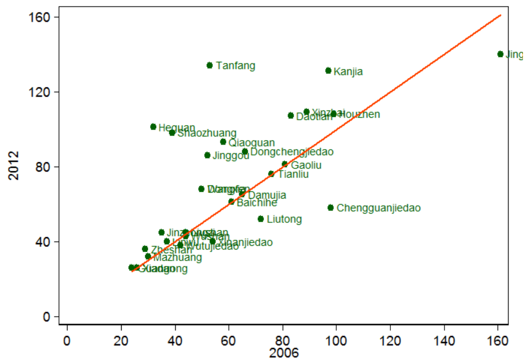


Figure 4.6: Evolution of the manning staff between 2006 and 2012

more discriminating. Several observations can be made:

- Endoscopes are the most rare equipment in the sample. In 2012, only four THCs out of 30 owned an endoscope: Xinzhai (since 2007), Daotian (since 2008), Houzhen (since 2009) and Baichihe (since 2006).

- Few THCs were not already equiped with a radiograph and an echograph in 2006 (5 for radiographs Jingzhi, Xinanjiedao, Zheshan, Dongchengjiedao, Baichihe, 7 for echographs Zheshan, Damujia, Dongchenjiedao, Heguan, Tanfang, Wangfen and Baichihe). The latest THCs to get a radiograph was Dongchengjiedao in 2012, and an echograph Damujia in 2011.

- On the contrary, the computed tomography machines (CT scans) are not that generalized. Only 5 THCs are equiped with it in 2006 (Jingzhi, Liutong, Xinzhai, Daotian and Houzhen), and 10 in 2012 only (Chaigou, Damujia, Dongchengjiedao, Heguan and Baichihe in more).

- Electrocardiograph, in 2006 was already common in the sample, as only Xinanjiedao and Jinggou did not have one. The number of machines varies very much, up to 6 in Kanjia (2012) and 8 in Daotian (2012). From 2007, the median number is 2 CT scans per THC in the sample.

- Concerning anesthetic machine, their presence grew over the period, from 12 in 2006 to 18 in 2012 (out of 30). THCs always own one only, except Damujia (2 since 2006) and Kanjia (4 since 2008), which are atypical in this respect.

- Lastly, ECG monitoring machines have been quite common in the sample since 2006 (18 equiped THCs, 25 in 2012). Nevertheless, a few THCs can be distinguished by the number of machines they own: 4 for Damujia, Kanjia and Houzhen, 6 for Gaoliu and 10 for Xinzhai.

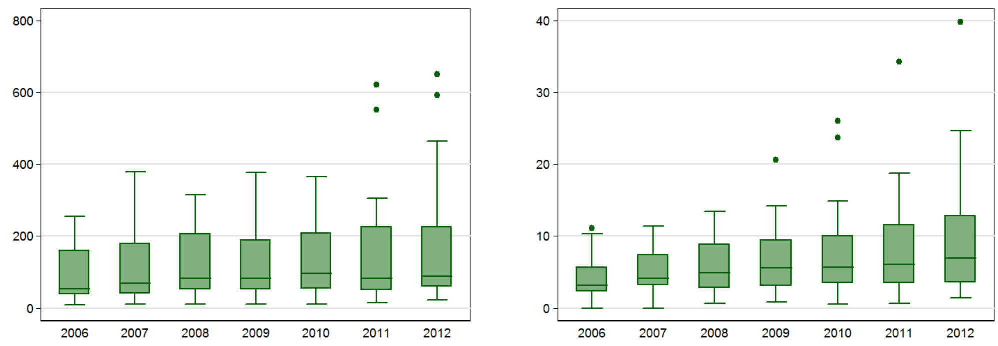
4.3.2.2 Increase of the activities of the THC of the sample

During the period under review, the level of activity increased. In average, the activity of the hospitals grew drastically regarding outpatients, inpatients, lab tests, medical examinations (radiology, etc.) (see Table 4.12). Yet, looking at the distribution of the variables (Figure 4.7), the increase of activity is concentrated on the highest half of the sample. This is confirmed by the stronger variation rate of the mean than of the median, for both inpatient admissions and outpatients. The mean number of daily inpatients went from 4.14 in 2006 to 9.34 in 2012 (+82%) while the median went from 3.19 to 6.98 (+63%). The mean daily number of outpatients went from 90 to 165 (+126%), while the median rose from 54 to 88 (+118%).

The evolutions of other THC activities are less spectacular, except for the number of surgeries, which more than doubled in mean and median over the period.

A few hospitals exhibit particularly high level of curative activities: Jingzhi and Kanjia for outpatients, Xinzhai and Kanjia for inpatients admissions.

The concept of Public Health, through the development of preventive activities (vaccinations, follow-up of riskier populations, informations campaigns) is at the core of the governement plan of 2009, especially in PHF like VHS. An increase in those activities is thus expected, and confirmed by Table 4.12: the average number of vaccinations rose by 43%, from 11 178 to 16 002, and the number of emergencies more than doubled, going from 605 to 1331 in average.



a. Daily inpatient admissions b. Daily oupatients

Figure 4.7: Evolution of activity between 2006 and 2012

4.3.2.3 Excessive increase of expensive tests and imaging ?

One of the consequences of the reform may be that THCs, as they cannot rely on drug sale income anymore, will increase the number of high-tech and expensive tests to increase their activity income. This behaviour has been observed in China since the liberalization of the health system, but the 2009 reform might spread it even more.

In the THCs of the sample several kinds of tests and imaging are provided:

- The first kind of tests are the blood, urine and stool tests. The number of those tests per inpatient admission ranges between 0 and 3 during the whole period, with the exception of Liutong where it ranges between 5 and 10 according to the years. THCs of the sample provide between 0 and 0.5 tests per outpatient visit. Liutong is on the top of the distribution, and the maximum value is 0.89 in Daotian (2007). No increasing trend is to be noticed between 2006 and 2012.

- The second kind of medical exams are imaging examinations: echography and radiology. Most of the sample provides less than 0.2 echography per outpatient (Wangfen exhibiting high features across the whole period) and less than 1.5 echographies per inpatient admission. The average number of echographies decreases for inpatient (0.43 to 0.31), and remains stable for outpatients (around 0.07). An inpatient receives less than 2 radiologies during his hospitalisation (a THC, Dongxia, exhibits higher scores, around 3). The average number of radiologies decreases for inpatient, from 0.9 to 0.64 over the period (not for outpatient, stable at 0.1).

- The third kind of tests provided by THCs are function tests, i.e. ECG and endoscopies. In the study sample, only 4 THCs practiced endoscopies in 2012: Xinzhai (since 2007), Daotian (since 2008), Houzhen (since 2009) and Baichihe (from 2006). ECG are far more common in the sample, and their use is stable for outpatients and inpatients over the study period. In average, an outpatient receives 0.05 ECG, with three atypical THCs: Liushan, Xinzhai and Wangfen. As for inpatient, an inpatient receives by average 0.4 ECG, and three THCs exhibit much higher figures: Chaiggou and Xinzhai in 2006 and 2007.

From the observations on the sample, there doesn't seem to be a clear stategy from THCs to rely on tests and exams to offset the drug sale loss. It is quite hard to analyze, since several scenarios are possible. The phenomenon of overprescrption of tests might be older than 2006, so there is no additionnal increase of tests linked to the reform. Or in Weifang, THCs rely mostly on drug sales, and the subsidies came to offset the loss so that the proportion of activity income remains stable.

4.3.3 Development of NCMS subsidies

4.3.3.1 NCMS enrollment and the benefit package

Over the period, the Weifang prefecture reached an almost universal enrollment of its rural population in the NCMS (see Figure 4.8). In 2010 every county exhibited a coverage rate of 99 or 100%. Looking at coverage rate at the THC level, the same evolution can be noticed. In the first years of the period, one THC only, Liushan, has a significantly low coverage rate (less than 40%), but it grows to more than 80% in 2012. Except Liushan, most THCs show high coverage rate, higher than 90% for most of them (Figure 4.9).

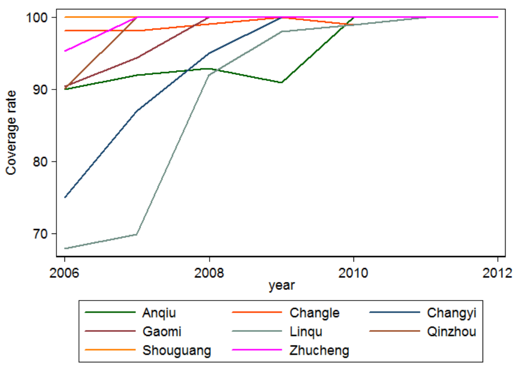
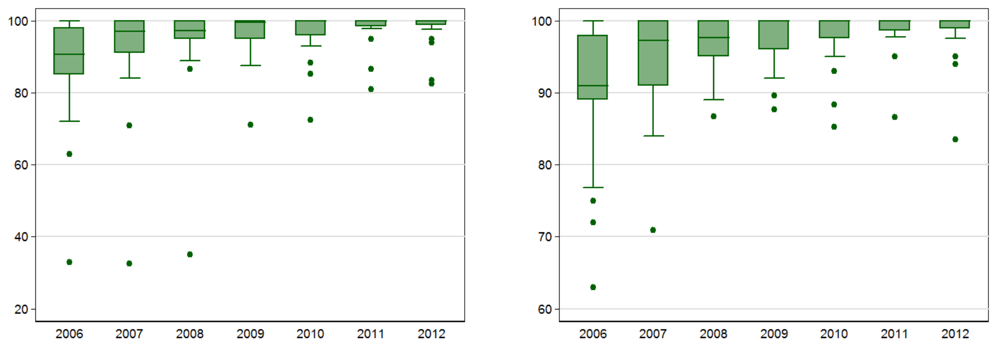


Figure 4.8: Enrollment in the NCMS bureaus of the study area

To this increase in the number of insured households is associated a deepening of the benefit package throughout the period. The NCMS in Weifang is typical of



Whole sample Without Liushan

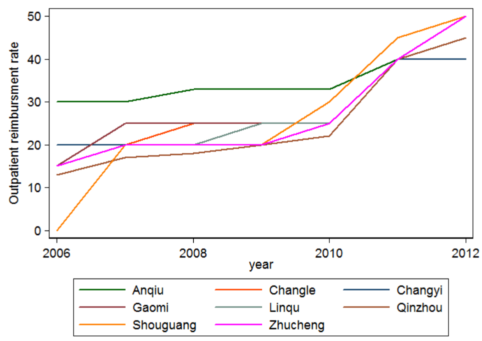
Figure 4.9: Evolution of NCMS enrollment in the THCs of the sample

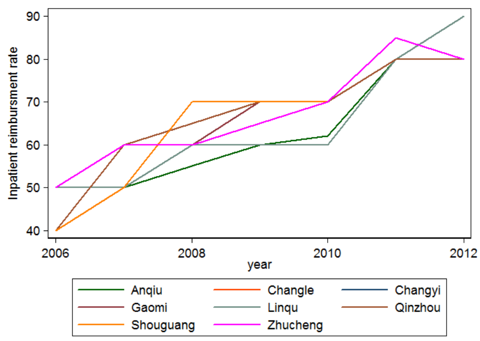
the priorities that have been given nationnally. In 2006, the reimbursment rate for outpatient care is very low (between 0 in Shouguang and 30% in Anqiu, Figure 4.10a), but with no deductible. An important increase of the reimbursment rate is observed between 2010 and 2011, illustrating the new focus on primary health care. On the contrary, the reimbursment rate of inpatient care has been at least 40% in every county in Weifang from 2006, with quite low deductibles, and the rate rised regularly to reach 80 to 90% in 2012 (Figure 4.10b). The difference between the reimbursment rate in the primary and in the second level health care facilities is not that important. 75 to 80% of expenses are reimbursed in every secondary health care facilities of Weifang in 2012. The financial incentives to respect the referral pyramid rely on the deductibles (between 100 and 200 yuans in PHF in 2012, and 500 to 800 in second level health care facilities).

This certainly reflects the fact that Weifang is a rich prefecture, relatively to China. NCMS bureaus have the financial means to propose a wide benefit package, and not to choose between protecting from health shocks and daily expenses as in many central and western regions.

As a result of the universal NCMS coverage and of the deepening of the benefit package, the expenditures of every NRCMS bureau in our sample soared, with a strong acceleration starting in 2010 (Figure 4.11). Regarding the inpatient reimbursment, the increase in NCMS expenditures is due to an increase of the reimbursed amount per inpatient case rather than to an increase of the number of cases (see Figure 4.29b and 4.30b).

Regarding outpatient reimbursment, except for Shouguang, the amount of reimbursment remained quite stable, like the number of outpatients cases, except for two atypical counties in this respect, Anqiu and Zhucheng (Figure 4.29a and 4.30a).





a. Outpatients cares in PHF

b. Inpatient cares in PHF

Figure 4.10: Evolution of reimbursment rates in Weifang

As expected, the mean reimbursment of an inpatient admission is lower in THCs than in all types of facilities, because of the higher unit cost in CHs and the more sophisticated cares provided there (surgeries for instance) (Figure 4.29c and 4.29d). The second financial aspect of the government investments in health is the amount of health expenditures from the county governements, dedicated to the several levels of health facilities. In addition to their spectacular increase, from 2010 the variety of their destination has reflected the new priority given to primary healthcare facilities : THCs are at the core of the expenses, but some subsidies are now expressively dedicated to VHSs (Figures 4.11 and 4.13).

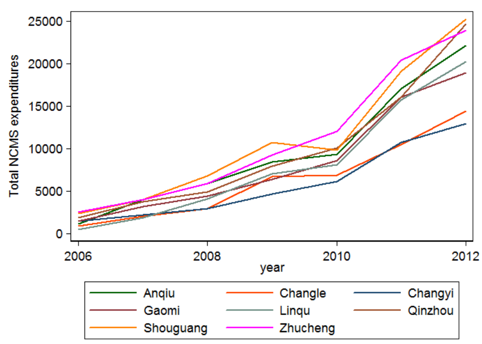


Figure 4.11: Global expenditures of the NCMS bureaus

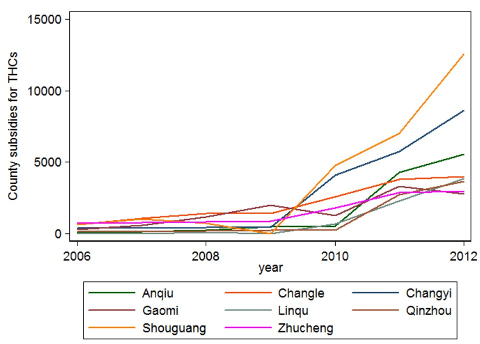
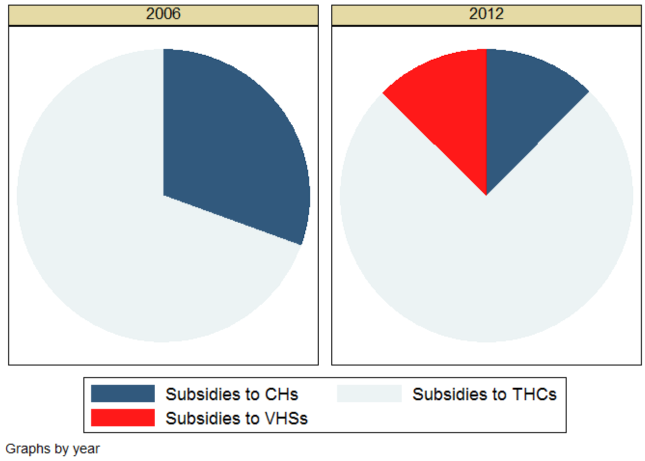


Figure 4.12: County subsidies allocated to THCs

Figure 4.13: Evolution of public health expenditures in Weifang



4.3.3.2 The refunding of THCs financing

As a result of the zero mark-up policy, the structure of THCs financing has been sharply reformed by the National Essential Drug Policy, and the THCs of this study are no exception.

Figure 4.14 shows the comparison of main sources of income for all observed THCs between 2006 and 2012. In 2006, over 80% of revenue is generated by hospital activities, including activity, drug and preventive income. Drug income is the primary source of income (about 50%). In contrast, income from activities, drug and prevention reward hospitals with up to 55% of their revenue in 2012. Particularly, drug income only represent 29% of THCs income. This huge decrease is due to the loss of drug prescription mark-ups, offset by governmental subsidies, (about one third of THC revenue). Those subsidies mainly come from the county level.

The increase in subsidies is much stronger than the drug income loss (Figure 4.15) even considering activity increase.

4.3.3.3 Prescription of injections and antibiotics

An aspect of the survey deals with the pharmaceutical stake. For every THC of the sample, the 15 drugs most sold by the THC pharmacy were listed, as well as their main characteristics, in 2011 and 2012: western or traditional, antibiotic or not, injected or not. This does not allow to produce any quantitative analysis about overprescription, but gives descriptive information about the nature of the most prescribed in Weifang THCs (see Table 4.14).

The mean number of antiobiotic drugs is 4.6 and 4.5 out of the 15 most sold drugs in THCs in 2011 and 2012 (median at 5). The mean number of drugs that have to

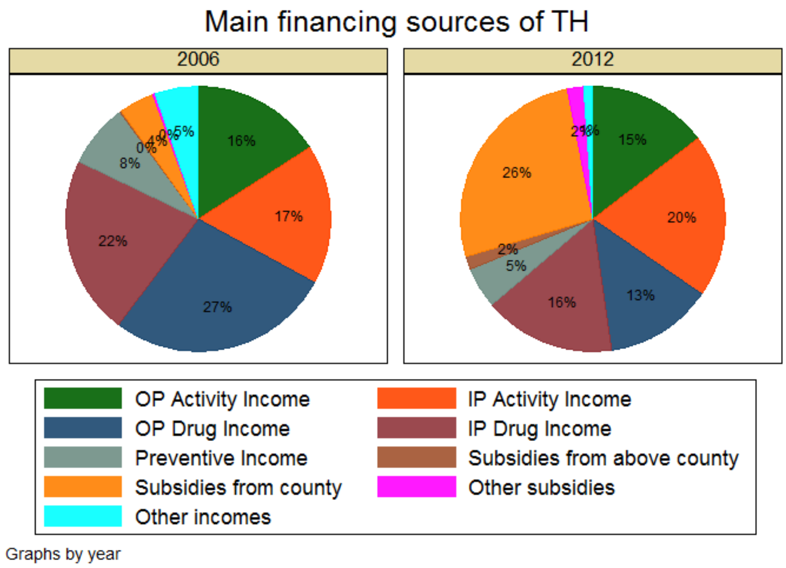


Figure 4.14: Evolution of the main source of THCS income 2006 and 2012

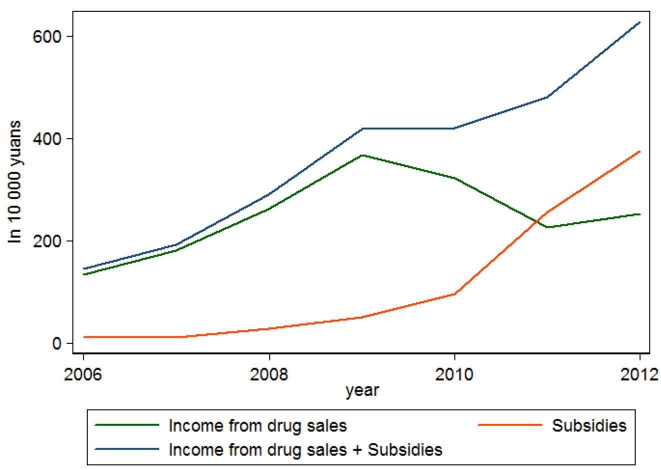


Figure 4.15: Evolution of subsidies and income from drug sales

be administered by injection is 10 out of 15 in 2011, and 9.7 in 2012 (the median went from 10 to 11). This high proportion of injections is certainly explained by the inpatient activity.

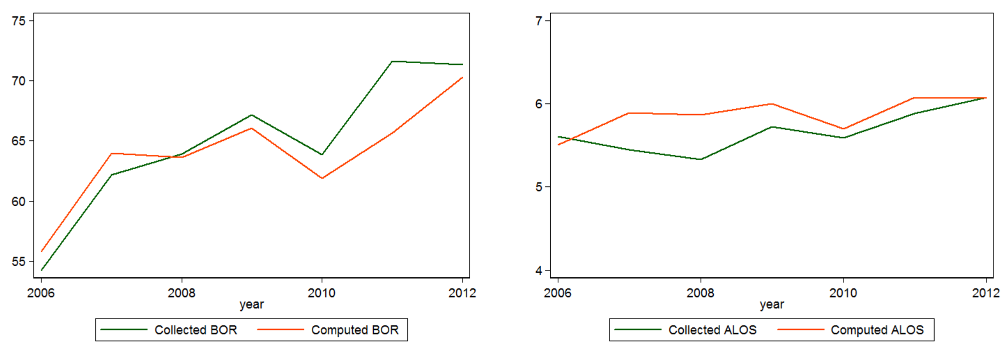
4.3.4 Evolution of main productivity indicators

4.3.4.1 The Bed Occupancy Rate and the Average Length Of Stay

The Bed Occupancy Rate (BOR) and the Average Length of Stay (ALOS) were collected in each THC, and they were also computed using the number of occupied bed days per year, the number of inpatient admissions and the number of available beds in the THC. For some THCs, we found important differences between the two figures, without any explanation. This issue concerns Xinzhai, Tianliu, Jinggou and Shaoguang for the BOR, and mostly Damujia and Wutujiedao for the ALOS. To compute them we used the following formulas:

$$BOR = \frac{\text{nb of occupied bed days} * 100}{\text{nb of available beds} * 365}$$
$$ALOS = \frac{\text{nb of occupied bed days}}{\text{nb of inpatient admissions}}$$

There a strong increasing trend of the BOR over the study period (Figure 4.16a), from a mean of 54% in 2006 to 71% in 2012 (the median went from 53 to 66%). This increase is not regular, it is concentrated on the beginning of the period (from 2006 to 2007 mainly). This figure is much higher than the national average quoted by Eggleston et al. (2008): 60% globally, 40 for THCs. On the contrary, there is no major change in the ALOS, which stays stable around 6 days (Figure 4.16b). This value is a little smaller than the ALOS found by Yang and Zeng (2014) in Shenzhen, around 7.



a. Bed Occupancy Rate b. Average Length of Stay

Figure 4.16: Evolution of global productivity

4.3.4.2 Staff productivity

Productivity indicators allow to have a first glance at trends concerning THCs activity and performances. Three indicators were computed: number of daily outpatients per medical staff, number of daily inpatient admissions per medical staff, and number of annual inpatient bed days per medical staff. Staff productivity is very low in the sample (2.2 daily outpatients per medical staff in 2012), even lower than the figures given in Eggleston et al. (2008) ("5 outpatients per doctor and 1.5 inpatient beddays per doctor for general hospitals in 2004", p.152). The difference may be partly due to the calculation, per medical staff or doctor only. Still, it remains very low: 2.2 daily outpatients per medical staff in 2012 and only 0.14 inpatient admissions (Figures 4.17 and 4.18). Yet, an increasing trend is observed from 2006 to 2012, especially concerning inpatient, and the annual number of bed days per medical staff went from 150 to 300, doubling between 2006 and 2012.

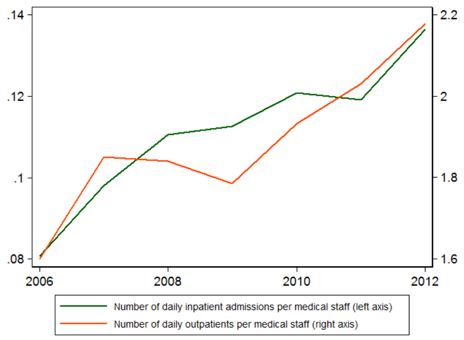


Figure 4.17: Evolution of daily outpatients and inpatient admissions

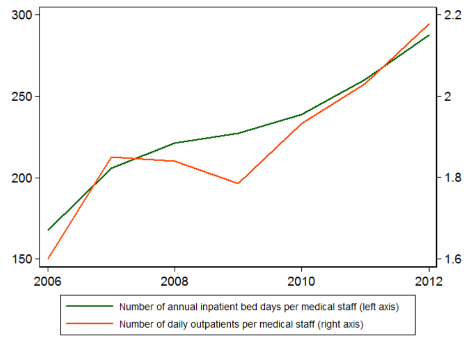


Figure 4.18: Evolution of daily outpatient and annual inpatient bed days

4.4 Estimating efficiency of Township Hospitals

This Section details the first stage of this case study, i.e. the efficiency assessment. It presents the production function, the empirical methodology, and discusses the efficiency scores.

4.4.1 The choice of the production function

The definition of the production function, i.e. the choice of the inputs and outputs is at the core of an efficiency analysis, as it reflects the relevance of the model to describe the activity of a Decision Making Unit (DMU). The several variables and parameters introduced in the following model were chosen with respect to the literature on efficiency in health care sector, and to our discussions with locals partners so that they could faithfully reflect THCs activity. As for the literature, several publication focus on the efficiency of health facilities in China .[3] Five articles contain an exhaustive description of the methodology, *detailed in Table 4.1*. They were used as starting points in the analysis.

4.4.1.1 Output of the production function : the Global Activity Index

The multi-output production function of a hospital has to be dealt with, regarding the diversity of a health facility activity. The selected outputs have to exhaustively represent the production of the DMUs. Usually, the literature only takes into account outpatient visits and inpatient days or admissions (see Table 4.1). Emergency vists are often associated to outpatients. Those three variables were selected in our study. The number of surgeries was also taken into account. This might be an exhaustive description of the activity of high-grade health facilities, since it is focused on curative services. However, in the case of THCs, preventive activities are central to production, especially in the context of the development of the Primary Health Care in the reform. Cheng et al. (2016) add the number of Electronic Health Records (EHR) under management and the number of chronic diseases patients under management to capture the preventive aspects of THCs production. In Weifang, in the interviews THCs managers and doctors emphasized the workload that the primary health care services represented. From our available data, the number of antenatal visits, vaccinations were added to the model.

An issue was the relevance of including the number of total tests realized in the THCs during the year. On the one hand, if we consider the quantity of activities in the outputs, it might be justified to include them. It takes more time and resources to practice an exam and examine its results than a "simple" consult. On the other hand, considering the Chinese context there a risk of overprescription of tests and imaging. Thus, including them in the outputs might lead to an artificial increase in the efficiency score of THCs that prescribe too many exams for financial reasons. Yet, this kind of behaviour concerns only expensive and high tech exams. Thus, blood, urine and stool tests were kept in the analysis and integrated in the "total tests" variable. The other tests were excluded from the list of the outputs. To see if the presence of the tests distorts the results, the same analysis was conducted with and without them.

However, the number of inputs and outputs has to be limited, for the estimator to converge quickly. Several "rules of thumbs" have been proposed to choose the

[3] In addition, several poster abstracts were published by The Lancet (Chen et al., 2016; Cheng et al., 2015b), but the production functions were not detailed.

optimal number of inputs and outputs, including the one of Boussofiane et al. (1991): the number of inputs multiplied by the number of outputs has to be smaller than the number of DMUs.

To limit the number of outputs, a solution is to compute a synthetic indicator, taking each one of the previously described variables into account. This choice was justified by the high correlation between most of the variables. All the outputs variables are significantly and positively correlated (except for tests and antenatal visits), which means that there is a strong common pattern among them (Table 4.13). The loss of information induced by the construction of a synthetic index is thus limited.

To get a synthetic index, one main option is to compute a weighted mean, as done in Audibert et al. (2013). The choice of the weights is then crucial. In Audibert et al. (2013), the weights were fixed by discussion with local actors. For this study, such information was not available. Thus, a Principal Component Analysis was conducted, to get the weights that would take into account the most common variability possible.

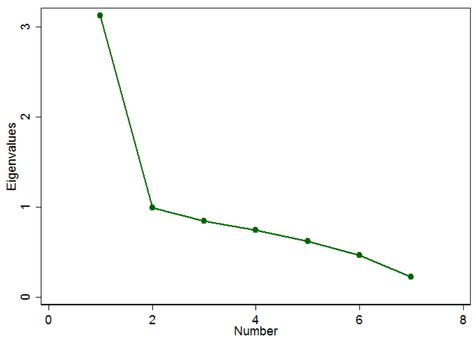


Figure 4.19: Eigenvalues for the first components

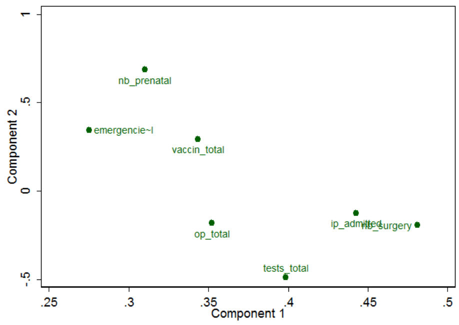


Figure 4.20: Weights for the first two components

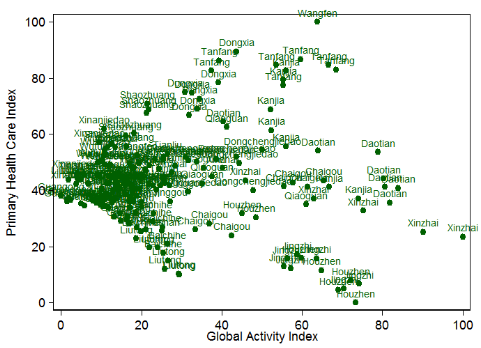


Figure 4.21: Relation between the first two axis

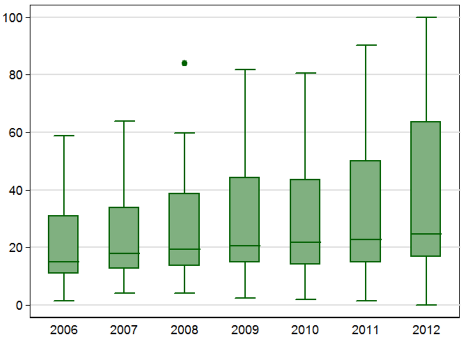


Figure 4.22: Evolution of the Global Activity Index

The Principal Component analysis results in two axis with Eigenvalue higher than

1 (Figure4.19), the first one being more higher than the second. This first axis explains 44% of the total variability and is positively correlated to each one of the variables introduced in the analysis. The positive signs of the weighting coefficients (Figure 4.20) suggest that this axis captures a global trend of the activity and is consistent with the characteristics of an output (an increase of any variable will increase the Global Activity Index). It is kept as the Global Activity Index and is used as the output of our efficiency model after being standardized so that the minimal observation is 0 and the maximal is 100.

The second axis explains 13% of the global variabilty, with an Eigenvalue of 1.09. It means that it can provide some information about a trend driving the data. Only three activities are positively correlated with this second axis, the others being negatively correlated (Figure 4.20). THCs that provide most antenatal visits, emergency visits and vaccinations provide the less outpatients and inpatients, tests and surgeries. The axis gives information about complementary and substituable activities: THCs which exhibit a large score in this axis are more focused than the others on Primary Health Care (Tanfang or Dongxia, Figure 4.21).

Figure 4.22 gives the distribution of the Global Activity Index across years. A clear increasing trend can be seen, but concentrated on the upper half of the sample The sample mean rose by 70% over the period, and the median by 63% (Table 4.3). Moreover, ther is no disruption in this increasing trend, it is regular across years.

Several alternative options were tested to judge the sensibility of the analysis and the Index to the chosen variables. First, the same analysis was conducted replacing the number of admitted inpatients by the number of occupied bed days during the year. A Spearman ranking test confirms that the two series are not independent, with a p-value of 0.00 and a Spearman'rho of 0.98. A unique difference can be found for Xinzhai, which has a better ranking with the number of admitted inpatients. Otherwise the two variables give the same serie. An Index close to the one used in Audibert et al. (2013) has also been tested. It is a weighted mean with weights equal to 1 for each variables but inpatient admissions (3) and emergency visits (3). The serie still has a strong ranking correlation with the one obtained with the PCA: the Spearman's rho is 0.87, and the p-value for the rejection of the null hypothesis of independance is 0.00. The PCA was also computed without including the total number of tests, and the results are almost the same. The Spearman rho is 0.96 between the two series of results, and the independence hypothesis is rejected with an error probability of 0.00.

4.4.1.2 Inputs of the production function

The choice of the inputs was made based on the existing literature, and on discussions with local actors. O'Neill et al. (2008) gives an almost exhaustive list of the possible inputs to be include in an efficiency analysis concerning health facilities. Three inputs were identified in this study, reflecting the human and physical capital of the THCs of Weifang: the global staff of a THC for the human capital, the number of available beds and an equipment index for the physical capital. Many studies consider only the medical staff(see Table 4.1), claiming that it was the only one to be productive in terms of health care, or, if they have enough information and observations, separate the different kinds of staff: doctors and and nurses and nonmedical staff . Here, given the number of DMUs (THCs), three inputs for human capital is a too much. Therefore, the global staff was kept as the nonmedical staff (mostly admnistrative and technical) is an essential resource for THCs activity. Alternative possibilities were tested, including the separation of medical and nonmedical staff, and the inclusion of the sole medical staff as input, with no major change in the ranking of the THCs. The separation between medical and nonmedical staff was not kept in the main results because three THCs in the sample declare employing no nonmedical staff (Xiadan, Liushan, Tianliu). As nonparametric methods do not accept the presence of nul values, it was not possible to get any score for those THCs with specifications that separate medical and nonmedical staff. Still, one alternative model was considered, with the medical staff only instead of the whole staff.

The Equipement Index was computed using a Principal Component Analysis (PCA), which included every kind of imaging and test machines found in a THC: radiology, echograph, ECG, endoscopes, computed tomography, anesthetic machines and ECG monitoring instrument. The first axis of the analysis is kept as the Equipement Index. It is positively correlated to every element of the analysis, and exhibits an Eigenvalue of 3.91.

4.4.2 Measurement of efficiency: the first stage

4.4.2.1 The adopted methodology

As the number of potential determinants of efficiency is rather important, it is hard to use conditional models. A two-stage procedure was thus chosen (see Section 2.3.2).

In the first stage, to deal with the presence of potential outliers and with the dimensionality issue (210 observations are available), robust frontier estimators were implemented (see Section 2.2.2.2).

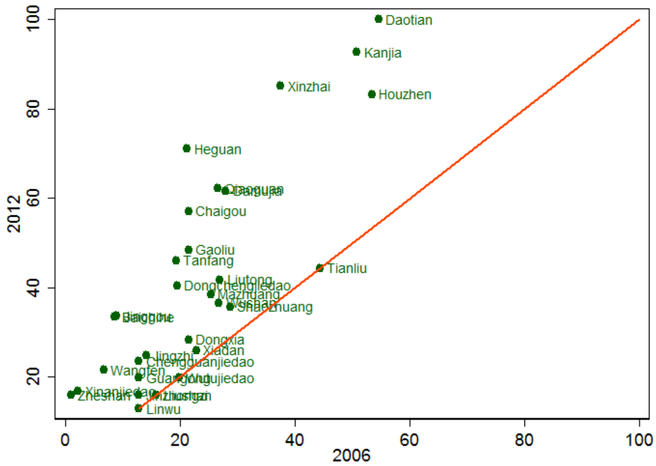


Figure 4.23: Evolution of the Equipment Index for the THCs of the sample

4.4.2.2 The empirical model

The orientation From the literature, and the context of THCs in Weifang, an output orientation was chosen. This orientation is the most common in the efficiency studies about Chinese health care facilities (see Table 4.1). Only Cheng et al. (2015a) uses a input orientation, claiming the County Hospitals do not have the control over the demand, and over their amount of activity, so that they can only minimize the quantity of resources to provide those activities.

Here, it is rather the quantity of inputs that THCs cannot choose: the number of available beds and staff (at least the manning staff) is fixed by the County Health Bureau, according to the catching area. Given those resources, THCs have to maximize their activity by providing good quality and affordable care to patients. In this perspective, an output orientation was used in this study.

The number of estimated production frontier There is a trade-off in the choice of the number of estimated production frontiers. If DMUs are evaluated with the same production frontier over the whole period, it implies that there is no technonlogical progress across the years. In the opposite case, if the production frontier changes over years or periods, and a sole frontier is considered over the period, then the estimated scores are biased because DMUs are not compared to the relevant frontier.

Estimating several frontiers also allows to compute a Malmquist Index, i.e. to decompose the change in Total Factor Productivity between the scale efficiency change, the technological change (the evolution of the production frontier) and the pure efficiency change (the evolution of the distance of a DMU to the current production frontier). In the review proposed in Table 4.1, only Audibert et al. (2013) do not use annual production frontier, because the NRCMS reform was implemented gradually, so there was a possibility of technological change over the study period. The limit of using several frontiers relies in the dimensionality issue: it lowers the number of DMUs per frontier and can make the comparison between them difficult. If the number of DMUs is low, it is crucial to have a very homogenous sample (as Li et al. (2014a), with 12 hospitals only). The other drawback is at the second stage of the analysis, while explaining the efficiency scores. Indeed, it means that scores (distances to production frontiers) related to different frontiers are to be explained in the same regression.

In this study, only one production frontier was considered for several reasons. First, it allows to raise the number of DMUs to 210. Second, from discussions with local actors, it appeared that that had not been any technological change across the period. Indeed, the core of THCs activity is basic health care that does not rely on very recent technologies. Thus, it is credible to rely on the assumption that there was no technological change between 2006 and 2012 in Weifang THCs. The fact that no disruption in the evolution of activity is observed confirms this idea.

The returns to scale As the partial frontier methods come from the Free Disposal Hull analysis, they do not require any assumption concerning the nature of the returns to scale (unlike the Data Envelop Analysis). For the altenative estimations and robustness checks, variable returns to scale were chosen to disantange the scale inefficiency from the technical inefficiency.

The choice of the m parameter In partial frontier model, the value of the m, seen as a "trimming parameter", is crucial. Simar and Wilson (2013) gives a method to choose the optimal value : "The final value of m can be chosen in terms of the desired level of robustness", i.e., of proportion of DMUs above the production frontier. The upper limit is the number of observations.

Here, all possible values from 10 to 200 were tested, with a range of 10, and with a narrower range around the threshold in the proportion of super-efficient units. As expected, the proportion of DMUs above the frontier (with a score higher than one) decreases as m increases, and so does the mean efficiency in the sample (Figure 4.24). From m=170, there are less than 10% of super-efficient DMUs in the sample, and from m=180, this proportion is equal to 8. The value of 190 was chosen for the rest of the final analysis , as from m equal to 190 the scores remained identical whatever the chosen value of m.

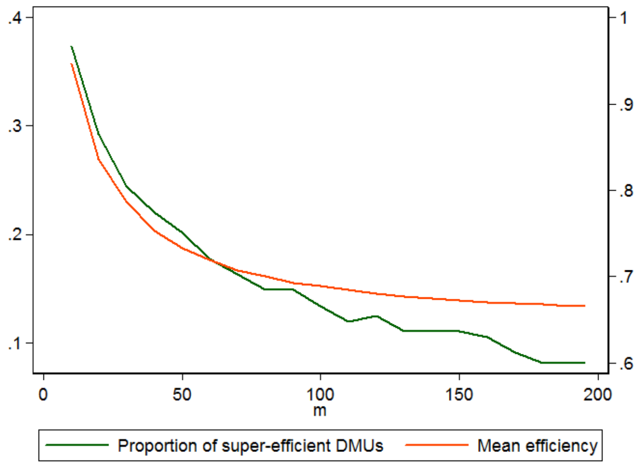


Figure 4.24: Proportion of super-efficient DMUs and mean efficiency according to the value of m

The choice of the *alpha* parameter Running the order-*alpha* analysis, the value of the parameter has to be chosen. This value gives the percentage of DMUs that will remain below the frontier: the higher the *alpha*, the closer the method is to the Free Disposal Hull. Contrary to the order-m method, the order-*alpha* method leaves a great part of the sample as super-efficient units, even for large values of *alpha*, and it is very sensitive to a slight change of the parameter when it is closing 1 (see Table 4.5). To get less than 10% of super-efficient THCs, only 1%, i.e. 2 DMUs of the sample, have to be left above the production frontier in its estimation, which is very few.

For this reason, the order-m scores will be used in priority in the second stage, as from 190 they do not vary with m. Nevertheless, it is necessary to check if the least and most efficent THCs are the same with the two methods, to see if the conclusions will remain the same whatever the estimations.

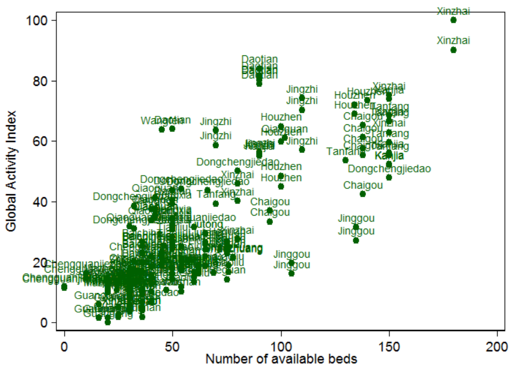
Value of *alpha*	95	97	98	99
Model 1	0.321	0.220	0.129	0.048
Model 5	0.3891	0.313	0.236	0.082
Model 7	0.316	0.215	0.129	0.048
Model 8	0.404	0.313	0.240	0.091

Table 4.5: Proportion of super-efficient DMUs according to the model and the value of *alpha*

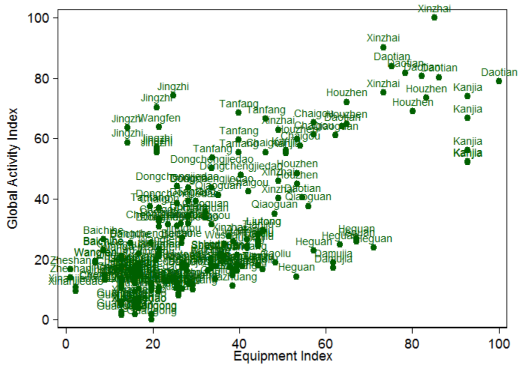
4.4.2.3 Detection of outliers

As stated in the previous paragraphs, the partial frontier analyses give some robustness to the estimation regarding the presence of outliers. Indeed, they will exhibit a score higher than 1 so that they are easily detectable and do not distord the production frontier (and bias the scores of the other DMUs).

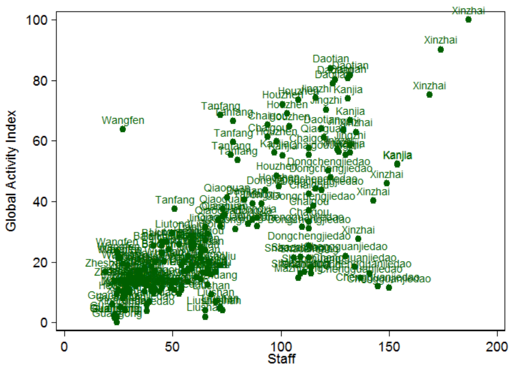
A first graphical analysis gives an idea of potentially atypical DMUs regarding the relation between the quantity of inputs and outputs. The limit of this analysis is that it can only handle two dimensions a time. However, it gives an idea of the sample homogeneity. The objective is to detect the potential super-efficient DMUs, that will distord the production frontier because they do not respond to the same production function.



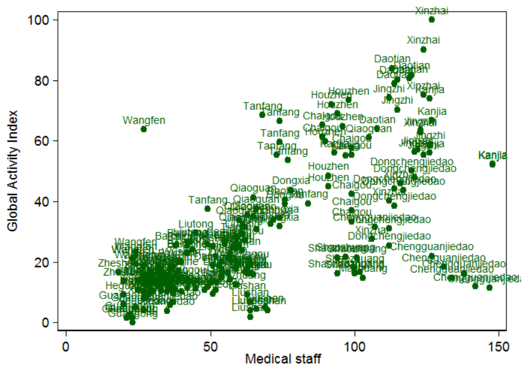
a. Number of available beds



b. Equipment Index



c. Number of total staff of the THC



d. Number of medical staff of the THC

Figure 4.25: Graphical detection of outliers

From Figure 4.25, several THCs seem specific. For instance, Xinzhai exhibits the highest level of activity, by far, with the highest number of available beds and total staff (Figures 4.25a and 4.25c). Wangfen, increased sharply its activity from 2011 to

2012, for a stable level of inputs, so that it becomes atypical. Daotian increased its quantity of available beds, and Jingzhi its equipment index (it seems over-equipied for its activity, Figure 4.25b). To a smaller extent, Tanfang also provides a high quantity of activity regarding its staff. On the contrary, Changguanjiedao provides very few activities for its medical staff.

To check whether those observations are actually outliers (i.e. if they distort the production frontier), their scores with the order-m method were analyzed, following the method developed by Simar (2003). They use the order-m scores, with various values of the m parameter, and define outliers as DMUs whose score remains much higher than 1 even when m increases. From Table 4.15, several THCs can be deleted from the list of potential outliers: Chengguanjiedao, Jingzhi, Daotian and Tanfang, because their scores decrease very quickly when m increases.

Finally, efficiency scores were computed leaving successively each one of potential outliers out of the analysis (leave-one-out strategy), in order to see if their presence affect to a large extent the scores of the others DMUs. A difference in the scores was only found when the observation of Wangfen in 2012 was left out of the sample (see Table 4.6). Indeed, both the Pearson correlation coefficients and the Spearman Rhos exhibit very high values (superior to 0.99). This means that the presence of those DMUs does not affect neither the value nor ranking of the scores. The only exception is Wangfen in 2012. When it is excluded from the sample, the Pearson correlation is around 0.86, and the Spearman aroud 0.85 with all the other models. This observation was thus defined as an outlier, and excluded for further analysis.

In the new efficiency analysis, the scores of super-efficient DMUs range between 1 and 1.089 only. They are about 28 observations in this case, but their score being close to 1, it is accepted that they are not ouliers.

4.4.2.4 Robustness of the production function

As stated in paragraph 4.4.2.1, as they are data-driven, non-parametric efficiency models are sensitive to the definition of the production function (Jacobs et al., 2006). To test the validity of the efficiency scores, several alternative scores, computed wih alternative inputs or outputs, were compared to the initial ones. All the tested models are summed up in Table 4.7. They mostly refer to the choices that were made in paragraph 4.4.1.1.

Moreover, there are two candidate methods, order-m and order-*alpha* frontier analysis, both of them with many possibilities regarding the value of the trimming parameter. It is essential to verify whether methodological choices had consequences on the scores, and on the ranking of the THCs.

Table 4.6: Pearson and Spearman correlations with a left-one-out strategy

Pearson	Whole	Xinzhai	Wangfen	Jingzhi	Chengguan.	Tanfang	Daotian
Whole	1						
Xinzhai	1	1					
Wangfen	0.870	0.867	1				
Jingzhi	1	1	0.865	1			
Chengguan.	1	1	0.869	0.999	1		
Tanfang	1	1	0.868	0.999	1	1	
Daotian	0.999	0.999	0.863	0.998	0.999	0.998	1

Spearman	Whole	Xinzhai	Wangfen	Jingzhi	Chengguan.	Tanfang	Daotian
Whole	1						
Xinzhai	1	1					
Wangfen	0.852	0.852	1				
Jingzhi	0.999	0.999	0.849	1			
Chengguan.	1	0.999	0.853	0.999	1		
Tanfang	0.999	0.998	0.853	0.998	0.999	1	
Daotian	0.997	0.998	0.848	0.997	0.997	0.996	1

[h!]

In this perspective, efficiency scores were computed for four production functions, with order-m and order-$alpha$ estimations. They were compared through a Pearson correlation matrix (Table 4.16), and through a matrix of Spearman rank correlation (Table 4.17). Tables are presented for $alpha = 98\%$, but there is no big differences with other the chosen values. The minimal value for the Pearson correlation is 0.530, between Model 2 (inpatient days instead of admissions) in order-m and Model 5 (Inpatient + Outpatients and emergencies together) in order-$alpha$, so their is a strong common trend between all those specifications. More than a third of bilateral correlations are higher than 0.8, so there is no big contradiction between the different options. The lowest correlations (around 0.6) are often found while comparing models with the Global Activity Index with models with the inpatient admissions and the outpatients as outputs. This was to be expected, as these production functions contain different information. As explained in paragraph 4.4.1.1, for the output to exhaustively describe THCs activities, the Global Activity Index will be chosen in the main analysis.

The analysis of the Spearman rank correlations leads to the same conclusion : the minimal Spearman's rho (0.552) is obtained comparing Model 2 (Global Activity Index, Global Staff) and 6 (Inpatients + Outpatients and emergencies, medical + nonmedical staff) in order-m. This is the only value below 0.6. This means that, whatever the efficiency model or the production function, the most efficient THCs are always the same, as well as the worst performers. This confirms the validity of the first stage of the analysis.

It is also to be noticed that the scores estimated with the same production function but with a different method are always highly correlated: the minimal Spearman's Rho is 0.938 between the two models 1. Moreover, the Pearson correlation ranges between 0.792 and 0.896. Therefore, the choice of one method over the other does not impact the scores to a great extent.

Table 4.7: Tested models for the production function

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Outputs								
Inpatient admissions			√		√	√		√
Outpatients			√					
Outpatients + emergencies					√	√		√
Global Activity Index (IP days)		√						
Global Activity Index (IP admissions)	√			√			√	
Inputs								
Total staff	√	√	√		√			
Medical Staff				√		√	√	√
Nonmedical staff				√		√		
Equipment Index	√	√	√	√	√	√	√	√
Number of beds	√	√	√	√	√	√	√	√

4.4.3 Efficiency scores

4.4.3.1 Large potential for efficiency improvement

The mean technical efficiency remained quite stable over the period, around 0.75 from 2006 to 2012 (Tables 4.8 and 4.9). More than 20% are efficient (efficiency score equal to or higher than 1), other 20% exhibit a score between 0.9 and 1 (see Figure 4.26). The median efficiency score is between 0.77 and 0.87, which is quite high. This means that there is a large gap between efficient or almost efficient THCs, and inefficient DMUs which exhibit very weak scores, around 0.2 for the least efficient each year (see Figure 4.26 and Table 4.8). Yet, the number of very inefficient THCs is very low. 6 THCs per year (less than 20% of the sample) are smaller than 0.5 (see Figure 4.26).

There is no clear evolotion across years, both in average and in distribution. Technical efficiency remains stable from 2006 to 2012 in the sample.

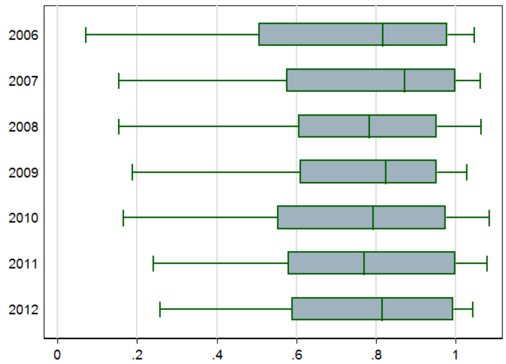


Figure 4.26: Distribution of efficiency scores across years

year	mean	N	min	max	p25	median	p75	sd
2006	0.734	30	0.071	1.056	0.510	0.816	0.980	0.268
2007	0.774	30	0.153	1.058	0.576	0.872	1.000	0.260
2008	0.745	30	0.154	1.089	0.606	0.783	0.952	0.243
2009	0.748	30	0.189	1.022	0.610	0.826	0.953	0.245
2010	0.752	30	0.166	1.061	0.553	0.797	0.976	0.245
2011	0.761	30	0.241	1.073	0.580	0.772	1.000	0.241
2012	0.769	28	0.257	1.052	0.590	0.816	0.990	0.228
Total	0.755	208	0.071	1.089	0.590	0.817	0.990	0.244

Table 4.8: Mean descriptives statistics about efficiency scores

Table 4.9: Mean descriptives statistics about efficiency scores, inefficient THCs only

year	mean	N	min	max	p25	median	p75	sd
2006	0.665	24	0.071	0.983	0.465	0.748	0.866	0.256
2007	0.619	18	0.153	0.989	0.481	0.627	0.834	0.228
2008	0.661	23	0.154	0.952	0.596	0.710	0.827	0.214
2009	0.683	24	0.189	0.990	0.533	0.715	0.862	0.230
2010	0.697	25	0.166	0.989	0.529	0.752	0.863	0.232
2011	0.668	22	0.241	0.989	0.496	0.686	0.831	0.214
2012	0.685	21	0.257	0.980	0.517	0.692	0.851	0.201
Total	0.670	157	0.071	0.990	0.501	0.696	0.850	0.223

4.4.3.2 No significant difference across counties

Most of decisions concerning THCs activity are made at the county level. Therefore some differences in the level of efficiency can be attributed to the county bureau (attribution of inputs for instance, or financial decisions). Table 4.10 suggests that 2 of the 8 counties are particularly efficient, Shouguang and Zhucheng with average efficiency of 0.87 and 0.84. However, the number of THCs per county is very variable, from 2 to 6, so it is not possible to directly read the means. Running statistical mean tests, only Shouguang exhibits an average efficiency significantly higher than the rest of the sample (Shouguang includes the townships of Daotian, Houzhen and Tianliu)[4].

Recalling that Changle and Linqu had a particular administrative status, this doesn't impact at all their THCs in terms of activity or efficiency.

Table 4.10: Efficiency scores across counties

year	mean	N	min	max	p25	median	p75	sd
Anqiu	0.744	48	0.215	1.036	0.590	0.805	0.973	0.249
Changle	0.797	14	0.553	1.015	0.613	0.803	1.000	0.190
Changyi	0.703	14	0.479	1.006	0.596	0.647	0.815	0.173
Gaomi	0.730	28	0.444	1.056	0.523	0.710	0.933	0.209
Linqu	0.710	28	0.071	1.073	0.393	0.840	1.000	0.342
Qingzhou	0.727	41	0.285	1.052	0.493	0.832	0.980	0.255
Shouguang	0.872	21	0.486	1.089	0.731	0.983	1.010	0.186
Zhucheng	0.844	14	0.383	1.000	0.795	0.855	1.000	0.167
Total	0.755	208	0.071	1.089	0.590	0.817	0.990	0.244

[4] Statistical mean tests were realized through the *ttest* Stata command

4.4.3.3 A link between efficiency and the size of THCs ?

A crucial issue in terms of policy is the optimal size of THCs, both in terms of available beds and staff. This is even more crucial in a context where the inputs increase as quick as the activity, even though efficiency was proven to be low at the beginning of the period.

A first graphical analysis, with Figure **??** does not allow to conclude in any way. Most THCs have less than 100 beds, and there is no common trend in their efficiency scores. Nevertheles, for the "big" THCs (with more than 100 available beds), the efficiency score is always higher than 0.6. As for the quantity of staff, the same analysis can be made. Nevertheless, some THCs employing more than 100 people exhibit low efficiency scores.

Statistical mean comparison tests confirm this idea: there is no statistical difference between small (less than 50) and medium (between 50 and 100) THCs, in terms of efficiency scores. On the contrary, efficiency scores of big hospitals are significantly higher than the one of small and medium THCs. There might be an effect of attraction coming from those well-equipied THCs, that could explain their higher scores.

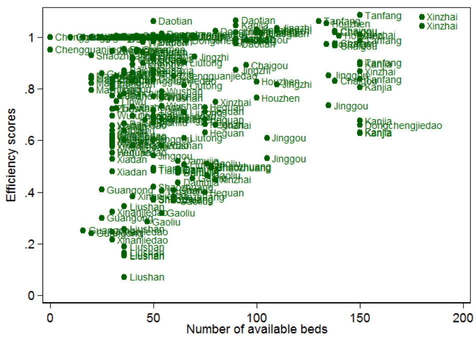


Figure 4.27: Efficiency scores and the number of available beds

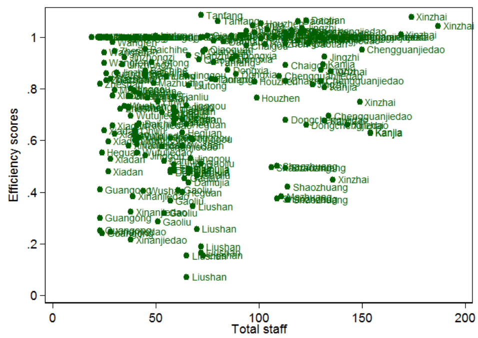


Figure 4.28: Efficiency scores and the total staff

4.5 Estimating the determinants of THs' efficiency

4.5.1 Empirical strategy

The two-stage approach normally uses the Tobit model to deal with the bounded nature of the DEA scores (Ng, 2011; Audibert et al., 2013; Cheng et al., 2015a). Nevertheless, due to skepticism regarding the censored nature of the boundary (Ramalho et al., 2010; Simar and Wilson, 2007), some alternative methods have been proposed (Simar and Wilson, 2007; Ramalho et al., 2010). Ramalho et al. (2010) proposes an instrumentalist approach, considering that efficiency scores are observed managerial performances regarding a best-observed practice, and not an estimate of the true efficiency of a score[5]. He proposes the second stage used here, based on fractional regression models (first developed by Wooldridge (2002)). As some of our THs exhibits order-m scores higher than 1, they have been normalized so that they would be bound between 0 and 1.

This second-stage procedure requires several hypotheses. First we must choose either a one-part or a two-part regression model, which would estimate different partial effects for efficient and inefficient firms. Here, as the proportion of efficient firms is small (20%), we considered a one-part model. Then, the link function (the distribution of the dependent variable conditionally to the environmental variable) for the model must be chosen, according to the results obtained by the p-tests. As robustness checks, regressions were tested with different link functions and with the DEA scores instead of order-m scores.

4.5.2 Results

Results of the second step are presented in Table 4.11, for normalized scores. The conclusions are robust to the use of DEA scores as dependent variable, and to other choices of link functions (Tables 4.18 and 4.19 in the Annex).

Most of the variables introduced to capture the effect of **the importance of subsidies** on efficiency exhibit a negative sign, giving evidence of a substantial phenomenon of soft budget constraint, especially when they are expressed in terms of ratios (proportion of subsidies in THCs income or expenditures). Subsidies induce perverse incentives, and no significant change is noticed between the two sub-periods. Thus, this situation did not change with the redefinition of THCs financing modalities. One of the goals of the reform has not been reached in this respect. Also, the net income per capita is never significant in the analysis.

Two main reasons to the effect of subsidies on efficiency can be proposed:

- the most inefficient hospitals relied the most upon subsidies before the reform. The subsidies were to offset financial losses in the first years (even for poor performers), so those THCs received substantial subsidies in the first years after the reform, as a compensation.

- the other explanation is the soft budget constraint phenomenon; THCs with poor performances expect to be bailed out at the end of the year, so they do

[5] True efficiency is here opposed to the efficiency *observed in the sample*

Table 4.11: Second stage of the analysis

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Covered population	0.069						
	-0.07						
Proportion of licensed staff	0.693	0.73	0.883	0.636	0.807	0.79	0.856
	-1.149	-1.162	-1.168	-1.188	-1.209	-1.164	-1.213
Number of VHS	-0.018***	-0.012**	-0.013**	-0.011**	-0.012**	-0.012**	-0.012*
	-0.005	-0.006	-0.006	-0.006	-0.006	-0.006	-0.007
Staff expenditures, per staff	0.130***	0.141***	0.109***	0.145***	0.129***	0.136***	0.110***
	-0.037	-0.036	-0.038	-0.039	-0.036	-0.035	-0.036
Test per outpatient	-2.520***	-2.702***	-2.500***	-2.756***	-2.792***	-2.708***	-2.306***
	-0.757	-0.777	-0.726	-0.772	-0.781	-0.769	-0.776
Inhabitants per bed	5.801***	6.670**	5.744***	5.896***	6.472***	6.022***	
	-1.582	-3.378	-1.945	-1.804	-1.947	-1.844	
2010/11/12	0.195	0.119	0.016	-0.197	0.052	0.158	-0.16
	-0.133	-0.296	-0.167	-0.157	-0.123	-0.13	-0.12
Prop. of subsidies in TH income	-1.047**	-1.035**	-2.094***				
	-0.423	-0.414	-0.739				
Inhabitants per bed * 2010/11/12		-0.675					
		-2.961					
Subsidies per inhabitant			0.010**	-0.001			
			-0.004	-0.002			
Proportion of Subs. * 2010/11/12			-0.789				
			-0.92				
Subs. Per Inhab. * 2010/11/12				-0.022***			
				-0.008			
Subs. Per medical staff					-0.024		
					-0.022		
Prop. of subs. in THC Expends						-0.921**	
						-0.403	
Subsidies per available bed							0
							0
Constant	1.118	1.296*	1.447**	1.383**	1.205*	1.266*	1.919***
	-0.716	-0.679	-0.718	-0.695	-0.703	-0.687	-0.683
Observations	206	206	206	206	206	206	208
R2	0.733	0.733	0.733	0.733	0.733	0.733	0.733

Logit functional form; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

not make any effort to maximize their activity. This negative effect of subsidies is also found in coastal Township Hospitals (Hu et al., 2012), confirming the perverse effects of inappropriate modalities of granting subsidies considering the overall context of the reform.

Two other internal variables are significantly associated with efficiency level: the **staff expenditures**, per medical staff (positive sign), and the **mean number of test per outpatient** (negative sign).

- The first one means that the economic incentives through remuneration can spur activity. This result can be qualified by the fact that we are not able to take the qualification and the experience of the staff into account. Still, it seems plausible and not surprising, as wages of health practionioners were proved to be very low (Qin et al., 2013), that higher wages (through bonus) can be a way to spur activity. Of course, for this system to be efficient, bonuses have to be connected to actual performances (and not to a THC bonus, itself coming from drug sales). This argument would require to check that the quality of care do not suffer from the increase inactivity, but no data is available on this matter.

- The number of tests per outpatient cases is significantly and negatively linked to the efficiency score. This variable was introduced to capture any overuse of imaging and testing in some THCs. Indeed, some inefficient THCs seem to rely on tests and imaging to increase their activity income.

Regarding the THCs environment on their performance, our results suggest that a low ratio of inhabitants per bed hinders the efficiency of THCs. This reflects an excess of inputs relatively to the current demand, as the THCs with the lower ratio are the least efficient. This results can be seen from two sides. From the demand size, the demand for health care is still not sufficient for the infrastructure, so there is a need to spur this demand through th development of the NRCMS benefit package for instance. From the supply side, there is a large potential in terms of productivity gain, that can be reached through proper incentives. Indeed the staff productivity is low

and to the increase in available beds in many THCs, though the BOR were not that high. There is a need for input adjustment to the demand in the catchment area. There is no significant correlation between efficiency and the covered population. Thus, the negative relation between inhabitants per bed and efficiency can be explained by the the density of health facilities here.

The number of VHS is also significantly related to efficiency scores, with a negative sign, suggesting that the competition effect is stronger than the referral effect for THCs in Weifang.
To sum up, the main results of the second stage are the following:

- the proportion of subsidies in the THCs income is negatively associated to their efficiency score.
- the amount of staff expenditures, per medical staff positively affects their efficiency
- the number of tests per outpatient is negativeley correlated to the efficiency scores
- the number of inhabitants per available bed affects positively the efficiency of the associated THC.

4.6 Discussion

Our results show that there is no improvement of THC efficiency between 2006 and 2012, contrary to what was expected of the pharmaceutical reform. A similar trend is found by Yang and Zeng (2014); Ng (2011), suggesting that health policy reform in Weifang prefecture shares some common characteristics with other areas. This feeds discussion about wider stakes of health policy reform in China.
First, as the core of the drug reform relies on reshaping TH financing from drug margins to subsidies, it is crucial that those subsidies should be allocated according to a clear formula. This formula should include clear performance incentives, linking subsidies to results, following a Result Based Financing (RBF) approach. THC managers should know at the beginning of the period how much their THCs will receive, rather than being easily bailed out at the end of the period.
Second, the optimal size of a Township Hospital is also a matter of concern. Our results suggest that the inputs of the THs are not fully utilized, since more inhabitants per bed lead to better performance. Yet, over the 2006-12 period, the number of available beds per TH grew drastically (from 44 to 77, on average, for the sample). In a situation of financial constraint, the priority should rather be the improvement of healthcare quality and affordability for households. Moreover, as pointed out by Yang and Zeng (2014), oversized hospitals may lead to quality concerns. In Weifang, several THCs were merged in recent years. In this study most of the biggest THCs are efficient ones, but the growing trend of the quantity of inputs lowered possible efficiency gains throughout the period.

Third, as the demand side is still important in explaining THCs efficiency, affordability of healthcare has to be improved, mainly by deepening the benefit packages of the NRCMS. In Weifang , outpatient activities were reimbursed only up to 30% in 2006, and many cases were not reimbursed at all. In 2012, user fees represent at least half the total costs of outpatient care, and from discussions with local NRCMS and Health Bureau officials, financial barriers to healthcare remain for some households.
Fourth, availability of essential drugs, periodic shortages have to be faced in several areas in China, despite the reform, is also a concern in some THCs in Weifang. At the national level, Li et al. (2013) point out that, because some ceiling prices have been fixed too low by the provincial government, manufacturers are reluctant to produce them, increasing the risk of shortages. A decline of quality is also a concern if cost is the crucial criteria to discriminate firms during the bidding process (Xiao et al., 2013). The lack of availability of some essential drugs can be challenging for THC daily activity and for healthcare costs. Indeed, since households have to get drugs at higher prices from places other than public facilities.
Finally, this study leads to the issue of provider payment, that has to be tackeld in Chinese Primary Healthcare Facilities. In the sample the fee-for-service method is still used, though it has been shown, in China and elsewhere, to be linked with higher costs than prospective payment methods (Yip et al., 2010). The introduction of prospective payment is in discussion in Weifang and appears to be an essential way to push the public healthcare system on the path of a better efficiency, particularly considering the multitasking nature of healthcare (Cheng et al., 2012; Eggleston, 2005; Robyn et al., 2014).

4.7 Conclusion

We use a two-stage partial frontier method to assess technical efficiency of THCs in Weifang, Shandong province, China. In a second stage, we use a fractional regression model in a second stage to explain THCS efficiency scores. Our results highlight the role of public subsidies and demand-side factors in the technical efficiency of Primary Healthcare Facilities in the rural prefecture of Weifang.
The technical efficiency remains constant over the period We identify a negative effect from the use of subsidies, showing that the soft budget constraint has not been efficiently tackled yet. Indeed, the subsidies allocation formula does not currently spur efficiency. More information must be collected regarding the real criteria of subsidies allocation, as several elements suggest that, in practice, actual processes may deviate from some of the criteria established in official texts. A greater under-

standing of this is necessary to assess how well the global scheme of incentives (in subsidies and elsewhere) is aligned, or not, with the health policy objectives.
At the same time, the effect of the covered population density per available bed, and of THCs being oversized in counties, highlights that focusing on the managerial aspects won't be sufficient to improve THC efficiency. Spurring activity needs outpatient healthcare to become more affordable and the NRCMS to enlarge its benefit packages, increasing reimbursement of outpatient care. Another crucial aspect is the perception of the quality of healthcare delivered in THC.

Bibliography

Audibert, M., Mathonnat, J., Pelissier, A., Huang, X. X., and Ma, A. (2013). Health insurance reform and efficiency of Township Hospitals in rural China: An analysis from survey data. *China Economic Review*, 27:326 – 338.

Bloom, G. and Xingyuan, G. (1997). Health sector reform: Lessons from China. *Social Science and Medicine*, 45(3):351–360.

Boussofiane, A., Dyson, R., and Thanassoulis, E. (1991). Applied Data Envelopment Analysis. *European Journal of Operational Research*, 52(1):1 – 15.

Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1):1–25.

Chen, M., Wang, L., Chen, W., Zhang, L., Jiang, H., and Mao, W. (2014). Does economic incentive matter for rational use of medicine? China's experience from the essential medicines program. *PharmacoEconomics*, 32(3):245–255.

Chen, S., Wu, Y., Chen, Y., Zhu, H., Wang, Z., Feng, D., Feng, Z., Yao, L., Xiang, L., yi Wong, E. L., Fung, H., kiong Yeoh, E., and Liu, Z. (2016). Analysis of operation performance of general hospitals in Shenzhen, China: a super-efficiency Data Envelopment Analysis. *The Lancet*, 388, Supplement 1:S57 –. The Lancet-CAMS Health Summit, 2016.

Cheng, S.-H., Chen, C.-C., and Tsai, S.-L. (2012). The impacts of DRG-based payments on health care provider behaviors under a universal coverage system: A population-based study. *Health Policy*, 107(2–3):202 – 208.

Cheng, Z., Cai, M., Tao, H., He, Z., Lin, X., Lin, H., and Zuo, Y. (2016). Efficiency and productivity measurement of rural Township Hospitals in China: a bootstrapping Data Envelopment Analysis. *BMJ Open*, 6(11):e0111911.

Cheng, Z., Tao, H., Cai, M., Lin, H., Lin, X., Shu, Q., and Zhang, R.-n. (2015a). Technical efficiency and productivity of chinese County Hospitals: an exploratory study in Henan province, China. *BMJ Open*, 5(9):e007267.

Cheng, Z.-H., Tao, H.-B., Cai, M., Lin, H.-F., Lin, X.-J., Shu, Q., and Zhang, R.-N. (2015b). Using a two-stage Data Envelopment Analysis to estimate the efficiency of county hospitals in China: a panel data study. *The Lancet*, 386, Supplement 1:S64 –. The Lancet-CAMS Health Summit Abstracts booklet.

Currie, J., Lin, W., and Meng, J. (2014). Addressing antibiotic abuse in China: An experimental audit study. *Journal of Development Economics*, 110(C):39–51.

Currie, J., Lin, W., and Zhang, W. (2011). Patient knowledge and antibiotic abuse: Evidence from an audit study in China. *Journal of Health Economics*, 30(5):933 – 949.

Daouia, A. and Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140(2):375–400.

Dong, K. (2009). Medical insurance system evolution in China. *China Economic Review*, 20(4):591–597.

Du, J. (2009). Economic reforms and health insurance in China. *Social Science & Medicine*, 69(3):387–395.

Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics*, 24(1):211 – 223.

Eggleston, K., Ling, L., Qingyue, M., Lindelow, M., and Wagstaff, A. (2008). Health service delivery in China: a literature review. *Health Economics*, 17(2):149–165.

Fang, Y., Wagner, A. K., Yang, S., Minghuan Jiang, M., Zhang, F., and Ross-Degnan, D. (2013). Access to affordable medicines after health reform: evidence from two cross-sectional surveys in Shaanxi province, western China. *Lancet Global Health*, 1:e227–e237.

Giuffrida, A. and Gravelle, H. (2001). Measuring performance in primary care: econometric analysis and DEA. *Applied Economics*, 33(2):163–175.

Henderson, G., Shuigao, J., Akin, J., Zhiming, L., Jianmin, W., Haijiang, M., Yunan, H., Xiping, Z., Ying, C., and Keyou, G. (1995). Distribution of medical insurance in China. *Social Science & Medicine*, 41(8):1119 – 1130.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, 17(10):1107–1128.

Hu, H.-H., Qi, Q., and Yang, C.-H. (2012). Analysis of hospital technical efficiency in China: Effect of health insurance reform. *China Economic Review*, 23(4):865–877.

Hu, T.-W., Ong, M., Lin, Z.-H., and Li, E. (1999). The effects of economic reform on health insurance and the financial burden for urban workers in China. *Health Economics*, 8(4):309–321.

Jacobs, R., Smith, P. C., and Street, A. (2006). *Measuring Efficiency in Health Care: Analytic Techniques and Health Policy*. Cambridge University Press.

Kornai, J. (2009). The soft budget constraint syndrome in the hospital sector. *International Journal of Health Care Finance and Economics*, 9(2):117–135.

Lei, X., Yin, N., and Zhao, Y. (2012). Socioeconomic status and chronic diseases: The case of hypertension in China. *China Economic Review*, 23(1):105 – 121.

Li, H., Dong, S., and Liu, T. (2014a). Relative efficiency and productivity: a preliminary exploration of public hospitals in Beijing, China. *BMC Health Services Research*, 14(158).

Li, M., Zhu, Y., Xue, C., Liu, Y., and Zhang, L. (2014b). The problem of unreasonably high pharmaceutical fees for patients in Chinese hospitals: A system dynamics simulation model. *Computers in Biology and Medicine*, 47:58 – 65.

Li, Y., Ying, C., Sufang, G., Brant, P., Bin, L., and Hipgrave, D. (2013). Evaluation, in three provinces, of the introduction and impact of China's National Essential Medicines Scheme. *Bulletin of the World Health Organization*, 91:184 – 194.

Lindelow, M. (2005). *Health Shocks In China : Are The Poor And Uninsured Less Protected ?* The World Bank.

Liu, G., Liu, X., and Meng, Q. (1994). Privatization of the medical market in socialist China: A historical approach. *Health Policy*, 27(2):157 – 174.

Liu, J. S., Lu, L. Y., Lu, W.-M., and Lin, B. J. (2013). A survey of DEA applications. *Omega*, 41(5):893 – 902.

Liu, X. and Hsiao, W. C. (1995). The cost escalation of social health insurance plans in China: Its implication for public policy. *Social Science & Medicine*, 41(8):1095 – 1101. Aspects of the Medical Care System in the People's of China.

Liu, X., Liu, Y., and Chen, N. (2000). The Chinese experience of hospital price regulation. *Health Policy and Planning*, 15(2):157–163.

Liu, X. and Mills, A. (2005). The effect of performance-related pay of hospital doctors on hospital behaviour: a case study from Shandong, China. *Hum Resour Health*, 3:11.

Liu, Y., Hsiao, W. C., Li, Q., Liu, X., and Ren, M. (1995). Transformation of Chinaś rural health care financing. *Social Science & Medicine*, 41(8):1085 – 1093.

Milcent, C. and Wu, B. (2015). How Do You Feel? The Effect of the New Cooperative Medical Scheme in China. *Journal of Development Studies*, 51(12):1585–1602.

Ng, Y. C. (2011). The productive efficiency of Chinese hospitals. *China Economic Review*, 22(3):428 – 439.

O'Neill, L., Rauner, M., Heidenberger, K., and Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences*, 42(3):158–189.

Pélissier, A. (2012). *Activités et efficience des établissements de santé dans le contexte de la couverture universelle de santé: Etudes sur données d'enquêtes au Cambodge et en Chine*. PhD thesis, Université d'Auvergne - CERDI.

Pélissier, A., Audibert, M., Mathonnat, J., Huan, X., and Ma, A. (2012). Curative Activities of Township Hospitals in Weifang Prefecture, China: An Analysis of Environmental and Supply-side Determinants. *Économie publique/Public economics*, (28-29):71–94.

Qin, X., Li, L., and Hsieh, C.-R. (2013). Too few doctors or too low wages? labor supply of health care professionals in China. *China Economic Review*, 24:150 – 164.

Ramalho, E., Ramalho, J., and Henriques, P. (2010). Fractional regression models for second stage DEA efficiency analyses. *Journal of Productivity Analysis*, 34(3):239–255.

Robyn, P. J., Bärnighausen, T., Souares, A., Traoré, A., Bicaba, B., Sié, A., and Sauerborn, R. (2014). Provider payment methods and health worker motivation in community-based health insurance: A mixed-methods study. *Social Science and Medicine*, 108(C):223–236.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.

Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1):31–64.

Simar, L. and Wilson, P. W. (2013). Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. *Foundations and Trends(R) in Econometrics*, 5(3–4):183–337.

Sun, Q., Liu, X., Meng, Q., Tang, S., Yu, B., and Tolhurst, R. (2009). Evaluating the financial protection of patients with chronic disease by health insurance in rural China. *International Journal for Equity in Health*, 8(1):42.

Sun, Q., Santoro, M. A., Meng, Q., Liu, C., and Eggleston, K. (2008). Pharmaceutical Policy In China. *Health Affairs*, 27(4):1042–1050.

Varabyova, Y. and Schreyögg, J. (2013). International comparisons of the technical efficiency of the hospital sector: Panel data analysis of {OECD} countries using parametric and non-parametric approaches. *Health Policy*, 112(1–2):70 – 79. Health System Performance Comparison: New Directions in Research and Policy.

Wagstaff, A. and Yu, S. (2007). Do health sector reforms have their intended impacts?: The World Bank's Health VIII project in Gansu province, China. *Journal of Health Economics*, 26(3):505–535.

Wang, L., Kong, L., Wu, F., Bai, Y., and Burton, R. (2005). Preventing chronic diseases in China. *The Lancet*, 366(9499):1821–1824.

WB (2010). The path to integrated insurance systems in China. China Health Policy Notes 3, World Bank, Washington, DC.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Xiao, Y., Zhao, K., Bishai, D., and Peters, D. (2013). Essential drugs policy in three rural counties in China: What does a complexity lens add? *Social Science and Medicine*, 93:220–228.

Xueshan, F., Shenglan, T., Bloom, G., Segall, M., and Xingyuan, G. (1995). Cooperative medical schemes in contemporary rural China. *Social Science & Medicine*, 41(8):1111 – 1118.

Yang, G., Lingzhi, K., Wenhua, Z., Xia, W., Yi, Z., C, C. L., and P., K. J. (2008). Emergence of chronic non-communicable diseases in China. *The Lancet*, 372(9650):1697–1705.

Yang, J. and Zeng, W. (2014). The trade-offs between efficiency and quality in the hospital production: Some evidence from shenzhen, China. *China Economic Review*, 31(C):166–184.

Yang, W. and Wu, X. (2015). Paying for outpatient care in rural China: cost escalation under China's New Cooperative Medical Scheme. *Health Policy and Planning*, 30(2):187–196.

Yip, W., Hsiao, W., Chen, W., Hu, S., Ma, J., and Maynard, A. (2012). Early appraisal of China's huge and complex health-care reforms. *The Lancet*, 379(9818):833–842.

Yip, W. and Hsiao, W. C. (2009). Non-evidence-based policy: How effective is China's New Cooperative Medical Scheme in reducing medical impoverishment? *Social Science and Medicine*, 68(2):201 – 209.

Yip, W. C.-M., Hsiao, W., Meng, Q., Chen, W., and Sun, X. (2010). Realignment of incentives for health-care providers in China. *Lancet (London, England)*, 375(9720):1120–1130.

You, X. and Kobayashi, Y. (2009). The New Cooperative Medical Scheme in China. *Health Policy (Amsterdam, Netherlands)*, 91(1):1–9.

Young, M. E. (1988). Impact of the rural reform in financing rural health services in China. Working Paper 31568, World Bank.

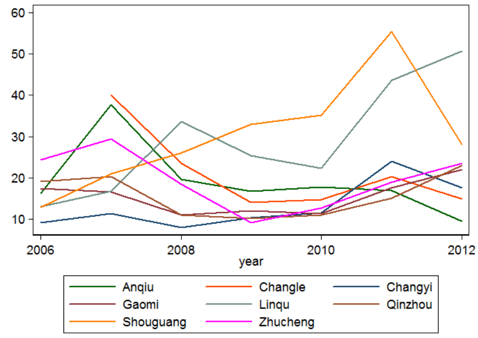
Yu, X., Li, C., Shi, Y., and Yu, M. (2010). Pharmaceutical supply chain in China: Current issues and implications for health system reform. *Health Policy*, 97(1):8 – 15.

Yuan, Q. and Tang, S. (2012). Comparative analysis of compensation model of essential drug system and its compensation mechanism research. *China Medical Herald*, 9(30):147–151.

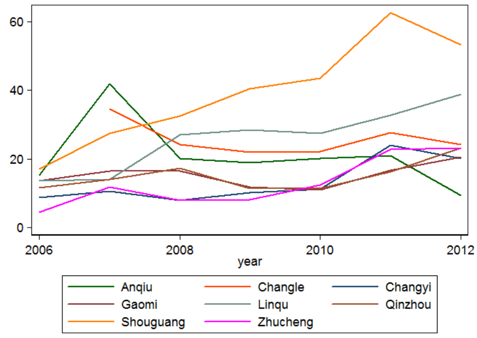
Zhu, N., Ling, Z., Shen, J., Lane, J. M., and Hu, S. (1989). Factors associated with the decline of the cooperative medical system and barefoot doctors in rural China. *Bulletin of the World Health Organization*, 67(4):431–441.

Zhuo, L. and Zou, L. (2012). Discussion on compensation mechanism of public hospital after cancelling drug price addition. *China Pharmacy 23*, 23(1).

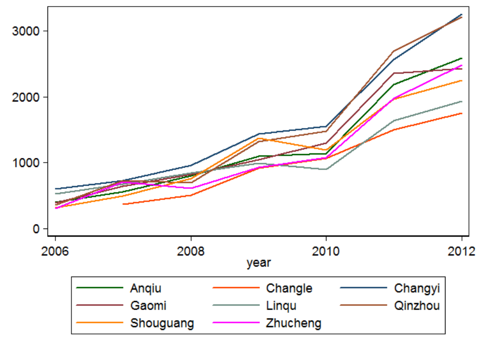
4.8 Appendix



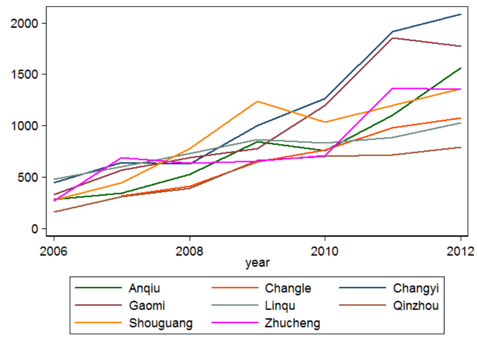
a. Outpatients



b. Outpatients in THC

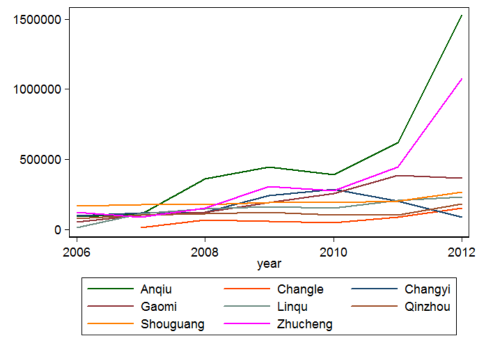


c. Inpatients

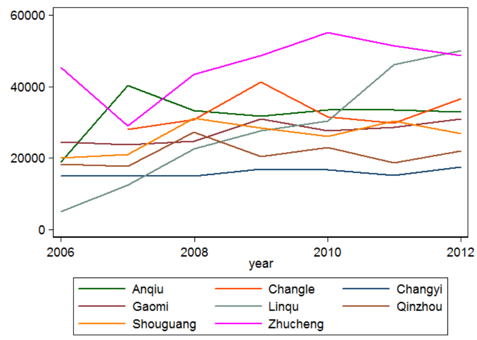


d. Inpatients in THC

Figure 4.29: Evolution of mean reimbursed amount per case in Weifang, in yuans



a. Outpatient cases in THC



b. Inpatient cases in THC

Figure 4.30: Evolution of the number of cases in Weifang

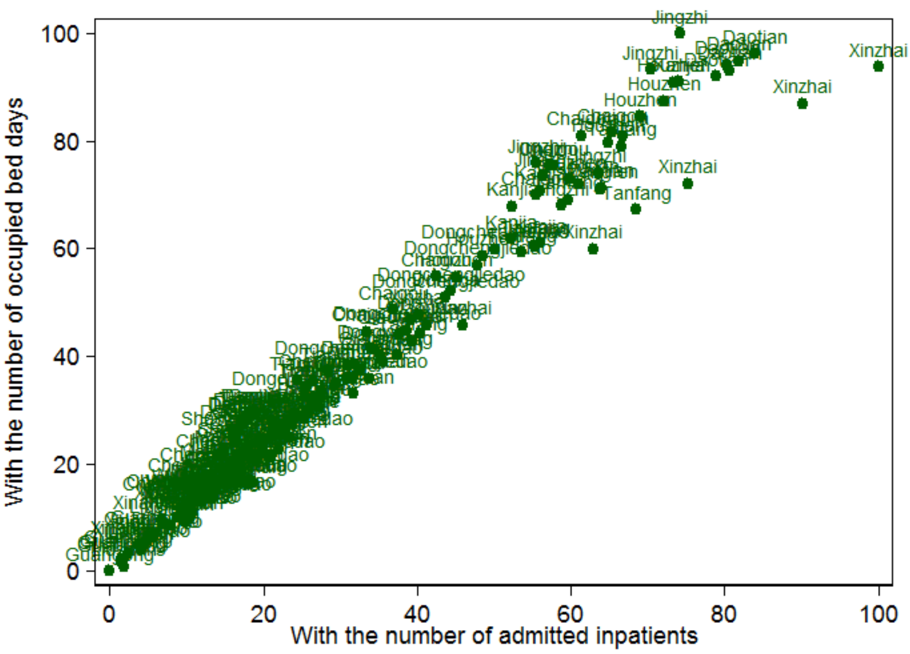


Figure 4.31: Comparison of the first axis scores

Table 4.12: Evolution of main activity and resources variables

	Mean			Median			Standard deviation		Relative std deviation	
Variable	**2006**	**Var(%)**	**2012**	**2006**	**Var(%)**	**2012**	**2006**	**2012**	**2006**	**2012**
Number of outpatients	33087	82.22	60290	19612	63.07	31982	27566	62230	0.83	1.03
Number of admitted inpatients	1510	125.85	3411	1164	118.9	2548	1024	2912	0.68	0.85
Number of tests	5417	75.27	9494	3320	105.29	6815	5629	8914	1.04	0.94
Number of surgeries	172	26.16	217	117	17.17	137	185	238	1.08	1.1
Number of vaccinations	11178	43.15	16002	11278	27.83	14416	7619	20391	0.68	1.27
Number of antenatal visits	1678	0.3	1683	1273	-24.19	965	1545	1561	0.92	0.93
Number of emergencies	605	119.95	1331	157	137.06	371	1144	2151	1.89	1.62
Global Activity Index (see Section 5.a)	21	69.87	36	15	63.98	25	15	27	0.69	0.75
Number of available beds	44	74.55	77	38	60.53	61	24	46	0.55	0.59
Staff of the TH	65	18.46	77	57	19.47	68	36	40	0.55	0.52
Equipment Index (see Section 5.a)	23	84.59	42	21	68.94	36	14	25	0.62	0.59

	Outpatients	Admitted inpatients	Emergencies	Tests	Surgeries	Vaccinations	Antenatal visits
Outpatients	1						
Admitted inpatients	0.420*	1					
Emergencies	0.220*	0.294*	1				
Tests	0.330*	0.467*	0.296*	1			
Surgeries	0.447*	0.628*	0.195*	0.641*	1		
Vaccinations	0.234*	0.342*	0.219*	0.305*	0.422*	1	
Antenatal visits	0.231*	0.331*	0.280*	0.084	0.415*	0.347*	1

**: significant at a 95% level of confidence*

Table 4.13: Correlation matrix between the outputs

Table 4.14: Nature of the 15 most sold drugs in the THCs of the sample

	Injections		Antibiotic	
	2011	2012	2011	2012
Guangong	6	13	0	2
Jingzhi	15	13	3	3
Jinzhongzi	8	6	2	1
Zheshan	5	1	1	7
Xinanjiedao	10	10	3	2
Wushan	8	9	4	3
Linwu	6	6	5	5
Qiaoguan	12	14	3	3
Wutujiedao	11	13	4	5
Liutong	10	8	2	0
Xiadan	13	11	7	4
Chaigou	11	13	4	5
Damujia	5	5	3	3
Jinggou	13	12	6	7
Kanjia	7	8	3	1
Chengguanjiedao	7	7	8	6
Dongchengjiedao	13	11	3	8
Liushan	13	12	7	4
Xinzhai	15	15	2	2
Dongxia	10	5	5	5
Gaoliu	11	9	5	5
Heguan	13	12	8	6
Shaozhuang	11	11	5	7
Tanfang	10	11	6	8
Wangfen	10	8	9	5
Daotian	5	5	5	5
Houzen	14	11	7	8
Tianlu	9	12	6	2
Baichihe	11	11	6	7
Mazhuang	9	9	6	6
Mean	10.0	9.7	4.6	4.5
Median	10	11	5	5

Table 4.15: Scores of the potential outliers according to the value of m

		Value of m																		
THC	year	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190
Jingzhi	2006	1.374	1.100	0.993	0.947	0.937	0.953	0.926	0.933	0.927	0.925	0.927	0.925	0.924	0.924	0.924	0.924	0.924	0.924	0.924
Jingzhi	2007	1.774	1.415	1.227	1.119	1.127	1.070	1.043	1.018	1.031	1.012	1.000	1.006	1.000	1.006	1.006	1.000	1.000	1.000	1.000
Jingzhi	2008	1.832	1.681	1.341	1.341	1.237	1.219	1.118	1.133	1.073	1.069	1.072	1.033	1.039	1.033	1.053	1.019	1.019	1.014	1.010
Jingzhi	2009	1.466	1.244	1.007	0.998	0.949	0.926	0.925	0.900	0.914	0.895	0.886	0.885	0.885	0.881	0.885	0.881	0.884	0.881	0.879
Jingzhi	2010	1.521	1.267	1.047	1.023	0.946	0.945	0.915	0.882	0.863	0.849	0.857	0.844	0.842	0.826	0.824	0.822	0.825	0.819	0.820
Jingzhi	2011	2.470	1.706	1.676	1.454	1.373	1.311	1.214	1.248	1.125	1.113	1.083	1.063	1.087	1.051	1.044	1.029	1.019	1.029	1.031
Jingzhi	2012	2.009	1.488	1.363	1.243	1.218	1.238	1.125	1.091	1.065	1.066	1.052	1.041	1.032	1.038	1.022	1.024	1.024	1.021	1.021
Chengguanjiedao	2006	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952
Chengguanjiedao	2007	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chengguanjiedao	2008	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chengguanjiedao	2009	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chengguanjiedao	2010	0.836	0.759	0.736	0.717	0.712	0.709	0.699	0.702	0.699	0.702	0.695	0.696	0.697	0.695	0.694	0.696	0.695	0.696	0.693
Chengguanjiedao	2011	0.989	0.898	0.872	0.861	0.854	0.849	0.838	0.841	0.839	0.839	0.832	0.832	0.833	0.833	0.832	0.832	0.832	0.832	0.831
Chengguanjiedao	2012	0.977	0.738	0.779	0.696	0.621	0.592	0.576	0.567	0.564	0.583	0.547	0.568	0.528	0.530	0.523	0.516	0.512	0.507	0.511
Xinzhai	2011	1.546	1.298	1.284	1.224	1.188	1.147	1.152	1.153	1.116	1.142	1.111	1.108	1.102	1.087	1.097	1.094	1.081	1.102	1.075
Xinzhai	2012	1.496	1.334	1.268	1.240	1.179	1.153	1.168	1.115	1.108	1.094	1.086	1.082	1.078	1.070	1.064	1.066	1.060	1.059	1.055
Tanfang	2010	2.091	1.761	1.444	1.334	1.232	1.186	1.149	1.151	1.152	1.106	1.082	1.062	1.060	1.047	1.042	1.029	1.034	1.038	1.017
Wangfen	2012	1.629	1.191	1.122	1.039	1.019	1.006	1.006	1.000	1.006	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Daotian	2008	1.868	1.523	1.477	1.358	1.329	1.255	1.261	1.200	1.170	1.184	1.179	1.136	1.115	1.089	1.094	1.125	1.089	1.095	1.054
Daotian	2009	1.654	1.384	1.281	1.220	1.111	1.122	1.081	1.095	1.059	1.072	1.038	1.050	1.030	1.031	1.024	1.021	1.010	1.008	1.021
Daotian	2010	1.630	1.400	1.226	1.219	1.144	1.117	1.072	1.069	1.047	1.042	1.026	1.031	1.013	1.015	1.005	1.010	0.990	0.998	0.995
Daotian	2011	1.725	1.455	1.305	1.215	1.156	1.126	1.119	1.074	1.042	1.035	1.047	1.029	1.020	0.999	1.005	1.009	0.981	0.986	0.989
Daotian	2012	1.680	1.406	1.309	1.151	1.114	1.087	1.098	1.058	1.043	1.040	1.024	1.002	0.988	1.005	0.994	0.993	0.983	0.979	0.981

Table 4.16: Correlation matrix between the scores of the different models

	M1	M2	M3	M4	M5	M6	M7	M8	A1	A5	A7	A8
M1	1											
M2	0.948	1										
M3	0.677	0.592	1									
M4	0.932	0.883	0.700	1								
M5	0.692	0.607	0.997	0.714	1							
M6	0.671	0.581	0.948	0.724	0.949	1						
M7	0.991	0.933	0.680	0.938	0.695	0.672	1					
M8	0.695	0.602	0.988	0.717	0.989	0.954	0.699	1				
A1	0.886	0.876	0.587	0.813	0.604	0.575	0.885	0.603	1			
A5	0.602	0.530	0.785	0.591	0.792	0.737	0.604	0.788	0.717	1		
A7	0.888	0.865	0.597	0.827	0.614	0.587	0.896	0.615	0.990	0.712	1	
A8	0.607	0.532	0.782	0.599	0.790	0.741	0.612	0.793	0.723	0.996	0.721	1

Note: M is for order-m (m=190), and A for order-*alpha* (*alpha* = 97%). The figures refer to those in Table 4.7
All the correlations are significantly different from 0 with a 95% degree of confidence.

Table 4.17: Spearman rank correlation matrix between the scores of the different models

	M1	M2	M3	M4	M5	M6	M7	M8	A1	A5	A7	A8
M1	1											
M2	0.941	1										
M3	0.659	0.588	1									
M4	0.921	0.852	0.611	1								
M5	0.676	0.603	0.994	0.628	1							
M6	0.624	0.552	0.931	0.644	0.930	1						
M7	0.980	0.916	0.654	0.928	0.670	0.615	1					
M8	0.673	0.600	0.989	0.631	0.991	0.927	0.675	1				
A1	0.938	0.925	0.647	0.872	0.658	0.622	0.933	0.664	1			
A5	0.669	0.609	0.951	0.638	0.950	0.895	0.674	0.953	0.707	1		
A7	0.931	0.907	0.641	0.884	0.653	0.618	0.947	0.661	0.985	0.703	1	
A8	0.672	0.609	0.943	0.644	0.941	0.887	0.681	0.952	0.712	0.995	0.712	1

Note: M is for order-m (m=190), and A for order-*alpha* (*alpha* = 97%). The figures refer to those in Table 4.7
All the p-values for the bilateral independence tests are 0.000.

Table 4.18: Robustness for the second stage: Probit link function

VARIABLES	Order-m score	Order-m score	Order-m score	Order-m score	Order-m score	Order-m score	Order-m score
Covered population	0.052						
	-0.041						
Proportion of licensed staff	0.437	0.469	0.546	0.416	0.514	0.502	0.553
	-0.69	-0.697	-0.701	-0.711	-0.721	-0.698	-0.727
Number of VHS	-0.011***	-0.008**	-0.008**	-0.007**	-0.007**	-0.008**	-0.007*
	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.004
Staff expenditures, per staff	0.073***	0.081***	0.063***	0.083***	0.074***	0.079***	0.063***
	-0.022	-0.021	-0.022	-0.023	-0.021	-0.02	-0.021
Test per outpatient	-1.360***	-1.481***	-1.360***	-1.519***	-1.536***	-1.487***	-1.256***
	-0.447	-0.464	-0.435	-0.458	-0.462	-0.456	-0.468
Inhabitants per bed	3.046***	3.488*	2.885***	3.025***	3.332***	3.099***	
	-0.684	-2.012	-0.811	-0.781	-0.841	-0.796	
2010/11/12	0.119	0.072	0.005	-0.116	0.031	0.095	-0.091
	-0.079	-0.176	-0.098	-0.092	-0.073	-0.077	-0.071
Prop. of subsidies in TH income	-0.629**	-0.617**	-1.208***				
	-0.248	-0.243	-0.43				
Inhabitants per bed * 2010/11/12		-0.386					
		-1.743					
Subsidies per inhabitant			0.006**	-0.001			
			-0.002	-0.001			
Proportion of Subs. * 2010/11/12			-0.526				
			-0.545				
Subs. Per Inhab. * 2010/11/12				-0.013***			
				-0.005			
Subs. Per medical staff					-0.014		
					-0.013		
Prop. of subs. in THC Expends						-0.554**	
						-0.234	
Subsidies per available bed							0
							0
Constant	0.653	0.792**	0.892**	0.848**	0.749*	0.780**	1.079***
	-0.412	-0.384	-0.404	-0.393	-0.398	-0.388	-0.4
Observations	206	206	206	206	206	206	208
R2	0.731	0.731	0.731	0.731	0.731	0.731	0.731

Probit functional form; *** p<0.01, ** p<0.05, * p<0.1

Table 4.19: Robustness of the second stage: Fractional Regression Model with DEA score as dependent variables

VARIABLES	DEA score	DEA score	DEA score	DEA score	DEA score	DEA score	DEA score
Covered population	0.07						
	-0.076						
Proportion of licensed staff	0.027	0.078	0.103	-0.115	0.087	0.123	0.178
	-1.172	-1.186	-1.205	-1.206	-1.246	-1.199	-1.276
Number of VHS	-0.015**	-0.010**	-0.010**	-0.009*	-0.009*	-0.010**	-0.007
	-0.006	-0.005	-0.005	-0.005	-0.005	-0.005	-0.006
Staff expenditures, per staff	0.151***	0.161***	0.135***	0.171***	0.139***	0.156***	0.125***
	-0.036	-0.037	-0.037	-0.037	-0.034	-0.034	-0.034
Test per outpatient	-2.073***	-2.235***	-2.063***	-2.377***	-2.418***	-2.296***	-1.989***
	-0.754	-0.775	-0.756	-0.783	-0.796	-0.784	-0.762
Inhabitants per bed	4.487***	4.547	4.139***	4.441***	4.986***	4.549***	
	-1.311	-3.264	-1.436	-1.389	-1.553	-1.448	
2010/11/12	0.117	0.096	-0.122	-0.308**	-0.106	0.056	-0.311***
	-0.115	-0.282	-0.149	-0.139	-0.102	-0.109	-0.105
Prop. of subsidies in TH income	-1.443***	-1.410***	-2.192***				
	-0.406	-0.409	-0.818				
Inhabitants per bed * 2010/11/12		-0.02					
		-2.794					
Subsidies per inhabitant			0.008*	-0.003			
			-0.005	-0.002			
Proportion of Subs. * 2010/11/12			-1.375				
			-1.005				
Subs. Per Inhab. * 2010/11/12				-0.024***			
				-0.009			
Subs. Per medical staff					-0.02		
					-0.02		
Prop. of subs. in THC Expends						-1.195***	
						-0.385	
Subsidies per available bed							0.001
							-0.001
Constant	1.167*	1.391**	1.592**	1.482**	1.334**	1.353**	1.742***
	-0.702	-0.621	-0.649	-0.628	-0.644	-0.626	-0.629
Observations	207	207	207	207	207	207	209
R2	0.729	0.729	0.729	0.729	0.729	0.729	0.729

Logit functional form; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Chapter 5

The efficiency of Ulan-Bataar Family Health Centers

[1]

This Chapter is a case study focused on Ulan-Bator primary healthcare facilities, the *Family Health Centers*, from 2011 to 2014, relying on survey data. The sample includes all the FHCs of Ulan-Bator. The study adopts a two-stage method, to estimate efficiency scores, and explain them by potential determinants. It concludes that there is a potential space for improvement, and gives some policy recommendations to spur activity ans performance.

5.1 Introduction

Universal access to primary health care is a major objective in most developing countries. Including prevention and diagnosis of the commonest diseases, it plays a crucial role in the fight against comunicable diseases, and in the improvement of maternal and child health. Moreover, it was proved to be more cost-efficient than more sophisticated care, as the unit costs are weak, and it allows to prevent more large expenses in case of disease. Reaching an entire population, at a reasonable cost requires an adequate health system, both in terms of organization and incentives. This chapter is a case study. It focuses on the activity and efficiency of primary

[1]This chapter comes from a research program on the efficiency of Family Health Centers of the city of Ulan-Bator, Mongolia, under the joint supervision of J. Mathonnat (CERDI) and Pr. L. Munk-Erdene, Head of the Department of Health Policy and Management, School of Public Health, Mongolian National University of Medical Sciences, Ulan-Bator. It benefited from funds from the National Reasearch Agency through the program "*Investissement d'Avenir*" (ANR-10-LABEX-14-10-01) to the CERDI and from financial support from the FERDI (*Fondation pour les Études et les Recherches sur le Développement International*). This program was conducted in collaboration with a team of researchers from the Mongolian National University of Medical Sciences. A preliminary version of the study is at the core of a research report whom I am co-author (M. Audibert, J. Mathonnat, L. Petitfour, L. Munkh-Erdene, A. Tevendorj, L. Undram, *Improving Productivity and Efficiency of Family Health Centers in Ulanbataar – A priority for health policy in Mongolia*, FERDI, september 2016, 59 p.). I was involved in its presentation and dsicussion in Ulan-Bator with the Health Department of Ulan-Bator city (Family Health Centers supervision). I remained solely responsible of this analysis presented here.

health care facilities (the so-called *Family Health Centers*, FHCs hereafter) in Ulan-Bator, capital of Mongolia. Mongolia is a former communist country, so since 1991 its health system has changed a lot, to meet the population needs while keeping the system affordable for users. It benefited from a favorable economic context for two decades, but now faces financial constraints.
Relying on survey data from all FHCs in Ulan-Bator drom 2012 to 2014, this study aims at estimating their efficiency, and explaining it by environmental and managerial factors through a two-stage procedure. Results show that there is a large variability in FHCs performances, meaning that there is some potential for improvement. In average, FHCs could increase their activity by 30% with the same level of resources.
Section 5.2 presents the Mongolian context in terms of health system, and the current stakes. Section 5.3 presents the survey data, and the main characteristics of the Family Health Centers. Section 5.4 details the first stage of the efficiency analysis, i.e. the calculation of the efficiency scores. Finally, Section 5.5 presents the second stage of the study, the identification of the determinants of efficiency.

5.2 The Mongolian Health System: background and context

The Section presents the specificities of the Mongolian health system, its pyramidal structure and the precise role of FHCs, our focus level. It also provides wider elements of context of the Mongolian economy.

5.2.1 The post-communist mutations

From its communist era (1941-1991), Mongolia inherited a state-managed health system, financed to a large extent by the USSR (up to 35% of Mongolian governement annual budget, (Manaseki, 1993). The health system became market-based, and the focus was put on prevention and primary healthcare thanks to a free access ensured by the Constitution of 1992 (WHO and MoH, 2012).
As a result, a noticable shift can be noticed in the composition of health expenditures (Figure 5.1). Until 1999, government expenditures represented nearly 80% of health expenditures. This proportion decreased to 50% approximately, and has stayed around this proportion since 2002. Figure 5.2 gives a similar piece of information: health expenditures are almost exclusively public until 2000, then private expenditures constitute most on the increase of health expenditures from 2000 to 2005. Figure 5.3 highlights the fact that between 2000 and 2005, the proportion

of health expenditures in GDP increased from 4 to 6 %, but only due to the increase in private expenditures, since the proportion of general health expenditures fell from 4 to 2%. Figure 5.4 underlines the explosion of total government expenditures from 2005 to 2015, from which general governement health expenditures are totally excluded. Indeed, their proportion fell from 11% in 2000 to 5% in 2014.

Since 2005, total and government health expenditures have followed the same evolution, suggesting that private expenditures have remained stable. In terms of amouts, total expenditures tripled between 1995 and 2005 (from 100 to 300 Int $) and then doubled from 2005 to 2015 (Figure 5.2)[2].

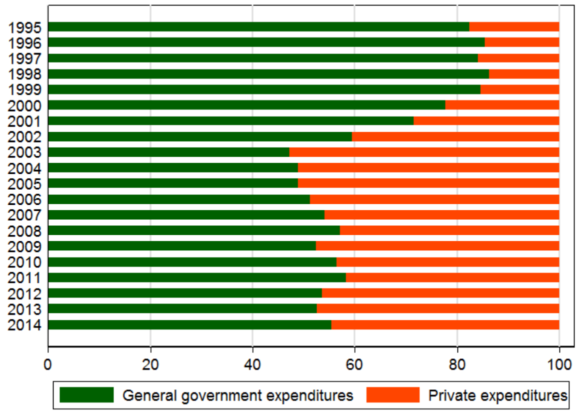


Figure 5.1: Evolution of the percentage of public and private health expenditures

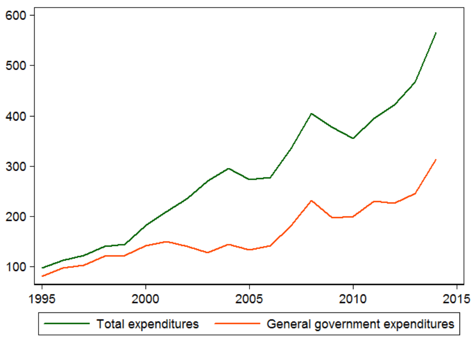


Figure 5.2: Evolution of health expenditures per capita, Int $

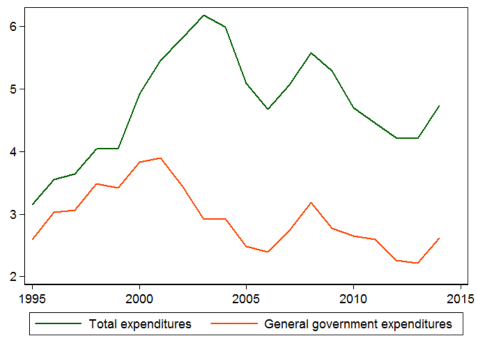


Figure 5.3: Evolution of the percentage of health expenditures in GDP

[2]Source of the data for Figures 5.1 to 5.5: World Health Organization Data Repository

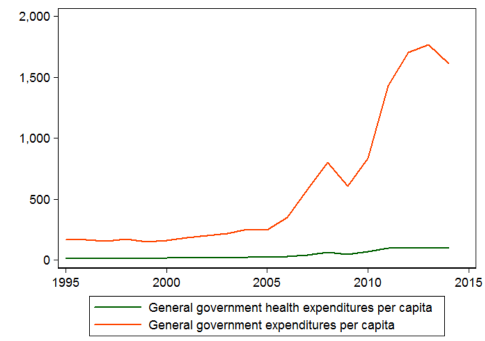


Figure 5.4: Evolution of general government health expenditures and general government expenditures per cap., Int $

Figure 5.5: Evolution of the percentage of general government health expenditures in general govenment expenditures

5.2.2 A three-tier system

As in China, The Mongolian Health system is organized as two three-level pyramids, following the administrative stratification of the country: one for rural areas, one for urban areas.

Mongolia is divided into 21 provinces (so-called *aimag*), subdivided in *soums* (and then in *baghs* in some areas). Apart from this organization, Ulan-Bator and big cities are divided into districts, and subdivided into *khoroos* or subdistricts.

A detailed description of the role of each type of facility is provided by WHO and MoH (2012).

Rural areas The healthcare system is composed of *Bagh feldshers*, *Soum health centres* (between 15 and 30 beds) and *Inter-Soum hospitals* at the primary level. The secondary level facility is the *Aimag hospital* (between 105 and 500 beds). At the tertiary level, there are four regional diagnostic and treatment center for specialized care.

Ulan-Bator and other big cities Primary level health care is provided at the Family Health Center (FHC hereafter), private facilities contracted with the city. Their missions are focused on basic and preventive care (monitoring, vaccinations, diagnostics tests). They refer patients to the district hospitals and central hospitals, i.e. secondary and tertiary level facilities, if needed.

FHCs are private entities, subjected to a three-peer contract with the disctrict Health department (municipal authority) and the State Health Department).

Their performances are supposed to be assessed to ensure the terms of the contract. The FHC earns some extra money if its performance is very good, and receive a warning that can lead to a break of the contract if the performance is very poor. But, according to our discussions with Ulan-Bator City Health department, this never or very rarely happens, the sanction is not really credible. This, associated to a lack of supervision, raises the issue of the incentives for FHCs to spur activity.

5.2.3 The central place of Ulan-Bator

A major characteristic of Mongolia is the importance of its capital, relatively to the rest of the country. Almost one half of the Mongolian population lives in Ulan-Bator, and this pregnance of the capital became stronger over the last two decades, due to a very fast urbanization, for several reasons.

First, hard living conditions in the rural areas (climatic conditions, lack of social services) bring nomadic populations to the city. Natural disasters, like *dzuds*, (particularly cold and snowy winters, when livestock cannot find any food) accelerate this phenomenon. They can induce huge livestock and income losses, forcing nomadic populations to come to the city, seeking for new economic opportunities.

Those nomadic populations mostly live in *gers*, so the urbanization of Ulan-Bator since the 1990 decade has been quite unorganized, and not managed by the city. New areas of *ger* housing appeared around the city, without sufficient infrastructures (sanitations, social services), leading to potential public health issues.

Moreover, many new households are not registered, therefore they are not reached by the Family Health Centers. Those latter are supposed to monitor the health of the households living in their *khoroos*. Risk of comunicable diseases are thus higher, and those population do not have access to basic health care. To deal with this major public health issue, several plans were implemented at the National or municipal level, including the "Reaching Every District" approach, studied by Lhamsuren et al. (2012). This program aims at reaching the most vulnerable in urban areas, and offering them a package of health services. As a result, FHCs have be found to be pro-poor unequal, contrary to the rest of urban health facilities (**??**).

Another major source of growing healh expenditures is the emergence of non-communicable diseases, including cardiovascular diseases and diabetes. They have become the leading cause of premature mortality (years of life lost). Diabetes, which was quite rare in Mongolia two or three decades ago, hit about one in seven persons. The diabetes rate increased by six and diabetes-related mortality by four between 2005 and 2014 (WHO, 2015). The government adopted preventive measures and put in place financial assistance schemes for some categories of patients. But the treatment is expensive for households and is not entirely covered by the aid measures. The government endorsed in 2014 the “Second National Programme on Prevention and Control of Diseases caused by Unhealthy Lifestyles 2014-2021" which requires significant funding.

Despite the implementation of such programs, out-of-pocket health expenditures rose in Mongolia from 12% in 1995 to 50% of total expenditures on healthin 2005. They have stayed above 40% ever since, and are far above the average (30%) of the upper-middle income countries (see Figure5.6). Global worldwide experience suggests that universal health coverage is difficult to achieve if out-of-pocket payments are greater than 30% of total health expenditure. They is a major risk of impoverishment for Mongolian households. With survey data from 2012, **?** find that 5.5% of Mongolian households spend more than 10% of their expenditures in healthcare. There is a need for better performance and equity of the global health system.

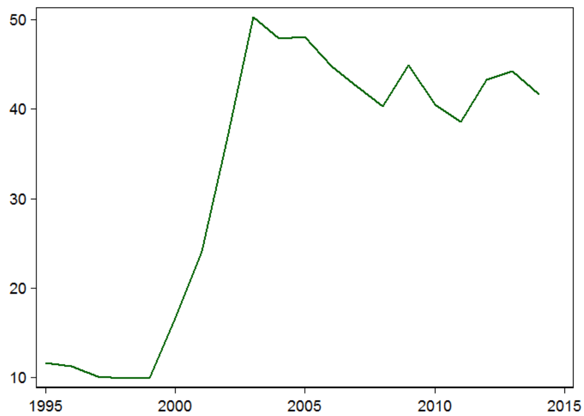


Figure 5.6: Evolution of the percentage of out-of-pocket payments in total health expenditures

5.2.4 A difficult economic context

GDP growth collapsed from 12.4% in 2012 to less than 1% in 2016, with 3% expected in 2017. External demand is weakening due to a continued dampening of the commodity market and slower growth in China, weakening Mongolian exports. Foreign direct investments have decreased, slowing investment and domestic demand.

Real household income growth has become sluggish, braking consumption and making households vulnerable to poverty. Socioeconomic factors have been shown to be strong predictors of health services utilization (Gan-Yadam et al., 2013). Several reports underlines that social safety nets should be strengthened and better targeted to the poor (WB, 2012; WHO, 2010), not to put excessive pressure on public finances.

Indeed, the fiscal situation has deteriorated over the past two years. The economy has faced sharp declines in Foreign Direct Investments (FDI) and coal exports. In response to these shocks, the authorities resorted to expansionary policies to tide the economy over until FDI and exports could recover. The fiscal expansion started in 2012 when the government implemented universal transfers to the population and the new Development Bank of Mongolia started to promote infrastructure development.

When FDI inflows tumbled and coal prices fell in 2013, the authorities maintained expansionary fiscal policy in an effort to sustain growth as well as they boosted infrastructure spending to realize the country's potential. The consolidated fiscal deficit (including spending of the DBM) oscillates between 9% and 10% of GDP since 2012. Therefore the government had to take stabilization measures. But the first quarter of 2016 revealed a more delicate situation than expected, showing the complexity of the situation. Revenue collections fell by 11% compared to the first quarter of 2015 (mainly due to falling mining revenues). Budget expenditures increased by 24% over the same comparison period, mainly due to the sharp increase in interest payments and social welfare transfers, an important program being the Child Money Program.

High vigilance is also required regarding the growing public debt, because the debt to GDP ratio is high (about 98 % in 2016 compared to 51% in 2012) because the external government debt structure has rapidly changed towards non-concessional conditions. The baseline scenario by the International Monetary Fund (2015) projections of the external debt sustainability of Mongolia suggests that debt service could absorb 49% of Government revenue in 2017, 15% in 2020 and 26% in 2025.

There is therefore an extremely challenging environment for public finances and for the public financing of health expenditure because there is no fiscal space to mobilize. Efficiency of health expenditures is thus a key to keep improving access to healthcare, particularly in a context of epidemiological transition with high costs.

5.3 The activity of Ulan-Bator Family Center Groups

This section presents the collected data, and the main information about Ulan-Bator FHCs: their activity, their resources. It also provides some productivity follow-up indicators.

5.3.1 The database: survey and study area

5.3.1.1 The survey

The survey was conducted jointly with the Department of Health Policy and Management, School of Public Health, and the Mongolian National University of Medical Sciences. A first part of the data was collected in Spring 2015 (activities and demographic data) from the FHCs registers and the Ulan-Bator city survey. It was followed by a visit of the CERDI team in Ulan-Bator, to discuss with different local actors of the health system, and work on the data. The second part (financial data) was collected in december 2015 from the accounting registers of FHCs.

5.3.1.2 Sample size and FHC catchment area

In Ulan-Bator, the urban area is divided into 8 districts and 145 sub-districts. The FHCs are at the sub-district level (one by sub-district), and the sample includes all of the 145 FHCs. Some of them have been divided in two FHCs, a main one, and a satellite during the period. As the two of them are managed together, it has been decided that only one Decision Making Unit would be considered, and that its output would the sum of the main and the satellite FHC. It was the case for the following pairs of sub-districts:

- 109 – 121
- 110 – 122
- 111 – 123
- 309 – 327
- 311 – 328
- 318 – 329
- 320 - 332
- 315 - 330
- 415 – 419
- 418 – 420
- 601 - 602 – 603
- 604 – 605

- 802 – 804
- 701 – 707
- 702 – 705

We also had data for 4 villages, but they provide more diversified healthcare than urban FHCs (including laboratory tests, inpatient activities for instance). Their production function is different from the one of the FHCs of Ulan-Bator, so they were excluded from the sample.

The biggest district is Bayan-Zurkh, the eastern district of Ulan-Bator, followed by Songinokhairhan, the north-western district. They both represent around 300,000 inhabitants, and are constituted of a large proportion of *ger* and mixed areas (18 out of 28 for Bayan-Zurkh, 19 out of 29 for Songinokhairhan, see Figure 5.1). The population increased in most districts. It remained stable in Nalaikh and Sukhbaatar (and the villages) and decreased, mainly from 2011 to 2012 in Sukhbaatar (see Figure 5.7). Bayan-Gol, Sukhbaatar, Chingeltei and Han-Uul all range between 100000 and 200000. They are mixed between *ger* and building areas, except for Bayan-Gol, where the housing is made of building in a large majority of sub-districts, as it is located in the very centre of the city. A last group if districts exhibits a stable population, around 5,000 people.

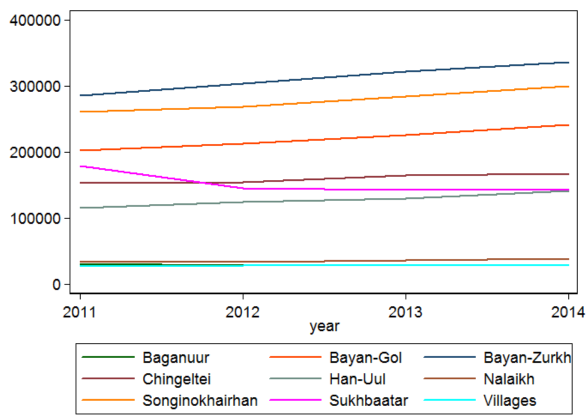


Figure 5.7: Evolution of the population of the districts from 2011 to 2014

The evolution of the activity and productivity of health establishments depends of the demand and hence of the size of the population of the catchment area and its demographic structure. Indeed, the youngest (under five year old) and the older (65 year old and above) are supposed to be more likely to use health facilities because of their health status. Different population sizes and demographic figure structures may induce more or less health facility activity.

District	*ger*	Building	Mixed	Total
Bayan-Gol	4	17	2	23
Bayan-Zurkh	12	10	6	28
Songinokhairhan	15	10	4	29
Sukhbaatar	2	8	4	14
Han-Uul	5	4	5	14
Chingeltei	13	3	2	18
Nalaikh	2	0	5	7
Baganuur	2	0	3	5
Total	55	52	31	138

Table 5.1: Number of subdistricts of the sample, by type, in 2014

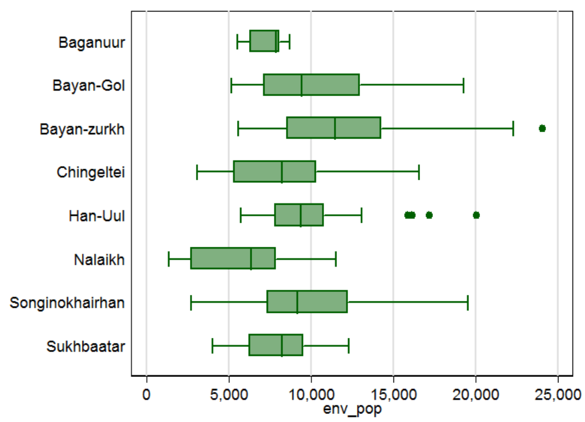


Figure 5.8: Distribution of the catchment area per subdistrict, by district

The proportion of population below 5 is between 0.05 and 0.2 for most of the subdistricts, and does not vary over the years (Figure 5.9). The proportion of elder people is smaller, with a 75^{th} percentile around 0.7 for every year. As a result, the dependency (inactive population over active population) is more important. Its median value is around 0.2 and the 75^{th} percentile around 0.3 for every year.

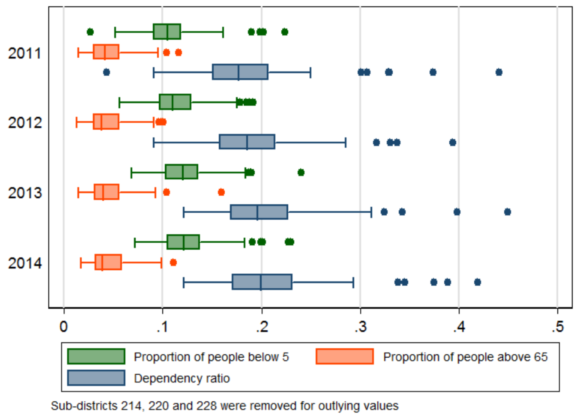


Figure 5.9: Evolution of dependency indicators from 2011 to 2014

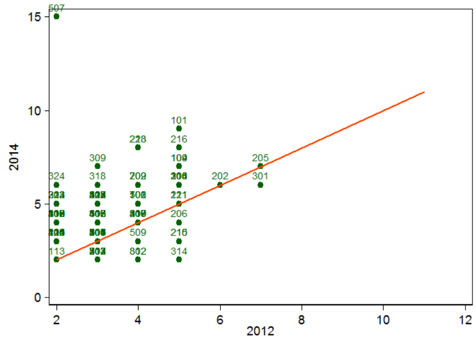
5.3.2 Evolution of activity

5.3.2.1 Evolution of staff endowment

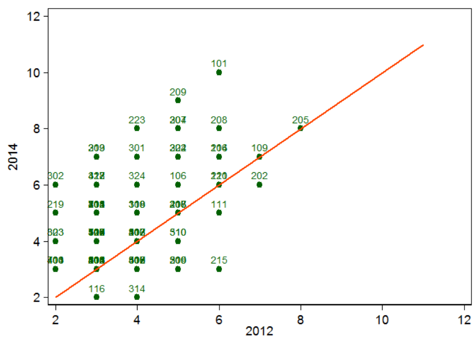
The staff of THCs includes doctors, nurses, and other (non medical staff). The number of medical staff (i.e. doctors and nurses) varies between 4 and 22, with an average value of 7 in 2012 and 9 in 2014 (see Table 5.11) and a median value increasing from 7 to 8. But the increase is far from being homogenous across the sample; the standard deviation sored over the study period. This idea is confirmed by Figure ??, as most of the FHCs are above the red line, meaning that their staff increased. Two FHC are atypical: FHCs 220 in 2012, due to its very large number of doctors and nurses in 2012 (11 and 11), and 507 and 511 for their large number of doctors in 2014 (15 and 10) compared to 2012.
Linking the staff endowment to the catchment area, the number of medical staff per 1000 inhabitants is mostly comprised between 0.5 and 1.5 for both of the two, with a slight increasing trend (Figures ?? and ??).

5.3.2.2 Evolution of FHCs activity

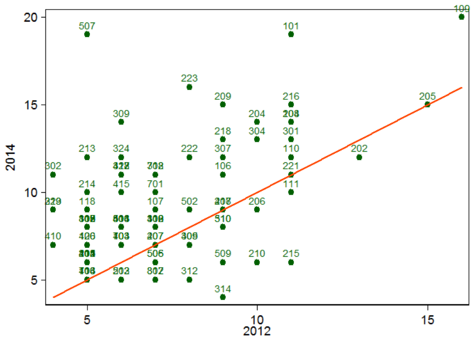
The activity of the FHCs relies both on curative (almost 52%, in numbers of acts) and preventive (47%) activities (Figure 5.12). Curative activity includes outpatient visits and outreach home visits which represent around 60% of the curative activity and 31% of the total activity. The staff of the FHC also has to call people of the catchment area at home in order to remember them to proceed to some prevention acts (immunization, diabetes analysis, etc.). This activity represents a very small part of their activity (1%).
All the FHCs have the same pattern of activity and it remained very stable over



a. Doctors

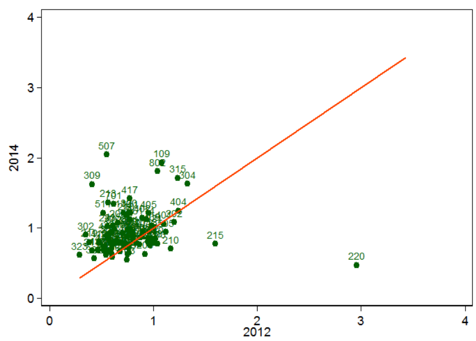


b. Nurses

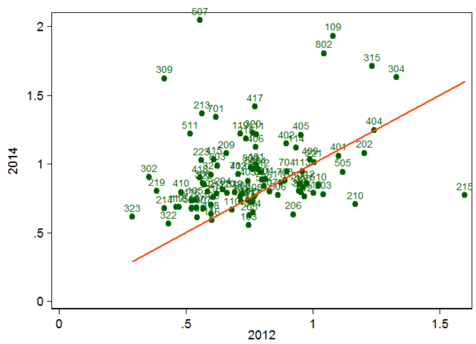


c. Medical staff

Figure 5.10: Evolution of the medical staff endowment between 2012 and 2014 (number of employees)



a. Whole sample



b. Without 220 and 601

Figure 5.11: Medical staff per 1000 inhabitants between 2012 and 2014,

the studied period . There is a strong correlation between the components of FHCs' activities (correlation coefficients are high, and all significant).

The average number of outpatient visits by FHC remains relatively stable over the

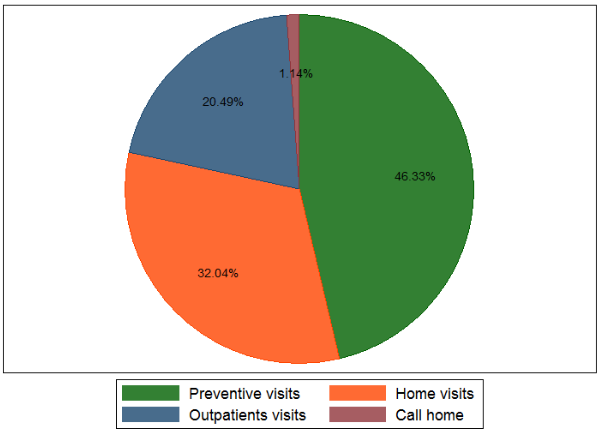
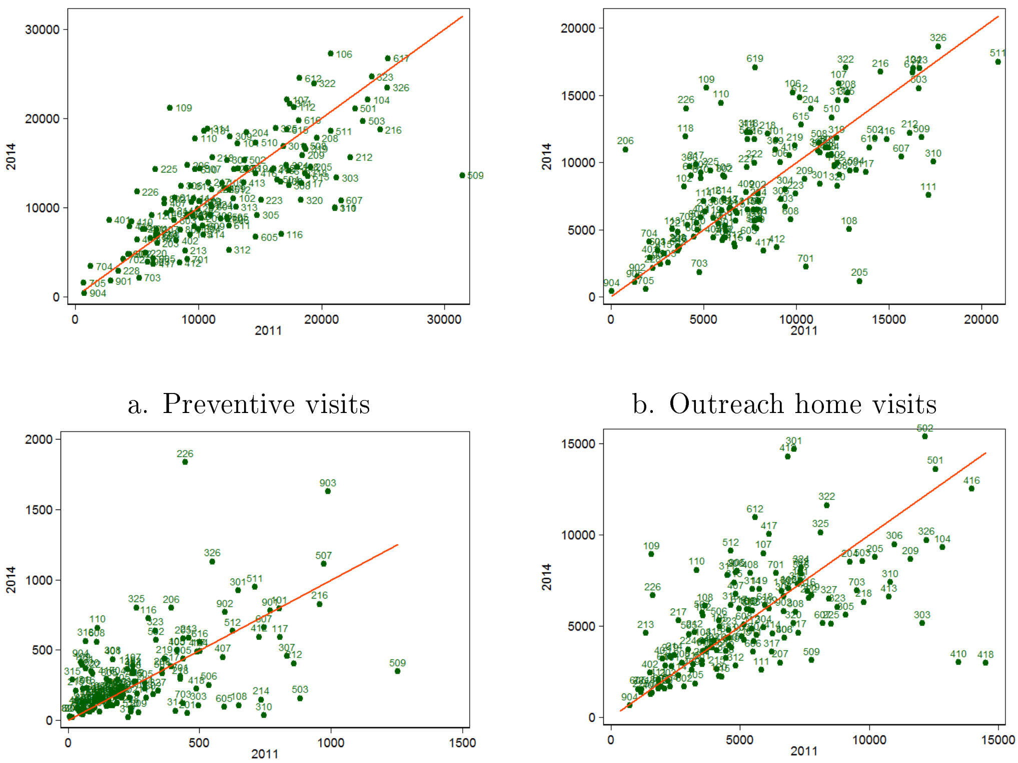


Figure 5.12: Composition of FHCs activity, in number of cases

studied period, raising in average from 5420 (2012) to 5853 (2014). The dispersion also increased, the standard deviation went from 2985 to 3201.

On the contrary, the average annual number of preventive visits decreased from 2012 to 2014 (13338 to 12242). This is raises some issues in terms of public health, as preventive visits are supposed to be the core of FHCs missions, and the catchment area of most FHCs remained stable or increased (see section 5.3.1.2). The same evolution is observed for outreach home visits, whose average annual number decreased from 8481 to 8181 (see Table 5.11). Those figure are confirmed by Figure 5.13a and 5.13b. The red lines represent the stability of the number of annual acts. Most FHCs stand below this line, which their number of preventive visits and outreach home visits decreased.

As the staff endowments remained exhibits a decreasing trend across the period, a decreasing efficiency is to be expected regarding descriptive statistics.



a. Preventive visits

b. Outreach home visits

c. Call home

d. Outpatient visits

Figure 5.13: Evolution of FHCs activity between 2011 and 2014, number of cases

5.3.3 Main follow-up indicators

5.3.3.1 Evolution of contact indicators

The evolution of activity has to be linked to the one of the cathment area, to check for the adequation of supply and demand in the primary healthcare sector. Contact indicators were computed for the main three components of FHCs activity: outpatient visits, preventive visits and outreach home visits.

The annual number of preventive visits per inhabitant ranges between 0.5 and 2, the same order of magnitude as the annual number of annual outreach home visits per inhabitant (Figures 5.14a and 5.14c). The number of outpatient visit per inhabitant is smaller, below 1 for most FHCs (Figure 5.14b also).

For two of the three indicators, the trend is decreasing between 2011 and 2014 (Figure 5.14a and 5.14c). The majority of them saw the annual number of preventive visits and outreach home visits per inhabitant decrease between 2011 and 2014. This was expected, considered the conclusions in Sections 5.3.2.2 and 5.3.1.2. The number of outpatient visit per oupatient (Figure **??**) does not decrease to the same extent,

it remains close to the red "stability" line for most FHCs.
Despite the increase of human resources (see Section 5.3.2.1), the contact indicators decreased, an alarming fact regarding universal access to health care. Moreover, those figures may be underestimated, because of unregistered households in the catchment area.
Even with only registered households, it is crucial to understand why FHCs are less frequented in 2014 than in 2011. From the literature and our discussion with local actors, several answers are possible.

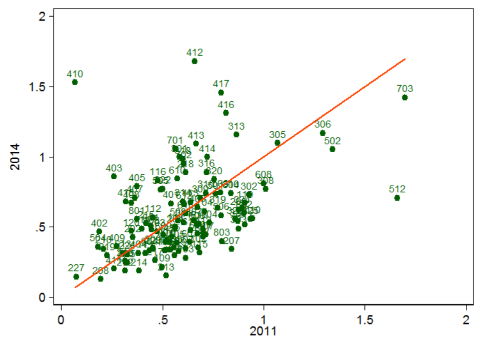
- financial constraints tightened across the study period, and some households can't afford anymore health care. This first option is not convincing here, as most care services delivered by FHCs are free of charge;
- the increase in the human resources may not be sufficient to answer the increase of the population. This scenario seems unlikely (Section 5.3.1.2 and 5.3.2.1);
- many households in the catchment area are not registered, so they have to pay to access healthcare, imposing a financial constraint on the demand side
- the medical staff workload has become very important because of administrative work, and represent a serious constraint on the time health workers can spend working with patients. This argument was put forward by many field partners we met. Administrative work means to fill health record for every patients, writing reports of all kind for health monitoring for instance;
- finally, there is a strong defiance facing the quality of healthcare delivered in FHCs. Sick people prefer to go to secondary facility and pay for better services than going to their FHC. This argument was presented as a crucial one by our local partners.

According to the reason, the answer are very different in terms of health policy: improving of staff skills and investment in their work, improving the bad reputation of FHCs in patients' minds, improving FHC management so that administrative work would not be a brake to "effective" healthcare. It is thus essential to further investigate the activity ans efficiency of Ulan-Bator FHCs.

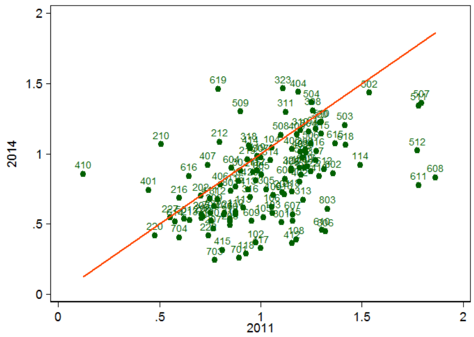
5.3.3.2 Evolution of productivity indicators

A first approach to efficiency analyses is the analysis of productivity indicators. They allow for comparison between different study areas, because they are expressed in

a. Preventive visits



b. Outpatient visits



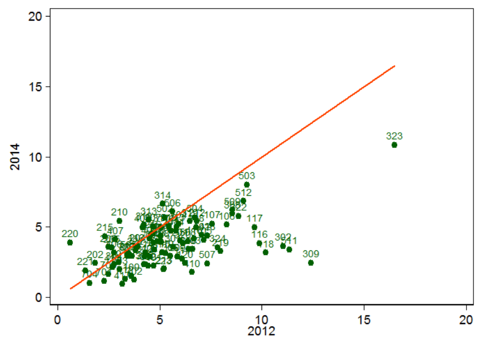
c. Outreach home visits

Figure 5.14: Evolution of contacts indicators between 2012 and 2014 (annual cases per inhabitant)

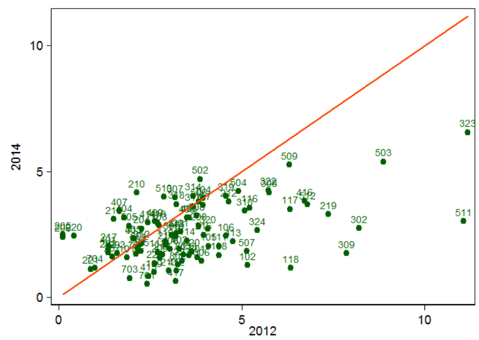
the same units to some extent. Their major weakness is that they can only link one input to one output.

Productivity indicators are expressed in number of daily medical acts (preventive, outpatient and outreach home visits) per medical staff. In 2012, medical staff practises in average 5.6 preventive visits, 2.3 outpatients visits, and 3.6 outreach home visits a day (respectively 5.1, 2 and 3.2 in median (Figure 5.11). The trend is decreasing for most FHC of the sample, regarding the three indicators (see Figure 5.15).

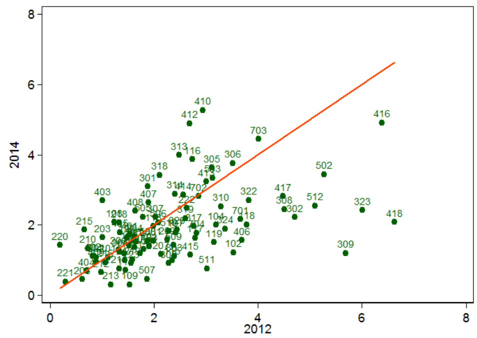
FHC 323 exhibits an atypical number of preventive and outreach home visits per staff in 2012 (17 and 12), due to its small number of staff (12) regarding its level of activity. From 2013 its medical staff rises to 23.



a. Preventive visits



b. Outreach home visits



c. Outpatient visits

Figure 5.15: Evolution of productivity indicators, in annual cases per medical staff between 2012 and 2014

5.4 The estimation of efficiency scores

This section deals with the first stage of the empirical analysis, the assessment of efficiency scores. It includes the definition of the production function, the adopted methodology and the interpretation of the scores. A specificity of this study was the way to take (or not) into account two main kinds of environment into the analysis : building areas and *ger* areas.

5.4.1 The empirical model

5.4.1.1 The production function

FHC are at the first level of the urban healthcare pyramid. Their activity is focused on public health and prevention (see Section 5.2.2). Three outputs were introduced in the analysis: the annual number of preventive visits, the annual number of outpatient visits and the annual number of home visits. They represent almost all the activity of FHCs.

The considered inputs are both human and physical resources of the FHCs. Human resources are the number of medical and non-medical staff. As they provide only basic healthcare, there is not much equipment in the FHCs: neither beds nor laboratory machines. As physical input, we considered the non-staff expenditures from the financing accounts files.

The production function therefore includes three outputs and three inputs.

Non-staff expenditures were only available from 2012 to 2014. To take them into account, year 2011 had to be excluded from the efficiency analysis, keeping only 2012, 2013 and 2014.

After the treatment of the satellites (see Section 5.3.1.2), the observations for which one variable of the production function or more was not available were removed from the sample. It then included 344 observations.

Finally, there were several observations with null values for the non-medical staff, or for the non-staff expenditures. This is not ideal in an efficiency analysis, particularly when it concerns inputs in an output-oriented analysis. Indeed, FHCs with a null value for an input will only be compared to other FHCs with a null value on this precise input. The number of peers is thus considerably reduced. For this reason, observation with at least one null value were also removed from the analysis.

The definitive sample was constituted of 311 observations, from 2012, 2013 and 2014.

5.4.1.2 The adopted methodology

The empirical model has to assess efficiency of FHCs, and the role of their environnement on their performances. The methodology is chosen according to the literature presented in Chapter **??**: partial frontiers models to get robust efficiency scores, and fractional regression model to explain those scores by potential determinants, managerial or external.

For all the variables introduced in the second stage, the separability condition is assumed: those variables have an impact on the distance between a FHC and the production frontier, not on the frontier itself.

For most of the potential determinants of efficiency, this condition was accepted. But following our discussions with local actors and the literature (Lhamsuren et al., 2012), the major type of housing in a sub-district (i.e. *ger* areas or building areas) was crucial for the FHC performance. Indeed, in *ger* sub-districts access to households is more difficult and requires more time, impacting the global activity. It takes more time for the medical staff to reach the patients for the outreach visits. The lack of infrastructure (sanitation for instance) in *ger* areas (mostly poor

suburbian sub-districts) is also a factor of difference, because comunicable diseases are more likely to spread in those areas. The case-mix may thus be different. For those reasons, it is possible that the type of housing does not only impact technical efficiency, but the production frontier itself.
Therefore, a possibility was to consider two different production functions and thus to estimate two separate production frontiers: one for the *ger* sub-districts and one for non-ger sub-districts. They both represent the maximal feasible production given their environnement. The efficiency scores, i.e. the distance to the production frontier - the maximal output they could produce given their environment- thus reflect mainly the managerial performance of each FHC.

5.4.1.3 The environment: ger and non-ger subdistricts

As two frontiers have to be estimated, every FHC has to be attributed to one or the other category. A great part of the sample is exclusively made of *ger* or building sub-districts. But for a certain number of other, the available information was a proportion of households or people living in *ger*, or even, for a few one, they were just described as "mixed". Estimating three production frontiers was not possible, so those sub-districts had to be assigned to the *ger* or the non-*ger* sub-sample.
The choice was made according to an assumption: to change the production function, the proportion of *ger* population has to be very high. Thus, the sub-districts just known as “mixed sub-districts” were associated to the non-*ger* sample.
When available, the attribution to one or another frontier was made according to the proportion of population living in *ger*.
As shown in Figure 5.16, there are two groups of FHCs where the proportion of *ger* population is neither 0 nor 1: a group around 20% of *ger* population, and another around 80%. The first one was associated to the non-*ger* frontier, following the assumption that this proportion of *ger* population was not sufficient to redefine the production function itself. On the contrary, the FHCs where around 80% of the population live in *ger* were associated to the *ger* frontier.
A last group of FHCs remained questionable, with no information at all about their environment. It is the case for FHC 402, 403, 405, 406 and 502.
To choose, they were alternatively associated to both of the two frontiers, to compare their scores in the two situations, and see if they distorted any production frontier. They did not impact the rest of the scores. We decided to put FHCs 402 and 403 in the *ger* frontier, and 405, 406 and 502 in the non-*ger* frontier, according to the frontier in which they had the higher score.
Finally, a *ger* frontier was defined with 99 observations, and a non-*ger* frontier with

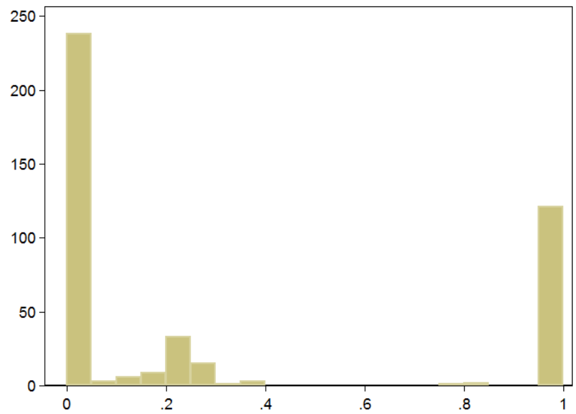


Figure 5.16: Distribution of the proportion of population living in *ger*, in frequency

212 observations (for the three years of the study). Table 5.12 gives the attribution of each FHC to its frontier.

5.4.1.4 One or two production function(s) ?

To confirm or not the hypothesis that *ger* and non-*ger* areas FHCs operate according to different production functions, several indicators are analyzed. If the access to households is more difficult in *ger* areas, then there should be some differences in the contact and productivity indicators.
Indeed, several scenarios can be considered.

- If the access is more difficult in *ger* areas than in building areas, the productivity and contact indicators are expected to be weaker in *ger* areas for two main reasons:
  - a doctor will provide less outreach home visits in a day in ger areas: negative effect on activity through outreach home visits, and on efficiency
  - fewer households will come to seek for health care because of accessibility of the FHC. This will have a negative effect on activity through preventive and outpatient visits, and on efficiency.
- If the health infrastructure is not sufficient, it can have an effect on the demand for healthcare, especially related to comunicable diseases. The effect is then positive on activity, and on efficiency.

Those first two scenarios have to be combined with the health authorities expectations and decisions. If those latter are aware of the differences between ger and non-ger areas and want to offset them, they may attribute more staff to ger areas. In this situation, no difference will be observed in the contact indicators, but the number of medical staff per 1000 inhabitants is likely to be greater in ger areas.

- The socio-economic differences between populations living in ger and in building areas also have to be considered:
  - If people living in ger are poorer than people living in building, than they should rely to a greater extent on free basic healthcare. They cannot afford, for instance, to bypass the pyramidal system and go directly to a secondary facility where they have to pay. The effect on FHC activity and efficiency is then positive.
  - If people living in ger are very poor nomadic population, it is likely that they are not registered. In this situation, either they go and seek for healthcare at the FHC anyway, and the effect on the FHC efficiency is positive, or they do not have the information that they can have free basic health care and do not go to the FHC. In this last situation the consequence is negatice on the FHC activity efficiency.

Table 5.2 gives average value for the ger and non-ger sub-districts, and the p-value to reject the null hypothesis of equal means, obtained through a mean significance test. It allows to confirm, in a descriptive way, the assumptions proposed above.
None of the three contact indicators is significantly different between *ger* and non-*ger* areas. This can mean either that the access difficulties are not that crucial and do not affect global activity to a large extent, or that there are more staff in *ger* areas, to offset the loss of time, so that no difference can be seen in the contact indicators.
There is not more medical staff per 1000 inhabitants, so this latter scenario does not seem to be verified. This could be qualified claiming that the data do not have reliable information on unregistered households, and that information concerning catchment areas are underestimated. Unfortunately, it is impossible to test this hypothesis.
The activity and productivity indicators are not significantly different between *ger* and non-*ger* areas. This suggests that the production functions are not different between the two type of sub-districts, or that the different effect are mutually offset. The only indicators for which a significant difference is observed are the staff variables, always greater in *ger* areas, even if the magnitude of the gap is not very large. This could mean that health authorities attribute more staff to *ger* sub-districts to offset their difficulties. But it is not important enough to be seen in the number of

Table 5.2: Average tests between *ger* and non-*ger* sub-districts

	Ger sub-districts	Non-ger sub-districts	Prob
Home vis. per inhab.	0.88	0.844	0.397
Prev. vis. per inhab.	1.288	1.289	0.988
Outpatient vis. per med. staff	0.565	0.62	0.142
Staff per 1000 inhab.	0.907	0.853	0.157
Preventive vis. per med staff	4.51	4.489	0.941
Home vis. per med staff	3.073	2.948	0.528
Outpatient vis. per med staff	1.947	2.119	0.219
Catchment area	10408.24	10108.29	0.501
Preventive visits	13365.61	12579.96	0.2624
Home visits	9051.11	8308.679	0.147
Outpatient visits	5661.293	5912.05	0.4961
Total Staff	14.374	12.575	0
Doctors	4.475	3.882	0.007
Nurses	4.545	4.207	0.068
Non medical staff	4.364	3.368	0
Medical staff	9.02	8.179	0.029
Prop. of above 65	442.229	552.102	0.198

The last column is the p-value with the inequality of averages as alternative assumption.

staff per 1000 inhabitants, so it is not sure that if justifies two distinct production frontiers.

Moreover, as there is no assumption on the nature of return to scale in partial frontier models, a difference of size is not a problem, FHCs will simply be compared to similar peers. There are also some technical considerations to take into account while choosing between one and two frontiers.

On the one hand, estimating two different frontiers, it is absolutely sure that FHCs will only be compared to similar peers.

On the other hand, this separate the sample in two subsamples, so it largely reduces the number of observations per frontier, potentially harming the statistical power of the scores. Moreover, it makes it impossible to compare the performances of FHCs, as they are assessed relatively to different frontiers. The scores become indicators of the dispersion of performances among the two subsamples. Lastly, in the second stage of the analysis, estimating two production frontiers requires to run two different regressions, one for the *ger* subsample, and one for the non-*ger* subsample. In terms of interpretation and discussion, those last two ideas are quite problematic.

Finally the two alternative have been computed in the first stage. The analysis of the efficiency scores suggests that there is no major diffeerence between the two options (i.e. the production frontier is not different for *ger* and non-*ger* FHCs), so scores computed on the whole sample were used in the second stage.

5.4.2 Measurement of efficiency: the first stage

5.4.2.1 The empirical model

The orientation and the number of frontiers As the FHC manager does not decide the number of staff in his/her facility (it is decided by the Health Bureau), an output orientation was chosen. For given resources, the objective of FHC is to provide as much health care as possible. This is in line with approaches like "Reaching Every District" (see Section 5.2.3).

FHCs provide only basic healthcare, so it is not likely that there were any technical progress between 2012 and 2014. An only production frontier was estimated for each of the two sub-samples.

As partial frontier models were used, it is not necessary to assume any nature about the returns to scale.

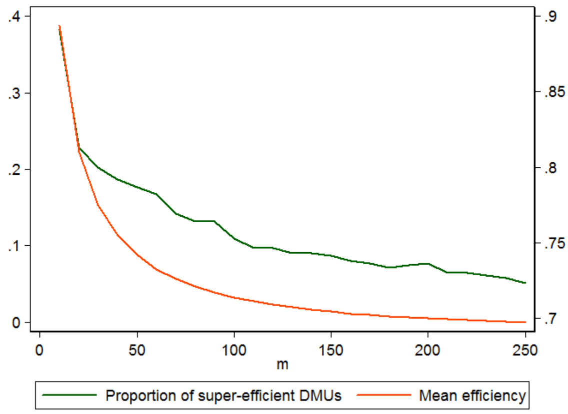
The choice of the m parameter In partial frontier models, the value of the m, seen as a "trimming parameter", is crucial. Simar and Wilson (2013) gives a method to choose the optimal value : "The final value of m can be chosen in terms of the desired level of robustness", i.e., of proportion of DMUs above the production frontier. The upper limit is the number of observations.

Here, all possible values from 10 to 90 for the *ger* frontier and from 100 to 200 for the non-*ger* frontier were tested, with a range of 10. As expected, the proportion of DMUs above the frontier (with a score higher than one) decreases as m increases, and so does the average efficiency in the sample (Figures **??**b and **??**c).

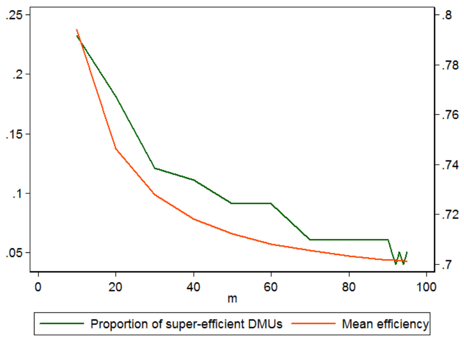
For the *ger* frontier, from $m = 70$ the proportion of super-efficient DMUs is stable around 0.05. This value was thus chosen. As for the non-*ger* frontier, the proportion of super-efficient DMUs is really stable from $m = 192$. This value was thus chosen. For each of the two frontiers, the precise value of m, starting from the chosen threshold, does not change a lot either the mean efficiency score, or the proportion of scores higher than 1.

For the whole sample model, values between 100 and 250 have been tested, with a range of 10 (see Figure 5.17a).

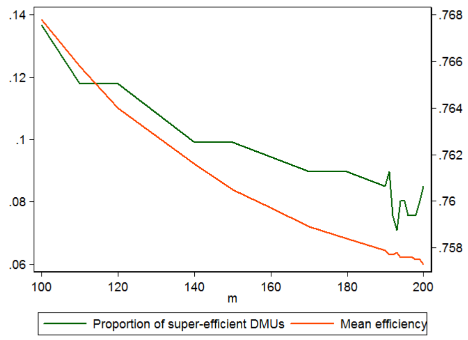
The choice of the *alpha* parameter Running the order-*alpha* analysis, the value of the parameter has to be chosen. This value gives the percentage of DMUs that will remain below the frontier: the higher the *alpha*, the closer the method is to the Free Disposal Hull. Contrary to the order-m method, the order-*alpha* method leaves a great part of the sample as super-efficient units, even for large values of *alpha*. It is very sensitive to a slight change of the parameter when it is closing the unit



a. Whole sample



b. *Ger* frontier



c. Non-*ger* frontier

Figure 5.17: Proportion of super-efficient DMUs and mean efficiency according to the value of m

(see Table 5.3). To get less than 10% of super-efficient THCs, 4%, of the sample, i.e. 6 observations approximately have to be left above the production frontier in its estimation.

Table 5.3: Proportion of super-efficient DMUs according to the model and the value of *alpha*, whole sample

Value of *alpha*	90	92	94	96	98
Model 1	0.392	0.341	0.257	0.170	0.071
Model 2	0.354	0.286	0.215	0.145	0.064
Model 3	0.447	0.408	0.341	0.222	0.122

Model 1,2 and 3 are detailed in section 5.4.2.3

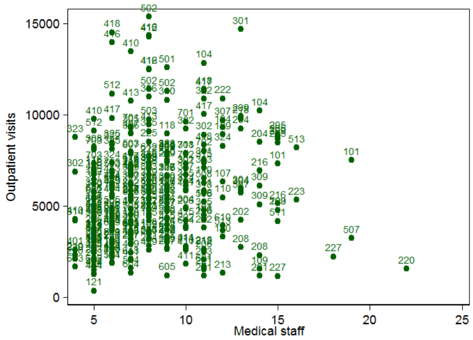
Order-m scores are used in priority in the second stage, because they are more stable. Nevertheless, it is necessary to check if the ranking of the FHCs is the same with the two methods, to see if the conclusions remain the same whatever the estimations.

5.4.2.2 Detection of outliers

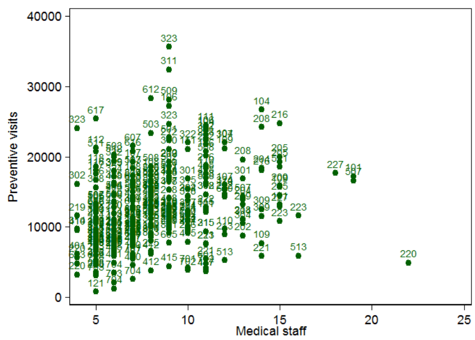
A first graphical analysis is run to detect some potential outlying FHC (Figure 5.20). They are presented here jointly for ger and non-ger frontier, but they were also analyzed separetely. There does seem to be any super-efficiency FHC in the sample, the only outlying values are particularly inefficient (220, 507) so they cannot distord the production frontier.
The efficiency scores were also analyzed, using the method of Simar (2003) already used in the previous chapter. With a threshold fixed at 1.05 several FHCs were detected as potential outliers:

- for the *ger* frontier, FHC 301(2014) and 323(2014)
- for the non-*ger* frontier, FHC 104 (2014) and 318(2013).

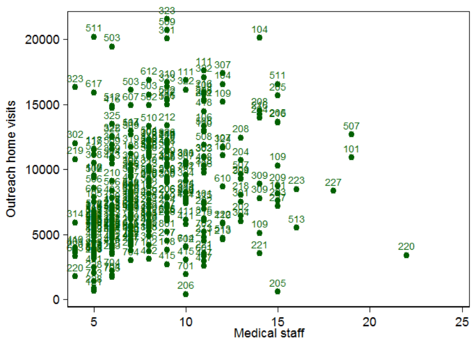
The same analyses were conducted, successively removing each of those observations of the samples. It appears that they did not affect the production frontiers, as the scores of the other FHCs did not change at all. The Pearson correlation was higher than 0.99 for each tested value of m. Thus, no outlying value had to be removed from the sample. The efficiency scores can be interpreted as such.



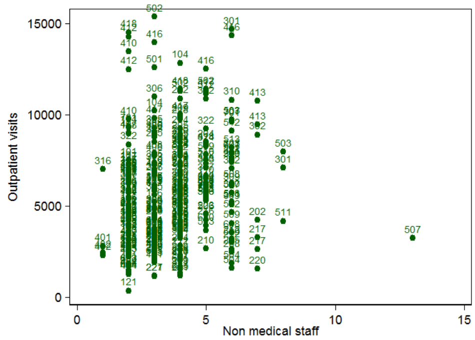
a. Outpatient visits/medical staff



b. Preventive visits/medical staff

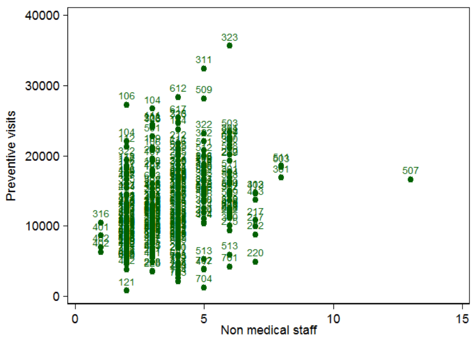


c. Outreach home visits/medical staff



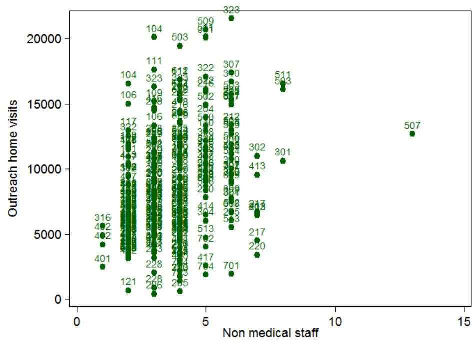
d. Outpatient visits/nonmedical staff

Figure 5.18: Outpatients visits



e. Preventive visits/nonmedical staff

Figure 5.19: Call home visits



f. Outreach home visits/nonmedical staff

Figure 5.20: Detection of potential outliers

5.4.2.3 Robustness of the model

Robustness of the production function Non-parametric models are data-driven, hence their sensitivity to the definition of the production function (Jacobs et al., 2006). Alternative models have to be tested, to check whether the efficiency scores are robust to the specification of the model.
In this study, there were no alternative for the outputs, as FHCs only provide basic health, so there are few sorts of medical acts. The definition of inputs was more flexible, so it was possible to test various alternative empirical models.

- Model 1 (the principal one) includes medical staff, non-medical staff and non-staff expenditures as inputs.
- Model 2 only non-staff expenditures and medical staff into consideration.
- Model 3 takes includes the non-staff expenditures and the global staff as inputs.

Tables 5.4 and 5.5 give the Pearson correlation matrix and the Spearman rank correlation matrix between the scores computed with the several models with one frontier.
The Pearson correlation coefficients are very high, the lowest is 0.798 between Model 2 in order-*alpha* and Model 1 in order-m for the *ger* frontier. The Spearman rank correlations are also very high, suggesting that the specification of the model does not affect to a great extent the ranking of FHCs performances.

Table 5.4: Pearson correlation matrix, whole sample

	M 1, m	M 2, m	M 3, m	M 1, α	M 2, α	M 3, α
Model 1, order-m	1					
Model 2, order-m	0.870	1				
Model 3, order-m	0.909	0.932	1			
Model 1, order-alpha	0.830	0.971	0.899	1		
Model 2, order-alpha	0.963	0.882	0.905	0.882	1	
Model 3, order-alpha	0.853	0.836	0.847	0.850	0.888	1

All Pearson correlations are significant at the 95% confidence level.
All α parameters are equal to 98%. The values of m are 210 for M1, 201 for M2 and 180 for M3.
They were chosen with the same methodology as in paragraph 5.4.2.1.

Robustness to the number of frontiers Finally, a choice has to be made between one or two production frontiers. Several criteria were used: are the two series highly correlated ? How many observations have different score according to the number of estimated frontiers ? Are they rather *ger* or non-*ger* observations ?

Table 5.5: Spearman rank correlation matrix, whole sample

	M 1, m	M 2, m	M 3, m	M 1, α	M 2, α	M 3, α
Model 1, order-m	1					
Model 2, order-m	0.876	1				
Model 3, order-m	0.915	0.936	1			
Model 1, order-alpha	0.840	0.979	0.907	1		
Model 2, order-alpha	0.969	0.888	0.914	0.879	1	
Model 3, order-alpha	0.881	0.848	0.863	0.855	0.911	1

All null hypothesis of independence between the variables are rejected with a confidence level of more than 99%.

To answer those questions, a graphical analysis is first conducted. Figure 5.21 plots the scores obtained with the two strategies. Most of FHCs are on the diagonal, so their score does not change at all. When the scores are different between the two methods, they are always higher with two distinct estimated frontiers. Four observations have very different scores: 219, 221, 401 and 402, but for the rest of the sample the gap is limited, both for *ger* and non-*ger* FHCs. This is confirmed by Figure 5.22 which gives the distribution of the difference between the two scores. There are some negative values (65 over 311) but their absolute value is not large (the minimal value is -0.027 over the whole sample).

The gap is more important for non-*ger* sub-districts than for *ger* ones, suggesting that there is no major difference in the production function due to environmental difficulties. For 56 FHCS,the difference between the two scores is greater than 0.1, and among them, only 16 are located in *ger* sub-districts. There is no reason to think that the performances of FHCs located in *ger* areas are underrated because they are compared to FHCs operating in non-*ger* areas.

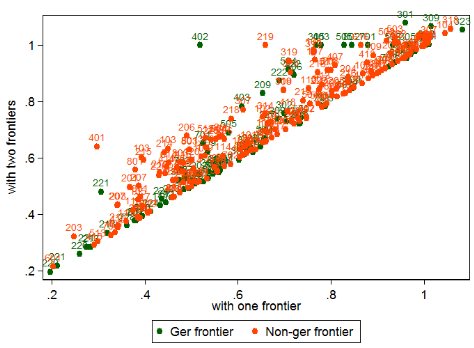


Figure 5.21: Comparison of scores obtained with one and two frontiers

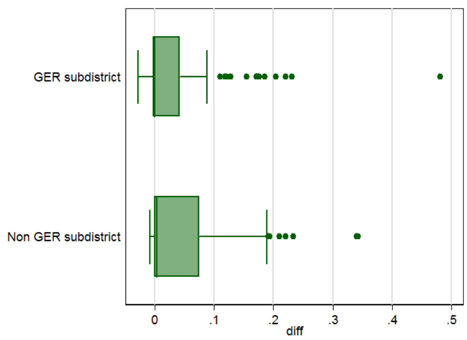


Figure 5.22: Distribution of the difference between two-frontier and one-frontier scores

The analysis of the Pearson and Spearman rank correlation matrices (Table 5.6 for the whole sample, and 5.15 and 5.16 by frontier) leads to the same conclusion. The correlation are always very high.

Finally, as efficiency scores are close with one or two production frontiers, suggesting that the production functions are not necessarily different, the scores obtained with a unique frontier are kept for the second stage analysis.

In terms of analysis, this allows to compare *ger* and non-*ger* areas FHCs, and to analyze all the FHCs in the regression in the second stage.

Table 5.6: Correlations between scores with one and two frontiers

Pearson correlation coefficients

	M1, 1 fr	M2, 1 fr	M3, 1 fr	M1, 2 fr	M2, 2 fr	M3, 2 fr
Model 1, 1 frontier	1					
Model 2, 1 frontier	0.870	1				
Model 3, 1 frontier	0.909	0.932	1			
Model 1, 2 frontiers	0.858	0.952	0.920	1		
Model 2, 2 frontiers	0.954	0.839	0.879	0.873	1	
Model 3, 2 frontiers	0.888	0.897	0.949	0.937	0.911	1

All Pearson correlations are significant at the 95% confidence level.

Spearman rank correlations

	M1, 1 fr	M2, 1 fr	M3, 1 fr	M1, 2 fr	M2, 2 fr	M3, 2 fr
Model 1, 1 frontier	1					
Model 2, 1 frontier	0.876	1				
Model 3, 1 frontier	0.915	0.936	1			
Model 1, 2 frontiers	0.860	0.953	0.921	1		
Model 2, 2 frontiers	0.950	0.845	0.887	0.875	1	
Model 3, 2 frontiers	0.890	0.895	0.949	0.930	0.913	1

All null hypothesis of independence between the variables are rejected
with a confidence level of more than 99%.

5.4.3 Efficiency scores

5.4.3.1 Decreasing efficiency over the study period

In average over the three years, FHCs exhibits a technical efficiency of 0.70. They could increase their activity by 30% with the same quantity of inputs. As expected in section 5.3.2.2 due to a decreasing trend in activity, FHCs experienced a decrease of their technical efficiency between 2012 and 2014. The average (median) score falls from 0.799 (0.864) to 0.643 (0.606) from 2012 to 2013, but remains stable from 2013 to 2014 (see Table 5.7). The same conclusion comes from Figure 5.23: the first quartile in 2012 is higher than the median in 2014. The distribution of the variable changed in one year. Only 3 FHCs in the sample experienced an increase higher than 0.2 in their efficiency scores, and the total number of THCs whose scores improved is 27.

Table 5.7: Evolution of efficiency scores between 2012 and 2014

Evolution of efficiency, whole sample

year	N	Average	Min	Max	p25	Median	p75	Std dev
2012	113	0.799	0.197	1.011	0.629	0.864	1.000	0.220
2013	99	0.643	0.212	1.057	0.479	0.606	0.803	0.218
2014	99	0.641	0.274	1.083	0.483	0.632	0.791	0.212
Total	311	0.699	0.197	1.083	0.505	0.699	0.938	0.229

Evolution of efficiency, inefficient FHCs only

year	N	Average	Min	Max	p25	Median	p75	Std dev
2012	76	0.701	0.197	0.996	0.561	0.706	0.872	0.207
2013	91	0.611	0.212	0.996	0.474	0.571	0.774	0.197
2014	91	0.608	0.274	0.999	0.477	0.562	0.747	0.187
Total	258	0.636	0.197	0.999	0.483	0.625	0.791	0.200

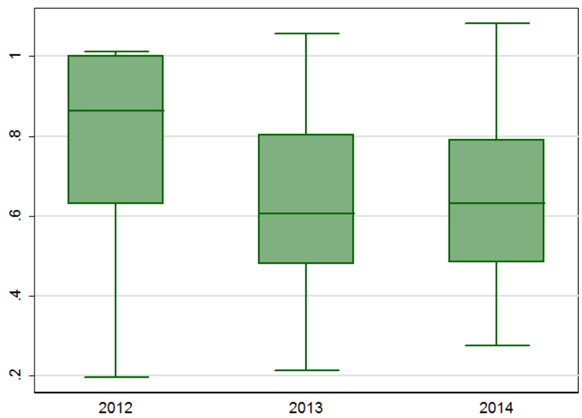


Figure 5.23: Evolution of efficiency scores

Testing the significance of the average difference, the 2012 average is significantly higher than 2013 and 2014 together. There is no significant difference between 2013 and 2014. This decrease is mainly due to a decrease in the number of cases for the three types of provided medical acts (Section 5.3.2.2). There are only 53 efficient observations in the sample, including 37 in 2012 (comparison between the two parts of Table 5.7).

The variability in FHCs performances is very large. Among inefficient ones, half of them exhibits a score lower than 0.57 in 2013 and 0.56 in 2012, which is very weak.

5.4.3.2 Differences between FHCs located in *ger* or non-*ger* areas

There is a slight difference in efficiency scores between FHCs located in *ger* areas (0.705 in average) and FHCs located in non-*ger* areas (0.758) (Figure 5.24 and Table 5.8). But this difference is not statistically significant. The evolution per year is almost the same in *ger* and non-*ger* areas (see Figure 5.25).
Larger disparities can be observed across districts, even if the number of observations per district is very unequal. The districts of Han-Uul and Songinokhairhan seem particularly efficient relatively to the other ones (0.769 and 0.794 of average efficiency), along with Bayan-Gol (0.735). On the contrary, Bayan-Zurkh exhibits a particularly low average and median efficiency (0.5658 and 0.517).

Table 5.8: Descriptive statistics about effciency scores, by frontier

frontiere	N	Average	Min	Max	p25	Median	p75	Std dev
Ger sub-districts	99	0.668	0.197	1.083	0.508	0.632	0.879	0.229
Non-ger sub-districts	212	0.714	0.204	1.057	0.503	0.718	0.950	0.228
Total	311	0.699	0.197	1.083	0.505	0.699	0.938	0.229

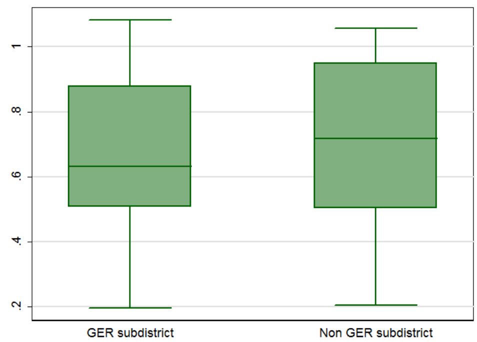


Figure 5.24: Distribution of efficiency scores across environments

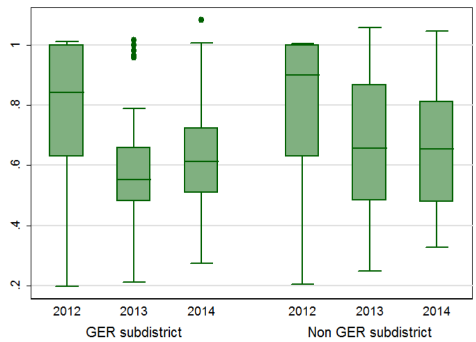


Figure 5.25: Evolution of efficiency scores, per environment

Table 5.9: Efficiency scores across districts

district	N	Average	Min	Max	p25	Median	p75	Std dev
Baganuur	9	0.542	0.379	0.845	0.389	0.496	0.614	0.177
Bayan-Gol	58	0.735	0.359	1.046	0.553	0.718	0.981	0.216
Bayan-zurkh	73	0.568	0.197	1	0.395	0.517	0.717	0.215
Chingeltei	13	0.709	0.204	1.001	0.507	0.671	1	0.262
Han-Uul	33	0.769	0.299	1.004	0.632	0.774	0.938	0.198
Nalaikh	12	0.682	0.327	1	0.502	0.688	0.898	0.225
Songinokhairhan	60	0.794	0.406	1.083	0.623	0.778	0.998	0.188
Sukhbaatar	53	0.718	0.292	1.001	0.500	0.747	0.969	0.242
Total	311	0.699	0.197	1.083	0.505	0.699	0.938	0.229

5.5 Estimating the determinants of FHCs efficiency

This section describes the second stage of the analysis, the estimation of the determinants and their interpretation, in the perspective of policy recommendations.

5.5.1 Potential determinants of FHCs efficiency

Two main categories are introduced in the second stage of the analysis to explain efficiency scores: environmental (external) one and managerial (internal) ones.

5.5.1.1 External variables

Several environemental variables were introduced to capture the difference in the potential demand for each FHC: the **logarithm of the catchment area**, and the **proportion of population below 5 and above 65 years old**. Both are expected to be positively associated to efficiency scores, as they spur demand for health care, resources being equal.
As a global production frontier was estimated, and not two distinct ones between FHCs operating in ger areas and FHCs operating in building areas, a dummy for the ger sub-districts.

5.5.1.2 Managerial variables

Managerial variables are the interest variable, because they can give information about relevant -or not- managerial practices in terms of activity and performances. They are of two main types:

- variables related to the **composition of the staff**. The tested hypothesis is that there might be a complementarity between the different staff in FHCs, and an excess of certain type of staff relatively to the other may lower their performance.

- variables about the **composition of the staff remuneration**, between salary and incentives. If incentives are allocated according to accurate criteria, they may spur the motivation of the staff, and thus spur efficiency.

Evolution of remuneration of the staff The proportion of incentives in the staff total remuneration changed a lot over the period. In 2012, it represented the same proportion for each FHC in the sample (around 25%, Figure 5.26). From 2013, every FHC applied its own remuneration formula between basic salary and incentives. Incentives can represent up to 50% of managers' remuneration. For other employees, the proportion of incentives is mostly below 30%. The situation did not change between 2013 and 2014. Regarding the evolution of the global remuneration of the staff, the change of the remuneration formula was globally at the advantage of the employees: the median remuneration rose for each professional position in the FHC (Figure 5.27). This increase is more important for managers, and can be qualified for the other staff (the first quartile of the other staff remuneration did not increase much over the period).

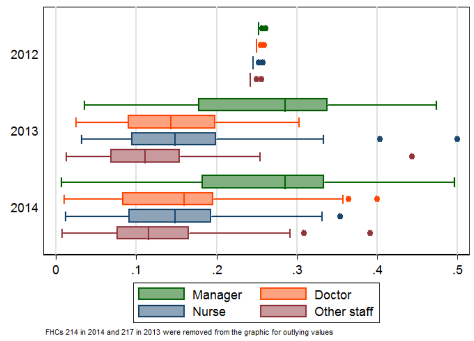


Figure 5.26: Distribution of the proportion of incentives in the staff remuneration

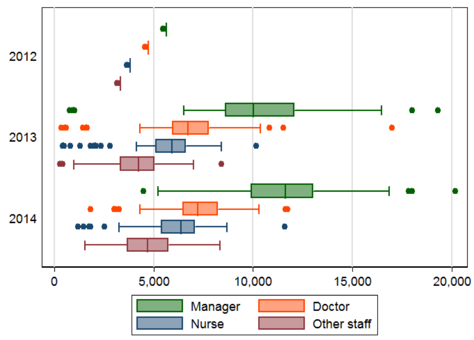


Figure 5.27: Distribution of annual staff remuneration, in thousands of Tugriks

5.5.2 Results

Results are presented in Table 5.10. The determinants of efficiency were estimated through a fractional regression model introduced by Ramalho et al. (2010) and Ramalho et al. (2011) .

As there only 53 over 311 efficient observations, a one-part model was chosen, with a Probit link function (other functions were tested in robustness checks). Dummy variables were introduced for the different districts, in order to capture unobservable characteristics. They highlight a significant and positive impact of the size of the catchment area, as a potential demand for healthcare.

On the contrary, the composition of the population in terms of age does not have any impact on efficiency contrary to what was expected. This variables suffers from a great number of missing values, so it reduced the sample from 311 to 210. It was thus chosen not to present it in the main result table, to keep the most observations possible in the sample.

The proportion of medical staff (i.e. doctors and nurses) has a negative and significant impact. This can be explained by a complementarity in the composition of the staff, particularly in front of the important administrative workload in FHCs. In our discussions with local actors, the burden of administrative work was often emphasized: reporting of several follow-up and monitoring indicators, constitution of health booklets. Those activities do not enter the medical production function as an output, thus they do not spur efficiency. Though, they take time. They work can be at least partially realized by non-medical staff, contrary to medical acts. In FHCs where there is no of few non-medical staff, doctors and nurses have to do all of it, therefore they see less patients, and are assessed as less efficient.

The proportion of incentives, whatever staff concerned has a global positive impact on efficiency. However, when we decompose this effect by type of staff, incentives become non-significant for doctors, and negative for managers and nurses. The positive sign is driven by the proportion of incentives in the remuneration of the other staff (administrative staff for example). This result can be explained by the nature of the "incentive" variable. It is more a salary complement, unconditionnally to the individual or collective performance of the staff. Thus it does not have any role of spurring activity through financial incentives.

5.5.3 Discussion

The results of the second stage highlights different phenomenon, and suggest several possible levers for action to a universal access to healthcare in Ulan-Bator.

From the demand side, the catchment area is a major determinant of FHCs efficiency. In the context of Ulan-Bator, this can have several levels of interpretation. This effect can represent the "classical" role of potential demand played by this catchment area. Second, this catchment area represents the registered population. Then, the positive sign means that the bigger the registered population of a subdistrict is, the higher the efficiency level is. It can suggest than the unregistered population does not go to FHCs. It does not represent a potential demand for them. In this case, it is crucial to understand why those populations do not seek for free healthcare: is it a problem of information (they do not know that they don not have to pay) ? Or a

Table 5.10: Determinants of efficiency scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Catchment area	0.878***	0.723***	0.837***	0.804***	0.878***	0.811***	0.878***
	(0.115)	(0.121)	(0.124)	(0.128)	(0.115)	(0.129)	(0.116)
Proportion of incentives	0.622				0.622	0.643	0.623
	(0.484)				(0.484)	(0.857)	(0.482)
Proportion of medical staff	-1.297***		-1.292***	-1.385***	-1.297***	-1.374***	-1.318***
	(0.438)		(0.484)	(0.480)	(0.438)	(0.479)	(0.492)
Ger area	-0.165**	-0.133	-0.139	-0.130	-0.165**	-0.131	-0.166**
	(0.0807)	(0.0882)	(0.0855)	(0.0869)	(0.0807)	(0.0876)	(0.0809)
year 2012	0.238	0.509***	0.477***	0.530***	0.238	0.286	0.239
	(0.226)	(0.0998)	(0.0854)	(0.0901)	(0.226)	(0.349)	(0.227)
Proportion of incentives, doctors		-0.651					
		(0.585)					
Proportion of doctors		-1.233*					0.0479
		(0.655)					(0.640)
Proportion of incentives, managers			-0.331				
			(0.443)				
Proportion of incentives, nurses				-0.686*		-0.968**	
				(0.411)		(0.376)	
Constant	-7.141***	-5.920***	-6.572***	-6.183***	-7.141***	-6.295***	-7.140***
	(0.973)	(1.097)	(1.025)	(1.073)	(0.973)	(1.099)	(0.973)
Observations	310	270	282	278	310	278	310
R2	0.339	0.286	0.302	0.310	0.339	0.310	0.339

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

problem of defiance regarding the perceived quality of healthcare in FHCs ?
From the supply side, it seems hard to give a unique answer but some issues emerge from the results of the second stage, and they are mainly linked to the staff composition and remuneration. FHCs are private entities, contracted by the city and the Ministry of Health to provide primary healthcare, and give some regular data about follow-up indicators. But according to our discussions with the Ulan-Bator Health Department, there is no real credible threat if they do not fulfill their missions, mostly because public authorities do not have any information about their efforts. This is a typical principal/agent situation, with a moral hazard at the benefice of the FHCs staff. Their staff is remunerated with a dotation from public authorities, according to the number and position of employees. Those remunerations are known to be low, which is a source of discontent, and they are not connected to any performance or activity indicators. Public authorities can only assess the completeness of the statistical data provided by the FHCs, but has no control on the effort of the medical staff.
As a consequence, spurring activity, and then efficiency in FHCs would require to increase the level of remuneration of their staff, and to partly connect them with their performance in terms of healthcare. The administrative workload also has to be rethinked, as it seems to be a burden not only for non-medical staff, but also for medical staff. This slows down the "effective" cring activity of FHCs.

5.6 Conclusion

This Chapter provides some information about the efficiency of urban primary healthcare facilities in Mongolia, the *Family Health Centers*. It relies on survey data from 2011 to 2014, and adopts a two-stage procedure. Performance declined in FHCs facilities during the study period, an expected result as human resources increased, contrary to the number of provided health services. In average, FHCs could increase by 30% their level of activity with the same level of resources. Contrary to what could have been expected, there is no major difference of activity or context between areas where people mostly live in *ger*, and the areas where they live in buildings.
It is crucial that the Mongolian health system heads to this objectives, as it has to face growing health needs with constant resources. Efficiency scores are explained both by the demand side (catchment areas), so it is necessary to improve information about primary health care, and preception of the quality of healthcare in FHCs. On the supply side, variables about the compositio and the remuneration of the staff are significant. A clear contract has to made between FHCs and the public authorities,

in their exercise of the *oversight* function described in the general introduction, so that the function of *service delivery* can be fullfilled optimally.

Bibliography

Gan-Yadam, A., Shinohara, R., , Sugisawa, Y., Tanaka, E., Watanabe, T., Hirano, M., Tomisaki, E., Morita, K., Onda, Y., Tokutake, K., Mochizuki, Y., Matsumoto, M., Sugita, C., and Anme, T. (2013). Factors associated with health service utilization in Ulaanbaatar, Mongolia: A population-based survey. *Journal of Epidemiology*, 23(5):320–328.

Jacobs, R., Smith, P. C., and Street, A. (2006). *Measuring Efficiency in Health Care: Analytic Techniques and Health Policy.* Cambridge University Press.

Lhamsuren, K., Choijiljav, T., Budbazar, E., Vanchinkhuu, S., Blanc, D. C., and Grundy, J. (2012). Taking action on the social determinants of health: improving health access for the urban poor in Mongolia. *International Journal for Equity in Healt*, 11(15).

Manaseki, S. (1993). Mongolia: a health system in transition. *BMJ*, 307(6919):1609–11.

Ramalho, E., Ramalho, J., and Henriques, P. (2010). Fractional regression models for second stage DEA efficiency analyses. *Journal of Productivity Analysis*, 34(3):239–255.

Ramalho, E. A., Ramalho, J. J., and Murteira, J. M. (2011). Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys*, 25(1):19–68.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.

Simar, L. and Wilson, P. W. (2013). Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. *Foundations and Trends(R) in Econometrics*, 5(3–4):183–337.

WB (2012). Health equity and financial protection report - Mongolia. Report 72450, World Bank, Washington, D.C.

WHO (2010). WHO country cooperation strategy for Mongolia 2010-2015. Technical report, WHO Mongolia.

WHO and MoH (2012). Health service delivery profile : Mongolia. Technical report.

5.7 Appendix

Table 5.11: General statistics about main indicators

	2012							2014						
	Average	Min.	Max.	p25	Median	p75	sd	Average	Min.	Max.	p25	Median	p75	sd
Catchment area	10004	3213	20027	7661	9208	11932	3422	10551	3855	24052	7779	9598	13050	3969
Prop. of above 65	0.048	0.013	0.322	0.030	0.038	0.055	0.041	0.052	0.018	0.312	0.033	0.041	0.061	0.043
Prop. of below5	0.113	0.061	0.201	0.099	0.110	0.129	0.026	0.134	0.073	0.679	0.106	0.125	0.145	0.062
Preventive visits	14212	3441	32417	9436	13716	17992	5939	12324	2577	35618	8205	11200	15783	6135
Home visits	9022	394	20678	5574	8215	11436	4456	8367	1389	22247	4931	7529	10983	4346
Outpatient visits	5681	1187	15375	3617	5421	7148	2887	5853	1269	14681	3322	4767	8111	3211
Calls	308	0	1441	119	236	403	256	280	16	1835	88	173	390	287
Nurses	4.026	2	11	3	4	5	1.399	4.612	2	10	3	4	6	2
Doctors	3.440	2	11	3	3	4	1.360	4.602	2	15	3	4	5	2
Non medical staff	3.500	1	7	2	3	4	1.436	3.821	0	13	3	4	5	2
Non staff expenditures	12.554	0	41.339	9.123	11.643	15.700	5.888	25800	0	148000	15600	23700	35000	19000
Staff per 1000 inhab.	0.820	0.288	3.424	0.592	0.760	0.954	0.405	0.884	0	2.044	0.727	0.851	1.024	0.304
OP per med. staff	2.257	0.194	6.628	1.400	1.987	2.817	1.285	1.875	0.305	5.272	1.153	1.664	2.419	1.027
Prev. per med. staff	5.591	0.610	16.497	3.635	5.097	6.785	2.702	3.763	0.943	10.843	2.591	3.563	4.899	1.612
Home v. per med. staff	3.573	0.106	11.168	2.211	3.210	4.380	2.023	2.565	0.525	6.556	1.687	2.429	3.431	1.174

Sample: every FHC with complete information for every variable of the efficiency analysis

Non staff expenditures in thousands of Tugriks

Table 5.12: Attribution of the subdistricts to the GER or non-GER frontier

Bayan-Gol		Bayan-zurkh		Songinokhairhan		Sukhbaatar		Han-Uul		Chingeltei	
101	Non-ger	201	Non-ger	301	Ger	401	Non-ger	501	Non-ger	601	Non-ger
102	Non-ger	202	Ger	302	Ger	402	Ger	502	Non-ger	604	Non-ger
103	Non-ger	203	Non-ger	303	Ger	403	Ger	503	Non-ger	606	Non-ger
104	Non-ger	204	Non-ger	304	Ger	404	Non-ger	504	Ger	607	Non-ger
105	Non-ger	205	Non-ger	305	Ger	405	Non-ger	505	Ger	608	Non-ger
106	Non-ger	206	Non-ger	306	Non-ger	406	Non-ger	506	Ger	610	Non-ger
107	Non-ger	207	Non-ger	307	Non-ger	407	Non-ger	507	Ger	611	Non-ger
108	Non-ger	208	Non-ger	308	Ger	408	Non-ger	508	Ger	612	Non-ger
109	Non-ger	209	Ger	309	Ger	409	Ger	509	Ger	613	Non-ger
110	Non-ger	210	Ger	310	Ger	410	Non-ger	510	Non-ger	614	Non-ger
111	Ger	212	Ger	311	Ger	412	Non-ger	511	Non-ger	615	Non-ger
112	Non-ger	213	Ger	312	Non-ger	413	Non-ger	512	Non-ger	616	Non-ger
113	Non-ger	214	Non-ger	313	Non-ger	414	Non-ger			617	Non-ger
114	Non-ger	215	Non-ger	314	Non-ger	415	Non-ger			Nalaikh	
115	Non-ger	216	Non-ger	315	Non-ger	416	Non-ger			701	Ger
116	Non-ger	217	Non-ger	317	Non-ger	417	Non-ger			702	Non-ger
117	Non-ger	218	Non-ger	318	Non-ger	418	Non-ger			703	Ger
118	Non-ger	219	Non-ger	319	Non-ger					704	Non-ger
119	Non-ger	220	Ger	320	Non-ger					Baganuur	
120	Non-ger	221	Ger	322	Ger					801	Non-ger
		222	Ger	323	Ger					802	Ger
		223	Ger	324	Ger					803	Non-ger

Table 5.13: Correlation between alternative models and methods, ger frontier

Pearson correlation coefficients

	M 1, m	M 2, m	M 3, m	M 1, α	M 2, α	M 3, α
Model 1, order-m	1					
Model 2, order-m	0.849	1				
Model 3, order-m	0.8775	0.9716	1			
Model 1, order-alpha	0.9478	0.8498	0.8703	1		
Model 2, order-alpha	0.7973	0.9511	0.918	0.8822	1	
Model 3, order-alpha	0.8262	0.9311	0.9489	0.9038	0.9749	1

All Pearson correlations are significant at the 95% confidence level.

Spearman rank correlations

	M 1, m	M 2, m	M 3, m	M 1, α	M 2, α	M 3, α
Model 1, order-m	1					
Model 2, order-m	0.844	1				
Model 3, order-m	0.874	0.971	1			
Model 1, order-alpha	0.804	0.965	0.930	1		
Model 2, order-alpha	0.963	0.844	0.869	0.855	1	
Model 3, order-alpha	0.836	0.940	0.961	0.969	0.885	1

All null hypothesis of independence between the variables are rejected
with a confidence level of more than 99%.

Table 5.14: Correlation between alternative models and methods, non-ger frontier

Pearson correlation coefficients

	M 1, m	M 2, m	M 3, m	M 1, α	M 2, α	M 3, α
Model 1, order-m	1					
Model 2, order-m	0.887	1				
Model 3, order-m	0.930	0.918	1			
Model 1, order-alpha	0.943	0.878	0.894	1		
Model 2, order-alpha	0.855	0.971	0.887	0.899	1	
Model 3, order-alpha	0.834	0.834	0.835	0.864	0.850	1

All Pearson correlations are significant at the 95% confidence level.

Spearman rank correlations

	M 1, m	M 2, m	M 3, m	M 1, α	M 2, α	M 3, α
Model 1, order-m	1					
Model 2, order-m	0.894	1				
Model 3, order-m	0.932	0.912	1			
Model 1, order-alpha	0.955	0.895	0.906	1		
Model 2, order-alpha	0.865	0.977	0.888	0.901	1	
Model 3, order-alpha	0.884	0.869	0.873	0.913	0.879	1

All null hypothesis of independence between the variables are rejected
with a confidence level of more than 99%.

Table 5.15: Correlations between scores with one and two production functions, ger frontier

Pearson correlation coefficients

	M1, 1 fr	M2, 1 fr	M3, 1 fr	M1, 2 fr	M2, 2 fr	M3, 2 fr
Model 1, 1 frontier	1					
Model 2, 1 frontier	0.870	1				
Model 3, 1 frontier	0.873	0.958	1			
Model 1, 2 frontiers	0.848	0.969	0.949	1		
Model 2, 2 frontiers	0.947	0.815	0.838	0.849	1	
Model 3, 2 frontiers	0.839	0.935	0.943	0.972	0.878	1

All Pearson correlations are significant at the 95% confidence level.

Spearman rank correlations

	M1, 1 fr	M2, 1 fr	M3, 1 fr	M1, 2 fr	M2, 2 fr	M3, 2 fr
Model 1, 1 frontier	1					
Model 2, 1 frontier	0.875	1				
Model 3, 1 frontier	0.881	0.966	1			
Model 1, 2 frontiers	0.852	0.974	0.957	1		
Model 2, 2 frontiers	0.949	0.820	0.849	0.844	1	
Model 3, 2 frontiers	0.841	0.939	0.951	0.971	0.874	1

All null hypothesis of independence between the variables are rejected with a confidence level of more than 99%.

Table 5.16: Correlations between scores with one and two production functions, non-ger frontier

Pearson correlation coefficients

	M1, 1 fr	M2, 1 fr	M3, 1 fr	M1, 2 fr	M2, 2 fr	M3, 2 fr
Model 1, 1 frontier	1					
Model 2, 1 frontier	0.872	1				
Model 3, 1 frontier	0.926	0.921	1			
Model 1, 2 frontiers	0.863	0.945	0.904	1		
Model 2, 2 frontiers	0.959	0.858	0.902	0.887	1	
Model 3, 2 frontiers	0.918	0.885	0.955	0.918	0.930	1

All Pearson correlations are significant at the 95% confidence level.

Spearman rank correlations

	M1, 1 fr	M2, 1 fr	M3, 1 fr	M1, 2 fr	M2, 2 fr	M3, 2 fr
Model 1, 1 frontier	1					
Model 2, 1 frontier	0.880	1				
Model 3, 1 frontier	0.932	0.924	1			
Model 1, 2 frontiers	0.864	0.946	0.902	1		
Model 2, 2 frontiers	0.957	0.872	0.915	0.894	1	
Model 3, 2 frontiers	0.915	0.886	0.955	0.912	0.932	1

All null hypothesis of independence between the variables are rejected
with a confidence level of more than 99%.

Chapter 6

General conclusion

Through a methodological review and three empirical studies, this thesis provides evidence on the issue of efficiency of health systems, as a means of heading towards Universal Health Coverage and the Sustainable Development Goal of Good Health and well-being.

Chapter 2 is a methodological review of the nonparametric measures of efficiency. The reference models, DEA and FDH, suffer from several flaws that have been dealt with in the last decade. Nevertheless, the introduction of environmental issue still raises some difficulties that have been questioned in empirical Chapters 3, 4 and 5.

Chapter 3 gives some empirical evidence about the macroeconomic efficiency of health systems in a sample of 120 low and middle income countries from 1997 to 2014. Countries are assumed to produce some maternal and juvenile survival with private and public health expenditures. Health expenditures increased over the study period (particularly private expenditures), and the mortality rates declined in most countries of the sample. As it is impossible to compare countries performance without considering the different contraints thay have to face, GDP per capita is introduced in the model as a non-discretionary input, i.e. an input that countries cannot control. GDP per capita is a proxy of economic developement and capacity, it is positively correlated to the level of education, to infrasructure for instance, factors that influence the production of health in a country. Technical efficiency is found to increase from 1997 to 2014, meaning the variability in health systems performances decreased. One of the result is that low income countries are mostly assessed as efficient, which is surprising economically, considered the poor performances of their health systems. This result is due to the way frontier models estimate frontier, looking for peers with worst or same environment, and it is a limitation od the implementation of frontier models in macroeconomic context.

Chapter 4 and 5 allow entering the "black box" of health system to provide some information about sources of inefficiencies. They both rely on survey data.
Chapter 4 sheds the light of the activity and efficiency of a sample of 30 Township

Health Centers in the prefecture of Weifang, Shandong province, China. In average, production of health services could increase by 30% for the same resources.
The interest of the chapter also consists in the study period, which includes the implementation of a major pharmaceutical reform. Before this reform, the medicine provision chain was highly inefficient, in the sense that it delivers some poor quality and very expenses drugs to Chinese households. This was due to strategic behaviors from healthcare providers and drug suppliers. The 2009 reform thus aimed at breaking those perverse incentives to increase the efficiency of the system. The results of our study suggest that it did not reach this objective, because drug revenue were replaced by subsides without any clear incentives mechanism. The asymmetry information went from the relation between health providers and patients (the patients were not sure that the drugs they were prescribed were necessary to their healing, and that they paid a fair price for those drugs) to the relation between health providers and financing authorities (if subsidies are not linked to THCs performance, there is no way to be sure that they maximize their efforts). The issue is still unsolved in this matter.
This chapter also highlights a possible oversizing of THCs, given the catchment area. Thinking in terms of optimal allocation of resources, deepening in benefit package of the insurance scheme (i.e. working of the deamand side of healthcare) appears to be a greater priority than adding human and physical resources to the system.

Chapter 5 focuses on the activity and efficiency of the Family Heath Centers of Ulan-Bator, Mongolia. The activities of those primary healthcare are limited to prevention and basic healthcare, that are provided for free to households. Given that out-of-pocket payments represent more than 40% of total health expenditures, it is crucial to spur activity of primary healthcare facilities. Yet, the assessment of the efficiency scores concludes the activity could increase, in average, by 30%, for the same inputs. Brokes to FHC activity, and efficiency seem to come from two different sides. First, the remuneration amount and composition (between fixed and incentives) have to be improved. In our discussion, the administrative burden and the low wages were a source of discontent, leading to demotivation to work and inefficiency.
Apart from the staff remuneration, the contract linking the FHC to health authorities (Ulan-Bator city and Ministry of Health) could also be reworked, in order to include a actual threat in the case where health is not provided in a satisfying way.

Both of our two studies highlight the crucial *stewardship* role of health authorities, to anticipate and avoid inefficiencies due to agency relation. If health reforms are not thought in a global way, strategic behaviors will lower their positive effects.

The demand size is also a determinant of health facilities efficiency in the two case studies. Patients' ability to pay, perceptions and expectations about health facilities are crucial components of healthcare facilities. In developing countries, a defiance is often observed towards primary healthcare facilities particularly. Patients prefer to bypass the referral system and pay for higher user fees than going to THCs in China, or FHCs in Ulan-Bator. In the perspective of *responsiveness*, the improvement of quality of healthcare is essential, for two reasons: first, for itself, and second, because if the quality of healthcare is poor, people will reject those facilities. Those who can afford it will go to other facilities, and the other will not seek for care at all.

The empirical works of this thesis also raise some questions in a methodological point of view, about the introduction of environmental variables in efficiency analysis. Chapter 2 presents the different available empirical strategies: one-, two- and three-stage models and conditional models.

From a theoretical point of view, conditional models are to be preferred, as they do not rely on the separability assumption, and they do not assume that the environmental variable has a positive or negative effect on performance. Moreover, they some efficiency scores purges from the effect of the environment. Empirically, they need a large data set, particularly to introduce several environmental variables to the model, that can make their implementation difficult.
Then, several options are available to the researcher. The choice between them depends upon several criteria. First, it depends upon the objective of the study. If there are several potential determinants to efficiency scores, and the researcher aims at determining which one(s) is(are) significantly associated to the scores, then the second-stage procedures are a useful tool. Moreover, they do not assume that environmental variable are positively or negatively related to efficiency, which is interesting in terms of exploratory study. It is the choice that has been made in Chapter 4 and 5 to estimate the determinants of efficiency scores.
If the researcher wants scores that are purged from the effect of environment, then one-stage models are an option widely used in the empirical literature, especially in macroeconomic studies. That is why it has been the choice made in Chapter 3. DMUs are compared to peers that are subjected to an environment at least as defavorable as them, *in the sample*. As a consequence, observations experiencing the worst environment (in Chapter 3, exhibiting very low values of GDP per capita) will automatically assessed as efficent. This is a limit of this method, because it implies

that subsaharian countries, with very poor health status and GDP per capita, are assessed as efficient. Yet, there is much to improve in those countries in terms of health system performance.

As a global conclusion, efficiency has become a major path to UHC, and the attainment of the third SDG, as a way to get "more health for the money". Efficiency can either be considered on a macroeconomic level, as the transformation of financial resources into health outcomes. Those performance assessment provides information for the policy makers, but there results have to be considered with precaution, because there multiple factors that affect the "health" production and no entirely satisfying way to take them into account, in terms of data and statistical tools.
Efficiency can also be considered at a lower level, assessing the production of health *outputs* and not *outcomes*. In this case, it can provide some evidence to the microeconomic phenomenon that prevent the system to be globally efficient.

Those studies lead to other issues to be dealt with. Due to a lack of data, this thesis does not tackle directly the question of the quality of care, though it is essential to the assessment of a health system performance. Further studies should increase indicators of quality (through mortality rate, readmissions for instance) and study more deeply their relation with efficiency. Assessment of the quality of care can also be done relatively to patient expectations. This is another path for extensions of this thesis: having a better knowledge of the demand side: what do people want ? Are they satisfied ? People are not only patients but also coproducers of health through their compliance to prescriptions, and global health bevahiors (Frenk, 2007). They can lower, or spur a policy effectiveness, so a better knowledge of their beliefs and expectations, put in relation with socio-economic characterictics give also give some useful pieces of information to policy makers.