Universitat Autònoma de Barcelona

Efficiency and Spatial Structure of the Public Healthcare System: The Ecuadorian Case

Juan Andrés Piedra Peña

Universitat Autònoma de Barcelona

Efficiency and Spatial Structure of the Public Healthcare System: The Ecuadorian Case

Juan Andrés Piedra Peña

Supervised by Rosella Nicolini and Diego Prior

Dissertation submitted to the
Department of Applied Economics
at the
Universitat Autònoma de Barcelona

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in the subject of
Applied Economics

September 2021

To my parents and grandparents, especially those that are not alive but will live ever after.

Acknowledgements

It has been a long journey since I started this new adventure that has finished with the completion of this dissertation. This means the end of a new stage of my life and I would like to express my gratitude to all the people that have been by my side and who have provided their support.

First of all, I am grateful to my thesis directors: Rosella Nicolini and Diego Prior. Diego, who has been supervising my research since my master thesis, from whom I have learnt about this fascinating world of efficiency and productivity and always found time to help and advise me. My most sincere gratitude to Rosella, who has provided her unconditional support and guidance ever since the first time we met. I do not have enough words to express how much I appreciate her advice and help throughout these four years, especially through what became the most difficult moments in my life.

To my family. My Mother and Father, who have supported me in every decision I have taken in my life. I dedicate this, what is one of the most important accomplishments in my life to my Father, who is no longer here but will be always be by my side. To Priscila, my partner and supporter; and my friends, who despite the distance have always been there.

I want to thank all the people from the Department of Applied Economics, who have provided a great environment to carry out my research. All the professors and researchers, who gave me advice and feedback during all the academic meetings and lectures, and my PhD mates.

I am also thankful to the people from the Department of Economics of Ca' Foscari University of Venice, for the opportunity to carry out my PhD visiting and providing the best environment possible, despite the Covid limitations. Especially to Francesco Moscone and his valuable advice and feedback.

I would like to thank my thesis jury professors: Emili Tortosa, Julie Le Gallo, José Luis Roig, Victor Gimenéz, Laura Lopéz and Claudio Thieme Jara for being part of this evaluation process. I am looking forward to receiving their valuable feedback and suggestions.

Finally, the financial support of the Pre-doctoral Trainee Research Scholarship (PIF) from the Universitat Autònoma de Barcelona is gratefully acknowledged.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This PhD dissertation aims to assess the efficiency performance of the public healthcare system of a developing country, under a context of policy reforms that moved towards higher equity and universal coverage. In so doing, we focus our attention on the public healthcare system of Ecuador.

The rapid increase in healthcare investment, as well as healthcare costs as a proportion of GDP in many countries all around the globe have placed a big emphasis on policies to improve hospital efficiency as a way to pursue health objectives, and at the same time, contain cost pressures (Bloom et al., 2015; Papanicolas and Smith, 2013). Attention to this matter has raised the interest in the academic literature, specially for developed countries. However, there is a scarce (although growing) literature applied to developing –and particularly Latin American– countries. As in many Latin American countries, Ecuador has faced a series of political reforms that have contributed to enlarge the existing regional income inequalities and led to profound territorial imbalances in the distribution of their healthcare resources. In 2008, the implementation of the new constitution led to new healthcare reforms to provide higher access to healthcare, specially for the marginal population. These reforms were accompanied by a significant public investment in health directed at increasing the supply of medical services, as well as improving the quality and performance of the public healthcare system. However, national reforms implemented in a country that suffers from deep territorial heterogeneities may have asymmetrical effects on the hospitals' performance, closely correlated with their geographical distribution. While, public investment aimed at increasing the healthcare system's performance may fail to attain the desired outcomes if these discrepancies are not properly taken into account. This is because hospitals located in developed regions deal with different conditions to treat patients (compared with hospitals in less-developed regions) that may affect their efficiency performance: they are likely to have more technology, capacity, resources, and higher competition from nearby hospitals, etc.

Throughout this thesis, we will explore relevant questions on the efficiency of public hospitals of Ecuador, and the effect that healthcare policies (particularly, after 2008) may have had on their performance. In our analysis, we address the territorial heterogeneities prevalent in the healthcare system in a framework were the geographical structure plays a major role. We start by defining efficiency as the optimal use of hospital inputs to attain a given set of healthcare outputs,[1] and use innovative optimization models to provide a single efficiency value (for every hospital) to measure it. Then, we use information on the geographic distribution of public hospitals to disentangle spillover effects among them and determine whether the spatial structure of the country's healthcare system plays a relevant role.

Therefore, with this dissertation we aim to contribute to the academic discussion and bring new evidence on the efficiency analysis of a developing country, that, as many others, has been suffering from continuous political changes and inequalities that have shaped the geographical distribution of its resources and modeled the behavior of its economic agents. The thesis consists on three essays developed in Chapter two, three, and four, respectively. Whereas, in Chapter one, we develop the motivation behind our study and a brief overview of the three essays. The main findings and policy implications are discussed in Chapter 5.

1.1 Motivation

As a determinant of the population's wellbeing, healthcare has commonly been the scope of economic debates and the public policy agenda for many countries around the globe. In developing countries, and particularly, Latin American ones, policy decisions have been aimed at reducing inequalities in healthcare access and attaining outcomes focused on the expansion of universal coverage due to their profound income inequalities and poor healthcare conditions. (Atun et al., 2015; Levy and Schady, 2013).

In Ecuador, many of these reforms have taken part in recent years, encouraged by the new constitution implemented in 2008. Some of them included free medical care to the uninsured population in hospitals belonging to the Public Ministry of Health and the mandatory enrollment of employees to the social security. These reforms were supported by a huge deployment of public investment by the central government, mostly targeted to the endowment of medical infrastructure and training (Granda and Jimenez, 2019).

However, lowering the barriers of access to medical treatment may have an effect on the way that public hospitals provide medical attention (i.e., on the efficiency of public medical provision), which is linked to the mobility of the demand. If policies that move towards higher equity promote an increase of demand for medical treatment in the short-run, then public hospitals face two potential scenarios. On the one hand, they can adapt to this higher demand by accommodating their spare resources and capacity, and hence,

[1] Conversely, efficiency can also be understood as the use of a given set of hospital inputs to attain an optimal level of healthcare outputs. These concepts will be explained in more detail throughout this dissertation.

improving their efficiency. On the other hand, this high inflow of patients may lead them to suffer from congestion effects, jeopardizing this efficiency.

Such a scenario raises many questions: when central policies are aimed at higher equity and universal coverage, what is happening to hospital efficiency? Does the increasing demand for medical treatment is affecting the efficiency performance of public hospitals? Are the best-performing hospitals those that are attracting more patients? May we observe a significant change in this performance after the new constitution was approved?

We tackle these questions throughout the following Chapters of this thesis. We intend efficiency in terms of the production process, which converts healthcare inputs (such as physicians or hospital beds) into health outputs (such as patients discharged or emergencies attended), being the efficient units those that either minimize inputs or maximize outputs. The importance to care about efficiency analysis has been widely addressed in the literature to monitor a country's healthcare system (see for example Hollingsworth, 2008; Cantor and Poh, 2018), identifying those under-performing hospitals in order to deliver tailored policies or to determine whether their efficiency has tended to increase in response to specific policy interventions. Thus, it can be used as a proper tool for governments and decision-makers to enhance the system's performance and promote the welfare state (Wagstaff, 1989). Its importance is emphasized in developing countries such as Ecuador, where the pressing need for proper resource allocation is imperative, given the restricted level of medical resources and healthcare budget (Hafidz et al., 2018; Kumbhakar, 2010).

In addition, we need to consider that Ecuador (as well as many other Latin American countries) suffers from deep territorial disparities, with a profound income concentration in certain regions (Mendieta Muñoz and Pontarollo, 2016). From the healthcare perspective, this heterogeneity derives in high concentration of hospitals and medical resources in developed areas, generating technological heterogeneities among public hospitals. This concentration suggests the existence of a spatial pattern where competition and learning effects among spatial clustered regions may lead to spatial clustering of healthcare behavior. That is, endogenous and exogenous spatial spillovers may be involved and the provision of medical services may entail an influence beyond the regional borderline (Bech and Lauridsen, 2008). If not taken into consideration in the economic models, this structure of spatial autocorrelation may produce statistically biased results, and hence, to misleading conclusions.

In this thesis, we take into consideration the regional heterogeneity and address it in a framework of analysis where the geographical structure plays a key role to determine the origin of the spatial heterogeneity, its determinants, and offers us a proper setting to disentangle potential endogenous and exogenous spillover effects. In this spirit, this dissertation addresses three different aspects. First, in Chapter 2, we evaluate the efficiency performance of the Ecuadorian public hospitals, determining whether we can observe significant differences on its evolution before and after the new constitution was implemented. Second, in Chapter 3, we tackle the questions of whether this efficiency is associated with significant strategic interactions among hospitals located within a bounded area, due to the existence of spillover effects, and whether this efficiency is affected by the increasing demand for medical attention. In particular, we explore whether this effect has varied after the new constitution. Finally, in Chapter 4, we focus on the patient mobility among the regions of the country and the hospitals located within those regions, and determine wether the performance of a hospital is driving patients from other regions to that hospital and to neighboring hospitals as well.

In what follows, we outline more precisely the content of the three core Chapters of the thesis. The first focus of our attention is on the development of an empirical methodology that allow us to measure a robust and single value of hospital efficiency that exploits a panel data setting.

1.2 Analyzing the effect of health reforms on the efficiency of Ecuadorian public hospitals[2]

In this Chapter, we investigate whether the Ecuadorian healthcare reforms carried out since 2008 have affected the efficiency performance of public hospitals. To tackle this question, we construct a database with hospital information coming from the Institute on National Statistics and Census (INEC) that covers the period from 2006 to 2014. To consider the technological heterogeneities in the hospitals' endowment, we use a two-stage approach. In the first stage we use factor and cluster analysis to obtain three clusters of hospitals: high-tech, intermediate-tech, and low-tech. In the second stage, we exploit a novel panel Data Envelopment Analysis proposed by Pérez-López et al. (2018) to estimate robust efficiency measures over time. With this approximation, we are able to consider the heterogeneity of healthcare institutions in the analysis of their efficiency performance. The results show a significant decrease in the average efficiency of low and intermediate technology hospitals after the new constitution was adopted in 2008. The decline in efficiency coincides with the two reforms of 2010 and 2011 that brought on higher social security coverage.

The decline in efficiency after the new constitution was adopted lead us to think that it might be related with the increase in the demand that public hospitals faced due to the new reforms. However, if hospitals were no making use of the spare resources and capacity in an efficient manner, we could be facing the opposite situation. That is, the increase in demand could be encouraging a better use of resources that fuels efficiency, whilst the decline in hospitals performance could be due to other external reasons. In fact, as mentioned, the spatial structure could be playing a key role in the hospitals' behavior. If spatial autocorrelation is significant, increases in demand could be affecting not just an observed hospital, but also those that surround it, through spillover effects. In this

[2]This Chapter has been published in the Working Paper series of the Graduate Program in Applied Economic Research (GEAR), 2020-01.

line, with significant spatial autocorrelation, we could also face a situation where changes in the efficiency performance of one hospital may be leading to changes in efficiency of neighboring hospitals as a response, suggesting strategic interactions among them, and, hence, the existence of spatial dependence. These questions are addressed in the next study.

1.3 Spatial dependence in hospitals efficiency: A spatial econometric approach for Ecuadorian public hospitals[3]

In this Chapter, we analyze whether the efficiency of Ecuadorian public hospitals experiences spatial dependence. We additionally investigate whether demand variations are affecting the public hospitals' efficiency through direct and spillover effects, in particular, whether this effect significantly changes after the adoption of the new constitutions in 2008. We take upon the efficiency results estimated in Chapter 2 as a first stage of our strategy. In the second stage, we use a spatial econometric framework to disentangle direct and spillover effects. The results confirm the existence of positive spatial interactions among public hospitals' efficiency. Following Longo et al. (2017) these results suggest the existence of complementary strategic interactions among public hospitals in terms of efficiency. Positive direct and spillover effects are found from demand increases, reinforced after 2008.

The positive direct and spillover effects of demand increases in hospital efficiency emphasize the importance of understanding the dynamics of patient mobility and the determinants behind the decision of selecting a given hospital. In a country characterized by deep spatial concentration, such as Ecuador, developed regions will concentrate big and best-performing hospitals. Hence, is basic to assume that patients will select those hospitals located in developed regions. If those best-performing hospitals are located in the developed areas, then patients from less-developed regions will seek the best possible medical attention, hence, traveling beyond regional borders to receive treatment in those high-performing hospitals. However, these movements may also present a spatial dependence in the form of patients inflows to neighboring hospitals. Conversely, spatial dependence in the form of outflows of patients from neighboring less-developed regions may also be observed. In the last study, we focus our attention on the analysis of patient migration flows and determine whether hospital efficiency performance could be used as a determinant to attract patient demand and whether spatial dependence is found in those patient migration flows.

[3]This Chapter has been published in the Working Paper series of the Graduate Program in Applied Economic Research (GEAR), 2020-05.

1.4 On the dynamics of patient migration flows: is efficiency performance explaining inflows for neighboring hospitals? An application to the Ecuadorian healthcare system[4]

In this Chapter, we analyze whether higher efficiency performance of public hospitals attracts more interregional patients and whether these inflows present spatial dependence in the form of larger inflows of patients from neighboring hospitals in the region. We apply a two-stage approach. In the first stage we use a conditional order-*m* model to estimate robust efficiency values for a sample of hospitals in 2014. We intend efficiency as the maximization of hospital outputs with a given set of inputs. The use of conditional order-*m* also allow us to take into consideration variables that proxy the level of development of a hospital's region that have an exogenous effect on the hospital's performance. In the second stage, we use a spatial interaction model to estimate the effect of hospital efficiency upon patient mobilization and the spillover effects on the migration dyad. The results show a positive and significant pulling effect of specialized hospital efficiency on patient flows. Furthermore, we find significant spillover effects on neighboring hospitals in the same region and from hospitals in the region neighboring the origin.

All in all, in this thesis we conclude that the increase in the demand for medical treatment had an overall positive effect on hospital performance, both through direct and spillover effects. Potential drivers of this effect refer to the inefficient use of the spare resources and capacity of public hospitals. The time that hospitals had to adapt to the forthcoming inflow of patients before the reforms and the public investment deployed may also have a significant participation in the effect. The results also provide evidence that the efficiency performance of specialized hospitals has a strong pulling effect of patients from less-developed regions. This inflow of patients is being captured by neighboring hospitals who are increasing their efficiency to attract this demand. This interactions among public hospitals are showing competition effects that enhance the regional performance and may translate into welfare improvements. However, these welfare gains may come in an asymmetrical manner (Brekke et al., 2014, 2016), as people living in less developed regions are not able to get medical attention in these best-performing hospitals. The results of this dissertation provide useful recommendations for policy decision-making. Policy implications drive the attention to the design of well planned healthcare strategies considering territorial externalities, technological endowment and specialization level as key features. Higher public investment can be targeted to increase the supply of specialized treatment in less-developed regions. In developed ones, decision-makers can take advantage of spillover effects to promote efficiency strengthening hospital reforms and well allocated public investment to enhance the regional healthcare system's performance.

[4]This Chapter has been published in the Working Paper series of the Graduate Program in Applied Economic Research (GEAR), 2021-01.

Chapter 2

Analyzing the effect of health reforms on the efficiency of Ecuadorian public hospitals*

Abstract

This study aims to assess whether Ecuadorian health reforms carried out since 2008 have affected the efficiency performance of public hospitals in the country. We contribute to the literature by shedding new light on the effects on public healthcare efficiency for developing countries when policies move toward health equity and universal coverage. We follow of a two-stage approach, wherein the first stage we make use of factor and cluster analysis to obtain three clusters of public hospitals based on their technological endowment; we exploit Data Envelopment Analysis for panel data in the second stage to estimate robust efficiency measures over time. Our innovative empirical strategy considers the heterogeneity of healthcare institutions in the analysis of their efficiency performance. The results show a significant decrease in the average efficiency of low and intermediate technology hospitals after the new constitution was adopted in 2008. The decline in efficiency coincides with the two reforms of 2010 and 2011 that brought on higher social security coverage.

Keywords: healthcare efficiency, health reforms, metafrontier, panel data DEA
JEL: I18, C14, H51

*We would like to thank the scientific committee and the participants of the Applied Lunch Seminar at UAB; the participants of the ISEG International Conference (Brasov, 2019) and the Meeting on Public Economics (Barcelona, 2020); and Núria Mas for their valuable comments. Any remaining errors are our own responsibility.

2.1 Introduction

As a determinant of population wellbeing and economic growth, improving health has become a major topic in economic debates and features high on the public policy agenda in many countries around the world. Healthcare is one of the main public policies implemented by most governments and improving the efficiency of its delivery is a crucial goal of health service providers globally (Au et al., 2014). Nevertheless, health systems in Latin America face specific challenges, including insufficient human resources and training, lack of evaluations of strategy outcomes, operating levels, and weaknesses in the public health system's response capacity, among others (Ruiz-Rodriguez et al., 2016).

In this context, the Ecuadorian health sector has undergone a continuous process of deterioration as a consequence of neoliberal reforms carried out in the 1990s (Homedes and Ugalde, 2005) and the crisis of 2000, which mainly affected the most deprived population. This deterioration meant a progressive reduction of the health budget, lack of infrastructure investment and shrinking human resources, and low quality and coverage of public services (López-Cevallos and Chi, 2010; Malo-Serrano and Malo-Corral, 2014).

In 2008, the government of former President Rafael Correa brought in a new constitution that guarantees health as a citizens' right and introduced a series of health reforms that moved toward universal coverage and free primary medical services. This change was accompanied by substantial public investment in the health sector in order to improve the quality and quantity of medical services (De Paepe et al., 2012; Hartmann, 2016).

There are two potential effects of these reforms on the efficient delivery of medical services to the population. On the one hand, the increased demand for health services by the newly insured population might encourage hospitals that were not using their spare capacity and/or medical resources correctly to take full advantage of them by optimizing their resources and delivering a more efficient service. On the other hand, in the desire to promote equal access to health, these policies might lead to over-demand for health services that hospitals are unable to cope with in the short term (Smith and Yip, 2016). The lowering of access barriers to medical services may cause the volume of patients to increase, thus raising input consumption and hence, healthcare costs. Even if there is an increase in capacity and personnel, the possibility remains that this will drive a reduction of efficiency if the system is not able to manage its resources properly (Cozad and Wichmann, 2013).

It is clear that improving the efficiency of resource use is a key issue in most health systems, and is particularly acute in developing countries where there is a pressing need for proper resource allocation given the limited level of overall infrastructure, resources and health budget (Kumbhakar, 2010; Hafidz et al., 2018).

In light of the above, a relevant question arises: when the objective of equity provides the rationale for governments' central involvement in healthcare, is healthcare efficiency negatively affected? This question has been addressed in the literature over the years (e.g.

Nord et al., 1995; Culyer, 2006), but the very nature of the healthcare market makes it difficult to define the concept of optimal allocation (Culyer, 1971). The degree and combination of certain healthcare industry characteristics distinguish it from any other sector (Arrow, 1963; Culyer, 1971); furthermore, no single health system model fits all countries (Smith and Yip, 2016), and the resulting scenario presents many challenges for health economists to define the right theoretical model that fits both equity and efficiency into the economic equilibrium (Leach, 2010). This context has led many authors to highlight the importance of carrying out empirical studies that have the potential to clarify, measure performance, and evaluate, particularly around specific policy decisions (Carr-Hill, 1994; Xingzhu, 2003; Hollingsworth, 2012; Smith and Yip, 2016). In this sense, healthcare efficiency studies can help to further understanding on the nature of institutional inefficiency in a particular economic scenario, allowing the public authority to improve health or to enhance the performance of their healthcare systems, not just by merely increasing the available resource, but by making a better use of them (Carrillo and Jorge, 2017).

We contribute to this topic by focusing on the Ecuadorian context, which offers a framework of analysis characterized by health reforms designed to bring in universal coverage and seeking the "well-living" of the population (López-Cevallos et al., 2014; Espinosa et al., 2017). We present an analysis with current information on the efficiency changes in a reality that is still adapting to these reforms, and must face potential problems arising in the short term. Thus, this study considers the public health reforms introduced since 2008 in order to assess whether they have negatively affected the efficiency performance of public hospitals in Ecuador.

To properly account for the situation in Ecuador, we need to consider the territorial heterogeneity of the Latin American context. In Latin American countries, the question of territorial disparities must be addressed in any economic analysis and the profound imbalances between regions characteristic of many Latin American countries are well documented in the literature (Cuadrado-Roura and Aroca, 2013). As a result of the unequal distribution of income in these countries, such disparities are higher than in other continents, reaching extremes where the wealthiest region has an income per capita almost ten times the poorest in some cases (Cuadrado-Roura and Aroca, 2013). In addition, Latin American countries have been subject to neoliberal reforms supported by the World Bank and the IMF that have aggravated the equity and efficiency of their health systems and widened existing urban-rural and inter-regional inequalities (Homedes and Ugalde, 2005). Given that urban dwellers exercise more political pressure than rural populations, and that large cities have more political clout than smaller ones, a disproportional amount of health resources is concentrated in large urban areas.

In Ecuador, these neoliberal reforms deeply exacerbated its existing regional disparities and led to a structural segmentation and fragmentation of its health system (Hartmann, 2016). One outcome of this process is the marked technological heterogeneity among its health institutions, with a greater concentration of hospitals with higher levels of technology in the most developed regions. In light of this evidence, the need to develop economic analysis tools that take into account these regional disparities has been widely called for in recent literature focusing on Latin America and particularly on the Ecuadorian economy (Mendieta Muñoz et al., 2015; Mendieta Muñoz and Pontarollo, 2016; Szeles and Mendieta Muñoz, 2016).

In order to consider the aforementioned technological differences in the Ecuadorian public health system, we introduce a methodological innovation in a two-stage analysis. In the first stage, we use multivariate factor analysis and clustering techniques to find homogeneous groups with uncorrelated characteristics of technological endowment. We compare them with a common frontier, similarly to the metafrontier approach (Battese and Rao, 2002; Battese et al., 2004; O'Donnell et al., 2008). Here, we face an issue that is commonly present in low and middle income countries, which relates to the lack of sufficient data to perform a meaningful efficiency analysis (Au et al., 2014) (this phenomenon is also described as the 'curse of dimensionality' in the literature on efficiency estimations. See Daraio and Simar (2007b)). The results show that when the data is scarce, and the system is heterogeneous, the metafrontier might not be representing the unrestricted production frontier identified by Battese and Rao (2002), and the cluster comparison might produce the wrong conclusions. In order to obtain a metafrontier that allows us to perform a good comparison between clusters, we propose an alternative approach to obtain a common frontier for developing countries, based on the seminal studies of Banker and Morey (1986, 1996) and Podinovski (2005).

Having identified these new clusters, in the second stage we combine the metafrontier and panel data envelopment analysis (panel data DEA) (Surroca et al., 2016; Pérez-López et al., 2018) to account for robust efficiency values over time. Considering an empirical methodology that allows for this heterogeneity enables us to obtain consistent efficiency values, which might otherwise be biased if we applied classical efficiency measurement techniques to the whole sample (Mitropoulos et al., 2015).

Our analysis covers the period from 2006 (starting before the new government came to power) until 2014. We use data from the Annual Survey of Hospital Beds and Discharges and the Survey of Health Activities and Resources provided by the Ecuadorian Institute of Statistics and Censuses (INEC). As far as we are aware, no published empirical literature has applied this methodology to assess efficiency performance in a Latin American country such as Ecuador. In this regard, this work can be of great utility for academics and policymakers and may be used to justify the implementation of public health policy and managerial health improvement reforms.

The paper begins with a brief contextualization of the Ecuadorian healthcare system, in Section 2.2. Section 2.3 then explains and reviews the methodological framework and the most recent cited empirical literature; the methodology is presented in Section 2.4. In Sections 2.5 and 2.6 we discuss the data and results obtained. Finally, in Section 2.7 we present the main conclusions.

2.2 Institutional context

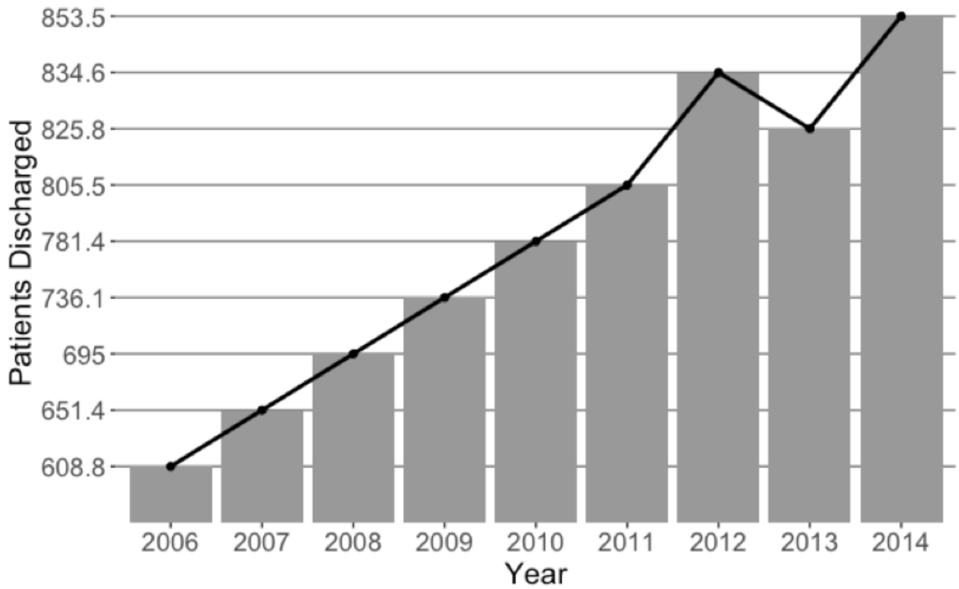
Ecuador's healthcare system combines the public and private sectors. The public sector comprises the Public Ministry of Health (MSP), the Ministry of Social and Economic Inclusion (MIES), the municipal health services and the social security institutions.[1] The MSP provides health services to the whole population, the MIES and municipality health programs supply medical care to those without insurance, while the social security institutions cover the affiliated working population (Lucio et al., 2011). Since Rafael Correa's government came to power in 2007 and implemented the new constitution in 2008, an unprecedented level of public investment has taken place in Ecuador, focusing on primary services such as education and health. This new government, called the "Citizen Revolution", marked the beginning of a stage of democratic stability that gave the State the central role that guarantees and promotes the enjoyment of rights for the entire population.

Additionally, the 2008 constitution also brought in significant changes, especially in access to health services and social security coverage. Articles 3 and 34 of the National Constitution state that health is a right guaranteed by the State and it shall ensure the full exercise of the right to social security. On this basis, several reforms have been implemented such as coverage of children under the age of 18 in 2010 (Article 102, Social Security Law) or deprivation of liberty for employers who do not affiliate workers within a maximum period of 30 days in 2011 (Art. 244, Organic Comprehensive Criminal Code), resulting in a significant increase in the number of active beneficiaries until 2014 (Orellana et al., 2017).

Some results can be drawn from the Annual Survey of Hospital Beds and Discharges and the Survey of Health Activities and Resources. Figure 2.1 shows that between 2006 and 2014 the total number of patients discharged from public hospitals rose from 608 thousand to over 853 thousand, representing a 40 percent increase in patients attended. The biggest jump in medical attention was seen in 2012, which coincides with the period following the above-mentioned reform of 2011.

[1] Ecuadorian Social Security Institute (IESS), Social Security Institute of the Armed Forces (ISSFA) and Social Security Institute of the National Police (ISSPOL).

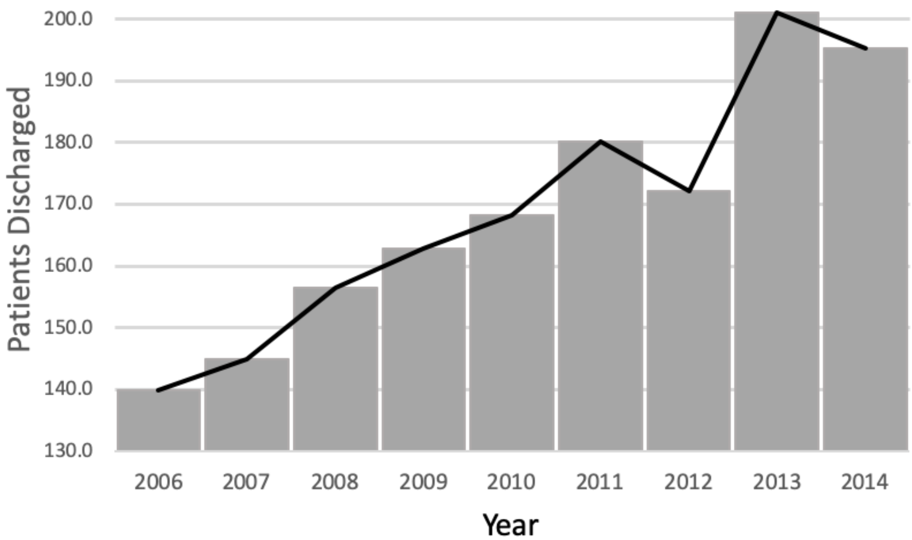
Figure 2.1: Number of discharged patients in public hospitals in the period 2006–2014 (in thousands)



The final goal of these reforms was to improve the wellbeing of the most deprived citizens, in pursuit of a more equal access to medical services. However, despite the new investment in infrastructure and human capital, little attention was paid to how these new health improvements would affect the performance of the hospitals. De Paepe et al. (2012), discusses how the introduction of these new free services and the increase in the insured population brought about a “demand crisis”, especially in larger cities. This higher demand meant that public hospitals could not cope with the influx of patients and drove the need to contract private services to stem public discontent. Some evidence of this measure can be found in Figure 2.2, where 2011 and 2013 present the biggest jump in patients attended in private clinics for three of Ecuador’s largest and most densely populated cities (Quito, Guayaquil and Cuenca). Unfortunately we do not have information on the patient referrals to private healthcare institutions, but the decrease in discharged patients from public hospitals in 2013 (Figure 2.1) might be signaling a small alleviation in the demand for public healthcare services that seems to have been referred to private clinics (Figure 2.2).

The public-private interface in healthcare can also represent a source of inefficiency for the public healthcare system; this phenomenon is known in the literature as *cream skimming* (Ellis, 1998). The concept relates to the selection of patients with lower expected cost of treatment by hospitals and healthcare providers, which can decrease their costs by selecting patients with less severe medical conditions (Cheng et al., 2015). In the Ecuadorian context, the increase in the volume of patients in the public sector derives in the referrals to private healthcare institutions, but the private sector will only be prepared––or willing––to treat the simplest less serious and complex cases, from which they

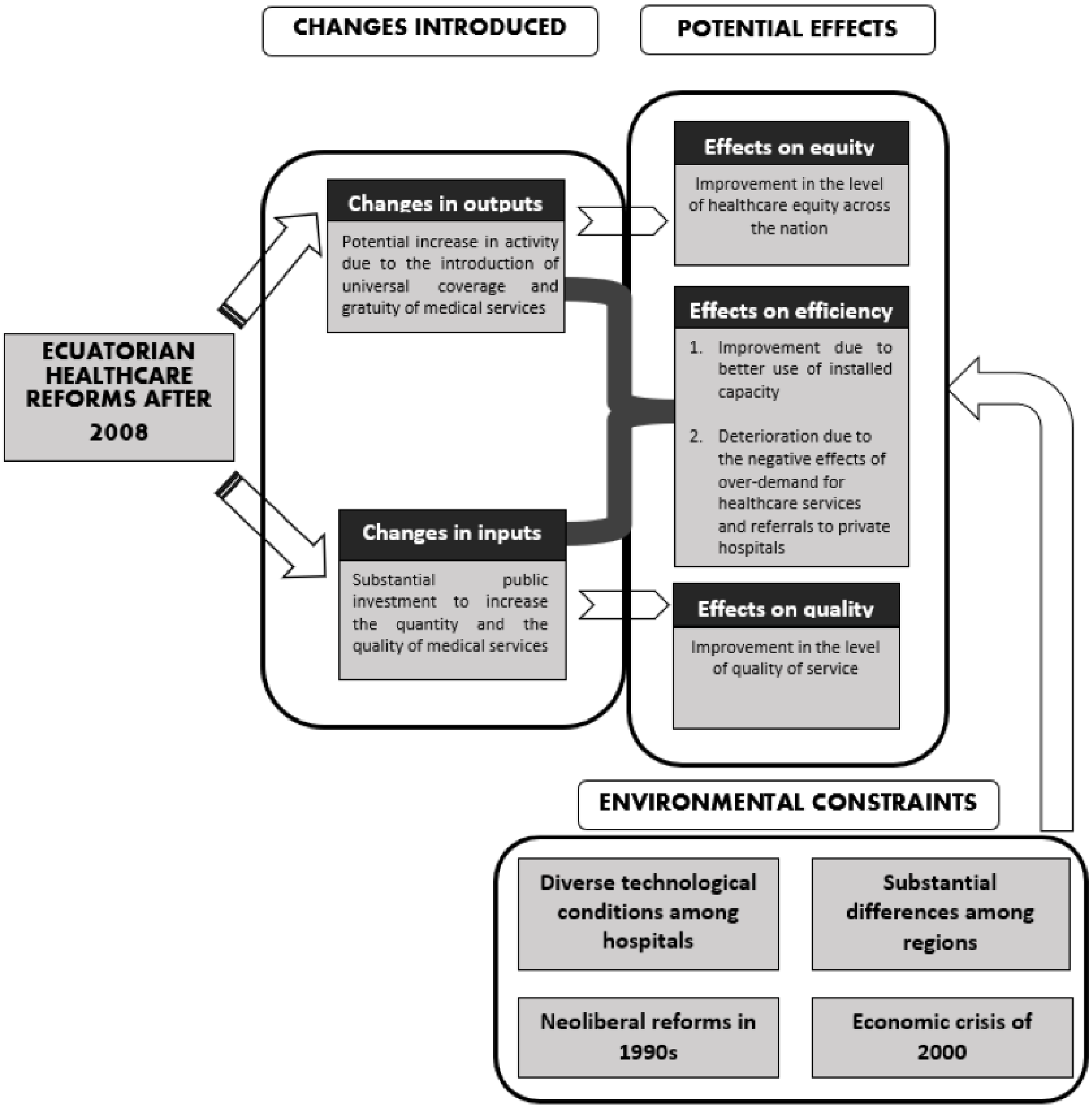
Figure 2.2: Number of discharged patients from private clinics in Quito, Guayaquil and Cuenca, 2006–2014 (in thousands)



derive their profitability. In these circumstances, the public sector deals with the most complex cases, which consume more inputs and are less profitable, and refers the simplest cases to the private hospitals, which consume fewer inputs and are more profitable.

The dynamic deriving from the above-mentioned changes introduced in the healthcare system is illustrated in Figure 2.3.

Figure 2.3: Dynamic in the Ecuadorian healthcare system



These facts reveal the need for an empirical strategy with which to measure the efficiency changes and the potential effects of the public health reforms. The following section reviews the methodological framework most commonly used in the literature to address this relationship.

2.3 Literature review

This paper takes its methodological framework from Production Theory (Debreu, 1951; Koopmans, 1951; Farrell, 1957) and the metafrontier production function (Battese et al., 2004; O'Donnell et al., 2008). The main idea of productive efficiency is linked to the concept of Pareto Efficiency Allocation, according to which a resource endowment is efficient when there is no other possible allocation that makes a Decision Making Unit (DMU) better off. [2] The efficiency analysis can be applied to any DMU, and we can distinguish between technical efficiency and allocative efficiency. The latter assumes that relative prices are known and are reasonably stable. Following Farrell (1957), an efficient unit would obtain a value of one, and it could take an input or an output orientation. The former focuses on minimizing the input use and the latter on maximizing the output obtained in the production process.

In turn, the metafrontier production function assumes that DMUs from different environmental conditions, regions, and/or countries face different production opportunities and have to make different choices taking into account variations in the feasibility of input-output combinations. These *technology sets* will therefore be different and difficult to compare. Battese et al. (2004) and O'Donnell et al. (2008) develop a way to make efficiency comparisons across groups of DMUs. They do this by measuring efficiency relative to a common *metafrontier* which is defined as a boundary of an unrestricted technology set, and they also define group frontiers to be boundaries of limited technology sets that are embedded in the common frontier.

The metafrontier envelops the group frontiers. Efficiencies that are measured with respect to the metafrontier can be decomposed into two components: one component measures the distance from an input-output vector to the group frontier, which is the common measure of technical efficiency; and a second component measures the distance between the group frontier and the metafrontier, which is defined as a *technological gap ratio* (TGR), and represents the restrictive nature of the production frontier.

In the empirical literature, healthcare efficiency measurement has attracted growing interest over the years. Most studies focus on measuring the efficiency and productivity of healthcare using parametric and non-parametric applications. Several authors offer extensive reviews of the published literature (Hollingsworth, 2003, 2008; Worthington, 2004; O'Neill et al., 2008; Cantor and Poh, 2018). However, more than half of these were applied in the US and Europe, while just a few have examined developing countries, although this number has been rapidly increasing over the last years (Hollingsworth, 2008).

Here, data envelopment analysis (DEA) has excelled over other techniques, as a non-parametric linear programming method for measuring relative efficiency of homogeneous DMUs. This approach is more consistent with economic theory as it locates technical

[2] Any unit of analysis can be labeled as a DMU, for example individuals, departments, firms, municipalities, or, in the case of this study, hospitals.

or Pareto inefficiencies instead of measuring efficiency based on averages (O'Neill et al., 2008; Cantor and Poh, 2018). It also allows a data driven assessment of the production process without strong assumptions about the functional form, which is a major advantage in the face of uncertainty (Staat, 2011).

In this regard, the published literature applied to Latin American countries has been somewhat scarce. De Castro Lobo et al. (2010a) use network DEA to assess the performance and integration of healthcare and teaching dimensions in Brazilian university hospitals. Keith and Prior (2014) measure technical efficiency using the DEA approach and evaluate the potential presence of scale and scope economies in Mexican private medical units. Ruiz-Rodriguez et al. (2016) also apply DEA analysis in a four-stage approach along with a series of Tobit regressions in order to estimate the technical efficiency of the three women's health promotion and disease prevention programs in Bucaramanga, Colombia. However, these studies aim to measure efficiency in a specific year analysis, and none of them attempted to identify the effects of health reforms.

Following this line, several authors have addressed research questions regarding the relationship between health reforms and performance (e.g. Linna, 1998; Maniadakis et al., 1999; Van Ineveld et al., 2016), but very few have focused on Latin American countries (Arocena and García-Prado, 2007; De Castro Lobo et al., 2010b). Most of them make use of non-parametric methods, like DEA models, to calculate efficiency scores and Malmquist productivity indices, subsequently decomposable on efficiency and technological change, which have been widely employed in the literature in part because they require neither relative price information nor restrictive behavioral assumptions for their estimation (Chowdhury et al., 2014).

Table 2.1 presents a summary of the most recent cited literature on health reforms and hospital performance. To answer questions regarding health reforms, the literature has mainly followed two approaches that rely largely on the availability of the data, and the results may depend on the context in which it took place and the type of reform implemented.

Table 2.1: Summary of the literature

Authors	Country	Year of the Reform	Study Period	Methodology	Conclusions
Linna (1998)	Finland	1993	1988–1994	Time-varying SFA, DEA, Malmquist Index	3-5% annual average increase in productivity due to cost efficiency and technological change. The state subsidy reform of 1993 did not seem to have any observable effects on hospital efficiency

Table 2.1 (continued)

Authors	Country	Year of the Reform	Study Period	Methodology	Conclusions
Maniadakis et al. (1999)	Scotland	1990	1991–1996	Malmquist Index	Overall net gain in productivity, although a slowdown was observed in the first year after the reform. The change was due to technological change as hospitals were relatively efficient at the time of the reform
Sommersguter-Reichmann (2000)	Austria	1997	1994–1989	DEA, Malmquist Index	Positive shift in technology between 1996 and 1998 as a result of the financing reform
Arocena and García-Prado (2007)	Costa Rica	2000	1997–2001	Generalized distance functions, Malmquist Index	Improvement in hospital performance mainly driven by quality changes and particularly significant for small hospitals. Productivity growth is mainly due to technical and scale efficiency rather than technological change
De Castro Lobo et al. (2010b)	Brazil	2004	2003–2006	DEA, Malmquist Index	Financial reform was a good stimulus for efficiency gains, but technology change was not able to take place
Van Ineveld et al. (2016)	The Netherlands	2005	2005–2010	DEA, Malmquist Index	Larger differences in efficiency among hospitals. In 2009–2010 the number of larger and more efficient hospitals decreased
Valdmanis et al. (2017)	Scotland	Series of reforms through the period of analysis	2003–2007	Malmquist index, time-series trend analysis	No steady movement with the use of the Malmquist index, but the time-series trend analysis revealed a trend of improvement in technological but not technical change
Xenos et al. (2017)	Greece	2008	2009–2012	DEA, Malmquist Index	Negative impact of the crisis in 2009 with 91% of the hospitals achieving a score lower than one. Improvement between 2010 and 2011 followed by stabilization in 2011–2012

Table 2.1 (continued)

Authors	Country	Year of the Reform	Study Period	Methodology	Conclusions
Giménez et al. (2019)	Colombia	1993	2009–2013	Mamlquist-Luenberger index	Total productivity worsened by 1% during the period of analysis, mainly due to technological backlash

Source: The authors

On the one hand, the hospitals' performance can be evaluated by considering their performance after a certain reform has taken place. For example, Maniadakis et al. (1999) use Malmquist indices of productivity and quality to evaluate the reforms of the UK National Health Service in the early 1990s in acute Scottish hospitals over the first five years of the reforms. Overall, they find that the hospitals showed a gain in productivity, although an initial regress was observed in the first year after the reform. The changes in productivity were led by technological rather than efficiency changes, given that hospitals were operating close to the industry boundary at the time of the reform and their position changed little over time. Van Ineveld et al. (2016) assess the productivity performance of Dutch hospitals since the health system reform of 2005. They use DEA based measures in a cross-sectional and longitudinal analysis as well as the Malmquist index; they find that the efficiency gap among hospitals has widened, benefiting some of the smaller hospitals but not some larger ones, which might be a consequence of the 2005 reform. Xenos et al. (2017) study the dynamics of efficiency and productivity in Greek public hospitals after the 2008 financial crisis, where in the period of study (2009–2012) hospital budgets were reduced by 40%. Using DEA and bootstrapping Malmquist analysis they find a negative impact in productivity due to the crisis in 2009 with a recovery in 2010 and a posterior stabilization. The latest study conducted in Latin America is that of Giménez et al. (2019). They analyzed the performance of level 1 Colombian hospitals for the period 2009–2013 to evaluate how the health system was performing after the 1993 reform. They also extended the analysis to find out whether the efficiency of high-level hospitals was affected by patient referrals from primary care centers. Using the Malmquist-Luenberger index, they found that productivity decreased by 1% during the period of analysis, providing evidence of a deficient performance in public hospital efficiency after the 1993 reform.

On the other hand, the approach can use a before-after design, which can further benefit the analysis, given that we can actually check the hospitals' behavior after the reform and gain an initial insight into its influence on their performance. Linna (1998) study the development of cost efficiency and productivity of Finnish hospitals before and after the 1993 state subsidy reform. This author uses panel data Stochastic Cost Frontier models as well as DEA and the Malmquist productivity index and finds that productivity progress was due to both technological change and cost efficiency change; however the state subsidy reform did not seem to have any observable effects on hospital efficiency since it appears to have been improving well before the reform. Sommersguter-Reichmann (2000) studies the 1997 hospital financing reform in Austria and evaluates the changes in productivity between 1994 and 1998. Using DEA and the Malmquist index, she finds a technological improvement was an immediate consequence of the financing reform. Arocena and García-Prado (2007) analyze how Costa Rican hospital efficiency and quality responded to the reforms carried out over the period 1997–2001. They use a generalized output distance function to obtain a Malmquist index that accounts for productivity changes while controlling for quality of care, and find an overall improvement in hospital performance following the reforms due to an increase in quality rather than a better use of resources, and more notably for small hospitals. De Castro Lobo et al. (2010b) evaluate the performance and productivity changes for Brazilian Federal University hospitals before and after the financing reform of 2004. Using DEA and the Malmquist index, they find that the financial reform gave the hospitals an opportunity to gain efficiency, but not for technological change. Valdmanis et al. (2017) applied the Malmquist index and time-series trend analysis to assess the shift in efficiency and technology in Scottish hospitals over the period 2003–2007 where health reforms required them to improve their services with fixed budget constraints. They did not find a consistent direction of either improvement or devolution; however, through the use of time-series analysis, they found a trend of growth in technological change.

To the best of our knowledge, most of the literature has relied on a Malmquist index analysis along with other parametrical and non-parametrical approaches. However, none of them has tried to account for technological heterogeneity that may arise when studying hospital performance; just a few studies have considered, at most, hospital size. In this sense, while the technical efficiencies of DMUs measured with respect to a given frontier are comparable, some problems might arise among hospitals that operate under different technologies (Mitropoulos et al., 2015). The efficiency of hospitals that work under a specific production technology cannot be comparable with those of different technology. This problem might be minimized in a context where the country is relatively centralized, and there are no significant regional differences; this might reinforce the homogeneity of the sample (Arocena and García-Prado, 2007). However, in a country like Ecuador, where regional heterogeneities have proved to be strong drivers of the socio-economic reality, the need to account for the potential heterogeneities is crucial.

Recent papers like Mitropoulos et al. (2015) and Chen et al. (2016) have tried to account for some technological differences in the health sector, but their grouping criterion is somewhat subjective and—in the first case—focused on a cross-sectional study. Authors like Carrillo and Jorge (2017) propose DEA-based procedures to classify the performance of intra-regional health systems. However, the application is —again— focused on a cross-sectional study and is aimed at providing an homogeneous ranking based on performance, which scapes the scope of this paper. The need to find a clear criterion to group the hospitals in our sample has an important relevance. There is no clearly established way to separate them into homogeneous groups. A priori, the units can be grouped on the basis of geographical, economic or political boundaries.

Here, our aim is to take into account technological heterogeneities among the hospitals by considering their resources and capacity. To this end we consider cluster analysis. O'Donnell et al. (2008) encouraged the use of multivariate techniques when natural boundaries are unavailable. This method has been previously adopted by Balaguer-Coll et al. (2013) to assess the provision of public services and facilities in Spanish municipalities, which they clustered according to output mix, environmental conditions and level of powers, although they followed a cross-sectional approach. Choi and Park (2019) also apply this method to classify Low and Middle income countries and assess the efficiency of governmental capacity to enhance social progress. Other authors like Villalobos-Cid et al. (2016) also use cluster analysis to obtain efficiency values based on the heterogeneous performances of Chilean hospitals when the diagnosis-related groups (DRG) weights are not available. Their approach is also based on cross-sectional data however, and they do not account for a common frontier to assess the technological gaps in the healthcare system.

2.4 Methodology

In this paper, we propose a new empirical approach based on the analysis of clusters of units relative to a common frontier, similar to metafrontier analysis (Battese et al., 2004; O'Donnell et al., 2008) and we combine it with panel data DEA (Surroca et al., 2016; Pérez-López et al., 2018). The problem that arises when we try to apply the metafrontier analysis is that it is a cross-sectional approach, so for all time periods there will be a time-specific frontier and time-specific efficiency coefficients; therefore, each time period is analyzed without any connection with the levels of activity of adjacent time periods.

To overcome this problem, we use panel data DEA proposed by Surroca et al. (2016) and Pérez-López et al. (2018). The advantage of this method over other methods proposed in the literature (like the Malmquist index) is that it enables us to estimate a single time-invariant coefficient of efficiency for the period of analysis, considering the inherent panel data structure. Also, the methodology proposed by Pérez-López et al. (2018) allows us to break down these time-invariant efficiencies into time-variant efficiency scores, obtained on a year-by-year basis. In consequence, we will not just be able to find a long-term average efficiency for the time period studied, but we can also calculate efficiency values for each year under evaluation. We extend the approach by accounting for technological asymmetries of the DMUs. As far as we are aware, this methodology has not previously been applied and represents a significant innovation in the current literature.

In this study we define a non-parametric technology set. We start by obtaining homogeneous clusters of hospitals using multivariate clustering techniques. Once the groups are estimated, the common frontiers and the group frontiers can be calculated using DEA (Charnes et al., 1978; Banker et al., 1984).

As a non-parametric frontier estimation method, DEA has significant limitations that have been highlighted in the literature; the curse of dimensionality, their lack of statistical properties, and the potential impact of outliers are among the most relevant (Simar and

Wilson, 2008; Cooper et al., 2006). In this respect, Pérez-López et al. (2018) state that one of the outstanding advantages of the panel data DEA is the robustness of the results to the presence of outliers and temporal random shocks; this provides a specific efficiency score, representative of the complete time period under analysis. Hence, the interpretation of the results is not far from what can be obtained from a fixed-effects parametric regression.

2.4.1 Estimation of (time-invariant) panel data efficiency values for public hospitals

As hospital managers and policymakers usually have more control over their inputs (O'Neill et al., 2008; Cozad and Wichmann, 2013), and our approach consists of assessing whether the hospitals have been able to make efficient use of their resources, we apply an input-oriented efficiency measurement. Also, we assume a variable return to scale (VRS) model as we are dealing with heterogeneous observations.[3] The efficiency frontier is developed by optimizing the weighted input/output ratio of each DMU, subject to the condition that this ratio can be equal to, but never exceed one for any other DMU in the data set (Charnes et al., 1978).

Let us introduce some notation. Assume that we have I DMUs (hospitals) ($i = 1, 2, \ldots, I$) classifiable in S clusters ($s = 1, 2, \ldots, S$); here are M outputs $[y_1^i, \ldots, y_m^i, \ldots, y_M^i \in \Re_M^+]$ produced by N inputs $[x_1^i, \ldots, x_n^i, \ldots, x_N^i \in \Re_N^+]$ in the common frontier; and $[y_1^{i,s}, \ldots, y_m^{i,s}, \ldots, y_M^{i,s} \in \Re_M^+]$ and $[x_1^{i,s}, \ldots, x_n^{i,s}, \ldots, x_N^{i,s} \in \Re_N^+]$ outputs and inputs for the s local frontier respectively. We denote
$[y_1^o, \ldots, y_m^o, \ldots, y_M^o \in \Re_M^+]$ and $[x_1^o, \ldots, x_n^o, \ldots, x_N^o \in \Re_N^+]$ as the observed units under analysis, and likewise for the observed units in the local frontiers. We define a time variable t ($t = 1, 2, \ldots, T$), so in the common frontier we have $[y_{1,t}^i, \ldots, y_{m,t}^i, \ldots, y_{M,T}^i \in \Re_M^+]$ outputs and $[x_{1,t}^i, \ldots, x_{n,t}^i, \ldots, x_{N,T}^i \in \Re_N^+]$ inputs; and likewise in the local frontiers. We define the following mathematical program using *contemporaneous technology* (Tulkens, 1986; Pérez-López et al., 2018), which estimates the VRS DEA (common frontier) efficiency values:

$$\max_{u_{0,,t}^c, u_{m,t}^c, v_{n,t}^c} \alpha_t^c = u_{o,t}^c + \sum_{m=1}^{M} u_{m,t}^c y_{m,t}^o$$

$$s.t. \sum_{n=1}^{N} v_{n,t}^c x_{n,t}^o = 1$$

$$u_{o,t}^c + \sum_{m=1}^{M} u_{m,t}^c y_{m,t}^i - \sum_{n=1}^{N} v_{n,t}^c x_{n,t}^i \leq 0; \quad i = 1, 2, \ldots, I$$

[3] This is also tested in the empirical application with the Simar and Wilson (2011) returns-to-scale test.

$$u_{m,t}^{c} \geq 0;\ v_{n,t}^{c} \geq 0;\ \ m = 1, 2,\ \ldots, M; \quad n = 1, 2, \ldots, N \tag{2.1}$$

Where, $u_{m,t}^{c}$ and $v_{n,t}^{c}$ are weights for the outputs and inputs, for the period t, corresponding to the unit under evaluation; and $u_{o,t}^{c}$ is a scalar that can take positive or negative values, depending on the prevailing returns to scale.[4] The problem arises when for every observed unit we obtain a time-specific frontier and a time-specific efficiency coefficient, so for every *ith* DMU we are obtaining T contemporaneous efficiency scores $\left(\alpha_1^c, \ldots, \alpha_t^c, \ldots, \alpha_T^c\right)$: $T \times M$ output weights and $T \times N$ input weights. This implies that each time period is analyzed without any connection with the levels of activity of adjacent time periods. Also, we are conducting just an efficiency measurement for the common frontier, meaning that we are considering all hospitals in the analysis without allowing for their heterogeneity.

To overcome these issues, Surroca et al. (2016) and Pérez-López et al. (2018) propose a time-invariant panel data DEA evaluation. This technique incorporates an intertemporal frontier, which assumes a single production function for all time periods, comprising all the observations during the period of analysis. Also, it establishes a common set of weights for the complete time period. We extend this application with the incorporation of the S clusters generated to account for technological asymmetries. The input-oriented VRS (time-invariant) program for panel data DEA can be extended in the following way:

$$\max_{u_0^{ti,s}, u_m^{ti,s}, v_n^{ti,s}} \widetilde{\propto}^{ti,s} = u_o^{ti,s} + \sum_{m=1}^{M} u_m^{ti,s} \tilde{y}_m^{o,s}$$

$$s.t. \ \sum_{n=1}^{N} v_n^{ti,s} \tilde{x}_n^{o,s} = 1$$

$$u_o^{ti,s} + \sum_{m=1}^{M} u_m^{ti,s} y_{m,t}^{i,s} - \sum_{n=1}^{N} v_n^{ti,s} x_{n,t}^{i,s} \leq 0; \quad i = 1, 2,\ \ldots,\ I; \quad s = 1, 2, \ldots, S$$

$$u_m^{ti,s} \geq 0;\ v_n^{ti,s} \geq 0;\ \ m = 1, 2,\ \ldots, M; \quad n = 1, 2, \ldots, N \tag{2.2}$$

Note that $\widetilde{\propto}^{ti,s}$ is an average value that represents the one time-invariant efficiency coefficient for hospital under observation '*o*' while comparing it with its respective cluster s; $\tilde{y}_m^{o,s} = \sum_{t=1}^{T} y_{m,t}^{o,s}/T$ is the average value, corresponding to output m in hospital '*o*' forming part of cluster s, for the complete time period T; and $\tilde{x}_n^{o,s} = \sum_{t=1}^{T} x_{n,t}^{o,s}/T$ is the average value, corresponding to input n in hospital '*o*' forming part of cluster s, for the complete time period T. By applying the programs for the S clusters, we obtain $M \times I$ output weights and $N \times I$ input weights corresponding to the I hospitals classified in the S clusters. According to Pérez-López et al. (2018), besides obtaining a time-invariant common set of weights for

[4] The output and input weights u and v can be obtained by solving the "primal" (or multiplier) form of the DEA program. They provide extra information in that they can be interpreted as normalized shadow prices (Coelli et al., 2005).

each hospital, program (2.2) has three additional properties: (1) it is less dependent on the specific values of the variables in one particular year; (2) it ensures that no changes in the valuation system (input and output weights) take place across time periods; and (3) the consideration of average values does not imply any loss of information.

To compare the time-invariant cluster efficiencies, relative to the time-invariant common frontier efficiencies, we need to define the technology reference for the entire sample of units. This way, we obtain the following program:

$$\max_{u_0^{ti}, u_m^{ti}, v_n^{ti}} \widetilde{\propto}^{ti} = u_o^{ti} + \sum_{m=1}^{M} u_m^{ti} \tilde{y}_m^o$$

$$s.t. \sum_{n=1}^{N} v_n^{ti} \tilde{x}_n^o = 1$$

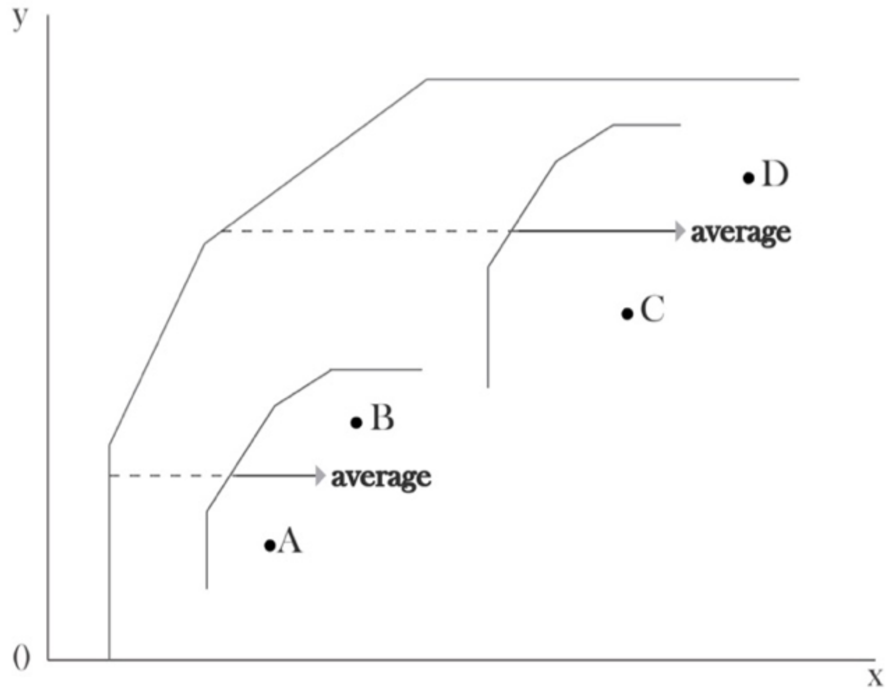
$$u_o^{ti} + \sum_{m=1}^{M} u_m^{ti} y_{m,t}^i - \sum_{n=1}^{N} v_n^{ti} x_{n,t}^i \leq 0; \quad i = 1, 2, \ldots, I$$

$$u_m^{ti} \geq 0; \; v_n^{ti} \geq 0; \; m = 1, 2, \ldots, M; \quad n = 1, 2, \ldots, N \tag{2.3}$$

Now we have $\widetilde{\propto}^{ti}$, which is an average value that represents the time-invariant efficiency coefficient for the hospital under observation;
$\tilde{y}_m^o = \sum_{t=1}^{T} y_{m,t}^o / T$ is the average value, referring to unit 'o' under observation, corresponding to output m, for the complete time period T; and $\tilde{x}_n^o = \sum_{t=1}^{T} x_{n,t}^o / T$ is the average value, corresponding to input n for unit 'o', for the complete time period.

For these efficiencies we are assessing the average level of efficiency of the complete time period with no isolated consideration of any specific time period in relation to the local and common frontier. The evaluation proposed here is depicted in Figure 2.4, where each average efficiency is evaluated relative to both the local frontier and the metafrontier.

Figure 2.4: Panel data time-invariant efficiencies, local frontier and metafrontier



Finally, the time-invariant TGR comes straightforwardly as:

$$TGR = \frac{\widetilde{\propto}^{ti}}{\widetilde{\propto}^{ti,s}} \tag{2.4}$$

The minimized value of $\widetilde{\propto}^{ti,s}$ that solves the cluster s linear program is no greater than the minimized value of $\widetilde{\propto}^{ti}$ that solves the metafrontier linear program, hence, the metafrontier will never lie below any of the group frontiers. This way, the TGR measures how close a group frontier is to the metafrontier, representing the restrictive nature of the production technology. The closer it gets to 1, the higher the efficiency in operations that can be achieved (Mitropoulos et al., 2015).

2.4.2 Estimation of (time-variant) panel data efficiency values for public hospitals

In their paper, Pérez-López et al. (2018) demonstrate that it is possible to derive time-variant efficiency scores from the previous time-invariant ones in order to obtain the variations in efficiency coefficients during the different time periods, maintaining the robustness of the values over time. If we consider one input example, under an input-oriented approach they demonstrate that:

$$\widetilde{\propto}^{ti} = \widetilde{\alpha}_1^{tv} \frac{x_{n,1}^{o}}{\sum_{t=1}^{T} x_{n,t}^{o}} + \cdots + \widetilde{\alpha}_t^{tv} \frac{x_{n,t}^{o}}{\sum_{t=1}^{T} x_{n,t}^{o}} + \cdots + \widetilde{\alpha}_T^{tv} \frac{x_{n,T}^{o}}{\sum_{t=1}^{T} x_{n,t}^{o}}$$

$$\widetilde{\propto}^{ti} = \sum_{t=1}^{T} \widetilde{\alpha}_t^{tv} w_t \tag{2.5}$$

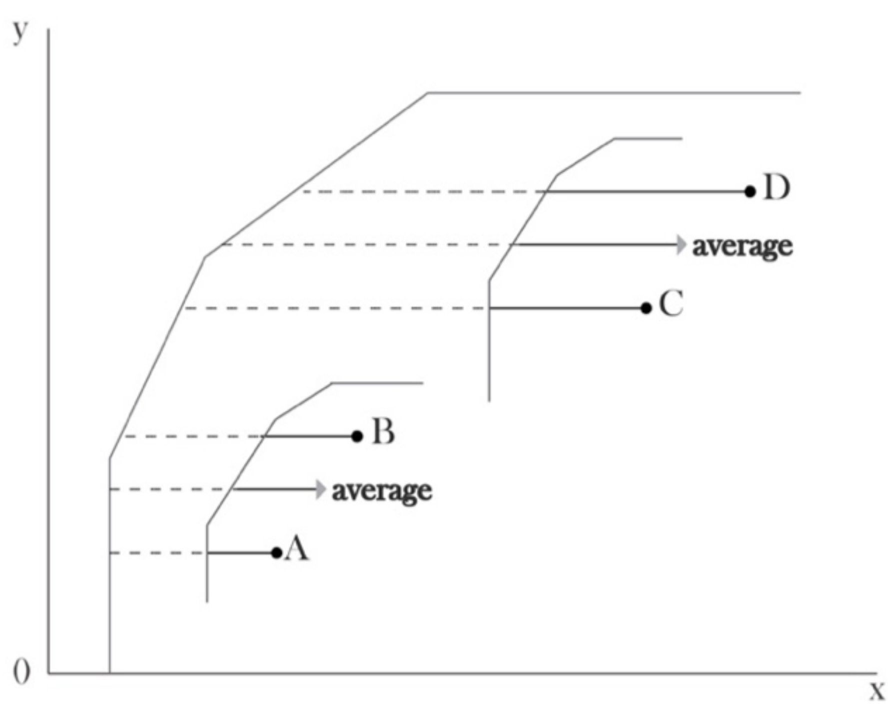
So that time-invariant panel data efficiencies are equal to the weighted average of the time-variant panel data efficiency coefficients. We can extend this same application to obtain time-variant panel data efficiency coefficients for every s cluster acquired. The mathematical representation is straightforward:

$$\widetilde{\propto}^{ti,s} = \widetilde{\alpha}_1^{tv,s} \frac{x_{n,1}^{o,s}}{\sum_{t=1}^{T} x_{n,t}^{o,s}} + \cdots + \widetilde{\alpha}_t^{tv,s} \frac{x_{n,t}^{o,s}}{\sum_{t=1}^{T} x_{n,t}^{o,s}} + \cdots + \widetilde{\alpha}_T^{tv,s} \frac{x_{n,T}^{o,s}}{\sum_{t=1}^{T} x_{n,t}^{o,s}}$$

$$\widetilde{\propto}^{ti,s} = \sum_{t=1}^{T} \widetilde{\alpha}_t^{tv,s} w_t^{s} \tag{2.6}$$

Figure 2.5 summarizes the evaluation of the time-variant and time-invariant efficiencies with respect to the common and local frontiers.

Figure 2.5: Panel data time-variant efficiencies, local frontier and metafrontier



2.4.3 Hypotheses

To answer the research question posed in this chapter, we need to look for a way to determine whether the healthcare efficiency performance of the public hospitals in Ecuador has undergone a significant change, which might be partly driven by the health reforms introduced under the new Correa government. In order to do so, we apply a before-after approach and divide the time period under study into two sub-periods. We estimate the time-variant and time-invariant efficiencies by applying the linear programs (2.2) and (2.3) to each sub-period. We consider 2008 as a potential turning point, when the new constitution was introduced and marked the beginning of several health reforms.

Thus, we define α_{p1}^{tv} as the time-variant efficiencies for the period 2006–2008, and α_{p2}^{tv} as the time-variant efficiencies for the period 2009–2014.[5] If the reforms that came after the new constitution affected the amount of inputs consumed in the health production process, for example, and if the new amount of patients attended caused an over-demand for healthcare services increasing the resources needed to treat them, then this would probably be reflected in a change in the average public hospital efficiency. Thus, if the health reforms negatively affected the efficient performance of public hospitals, then we should see a significant decrease in their average efficiency ($\widetilde{\alpha}$), so $\widetilde{\alpha}_{p1}^{tv} > \widetilde{\alpha}_{p2}^{tv}$. We will test this hypothesis by means of two statistical tests. The first one is the Wilcoxon signed rank test for dependent samples, which is a non-parametric test that does not need the assumption of normal distributions and has often been used in the literature to test significant differences in ordinal variables (O'Neill et al., 2008; Prior and Surroca, 2010). For the second test, we consider a method that provides us with more accurate information, namely the Li (1996) test for unknown distributions.

The Li (1996) method relies on kernel smoothing to non-parametrically estimate the density functions corresponding to α_{p1}^{tv} and α_{p2}^{tv} indices. However, Simar and Zelenyuk (2006) argue that in order to test the efficiency values estimated, the Li (1996) method has to be modified in several ways (see Simar and Zelenyuk, 2006). They provide consistent bootstrap estimates of the ρ values of the Li (1996) test and encourage its empirical application in efficiency measurement research.

Finally, the estimation of the time-variant efficiency values will help us to back up these hypotheses and find the trends in healthcare efficiency over the years.

2.5 Data

For the purpose of the study, we use the Annual Survey of Hospital Beds and Discharges and the Survey of Health Activities and Resources provided by INEC for the years 2006–2014. We consider the information on public hospitals excluding from the sample psychiatric,

[5] We apply the same procedure to the time-invariant efficiencies of each cluster obtained.

dermatology and geriatric hospitals.[6]

In a first stage, we use factor analysis with various correlated input variables available for all time periods in our dataset that can best approximate the health resources which contribute to the production of health. Based on the interdependencies of these variables we obtain a reduced set of uncorrelated variables called factors. With these new factors we run a hierarchical cluster analysis, which is a multivariate technique that seeks to cluster a set of I units into S groups depending on the similarities between them, so that (1) each unit is in one and only one of the groups; (2) every unit is classified, and (3) each group is internally homogeneous. The advantage of running a factor analysis previous to the clustering technique is that we can eliminate the dimensions which we are (practically) sure are only noise. Therefore, we retain the components responsible for a very high percentage of the inertia; thus, the hierarchy obtained is considered to be more stable and clearer (Husson et al., 2010).

In order to define the number of clusters to be constructed, we need to consider both the measure of similarity and the clustering method. We use the Euclidean distance as it is the most commonly used method in the literature to measure similarity, and the Ward hierarchical clustering method, which has the advantage of maximizing intra-group homogeneity and inter-group heterogeneity. Additionally, it is robust to outliers and groups are not too dissimilar in size (Balaguer-Coll et al., 2013). Finally, the Caliñski and Harabasz (1974) stopping rule was used in order to determine the number of clusters.

With this approach, we improve the standard applications used so far in the healthcare efficiency literature.[7] Thus, we use multivariate statistical analysis to generate specific clusters, differentiated by their technological endowment rather than size, and by applying factor analysis in the first stage, we make the variables independent of each other, avoiding potential correlation problems in the following analyses. The variables used to obtain the technology clusters are described and summarized in Tables 2.6 and 2.7 of the Appendix 2.8.

Given that factor and clustering analyses are cross-sectional techniques, we face the problem of an inconsistent grouping of DMUs for each year, which makes the efficiency values challenging to obtain for each group over the years. To overcome this problem, we take the average values of each variable over time and perform the multivariate analyses. This yields $\tilde{S}_1, \tilde{S}_2, \ldots, \tilde{S}_c$ groups shaped by the average technological endowment of each hospital. Despite some limitations that it could bring to the analysis, our goal is to obtain average efficiency estimations with the programs (2.2) and (2.3), making this approach the best fit to our empirical application.

The second stage of the analysis measures the average efficiency of hospitals over the years using programs (2.2) and (2.3) in both the group frontier and metafrontier, but

[6] Given some irregularities in the information for some hospitals and missing data for some years, we retrieve a non-balanced panel data.

[7] The literature has mostly used hospital size as a simple grouping criteria, proxied by variables like total beds or patients attended (see for example, Arocena and García-Prado, 2007; Mitropoulos et al., 2015)

before turning to the results, we must define the inputs and outputs to be used.

2.5.1 Inputs

There is common agreement on the use of inputs in the literature (O'Neill et al., 2008). To avoid potential problems of dimensionality, we aggregated the different health resources (described in Appendix 2.8.1) into four input variables: the total number of beds (totcam), hospital equipment and infrastructure (variables eq2 to eq8), the number of physicians (variables m1 to m7) and the number of professional healthcare personnel other than physicians that work in the hospital (proftit to p2).

The number of hospital beds has been widely used in the literature as a proxy for hospital size and capital investment (O'Neill et al., 2008); we also include variables that describe equipment and infrastructure of hospitals to account for this. The majority of studies include the number of clinical staff as a proxy for labor costs (O'Neill et al., 2008; Cantor and Poh, 2018).

2.5.2 Outputs

Most published research uses some variant of intermediate outputs in terms of patients treated or number of inpatient days hospitalized (Hollingsworth, 2008). To measure the final production of the health of public hospitals we use the number of discharges as an output variable. However, we need a method to adjust outputs for patient heterogeneity (i.e. case mix) as not all conditions can be treated with the same amount of resources and not all hospitals have the means nor the capacity to treat serious illnesses; therefore, if not taken into account, hospitals with a more complex case mix are likely to receive lower efficiency scores. By including a case-mix weight, we are explicitly designing groups that provide comparable resource intensity care, and we can also distinguish the hospitals treating more severely ill patients, requiring more inputs from hospitals treating less resource-intensive patients (Valdmanis et al., 2017). Perhaps the most successful approach is the use of DRG classification which categorizes patients according to diagnosis, treatment and length of stay. However, in developing countries this tool is not fully (or even partially) implemented, which limits the efficiency of the evaluation (Villalobos-Cid et al., 2016). This constraint holds true for the Ecuadorian case, which leads us to apply alternative approaches based on the available data, and which have been previously applied in the literature.

Therefore, to treat the severity of cases in this study, we use the three-digit *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) to construct the case-mix weight, following the approach developed by Herr (2008).[8] This approach

[8]Refer to Appendix 2.8.2 for a description on Herr's (2008) case-mix index.

Table 2.2: Inputs and Outputs, descriptive statistics

	2006						2014					
	Cluster 1 (High)		Cluster 2 (Intermed.)		Cluster 3 (Low)		Cluster 1 (High)		Cluster 2 (Intermed.)		Cluster 3 (Low)	
	N=8		N=14		N=143		N=6		N=15		N=148	
	Median	SD	Median	SD	Median	SD	Median	SD	Median	SD	Median	SD
Output												
Weighted Discharges	4505	(2623.6)	782	(3053.9)	814	(2076.5)	6344	(5372.8)	977	(3943)	984	(2547.1)
Inputs												
Physicians	134	(76.2)	17	(82.7)	16	(34)	275	(198.5)	27	(196.4)	29	(80.3)
Beds	263	(118.8)	33	(165.8)	21	(101.6)	277	(232.9)	32	(213.5)	29	(100.4)
hospital personnel	319	(210)	22	(188.9)	23	(108.7)	661	(344.4)	51	(365.5)	51	(169.5)
Equipment and Infrastructure	152	(56.5)	40	(65.6)	32	(57.8)	358	(188.2)	52	(69.9)	48	(104.4)

Note: MANOVA tests indicated that the differences between groups are statistically significant. The observations for the clusters over the years differ slightly due to some missing information in the dataset. However, the methodology applied is robust to these discrepancies.
Source: The authors

relies on the assumption of a correlation between the length of stay and the severity of illness, so the idea is that the more days of patient stay, the more severe the disease and the more resources are used.[9] Other authors such as Herr et al. (2011), Herwartz and Strumann (2012, 2014) and, Varabyova and Schreyögg (2013) suggest using this approach in the absence of the DRG classification.

The descriptive statistics for the years 2006 and 2014 for each cluster are presented in Table 2.2. We ran two outlier detection methods on the data. The first was proposed by Prior and Surroca (2010), and the second one is based on Andrews and Pregibon (1978) and Wilson (1993).[10]

Comparing the levels of input mix across the clusters presented in Table 2.2, we can observe that the first cluster accounts for high levels of technological endowment. As expected, the hospitals belonging to this cluster attend to a much broader share of patients in the country, even though their number is remarkably lower than the other clusters. The second cluster is shaped by hospitals with an intermediate level of technology, and on average, is not very far from the final cluster. Finally, the last cluster comprises hospitals with a low level of technological endowment. It is important to note the marked difference

[9]The potential problem that could arise with the inclusion of this variable is that if hospitals are reimbursed according to the number of patient stays, we could expect hospital managers to behave opportunistically in order to increase their hospitals' revenues.

[10]Although the methodology used in this study to obtain the efficiency values has proved to be robust to outliers, we still run some outlier detection procedures to avoid potential bias in the estimations. We identified nine hospitals that were behaving as outliers, on average.

in the number of hospitals in this cluster, relative to those in the high and intermediate technology clusters; this difference highlights the profound technological heterogeneity, not just in terms of the large asymmetries present in the public healthcare system, but also in its notable share of technologically lagging hospitals.

2.6 Results and discussion

Table 2.3 shows the time-invariant efficiencies resulting from programs (2.2) and (2.3), both for the metafrontier and for the respective local frontiers. Looking at the metafrontier average efficiency, we can see that overall the hospitals present very low efficiency scores. The value of 0.3894 would mean that to be fully efficient Ecuadorian hospitals have to reduce their input consumption by 61.06%, representing more than half of resource consumption. Additionally, some hospitals present a minimum level of inefficiency as low as 0.1088, which represents a severe problem of inefficiency in the system. However, we cannot draw hasty conclusions in this manner, as other hospitals present high levels of efficiency, showing profound asymmetries inherent in the system.

Table 2.3: Time-invariant efficiencies, summary statistics

	Mean	Median	SD	Min	Max	TGR
Metafrontier	0.3894	0.3838	0.1483	0.1088	0.9318	
Cluster 1 (High)	0.6479	0.6371	0.1926	0.4135	0.9095	0.5433
Cluster 2 (Intermediate)	0.5623	0.5781	0.2434	0.2269	0.9339	0.7183
Cluster 3 (Low)	0.4269	0.4282	0.1502	0.113	1	0.9176

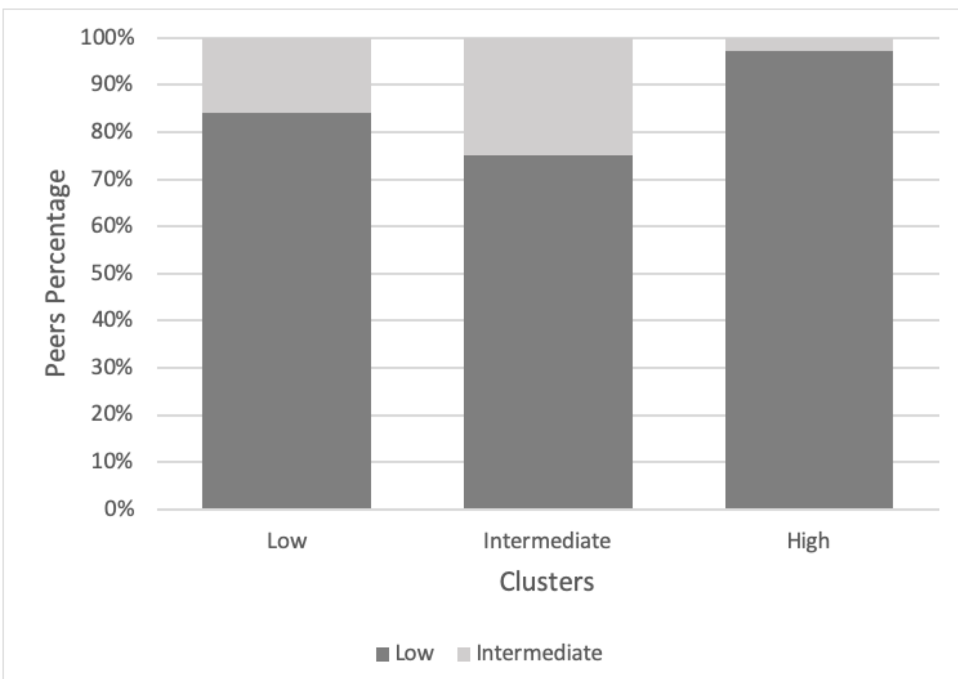
Source: The authors

When we look at the efficiencies obtained for each cluster, the results are quite different. Overall, the average (group) efficiencies in cluster 1 and cluster 2 are much closer to the frontier than those of cluster 3. Hence, when considering the technological differences between hospitals, on average, high-technology ones are making better use of inputs than low-technology ones. The differences compared with a single frontier estimation are remarkable, highlighting the importance of accounting for the heterogeneity of the system. Assessing public healthcare systems within a homogeneous framework of comparable hospitals is therefore questioned, especially Ecuadorian public hospitals whose differences have been worsened by its historical economic and political situation.

Regarding the TGR, results seem to be counterintuitive. Cluster 1 and cluster 2 present the widest gap between group efficiencies and the metafrontier, suggesting that they are more constrained by the nature of their production environment, and efficiencies in these clusters are further away from the metafrontier than those of cluster 3, whose efficiencies are very similar to the metafrontier. It would appear that the output complexity is not properly captured by our data, which is leading the metafrontier to fail in the measurement of the unconstrained production defined by Battese et al. (2004).

To investigate this question, we plotted the hospitals' peer participation for each cluster with respect to the metafrontier in Figure 2.6.[11] The figure shows that the hospitals belonging to each group take as reference hospitals for production (in the metafrontier) those that belong to the low and intermediate technology hospitals. For example, 97% of the reference units for the inefficient high-tech hospitals belong to the low-tech cluster, while the remaining 3% belong to the intermediate technology cluster. These results show that, even though we tried to capture the complexity of the cases through length of stay, we are still not able to find a proper case mix that accounts for the full complexity of the patients treated. In addition to this, the profound heterogeneity in the system is also playing a significant role in these results. This is because the few high-tech hospitals treat more than five times as many patients as the low-tech hospitals, leading to saturation of their resources, and as a result, jeopardizing their performance.

Figure 2.6: Percentage of peer participation relative to the metafrontier



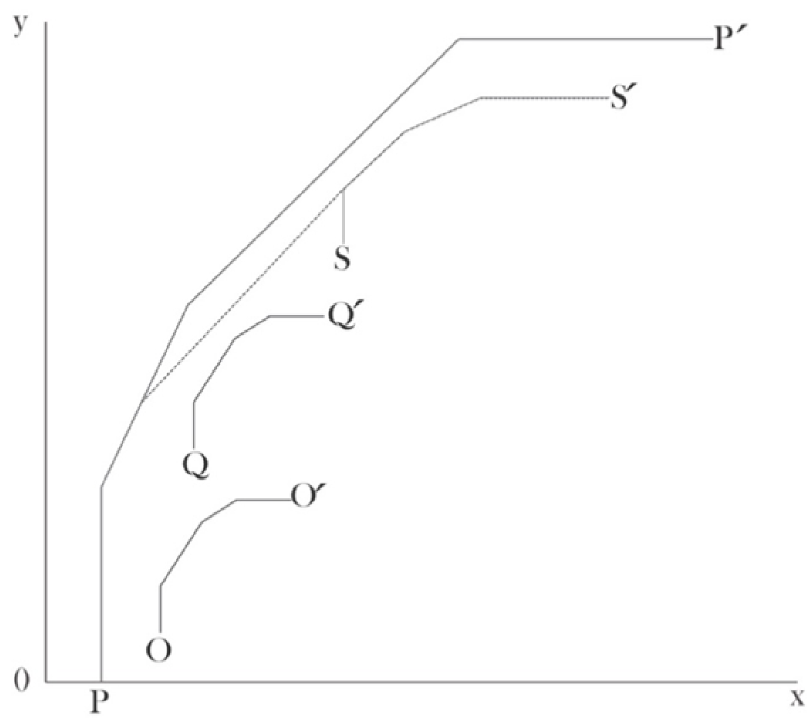
Thus, in a situation of limited data availability, and where the marked heterogeneity in the system is preventing a proper metafrontier evaluation, we need a method that considers the technological differences of the system to avoid misleading conclusions. In this context, it is reasonable to assume that the hospitals that have the higher technology can only be compared with each other, meaning that the only reference units they will be

[11] The peer observations are those efficient DMUs with which the inefficient units are directly compared, in order to be fully efficient; that is, they are those reference units that define the efficient production for every inefficient hospital (Coelli et al., 2005; El-Mahgary and Lahdelma, 1995)

compared with are those that have the same level of technology. It would be unreasonable to think that low-tech hospitals that do not have the same resources can be a reference for those in the high-tech group. Similarly, the intermediate technology hospitals would be compared with each other and with the high-tech hospitals but cannot be compared with the low-tech hospitals. This idea leads us to construct a new metafrontier, taking into account what Banker and Morey (1996) call a shift in the production frontier. Banker and Morey (1996) state that different hospitals have different characteristics that need to be considered in the efficiency analysis; unfortunately, these characteristics cannot always be observed in practice. This is especially important when the impact of a factor (like technology endowment in our case) varies substantially across demographic, competitive or other contingent environments.

Based on the idea of Banker and Morey (1986, 1996), in this research we propose a method to construct a new metafrontier. We will assume that the hospitals studied here cannot be compared with hospitals endowed with less technology. The shift of the production frontier leads us to define a new metafrontier. The intuition underlying our approach is displayed in Figure 2.7, where the metafrontier PP' depicts the prior metafrontier where all hospitals are benchmarked against each other, meaning that we take all clusters into consideration ($s = 1, 2, \ldots, S$) The shift is produced when we constrain the metafrontier to be benchmarked only against those hospitals with similar or higher technology. Here, we imply that the high-tech hospitals, represented in the frontier SS', will only be compared with each other, so they do not present a TGR. The new metafrontier is depicted by PS'.

Figure 2.7: Newly constrained metafrontier



This idea also follows the line of selective convexity developed by Podinovski (2005). This concept states that the DMUs can be used to form convex combinations provided that they are different only in the inputs and outputs for which the convexity assumption is accepted. Additionally, this author demonstrates that under the free disposability assumption, selective convexity generalizes the inequalities stated by Banker and Morey (1986).

Using this approach, we solve the new metafrontier applying the linear program (2.3) under three scenarios where the three hospital groups analyzed are benchmarked according to their level of technology: $s = 1; s = 1,\ 2$; and $s = 1, 2, 3$. The sum of these three estimations together result in the new metafrontier depicted by PS'.

The results obtained with the new metafrontier are presented in Table 2.4. This new estimation of a metafrontier allows us to determine the technological gap of the three clusters of hospitals, considering the asymmetries in the system not just in the local frontiers, but also in the common frontier. As we can see, now cluster 3 accounts for the highest technological gap relative to cluster 2, which shows a shorter distance from the metafrontier. The difference in TGR for the low-tech hospitals shows that, given their limited levels of technology, they can achieve a maximum efficiency of 90% of what is feasible with the highest level of technology available. We have to be aware that overall, the level of efficiency in the system is rather low, which can be explaining the short distances in the TGR.

Table 2.4: Time-invariant efficiencies, summary statistics (new constrained metafrontier)

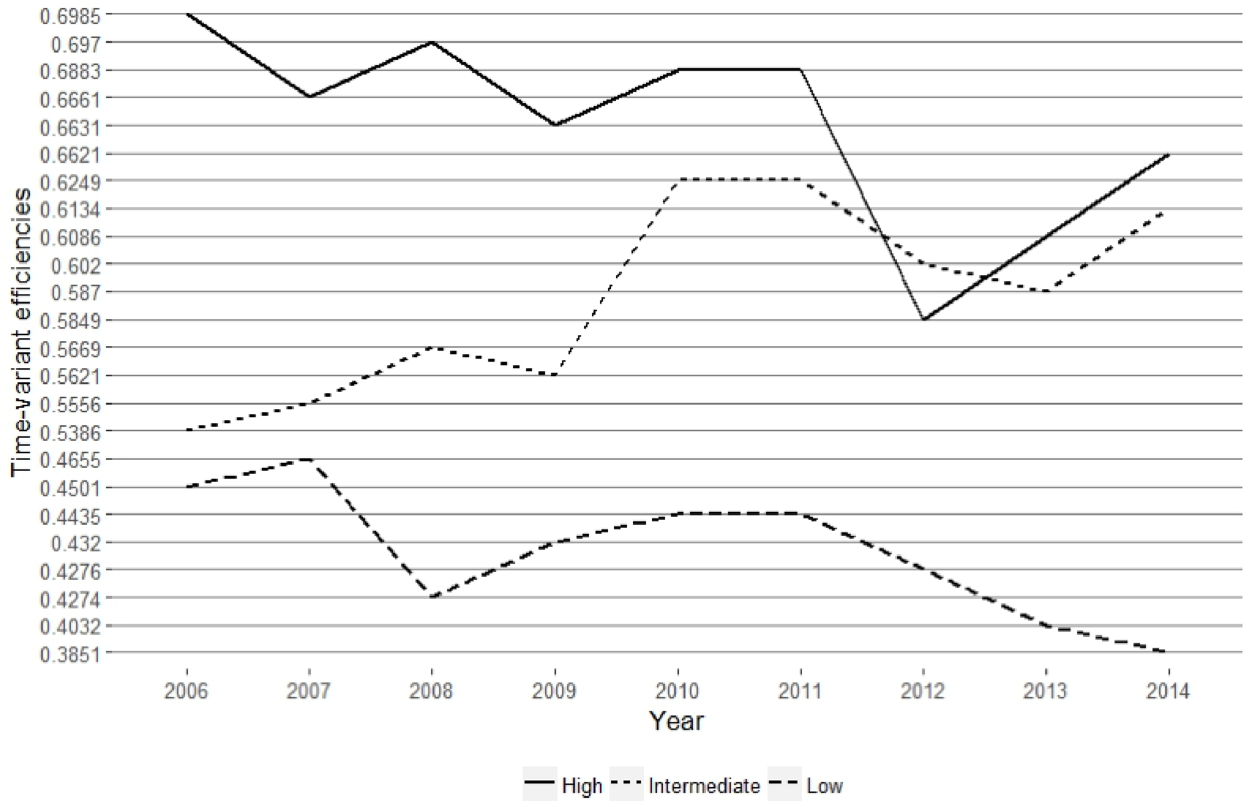
	Mean	Median	SD	Min	Max	TGR
Metafrontier	0.4215	0.3962	0.1731	0.113	0.9339	
Cluster 1 (High)	0.6479	0.6371	0.1926	0.4135	0.9095	1
Cluster 2 (Intermediate)	0.5623	0.5781	0.2434	0.2269	0.9339	0.9994
Cluster 3 (Low)	0.4269	0.4282	0.1502	0.113	1	0.9176

Source: The authors

The advantage of applying the panel data DEA technique in our analysis is that it provides an additional tool with the time-variant efficiencies, allowing us to obtain the trends in efficiency of the hospitals analyzed without losing the robustness of the previous results. This enables us to shed some light on the exact year when the efficiencies started to decrease and have a clearer idea of whether this could have been a direct result of the healthcare policies implemented.

Following this idea, Figure 2.8 shows the time-variant efficiencies obtained for the group frontiers in the period under analysis. Some interesting facts can be garnered from this figure. First, the efficiencies of the high-technology hospitals show peaks of performance in the first period of analysis, but their behavior does not seem to change immediately after 2008. This could be because historically there has always been a limited

Figure 2.8: Evolution of time-variant panel data efficiencies



number of high-technology public hospitals in Ecuador and they attend to most of the patients in the Ecuadorian health network, which has not changed over the years.

In contrast, low and intermediate technology hospitals show an increase in 2008 and 2009. This improvement might be reflecting the positive effect on efficiency due to the optimization of spare resources and capacity, likely misused prior to the reforms. The investment deployed in the health sector could also be a potential driver of this rise in efficiency. In the short run, the increase in the health budget could have triggered higher productivity in the system; for example, physicians, managers or general health personnel may have been motivated by potential salary raises.

There seem to be two particular years when the behavior of the three clusters changes. The first one goes from 2010 to 2011, when all groups shifted from an increasing to a constant efficiency. The second one relates to the year 2011, a year in which efficiency declined severely. Two facts are worth noting in these two periods. In 2010 there was a social security reform that allowed the insured population to extend insurance to their children under the age of 18, which might have halted the increase in performance of all clusters of hospitals due to the sudden rise in patients. Moreover, in 2011 Ecuador held a referendum, which (among other matters) included the approval of a law for the deprivation of

Table 2.5: Time-invariant efficiencies for each sub-period, summary statistics

		Mean	Median	SD	Min	Max
2006-2008	Cluster 1 (High)	0.7694	0.8942	0.2190	0.4415	1
	Cluster 2 (Intermed.)	0.6991	0.7900	0.2701	0.2353	1
	Cluster 3 (Low)	0.5226	0.5069	0.1878	0.1519	1
2009-2014	Cluster 1 (High)	0.6985	0.7190	0.2178	0.3921	0.9656
	Cluster 2 (Intermed.)	0.5816	0.6003	0.2522	0.2340	0.9664
	Cluster 3 (Low)	0.5144	0.5299	0.1650	0.1178	1

Source: The authors

liberty for employers who do not affiliate workers within a maximum period of 30 days (Orellana et al., 2017). This new law, added to the new free services, caused an increase in demand that De Paepe et al. (2012) refer to as a "demand crisis" due to the sudden rise of the insured population, especially in larger cities. This increased demand might be a substantial cause of the pronounced decline seen in all three groups of hospitals.

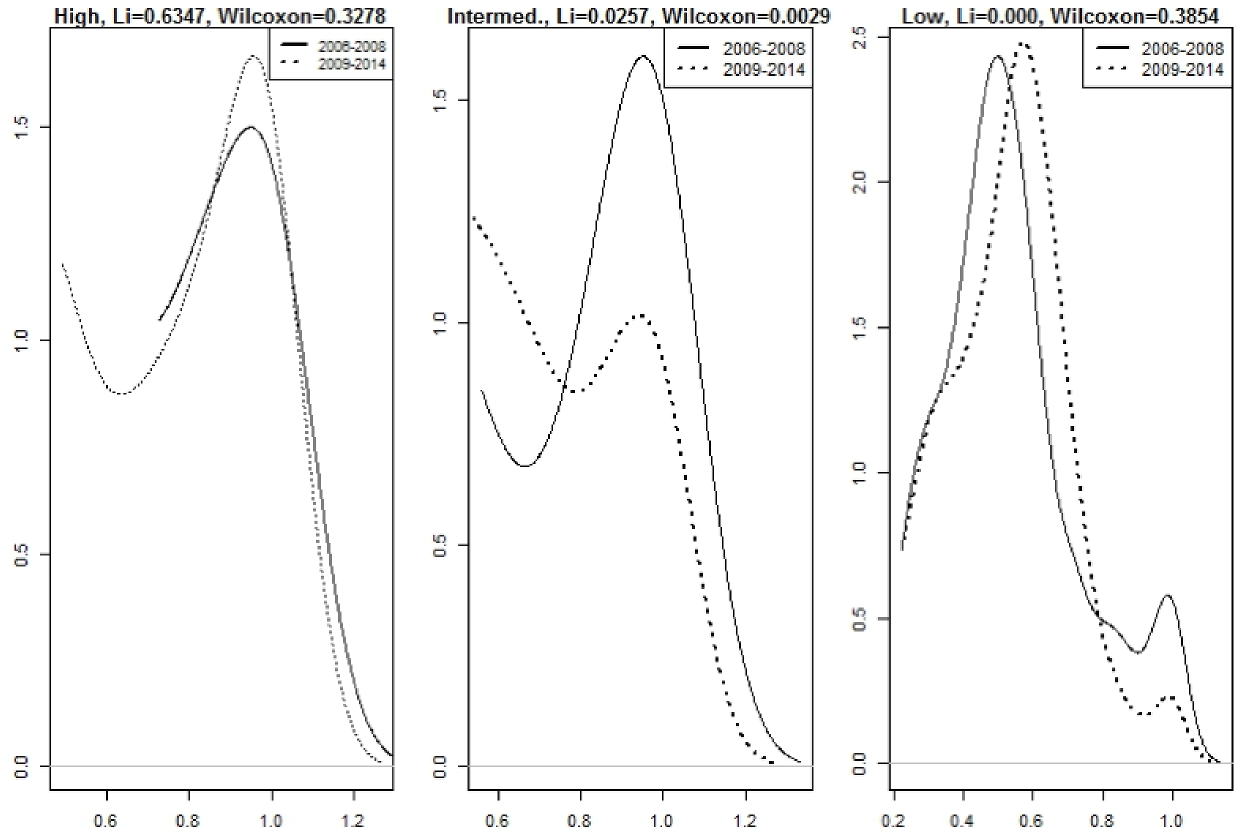
It should be noted that we are not claiming that this decline in efficiency was actually caused by the increase in demand, but in this study we offer strong empirical evidence to suggest that this could be a strong driver.

2.6.1 Hypotheses tests

In order to corroborate the hypothesis of a significative change in the average efficiency performance after 2008, Table 2.5 shows the descriptive statistics of the time-invariant efficiency values for the sub-periods p_1 (2006 – 2008) and p_2 (2009 – 2014). The average values for both the group frontiers and the metafrontiers seem to have decreased. This decline could be signaling that the government's policies indeed had a negative effect on the efficiencies of all groups of hospitals in the public health system.

We provide evidence of this possibility in Figure 2.9, where we plot the smoothed densities of the group efficiencies. The Li (1996) and Simar and Zelenyuk (2006) test, and the Wilcoxon signed rank test p-values for the efficiency scores of the group frontiers in both sub-periods are depicted along with the graph. Based on the information provided by the extended Li (1996) and Simar and Zelenyuk (2006) test––which has proved to provide more accurate and reliable results in several fields applied to efficiency measurement (Pastor and Tortosa-Ausina, 2008; Li et al., 2009; Balaguer-Coll et al., 2013) ––the null hypothesis of equal distributions is rejected for all groups except the high-tech hospitals. The evidence presented here falls in line with the above-mentioned results. Apparently, the high-tech hospitals have not experienced a significant decrease in efficiency since 2008, but rather the low and intermediate technology hospitals are the most affected. Although we cannot say that this decrease in efficiency happened solely due

Figure 2.9: Density plots 2006–2008 vs 2009–2014



Note: The Li and Wilcoxon scores correspond to Simar and Zelenyuk's extension of the Li test and the Wilcoxon signed rank test p-values, respectively.

to the health reforms implemented since the new constitution, we now have substantial evidence that they could have played a significant role in this decline.

The evidence provides an initial picture of the performance of the public healthcare system in Ecuador. Despite the (overall) low levels of efficiency, there seem to be some potential factors that are causing the decline in performance, and it appears to be affecting mainly the less technological hospitals. The literature on healthcare efficiency measurement offers some explanations in this matter. For example, the non-significant change in efficiency of high-tech hospitals might reflect their capacity to treat complex cases in a more efficient manner; the concentration of specialized physicians and equipment in these hospitals might be allowing them to cope better with the increasing volume of patients than low and intermediate technology hospitals would be able to, suggesting the existence of a process of learning-by-doing in high-tech hospitals (Gobillon and Milcent, 2013).

Cream skimming could also be playing an important role in this difference. The referrals to private institutions might not be alleviating the consumption of inputs in public hospitals, given that complex cases remain in the public sector and tend to stay for a longer time (Cheng et al., 2015) and demand more health services than necessary if they are covered by public insurance (Orellana et al., 2017). Given that high-tech hospitals could be showing a process of learning-by-doing, the low-tech hospitals could be more affected.

These results suggest some recommendations for public authorities and policymakers. The drop in efficiency coincides with two of the most far-reaching reforms in social security and promotion of universal coverage. The authorities should keep in mind that public hospitals need to have the necessary means in the short term with which to adapt to a sudden increase in the insured population, and that the effect on their efficiency can depend on the type of hospital where the resources are allocated. Our findings provide an initial motivation to look deeper into this matter and formulate focused policies that encourage better allocation of resources to hospitals that might be suffering most from these negative effects. The study also reveals a positive effect on efficiency in 2008 and 2009. Academics and authorities should further explore this effect and identify the sources of this improvement, which can also bring strong policy recommendations to enhance the healthcare system. Although the potential causes of efficiency variation do not fall within the scope of this study, we offer readers a wide range of unexplored research ideas in this field, and strongly encourage further investigation.

2.7 Conclusions

The present study aimed to analyze whether the public health reforms introduced in Ecuador since 2008 have had a significant effect on the efficiency of its hospitals. To take into account the technological differences of Ecuadorian hospitals, we use a two-stage analysis, wherein the first stage we apply a multivariate factor analysis and clustering techniques to obtain homogeneous groups characterized by their technological endowment. In the second stage, we propose a combined metafrontier panel data DEA method that yields robust efficiency scores, representative of the complete time period.

The results show considerable inefficiency in the whole period when we contemplate all hospitals in a common frontier. However, when they are disentangled into technologically different groups, the difference is remarkable. Compared to their respective local frontiers, high and intermediate technology hospitals seem to be performing rather better than low-tech hospitals, which present an average efficiency very similar to the metafrontier. These results highlight the importance of considering the heterogeneities inherent in the system; if not taken into account, these heterogeneities can bias the results and lead to misleading conclusions.

The TGR for the respective groups seems to be counterintuitive as low-technology hospitals show a shorter distance from the metafrontier. However, conventional methods in the literature applied to developed economies cannot be simply translated to developing countries, which have different economic and social structures. The lack of good quality data in developing countries such as Ecuador, and the deep heterogeneity of their systems, complicates the application of conventional models such as the metafrontier production function. In this study, we propose an approach to re-define the metafrontier function by means of concepts such as frontier shift and selective convexity, introduced by Banker and Morey (1986, 1996) and Podinovski (2005).

Our approach assumes that in such a technologically heterogeneous context, we cannot compare groups of hospitals, as those with lower technology will not be able to perform at the same level as those with much higher technology. Hence, high-tech hospitals can simply be benchmarked against each other. This way, we find that given these constraints, the lowest technology hospitals can perform with a maximum efficiency of 90% of what they would be able if they had the maximum technology available.

The empirical results of the efficiencies run before and after the new constitution in 2008 show an evident decline in the average efficiency of the public hospitals. Moreover, we find that 2008 had no significant effect on the trend of high-technology hospitals, whereas a statistically significant decrease in efficiency is found in low and intermediate technology hospitals. A short-run effect, shown as an increase in efficiency, is also observed among low and intermediate technology hospitals. This improvement may be due to the public investment made in Ecuador since the beginning of Rafael Correa's mandate, which might have had an immediate effect in the system. With the immediate increase in their budget, hospital managers or medical personnel could have been motivated to increase their productivity. Additionally, the slight increase in demand prior to the most far-reaching health reforms in social insurance could have allowed some hospitals to make better use of spare capacity and medical resources, which may have been inefficiently utilized. Nonetheless, this effect was interrupted in 2010 and further reversed in 2012, coinciding with the Ecuadorian health reform that guaranteed social insurance for all workers in a dependency relationship with their employer. The evidence suggests that the sudden influx of patients generated by this reform could have had a direct effect on the observed drop in efficiency. These hypotheses are not firm conclusions, but they open up new research questions and encourage future inquiry in this field.

This study can be considered as a first step to further research to more deeply explore the potential determinants of the efficiency behavior in Ecuador's healthcare system. This strand of research can be of significant relevance to implement focused healthcare policy better able to allocate resources in the system and alleviate the saturation that might be occurring in a limited number of hospitals that receive more than half the demand for medical services in the country.

The methodology implemented can also serve as a reference to apply to other heterogeneous realities such as that of Ecuador, where good quality data may not be available to implement classical efficiency measurement approaches. In this regard, further work can be conducted in Latin American countries. The literature has found that high territorial heterogeneity in developing countries, particularly in Latin America, shapes economic and social inequalities that have characterized the region over the years (Cuadrado-Roura and Aroca, 2013) and many of them lack good healthcare data (Villalobos-Cid et al., 2016).

Further methodological innovations can also be implemented. For example, we can consider a similar approximation by adapting a stochastic frontier analysis (SFA) to our dataset. However, the main setback of SFA approaches is that they rely on a production function that has to be defined a priori (O'Neill et al., 2008) and that cannot be simply proposed in the context of a developing country. Technical efficiency measures are very sensitive to the choice of functional specification (Giannakas et al., 2003), which can be misleading if not correctly specified. Future work should focus on defining the theoretical framework of a proper production function to provide the background for empirical applications.

Finally, some limitations of this work should be noted. First, the limited quality and availability of data has constrained the sample to the years addressed here and necessitated alternative data treatment approaches. The need to take into account a wider time period is highlighted, which would provide useful information on how the country has been adapting to these relatively new reforms over the years. Second, the main findings of this research apply to this context of analysis. The findings in terms of efficiency cannot be extrapolated to other health reforms in other countries where such heterogeneity does not exist.

2.8 Appendix

2.8.1 Variables description

Table 2.6: Variable description

Variable	Description
Totcam	Total number of hospital beds
m1	Total number of general physicians
m2	Total number of surgeons
m3	Total number of plastic surgeons
m4	Total number of specialized physicians
m5	Total number of resident physicians
m6	Total number of rural physicians
m7	Total number of other physicians
proftit	Health personnel
p1	graduates and technologists
p2	Nursery auxiliary
p3	Administrative personnel
eq1	Stomatology equipment
p4	Stomatology personnel
eq2	Imaging equipment
eq3	Diagnostic equipment
eq4	Treatment equipment
eq5	Physical infrastructure for surgery, obstetrics and intensive care
eq6	Equipment for surgery, obstetrics and intensive care
eq7	Sterilization equipment
eq8	Other equipment

Source: The authors

Table 2.7: Technological endowment variables, means and SD

	2006	2007	2008	2009	2010	2011	2012	2013	2014
totcam	74.93	74.03	80.88	75.97	82.12	83.10	86.44	87.65	85.04
	(112.93)	(110.68)	(121.41)	(112.10)	(126.47)	(120.47)	(116.18)	(119.57)	(126.61)
m1	3.20	3.18	4.03	4.79	5.40	5.35	7.70	8.39	10.97
	(4.26)	(2.94)	(6.79)	(8.02)	(8.70)	(7.47)	(16.68)	(16.38)	(24.02)
m2	2.43	2.69	2.70	2.76	2.94	3.08	3.21	3.16	3.07
	(2.72)	(3.14)	(2.97)	(2.94)	(3.45)	(3.88)	(4.26)	(4.26)	(4.86)
m3	0.40	0.36	0.39	0.37	0.50	0.40	0.51	0.46	0.39
	(1.00)	(0.93)	(0.97)	(1.00)	(1.32)	(1.18)	(1.43)	(1.13)	(0.97)
m4	18.39	19.13	24.11	24.92	28.08	27.02	29.50	30.96	32.74
	(27.60)	(27.90)	(39.49)	(41.39)	(46.53)	(44.75)	(47.90)	(52.57)	(59.12)
m5	10.14	9.91	9.10	9.73	11.65	14.15	17.24	16.88	19.47
	(19.72)	(18.23)	(15.75)	(14.48)	(17.26)	(22.75)	(28.38)	(25.54)	(32.67)
m6	1.96	1.80	2.05	2.68	2.65	3.01	2.68	3.82	1.05
	(2.53)	(2.39)	(2.39)	(3.81)	(3.42)	(5.19)	(4.83)	(17.10)	(5.33)
m7	1.16	1.41	1.36	1.20	1.64	1.59	1.29	1.62	2.81
	(4.35)	(4.09)	(7.38)	(3.75)	(7.56)	(6.95)	(6.36)	(5.51)	(10.94)
proftit	31.32	32.94	36.42	39.16	48.15	51.12	60.26	63.63	65.53
	(54.41)	(55.59)	(53.99)	(53.66)	(76.70)	(83.66)	(92.52)	(92.00)	(100.16)
p1	3.49	3.43	11.42	12.09	14.95	14.91	18.65	19.63	18.34
	(8.72)	(8.41)	(19.48)	(19.94)	(24.65)	(25.94)	(30.65)	(30.39)	(31.69)
p2	46.14	43.85	47.74	45.78	53.07	50.98	55.20	58.63	56.29
	(81.60)	(69.78)	(81.85)	(81.02)	(91.79)	(89.42)	(97.93)	(101.01)	(98.75)
p3	20.51	20.71	22.88	22.53	25.70	28.85	37.02	35.22	35.29
	(27.22)	(28.38)	(28.87)	(27.08)	(29.98)	(38.43)	(48.45)	(44.86)	(51.46)
p4	3.90	3.86	4.02	3.88	4.22	4.45	4.34	4.50	3.20
	(3.86)	(3.69)	(4.52)	(2.87)	(3.38)	(3.97)	(4.24)	(4.93)	(4.71)
eq1	16.35	17.26	17.63	18.47	19.26	20.33	20.55	46.07	38.43
	(11.00)	(15.52)	(14.48)	(14.96)	(17.01)	(15.90)	(16.47)	(58.31)	(58.46)
eq2	4.60	3.52	3.78	3.83	4.14	4.44	4.62	4.95	5.13
	(16.06)	(3.23)	(3.41)	(3.30)	(4.10)	(4.46)	(4.39)	(5.28)	(5.30)
eq3	4.65	3.53	3.26	3.18	3.91	4.31	4.99	5.19	6.34
	(16.69)	(4.66)	(5.02)	(5.21)	(6.66)	(6.74)	(7.95)	(8.42)	(9.90)
eq4	4.98	6.15	6.71	6.42	7.14	7.54	7.86	8.37	8.21
	(12.49)	(16.95)	(21.29)	(15.76)	(18.41)	(18.59)	(17.40)	(17.06)	(17.63)
eq5	3.63	3.73	4.82	4.72	5.00	5.53	4.95	5.05	4.75
	(3.48)	(3.32)	(5.17)	(4.34)	(4.95)	(8.21)	(6.10)	(5.33)	(4.30)
eq6	29.24	28.29	38.84	44.81	48.09	50.05	55.79	61.10	61.61
	(39.15)	(28.85)	(50.32)	(63.80)	(71.12)	(69.36)	(80.75)	(89.87)	(87.63)
eq7	3.90	3.87	4.08	3.99	4.46	4.44	4.37	4.35	4.25
	(3.21)	(3.06)	(2.93)	(2.90)	(4.93)	(5.11)	(3.21)	(3.25)	(3.42)
eq8	6.13	6.24	4.28	3.97	4.35	4.62	4.83	5.17	4.50
	(7.33)	(7.27)	(3.77)	(2.85)	(2.90)	(3.67)	(3.93)	(3.87)	(2.79)

Note: Standard deviations in parentheses

Source: The authors, based on the Annual Survey of Hospital Beds and Discharges and the Survey of Health Activities and Resources 2006–2014.

2.8.2 Case-mix weights

To control for the severity of cases in this study, we construct the case-mix weight following the approach developed by Herr (2008). These weights are based on the across-hospital average length of stay (LOS) of each diagnosis relative to the overall length of stay. In developing a list of diagnostic categories (cases), we use the three-digit *International Statistical Classification of Diseases and Related Health Problems* (ICD-10).

The weights are then constructed as follows. A mean of LOS by year and main diagnosis $m = 1, \ldots, M$ over N hospitals is calculated using the following formula:

$$LOS_m = \frac{1}{N} \sum_{i=1}^{N} \frac{days_{mi}}{cases_{mi}} \tag{2.7}$$

Where *cases* represent the severity of the illness. The mean LOS over all diagnoses and all hospitals is then denoted by LOS_G and the final weights π_m are obtained by:

$$\pi_m = \frac{LOS_m}{LOS_G} \tag{2.8}$$

The weights π_m will be bigger (smaller) than one if the treatment of diagnosis m takes more (less) time than the overall average LOS. These weights rely on the assumption of a correlation between the length of stay and the severity of illness, so the idea is that the great number of days a patient stays in hospital, the more severe the disease and the more resources are used.

Finally, the weighted numbers of discharges are obtained by multiplying the number of discharges of each case times π_m and adding them up for every hospital.

Chapter 3

Spatial dependence in hospitals efficiency: A spatial econometric approach for Ecuadorian public hospitals.*

Abstract

This study aims to analyze whether the efficiency of Ecuadorian public hospitals experiences spatial dependence. The paper explores the question of whether demand variations are affecting public hospitals' efficiency performance through direct and spillover effects, especially since the adoption of the new constitution in 2008. We exploit a two-stage approach, wherein we use an innovative panel-data DEA to estimate the hospital efficiency in the first stage and then apply a spatial econometric framework to disentangle direct and spillover effects in the second. The results confirm positive spatial interactions among public hospitals' efficiency, as well as positive direct and spillover effects coming from demand increases, which have been reinforced since 2008.

Keywords: healthcare efficiency, healthcare reforms, spatial dependence.
JEL: C21, D61, I11, I18.

3.1 Introduction

In recent years, the assessment of the ability of public hospitals to optimally utilize their resources for the provision of healthcare (i.e. how efficiently they are performing) has

*We want to thank the participants of the Applied Lunch at UAB and at University of Barcelona, the scientific committee of the PhD on Applied Economics, Nicola Pontarollo and Judit Vall for their valuable comments. Any remaining errors are our own responsibility.

become a topic of interest that has driven the attention of academics, healthcare managers, and policymakers on measures to contain healthcare costs. Attention to this matter has gained relevance as spending in healthcare continues to rise exponentially, which drives policymakers to seek ways to pursue health objectives and at the same time contain cost pressures (Papanicolas and Smith, 2013). The increase in hospital expenditures has led to a series of reforms in developed economies to induce hospital efficiency improvements, e.g. the introduction of activity-based hospital budgets (Pross et al., 2018).

However, healthcare efficiency improvement is not just a concern of developed economies. The efficiency of public hospitals' resource use is crucial in developing countries, given the pressing need for their proper allocation due to their scarcity and limited healthcare budgets (Hafidz et al., 2018; Kumbhakar, 2010). The importance of this is highlighted by the World Health Organization (2000) as a measure that could decrease the gap of mortality rates between rich and poor countries, and within countries. It is also important to ensure that resources are well allocated to promote the goal of universal health coverage (UHC) and ensure equity of access to medical services (Hafidz et al., 2018; World Health Organization, 2000).

Despite its importance, studies of healthcare efficiency have been mainly performed in developed economies (Hafidz et al., 2018), with a small but growing number of literature being applied to developing countries (Hollingsworth, 2008). However, the methods used to study healthcare efficiency in developing economies have shown little consideration for other variables that are specific to their local setting (Au et al., 2014). From this perspective, healthcare efficiency can be influenced by different factors that vary from socio-economic, environmental, political, structural, and geographical (Hafidz et al., 2018). One important and common evidence found throughout the literature is the relevance of the spatial dimension as a catalyst for the effectiveness of selected determinants in shaping the degree of efficiency achieved by different healthcare providers. This spatial dependence can impact on the needs of the population and the behavior of healthcare providers across a wide geographical area, causing geographical concentration of needs and risk factors, as well as the rise of network effects that are often detected in the data (Tosetti et al., 2018) and which translate into a structure correlation, also known as spatial dependence (Anselin, 2010).

In developing economies, this spatial structure can take the form of heavy territorial concentration that can result in agglomeration economies.[1] The presence of agglomeration economies would lead to interactions in the health system that could be related to this spatial pattern, generating some complementarities and indivisibilities such as spillover effects (Behrens and Robert-Nicoud, 2015) that shape the healthcare behavior and efficiency performance of the system if they are proven to be significant (Bhattacharjee et al., 2014; Kinfu and Sawhney, 2015). But, once again, the literature on public healthcare efficiency that accommodates the analysis to include spatial structure in the data for developing countries has been rather limited (Kinfu and Sawhney, 2015). In this respect,

[1] Here we will associate the concept of spatial unit to a region, or an area or a territory in an alternative manner.

one of the contributions of this paper is to fill the existing gap in the literature on public healthcare efficiency for developing countries. To do this, we focus our analysis on hospital efficiency and apply it to the Ecuadorian context.[2]

The Ecuadorian case represents a suitable context of analysis, since it is characterized by significant territorial disparities and spatial dependence that arises due to the existence of spillovers effects, as has been pointed out in recent studies (Mendieta Muñoz and Pontarollo, 2016; Szeles and Mendieta Muñoz, 2016). Along with other Latin American countries, Ecuador has been facing a process of continuous deterioration of its public healthcare system due to neoliberal reforms carried out in the 1990s (Homedes and Ugalde, 2005) and the crisis of 2000. As a consequence, Ecuador suffered from a deep decline in its healthcare equity and efficiency with an increase in preexisting urban-rural and inter-regional inequalities (De Paepe et al., 2012), and in the structural segmentation and fragmentation of the healthcare system (Hartmann, 2016). These effects resulted in significant technological heterogeneity between public healthcare institutions, [3] in which the hospitals with higher technology are concentrated in the most developed cantons. [4]

Given the deteriorated condition of the healthcare sector, the government of Rafael Correa carried out a series of political reforms which introduced many changes with respect to equal access to medical attention. These reforms started with the new Constitution in 2008, which established healthcare access as a right guaranteed by the state. The free healthcare provided by the Ministry of Health's hospitals (widely advertised by government campaigns), jointly with new social security and criminal-code laws that made insurance coverage compulsory, are among the country's most salient policies (De Paepe et al., 2012). The new access to medical attention resulted in a higher inflow of patients to public hospitals. According to the Public Ministry of Health (MSP), between 2006 and 2010 the number of surgeries increased by 47% hospital discharges 43% (Ministerio de Salud Pública, 2012).

In light of this evidence, we can expect that the potential increase in demand has an effect on hospital efficiency in the short-run. The rationale is the following: a higher number of treated patients can lead to better use of hospital resources, which are usually well endowed but inefficiently exploited in developing economies (Hafidz et al., 2018). In other words, these hospitals have spare resources that are not used to provide medical treatment. The increase in the number of patients would force the hospital managers to

[2]In this study, we intend "hospital efficiency" as the optimal use of a hospital's inputs in order to produce a given healthcare output. This is commonly understood in the healthcare efficiency measurement literature as technical efficiency (for a survey of the literature see Hollingsworth, 2008). Additionally, "hospital inputs" means hospital resources that are frequently measured, such as the number of physicians, beds, medical equipment, etc. "Hospital outputs", on the other hand, are viewed as the units of delivery of hospital services, and are usually measured as the number of discharges or procedures carried out.

[3]Here we consider technology as the set of constraints defining how one can combine or convert inputs into outputs in the production process. In this particular context, this can relate to the availability of human capital, infrastructure, etc.

[4]In Ecuador, cantons are the second level administrative divisions. The Republic of Ecuador is divided into 24 provinces, which in turn are divided into 221 cantons. The cantons in turn are subdivided into parishes.

make use of these unexploited resources and therefore increase hospital efficiency would also rise. However, the increase (decrease) of efficiency might not just affect a given hospital, but also those surrounding it, given that hospitals can have strategic interactions in terms of quality and efficiency (Longo et al., 2017) that are linked to the mobility of the demand.[5,6] The idea of how these interaction effects work is the following: when the new reforms decrease the barriers of access to healthcare, patients seek treatment in hospitals where they believe they will benefit from higher quality services (which can include the high-tech hospitals) or they could also be referred from low-tech hospitals to receive treatment for a complex pathology. In Ecuador, the criteria for the distribution of public funding for healthcare services are based on the healthcare needs and the size of the served population (Villacrés and Mena, 2017). Hence, this system generates incentives for hospitals to attract more patients. As a consequence, within a bounded area, surrounding hospitals can perceive how bigger hospitals are behaving and adapting to a changing reality and can react by trying to capture some of this newly-created demand by increasing their own quality (which will be constrained by their technological endowment). If the costs of providing more quality are increasing, then higher costs stemming from higher demand will reduce the incentives for cost control, thus reducing hospital efficiency.[7] Given that hospitals have to make a decision about their efficiency, they can also react by increasing or decreasing (strategic complements and strategic substitutes, respectively) their efficiency in reaction to the changes in the efficiency of neighboring hospitals.

Moreover, taking into account the technological differences of the healthcare system, an increase in demand can lead to a congestion effect for high-technology hospitals, which is the case for the vast majority of patients treated in Ecuador. If these hospitals cannot manage their resources efficiently, the increase in the number of patients can lead to a decrease in their performance (Cozad and Wichmann, 2013). Thus, surrounding hospitals could increase their quality to capture some of the demand that cannot be met by high-tech ones. This reaction, in turn, can affect their efficiency in the same manner mentioned earlier.

In light of this evidence, the aim of this study is to analyze whether public hospitals in the Ecuadorian healthcare system adapt their efficiency in response to changes in the efficiency of neighboring hospitals. We tackle the question of whether demand variations are affecting the efficiency of public hospitals through direct and spillover effects, and whether that level of efficiency has significantly changed since the new constitution came into force in 2008. We make use of the hospital occupancy rate to measure the demand.

[5]The term "strategic interactions" is used in the literature to refer to the interdependence among features or actions of selected units stemming from competition between those units. Strategic interactions arise due to the existence of spillover effects (Brueckner, 2003) that cause the levels of the variables of one unit to be affected by the levels of the same variables of neighbouring units.

[6]We make use of the hospital occupancy rate to measure demand. The occupancy rate has been widely used as an index to show the actual utilization of an inpatient health facility for a given time period, and is commonly applied in the literature to proxy medical resource utilization (Herwartz and Strumann, 2014; Town and Vistnes, 2001).

[7]In fact, according to Villacrés and Mena (2017) the current funding scheme of the country can generate inefficiencies given that the hospitals have an incentive to attract patients and inflate the costs.

The occupancy rate has been widely used as an index to show the actual utilization of an inpatient health facility for a given time period, and is commonly applied in the literature to proxy medical resource utilization (Herwartz and Strumann, 2014; Town and Vistnes, 2001).

Our research covers the period of 2006–2014 and uses hospital and cantonal data gathered from the public statistics of the Ecuadorian Institute of Statistics and Censuses (INEC) and the Ecuadorian Central Bank (BCE). We contribute to the existing literature by generalizing the approach by Longo et al. (2017) by means of the non-parametric efficiency measurement analysis that accounts for both the panel structure of the data and the technological differences of the healthcare system developed in Chapter 2 to obtain robust time-varying efficiency scores. By adopting efficiency measurement techniques, we can account for one efficiency measure that considers the use of multiple inputs to produce a given level of healthcare output, rather than relying on different productivity ratios that might produce mixed results. Also, we adopt spatial panel econometric techniques as a framework of analysis for performing our second part of the empirical analysis by taking into account the spatial dependence of the data and disentangling direct and spillover effects that can affect the hospitals' efficiency performance.

By doing this, we combine two strands of literature that have been little exploited jointly to implement our empirical framework referring to developing economies (Kinfu and Sawhney, 2015). If spatial autocorrelation in hospital efficiency is found, then the relevance of being able to assess spatial dependence stands to be an important consideration in planning public policies. If so, hence when spatial dependence is identified, policymakers cannot neglect the existence of spillover effects for achieving pre-established levels of efficiency when implementing new healthcare public policies (Mobley et al., 2009). In this study, we bring new evidence to understand the way in which the spatial dimension may contribute to shaping more effective actions for fueling territorial healthcare access and resource allocation, especially when dealing with very heterogeneous settings such as those found in developing countries.

Our main results identify a significant positive spatial dependence among hospitals in Ecuador, suggesting that their healthcare services are perceived as complementary in terms of efficiency. Also, the higher demand for medical treatments reflects a positive association with efficiency, regardless of the technological group; in addition, this demand is affecting the efficiency of surrounding hospitals as well, providing evidence of spillover effects. Both direct and spillover effects have significantly increased since 2008. This result suggests that reforms carried out after the constitution boosted the efficiency of the public healthcare system.

The organization of this chapter is as follows. In Section 3.2, we outline a short description of the institutional setting in Ecuador relevant to learning about the local healthcare system. A literature review is presented in Section 3.3. Section 3.4 introduces the theoretical framework as developed by Longo et al. (2017) and the empirical strategy is discussed in Section 3.5. Section 3.6 describes our dataset, while estimation results and conclusions are presented in Section 3.7 and Section 3.8, respectively.

3.2 Institutional setting

The Ecuadorian healthcare system includes the public and private service sectors, the former being the sector used by most of the insured population. According to the Survey of Life Conditions (ECV) of INEC, around 66% of the population was covered by public insurance in 2014, while private insurance accounts for only 6%.

The public healthcare sector is the result of the actions supported by the Ministry of Public Health (MSP), the Ministry of Social and Economic Inclusion (MIES), and municipal health services and social security institutions.[8] The MSP provides healthcare for the whole population. The MIES and the municipalities establish and finance healthcare programs to guarantee medical treatment services to uninsured citizens, which by 2014 represented around 33% of the national population, according to the ECV. Finally, social security institutions sponsor medical services to those covered by social insurance (Lucio et al., 2011).

As for funding sources, public services are financed mainly through the general public budget, but they also receive funding from extra-budgetary sources, emergency and contingency funds, and other contributions from national and international projects. The social security services for employees works on a contributive base and is financed by the contributions of affiliated workers. They are secured by the Social Security Law, as a right of protection for Ecuadorian workers (Organización Panamericana de la Salud, 2008).

Since the ratification of the new constitution in 2008, many reforms have been carried out to promote higher access to medical treatment for uninsured citizens such as the provision of free medical services by the MSP in 2008, coverage for children under 18 years old in 2010, and the civil responsibility with penal charges for employers who fail to affiliate their employees within a maximum period of 30 days in 2011. After the implementation of these policies, there has been an increase in the annual growth rate of active beneficiaries,[9] while the number of patients seen in public hospitals increased around 40% between 2006 and 2014 (see Chapter 2).

[8] The Ecuadorian Social Security Institute (IESS), the Social Security Institute of the Armed Forces (ISSFA) and the Social Security Institute of the National Police (ISSPOL).

[9] Orellana et al. (2017) present descriptive data of social insurance beneficiaries and describe an annual growth rate of 10% after 2010, compared with a 7% growth rate in previous years.

3.3 Literature review

The importance of healthcare services around the globe is widely recognized. Investment in healthcare has been rising rapidly, as have healthcare costs as a proportion of GDP; as a result, there is great policy emphasis on improving efficiency (Bloom et al., 2015). The territorial assessment of healthcare services is a key aspect of this, as there may be many sources of geographic variation that can produce different health outcomes according to the area of study (Allin et al., 2016; Chandra and Staiger, 2007; Williams et al., 2016). Also, the recognition of significant geographical concentration for many health indicators has motivated an extensive use of spatial methods to analyze health economic issues (Moscone and Tosetti, 2014).

In this strand of literature, there have been many applications related to different topics in health economics that address a spatial perspective; a complete review of most of this literature can be found in Moscone and Tosetti (2014); Baltagi et al. (2018) and Tosetti et al. (2018).

A wide body of this literature focuses on knowledge spillovers, hospital competition, and agglomeration. Common findings suggest that agglomeration economies in healthcare markets promote a quicker adoption of a new innovation among firms, mainly hospitals (Baicker et al., 2013; Chandra and Staiger, 2007; Cohen and Morrison Paul, 2008; Goodman and Smith, 2018). It is the interaction and competition between these hospitals that impact some market variables such as prices (Mobley, 2003; Mobley et al., 2009) or the quality and efficiency of services (Gravelle et al., 2014; Longo et al., 2017, 2019).

Although spatial economic methods have been applied in much of the literature, there is a lack of empirical research that addresses spatial dependence in healthcare efficiency analysis. The consideration of efficiency analysis using a spatial approach can provide several benefits to health providers, planners, and policymakers alike. It can help decision-makers to identify geographic units that can attain a better outcome without increasing the allocation of resources. Also, it can provide information on the exogenous factors whose presence (or absence) affects the performance of services and hence health outcomes in the country (Kinfu and Sawhney, 2015).

There are few and very recent papers that address a joint study of healthcare efficiency analysis from a spatial perspective. Herwartz and Strumann (2012) study whether the introduction of prospective hospital reimbursements based on diagnosis-related groups (DRG) has caused an increase in the negative spatial autocorrelation of hospitals' efficiency due to the competition for low-cost patients. Using Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) methods to measure hospitals efficiency in a first stage, and Spatial Autoregressive Models with Autoregressive Disturbances (SARAR) in a second stage, they find a statistically significant presence of negative spatial autocorrelation among hospitals in Germany, which significantly increased after the financial reform. Herwartz and Strumann (2014) extend the analysis in Germany in order to identify efficiency gains as a consequence of the same financial reform. They follow two different approaches. First, they consider a two-stage approach, starting with the decomposition of Malmquist index technical efficiency, and then complete the analysis with a SARAR model. In the second approach, they use a one-step fixed effects SFA model, accounting for technological change and spatial dependence. Both methods fail to find any efficiency gains from the new incentive structure in Germany. Felder and Tauchmann (2013) also study the efficiency of healthcare provision in Germany, considering the spatial perspective, which they state is important due to the regional competition and patient migration. They adopt a longitudinal approach for Germany's regions utilizing an order-*m* DEA method to measure regional efficiency and a spatial autoregressive model in a second step. Their findings show that accounting for spatial dependence increases the estimated effects of federal states on district efficiency. This may be a way to understanding why more efficient states are less affected by spillovers. As can be seen from these results, introducing spatial dependence in the economic analysis clarifies the importance of health policy at the state level. Herwartz and Schley (2018) depart from these findings and consider socio-economic characteristics that influence the regional efficiency in the provision of healthcare services in Germany. By means of the SFA approach, they identify that income, unemployment, the proportion of immigrants, and educational level have an effect in shaping the efficient provision of healthcare services in German districts.

Martini et al. (2014) analyze the trade-offs between hospital health outcomes (such as mortality) and efficiency using a ward-level set of hospitals in Lombardy, Italy. Their findings support the existence of a trade-off between mortality rates and efficiency, where more efficient hospitals have higher mortality rates but lower readmission rates. They also point out the role of the spatial dimension, since mortality rates are higher for hospitals subject to a high degree of horizontal competition but lower for those hospitals having strong competition but high efficiency. [10]

3.4 Theoretical framework

The building blocks of the theoretical model we refer to in this analysis were developed by Gravelle et al. (2014) and Longo et al. (2017). Their theoretical models considers strategic interactions in hospital quality and efficiency arising from spillover effects within a geographical area. The idea is that if hospitals compete within a given area, they will attract patients by increasing their quality. If neighboring hospitals react by increasing (or decreasing) their own quality, then we identify that hospitals are strategic complements (substitutes) in their quality. Furthermore, the reduction in a hospital's demand that follows from an increase in their closest neighbor's quality also has an effect on its efficiency. The cost of increasing quality to attract higher demand might also reduce incentives to control costs and thus reduce efficiency. In the way, hospitals can be strategic

[10]In their analysis, Martini et al. (2014) claim that when there is national health insurance, the cost of service are irrelevant for the consumer hospital choice. This makes competition among hospitals mainly focused on location, which they refer to as horizontal competition (Tay, 2003).

complements (or substitutes) in their efficiency and a neighboring hospital's increase in efficiency can induce an increase or decrease in its own efficiency.

In order to present the framework in terms of the strategic interaction in efficiency used by Longo et al. (2017) we consider a two-provider model of quality competition (q) and cost-reduction effort (e).[11] Let us assume q_i as the healthcare quality of hospital i and q_j the healthcare quality of hospital j, with $i \neq j$. The demand function for hospital i is given by $D_i = (q_i, q_j)$, such that $D_{iq_i} = \frac{\partial D_i}{\partial q_i} > 0$ and $D_{iq_j} = \frac{\partial D_i}{\partial q_j} < 0$, so it is increasing in its own quality but decreasing in the quality of hospital j. This assumption implies that hospitals are demand (imperfect) substitutes: patients switch from one hospital to another in accordance with the variation in the quality of the healthcare of the two hospitals. However, switching from one hospital to another entails costs in terms of time and transfer costs. Here, we define the objective function of hospital i as:

$$\pi_i = [p - c_i(q_i, e_i; \theta_i)] D_i\left(q_i, q_j; \theta_i\right) - G_i(q_i, e_i; \theta_i) \tag{3.1}$$

Where p is a fixed price per treatment that the hospital i receives from a third-party payer, such as the government in our case, $c_i(q_i, e_i)$ are the variable treatment costs, given that $c_{iq_i} = \frac{\partial c_i}{\partial q_i} > 0$ and $c_{ie_i} = \frac{\partial c_i}{\partial e_i} < 0$, they are increasing in quality and decreasing in efficiency, e_i. $G_i(q_i, e_i)$ are monetary and non-monetary fixed costs, with $G_{iq_i} = \frac{\partial G_i}{\partial q_i} > 0$ and $G_{ie_i} = \frac{\partial G_i}{\partial e_i} > 0$, whereas θ_i is a vector of shift parameters, such as location of patients and other hospitals, input prices, demographics, central policies, type of hospital, etc. The authors assume that quality and efficiency are substitutes $\left(G_{iq_i,e_i} = \frac{\partial^2 G_i}{\partial q_i \partial e_i} > 0\right)$, meaning that an increase in quality would require a decrease in cost-reduction effort. Also, for sake of simplicity, Longo et al., (2017) make the assumption of independence in variable costs, that is $c_{iq_i,e_i} = \frac{\partial^2 C_i}{\partial q_i \partial e_i} = 0$. The first order conditions to the equation (3.1), by which hospital i maximizes its profit with respect to quality and efficiency, is as follows:

$$\begin{aligned}\pi_{iq_i} = \frac{\partial \pi_i}{\partial q_i} &= [p - c_i(q_i, e_i; \theta_i)] D_{iq_i}\left(q_i, q_j; \theta_i\right) - c_{iq_i}(q_i, e_i; \theta_i) D_i\left(q_i, q_j; \theta_i\right) \\ &\quad - G_{iq_i}(q_i, e_i; \theta_i) = 0\end{aligned} \tag{3.2}$$

$$\pi_{ie_i} = \frac{\partial \pi_i}{\partial e_i} = -c_{ie_i}(q_i, e_i; \theta_i) D_i\left(q_i, q_j; \theta_i\right) - G_{ie_i}(q_i, e_i; \theta_i) = 0 \tag{3.3}$$

With $D_{iq_i} > 0$, $c_{iq_i} > 0$ and $G_{iq_i} > 0$. Optimal quality is achieved when the marginal profit from one additional unit of demand is equal to the correspondent marginal cost.

[11] The cost-reduction effort is interpreted as an efficiency improvement. As the more efficiently the resources are used to obtain a given output, the fewer costs there are for the hospital.

On the other hand, the optimal level of efficiency is such that the marginal benefit from lower costs and higher profits are equal to the marginal disutility from efficiency.

Since the scope of Longo et al. (2017) is to propose a model to examine hospitals' strategic interactions, they find the interaction functions of hospital i's quality (q_i) and efficiency (e_i) as a function of the choice of quality by hospital j. The reaction functions defined by the first-order conditions (3.2) and (3.3) satisfy:

$$q_i = q_i^R(q_j; \theta_i) \tag{3.4}$$

$$e_i = e_i^R(q_j; \theta_i) \tag{3.5}$$

Here, it would seem that the quality and efficiency of hospital i are independent from the efficiency of hospital j because neither of the first order conditions of hospital i depends on the efficiency of hospital j. But the total differentiation of the first-order conditions yields:

$$\begin{aligned} \frac{\partial q_i^R}{\partial q_j} &= \left\{-\pi_{iq_i,q_j}\pi_{ie_i,e_i} + \pi_{ie_i,q_j}\pi_{iq_i,e_i}\right\}\Delta^{-1} \\ &= \left\{-\left[(p-c_i)D_{iq_iq_j} - c_{iq_i}D_{iq_j}\right]\pi_{ie_i,e_i} - c_{ie_i}D_{iq_j}\pi_{iq_ie_i}\right\}\Delta^{-1} \end{aligned} \tag{3.6}$$

With $\Delta = \pi_{iq_i,q_i}\pi_{ie_i,e_i} - \pi_{iq_i,e_i}{}^2 > 0$. The first term in the square brackets is the direct effect of the neighbor's quality on the marginal profit from higher quality. It is not clear whether an increase in hospital j's quality increases or decreases the marginal demand of hospital i, so the sign of $D_{iq_iq_j}$ is unknown. For the sake of simplicity, if we assume that $D_{iq_iq_j} = 0$, this will lead to a reduction in the variable costs (second term in the square brackets), because the increase in the neighbor's quality reduces demand and the marginal cost of output of hospital i, which will respond with an increase in quality. However, the second term in the curly brackets also emphasizes another effect. Lower demand will also reduce incentives to control for costs (lowering efficiency), and so variable costs may increase.

Hospitals, then, can be affected by the patients' perception in their quality; if the quality of a hospital is perceived to be high, this will end in an increase in patients' demand for this hospital, switching from its neighbors and yielding less efficiency. However, this is conditioned to the spatial structure. The strategic interaction will be stronger for hospitals that are closer to one another. Changes in quality and efficiency will matter because of hospitals' proximity, and because of the decay effect of spillovers.

In our case, the healthcare reforms that have been implemented relax some barriers to access to medical services, allowing citizens to select between different hospitals. In the short run, a hospital gains more patients when it increases its quality since the patients have the opportunity to choose those hospitals which they perceive as better qualified. But the effect that the reforms can have on the demand for a particular hospital is ambiguous. It will depend on the quality of the other hospitals and the geographical distribution of the patients and hospitals (Gravelle et al., 2014). So, patients will decide to switch from one hospital to another depending on the travel distance and transfer costs. Neighboring hospitals can react to the increase in quality of a hospital by either increasing or decreasing their own quality. This affects the final demand and therefore the hospitals' efficiency.

Therefore, in order to test the spatial interaction in hospital efficiency we use the following function:

$$e_i = f(e_{i-1}, Z_i, \varepsilon_i) \tag{3.7}$$

With e_i being the efficiency of hospital $i = (i, \ldots, I)$, e_{i-1} is the efficiency of hospital i's neighbor, Z_i is the vector of covariates, including hospital variables (e.g. occupancy and mortality rate, market share, etc.), and cantonal variables (e.g. GVA, density, etc.).

3.5 Empirical strategy

The first stage of our empirical strategy involves defining a measure of efficiency. We make use of the efficiency scores obtained in Chapter 2. As explained, these efficiency scores are mainly based on the panel Data Envelopment Analysis (panel-data DEA) proposed by Surroca et al. (2016), and Pérez-López et al. (2018). The advantage of this approach over other efficiency measurement analyses such as classical DEA or other dynamical approaches such as the Malmquist index is that it allows one to estimate time-invariant coefficients of efficiency for the period of analysis, considering the inherent panel data structure. Additionally, these time-invariant efficiencies can be broken down into time-variant ones, calculating efficiency values for each year under evaluation. One of the principal advantages of this approach is that the results are robust to outliers and temporal random shock, which provides efficiency scores representative of the complete period.

In Chapter 2, we extend this approach to account for technological heterogeneities of Ecuadorian public hospitals by applying multivariate techniques (factor analysis in combination with clustering methods) to obtain panel data-DEA efficiency scores for three different groups (clusters): high-tech, intermediate-tech and low-tech.

In this Chapter, we follow an input-oriented efficiency measurement. We assume a variable return to scale (VRS) model to deal with heterogeneous observations.[12] The efficiency frontier is developed by optimizing the weighted input/output ratio of each Deci-

[12]This is also tested in the empirical application with the Simar and Wilson (2002, 2011) returns-to-scale test.

sion Making Unit (DMU),[13] subject to the condition that this ratio can be equal, but never exceed one for any other DMU in the data set (Charnes et al., 1978).

The second step of our strategy defines a convenient spatial model whose main idea is to assess whether hospitals' efficiency is associated with the efficiency of nearby hospitals and with other observed and unobserved variables. For this, the spatial econometrics literature has developed models that treat three different types of interaction effects among units of analysis (Halleck Vega and Elhorst, 2015). These interaction effects account for (i) endogenous interaction effects among the dependent variable; (ii) exogenous interaction effects among the explanatory variables; and (iii) interaction effects among the error terms.

The identification of the source of spatial autocorrelation needs to be carried out in order to avoid model misspecifications and omitted variable bias. Following the strategy described in LeSage and Pace (2009) and Elhorst (2010), we begin with a with a Spatial Durbin Model (SDM) setting as a general specification and, then test for alternatives. The process of model selection can be found in Appendix 3.9.2. We also provide Lagrange Multiplier (LM) lag and error tests for spatial panel models (Anselin et al., 2006) and their robust counterparts (Elhorst, 2010), which are commonly used in the literature to make inferences for spatial interaction effects.

To select between random and fixed effects models, we ran the robust Hausman test (Hausman, 1978) and found robust evidence for the fixed effects model. Elhorst (2014) also recommends the selection of the fixed effects in spatial panel models when space-time data of adjacent spatial units are located in unbroken study areas. Also, given the assumption of orthogonality between the individual-specific component and the explanatory variables, this assumption is particularly restrictive and difficult to hold in empirical applications (Baltagi, 2013; Baltagi et al., 2018).

The model selection points out a SAC model as the appropriate framework of analysis.[14] This is consistent with similar applications in the existing literature (Felder and Tauchmann, 2013; Herwartz and Strumann, 2012, 2014), suggesting that the sources of autocorrelation occur in the efficiency performance of hospitals and unobservable factors that we cannot measure. Thus, from equation (3.7) we specify the following spatial panel data SAC model estimated by Quasi-Maximum Likelihood (QML):

$$\log(e_{it}) = \rho \sum_{j \neq i} w_{ij} \log(e_{jt}) + \beta' \log(Z_{it}) + \emptyset_i + \gamma_t + \varepsilon_{it}$$

[13] We can call DMU to any unit of analysis, such as, individuals, departments, firms, municipalities, or, in the case of this study, hospitals.

[14] The acronym SAC is consistent with the terminology of LeSage and Pace (2009), but other authors give this model the acronym SARAR, which stands for Spatial Autoregressive Models with Autoregressive Disturbances.

$$\text{with } \varepsilon_{it} = \lambda \sum_{j} w_{ij} \varepsilon_{jt} + \epsilon_{it} \tag{3.8}$$

The variable e_{it} is the logarithm of the efficiency of the hospital i at time t and w_{ij} (with$j \neq i$) are the spatial weights that capture the pattern of spatial dependence and the strength of potential interaction between units i and j. The variable Z_{it} is the vector including variables such as occupancy rate, market share, mortality rate, and regional demographics that can affect the efficiency of the hospital. The variable $\emptyset_i$ captures the hospital fixed effects and γ_t is the time effect. Finally, ε_{it} is the error term. We define equation (3.8) in matrix form as:

$$\begin{gathered} \boldsymbol{e}_t = \rho \boldsymbol{W} \boldsymbol{e}_t + \boldsymbol{Z}_t \boldsymbol{\beta} + \boldsymbol{\emptyset} + \boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t \\ with\ \boldsymbol{\varepsilon}_t = \lambda \boldsymbol{W} \boldsymbol{\varepsilon}_t + \boldsymbol{\epsilon}_t \end{gathered} \tag{3.9}$$

As for the specification of the components of the weight matrix $\boldsymbol{W}$, we use two different specifications. The former (hereinafter $\boldsymbol{W}_d$) is the inverse of the shortest Euclidean distance between any pair of spatial units (i and j), which has been commonly used in the literature when the data covers healthcare providers (Tosetti et al., 2018). The latter (hereinafter $\boldsymbol{W}_v$) uses the inverse shortest time travel distance by car still between any pair of locations (i and j), as in Gravelle et al. (2014).

The key parameters to be estimated for the spatial autocorrelation are the coefficients ρ and λ. These measure the strength of the spatial dependence due to efficiency changes and to unobservable factors in neighboring hospitals respectively, conditional on the vectors of explanatory variables. If $\rho > 0$ then a positive autocorrelation is found in the efficiency of hospital i and the efficiency of their neighboring hospitals, and similarly for λ.

One of the main advantages of using spatial econometrics is its capacity to empirically assess the magnitude and significance of spillover effects (Elhorst, 2014). In this sense, spatial regression models exploit the dependence structure among hospitals: the effect of the change of an explanatory variable for a specific hospital will affect the hospital itself, and, potentially, all other neighboring hospitals indirectly. This implies the existence of direct, indirect (spillover) and total effects. We can estimate these effects by obtaining the matrix of partial derivatives of the expected values of e_{it}, as proposed by LeSage and Pace (2009). So far, the literature on spatial healthcare economics has identified the existence of spatial spillovers based on coefficient estimates (Baltagi et al., 2018). We improve the empirical approach by accounting for the direct, indirect, and total effects of independent variables. As stated by LeSage and Pace (2009), the partial derivative interpretation of the impacts coming from changes in the independent variables provides a more valid basis for testing the existence of spillover effects. Here, we are also interested in measuring the effects of the hospitals' occupancy rates, which can bring tangible evidence of how the demand for medical services is affecting the efficiency of a given hospital and whether this is also affecting neighboring hospitals due to spillover effects. In addition, we carry out the LeSage and Pace (2009) partitioning analysis of the spatial multiplier.[15] With this, we are able to trace the effect of the linkages between demand levels of neighboring hospitals. Thus, we not only concentrate on analyzing the direct, spillover, and total effects, but also determine the impacts that the demand itself has over the higher order of contiguity. In other words, we are able to examine how the impact of hospital demand manifests itself over space (Jensen and Lacombe, 2012). Finally, by means of hypotheses testing, we can check for its significant increase (or decrease) of the direct and indirect effects since 2008.

To test the statistical variations of the healthcare demand upon the hospitals' efficiency before and after 2008, we interact the logarithm of the occupancy rate with time dummies ($ocrate_t$). Specifically, we have built the following test: $H_o : ocrate_1 = ocrate_2$, where $ocrate_1 = 1/2\sum_{t=2007}^{2008} ocrate_t$ and represents the subperiod before the constitution,[16] while $ocrate_2 = 1/6\sum_{t=2009}^{2014} ocrate_t$ constitutes the subperiod after the constitution. These hypotheses are tested by means of a two-sided t-test.[17]

3.6 Data and variables

The database we have used covers the period from 2006 (two years before the new constitution was approved) to 2014. We make use of the same information collected for Chapter 2 to carry out the first stage of our strategy. The hospital information was collected from the Annual Survey of Hospital Beds and Discharges and the Survey of Health Activities and Resources provided by the INEC. We excluded the psychiatric, dermatologic, and geriatric hospitals, and removed outliers from the sample.[18] We retrieved a panel data of 186 hospitals for which an average of 21 hospitals per year had missing values that were imputed by means of Predictive Mean Matching imputation (Rubin, 1986).[19] Cantonal economic and demographic variables were retrieved from the BCE and INEC's public statistics respectively. A description of all the variables is presented in Appendix 3.9.1.

[15]Refer to Appendix 3.9.3 for an explanation of LeSage and Pace (2009) spatial effects and its respective partitioning analysis.

[16]The constitution came into force in October 2008.

[17]The logarithmic transformation of the efficiency scores ensures an unbounded dependent variable and thus enables a consistent maximum likelihood estimation (Simar and Wilson, 2007).

[18]We excluded psychiatric, dermatologic, and geriatric hospitals as they focus on specific illness and patients that require different treatments that could bias the efficiency values. For example, psychiatric hospitals often require inpatients to stay for long periods of time, which our analysis would consider as a criteria for less efficiency.

[19]The imputation results were diagnosed by means of displays of completed data, distribution comparison, and checks for fit of the data suggested by Abayomi et al. (2008).

3.6.1 Variables for the efficiency measurement

As was previously mentioned, we employed the efficiency estimations granted from Chapter 2. The selection for both input and output variables was related to the existing literature on hospital efficiency measurement. A complete overview is proposed by Hollingsworth (2008); O'Neill et al. (2008), and Cantor and Poh (2018).

In our study, the input variables (controlled by the hospitals) are the number of beds, the medical equipment, and the availability of the infrastructure that is widely used as a proxy for hospital size and capital investment (O'Neill et al., 2008). To proxy labor costs, clinical staff were usually included (Hollingsworth, 2003, 2008). To this end, we included the number of physicians and healthcare professionals beyond the number of physicians of the hospital. To measure public hospitals' final production of health, the number of hospital discharges was employed. This variable is weighted with the case-mix index proposed by Herr (2008).

3.6.2 Variables for the spatial econometric model

To account for the changes in the number of treated patients we used the logarithm of the hospital occupancy rate.[20] Herwartz and Strumann (2012, 2014) point out that the importance of this variable in relation to healthcare efficiency. It serves as a proxy to determine whether hospitals promptly adjust their working staff to the increase in treated patients. Thus, hospitals with a relatively low occupancy rate can be interpreted as having an oversized staff, and thus as being unlikely to meet the demand of patient care efficiently. This issue has recently been highlighted for low- and middle-income countries, which present an occupancy rate well below that recommended by the WHO (Hafidz et al., 2018).

To provide a proxy for market structure in the hospitals' respective cantons, we used the logarithm of the hospital's market share. Market share has often been used as an explanatory variable in research regarding healthcare efficiency in developed economies to provide a measure of concentration (or competition). For example, Longo et al. (2019) identify hospitals that compete in a given district to proxy the patients' choice of provider. Hence, the higher the competition in a district, the wider the range of healthcare providers that the patients can choose from. This reaction is expected to drive hospitals to compete in quality and increase incentives to increase efficiency to contain costs. Despite its importance, few studies have embedded this variable in the case of setting involving developing economies (Hafidz et al., 2018). In developing economies experiencing marked healthcare heterogeneities, market share might also have an additional implication, considering that there would be just a few hospitals in which the patients believe they will be able to get quality treatment for their disease. Therefore, higher market share could also to a certain extent be a proxy for patients' perception of the quality of a hospital. In our context, we

[20] All the variables expressed as percentages were on a 0–100 scale prior to obtaining the logarithms in order to facilitate the estimations and interpretation of the results.

envisage two scenarios. In the first case, larger market shares could be related to larger hospitals, which are often located in more developed cantons. In Chapter 2, we find that these types of hospitals are those with better technology and better performance (hence, the most efficient). The second case would represent those hospitals located in less developed cantons (hence, with lower technology and efficiency) which do not have to deal with many close competitors.

One of the limitations that we faced was finding appropriate variables in the dataset that could properly measure the quality of the hospitals. To address hospital quality, the variables commonly used in the literature range from mortality, readmission or health satisfaction rates (Hafidz et al., 2018; Hollingsworth, 2008). Unfortunately, these were not available in our data. For this reason, we decided to take into account hospitals' quality by including the logarithm of the hospital and cantonal mortality rates. Other morbidity variables were also included, such as the number of disease-specific treated patients, to provide additional controls on the complexity of cases treated. Hospitals whose performance displays a significant positive relationship with these morbidity variables may suggest not only a higher quality in the treatment of the disease, but also a process of learning-by-doing (Gobillon and Milcent, 2013), as they would show increasing experience in the treatment of these diseases over time.

The technological differences were included as a dummy interacting with different hospital independent variables to estimate their differential effect on the hospitals' efficiency scores.

As for canton specific variables, we included the logarithm of the density and gross value added (GVA) to control for the canton's level of urbanization and proxy some exogenous socio-economic factors respectively (Herwartz and Strumann, 2012, 2014). Many have addressed the influence of the elderly population on hospital efficiency (e.g. Herr, 2008; Longo et al., 2017), as they are likely to be more cost and resource intensive and present more complications in treatment. In addition, Orellana et al. (2017) provide evidence of over-utilization of medical treatment in the Ecuadorian public health system for people over 60 years old, which can negatively affect the systems' performance, as they might be using medical resources that could be employed for higher priority or more severe cases. Here, we used the logarithm of the population over 65 years old to control for this effect.

Finally, we used the logarithm of cantonal patient migration measured as the number of patients treated in cantons different from the ones of their place of residence. Felder and Tauchmann (2013) state the importance of accounting for regional patient migration as it can be potentially correlated with inefficiency. Patient migration can explain efficiency differences between territories, as it could be capturing deprivation effects (Herwartz and Schley, 2018). Bigger hospitals located in the developed regions are very likely to treat patients from outer regions, as patients in less-developed regions have access restrictions to good healthcare quality and perceive these bigger hospitals to have higher quality than those located in their residence area (Martini et al., 2014). In this way, smaller hospitals –likely to be located in less-developed areas – can present higher efficiencies that are not due to more efficient use of their inputs, but rather a lower local demand due to patient migration (Herwartz and Schley, 2018).

The descriptive statistics of our data are presented in Table 3.1. We split the sample in technology cluster according to the criterion proposed in Chapter 2 (low-tech, intermediate-tech, and high-tech). At first sight, this table emphasizes the important heterogeneity in the Ecuadorian public healthcare system. Low-tech hospitals are the majority in the system, but they have a much lower number of healthcare inputs on average than their high-tech counterparts. However, these high-tech hospitals treat more than 14 times the number of patients attended in the low-tech hospitals.

Table 3.1: Summary statistics of the variables

Variable	Cluster 1 (Low) n=156				Cluster 2 (Intermediate) n=21				Cluster 3 (High) n=9			
		SD				SD				SD		
	Mean	Overall	Between	Within	Mean	Overall	Between	Within	Mean	Overall	Between	Within
Output												
Number of discharges (weighted)	15034	439416.6	141543.5	414089.6	2006	3348.9	3350.2	641.78	221772	1827255	647701.2	1720670
Inputs												
Number of physicians	44	56.59	50.76	23.64	47	87.52	81.39	35.78	213	126.16	105.72	76.43
Number of beds	71	103.96	100.93	17.28	81	146.1	146.41	26.65	273	136.24	137.05	40.38
Number of hospital staff	96	144.79	137.84	38.1	98	218.03	207.83	77.1	445	242.12	226.42	111.4
Number of equipment and infrastructure	68	81.92	74.5	34.44	64	60.8	59.98	15.57	255	137.41	106.81	92.73
Explanatory Variables												
Occupancy rate (%)	57.91	26.13	19.75	17.17	45.89	28.95	21.42	19.98	73.8	20.63	18.27	11.2
Market share (%)	67.23	40.55	37.8	14.95	45.86	41.76	38.02	18.98	18.48	17.31	16.07	8.2
Mortality rate (% hospital)	0.84	1.49	1.32	0.7	0.61	0.67	0.54	0.42	2.65	1.42	1.05	1.02
Number of disease 1	224	377.24	331.29	182.16	195	255.86	225.66	129.25	800.53	1293.1	1056.21	817.36
Number of disease 2	152	504.12	491.79	116.86	381	1043.97	960.01	455.48	1019	895.49	723.62	575.01
Number of disease 3	25	48.42	44.71	18.89	33	76.26	67.71	37.76	217	170.6	146.48	98.96
Number of disease 4	253	384.47	352.57	155.63	255	394.24	341.21	209.66	998	834.42	763.3	414.59
Number of disease 5	50	69.1	62.75	29.32	55	74.36	58.67	47.27	198	171.79	162.85	75.12
Number of disease 6	1490	3188.42	3113.28	727.17	1010	1507.23	1359.28	709.04	1845	2115.75	2015.82	905.11
Number of disease 7	205	707.51	627.11	330.98	125	240.65	197.48	143.44	387	503.81	418.82	309.76
Number of disease 8	34	114.34	109.11	35.19	30	69.89	62.42	33.97	365	493.5	470.79	209.91
Number of disease 9	291	536.01	493.68	212.07	285	520.75	486.98	210.06	1572	1599.99	1375.53	925.8
GVA (thousand $)	2.00E28	4188748	4154598	619145.2	2.00E28	3638626	3668601	594701.1	8.00E28	5342403	5445905	1359843
Density (population per km2)	264.25	485.06	485.16	35.32	202.25	205.92	209.61	18.39	465.8	142.19	146.47	30.16
Mortality rate (% cantonal)	0.41	0.12	0.1	0.05	0.39	0.12	0.12	0.04	0.48	0.06	0.05	0.03
Total population over 65	24969	43875.11	43949.93	2107.82	19921	36822.14	37517.95	2858.56	91658	51679.5	54191.08	5269.4
Total patient migration	7684	15001.32	14963.12	1556.07	4530	6034.01	6050.24	1166.84	25511	18520.7	19297.82	2802.24

Source: The authors

Regarding hospital demand, we see a higher occupancy rate for the high-tech group (73.80%). Despite presenting higher demand, this occupancy rate suggests an inefficient utilization of hospital resources: there seem to be spare hospital inputs that are not currently used for treatment, implying that in general there is still room for improvement for public hospitals. Furthermore, high-tech hospitals settle in regions that concentrate a larger amount of the population and economic production. The lower market share (18.48%) shows that there is more competition in these areas with respect to the low-tech hospitals' regions, which also present a lower level of patient migration. This preliminary evidence anticipates the need to adjust the hospitals' efficiency performance to the patients' needs with strategies tailored accordingly to the technological groups.

3.6.3 Exploratory spatial data analysis

Before performing the more quantitative analysis, it is important to assess the true existence of spatial dependence in the distribution of the health resources in the Ecuadorian territory. Hence, we perform an exploratory spatial data analysis (ESDA) to identify different patterns of spatial association and regional clusters or atypical locations of our observations (Anselin et al., 2006) and gain a better understanding of the spatial structure of the data.

The aim of our spatial data analysis is to test whether strategic interaction between hospitals is occurring. This interaction can arise from the concentration of health resources in selected areas that can yield similar patterns of efficiency (Longo et al., 2017). We test for the spatial autocorrelation and proximity of the data by means of Moran's I-statistic (Moran, 1948). Moran's I has been widely used in the literature to test for spatial dependence (LeSage and Pace, 2009). If the statistic is positive and significant, this means that hospitals with high amounts of healthcare resources are clustered.

Figure 3.1 depicts the Moran's map and scatterplot for the mean value of four different hospital features between 2006 and 2014: numbers of physicians, beds, medical equipment, and hospital personnel (not including physicians).[21] Table 3.2 reports the Moran's I test results using the weight matrix $\boldsymbol{W}_d$ based on the inverse Euclidean (shortest) distance between hospitals.[22]

[21] A more detailed description of this is presented in the data section.

[22] We also used different weight matrices such as the inverse of the shortest time travel distance, and the inverse of the squared distance and time travel distance. The results are similar in all cases.

Figure 3.1: Moran's map and Moran's Scatterplot. (a) Physicians, (b) beds, (c) personnel, (d) equipment

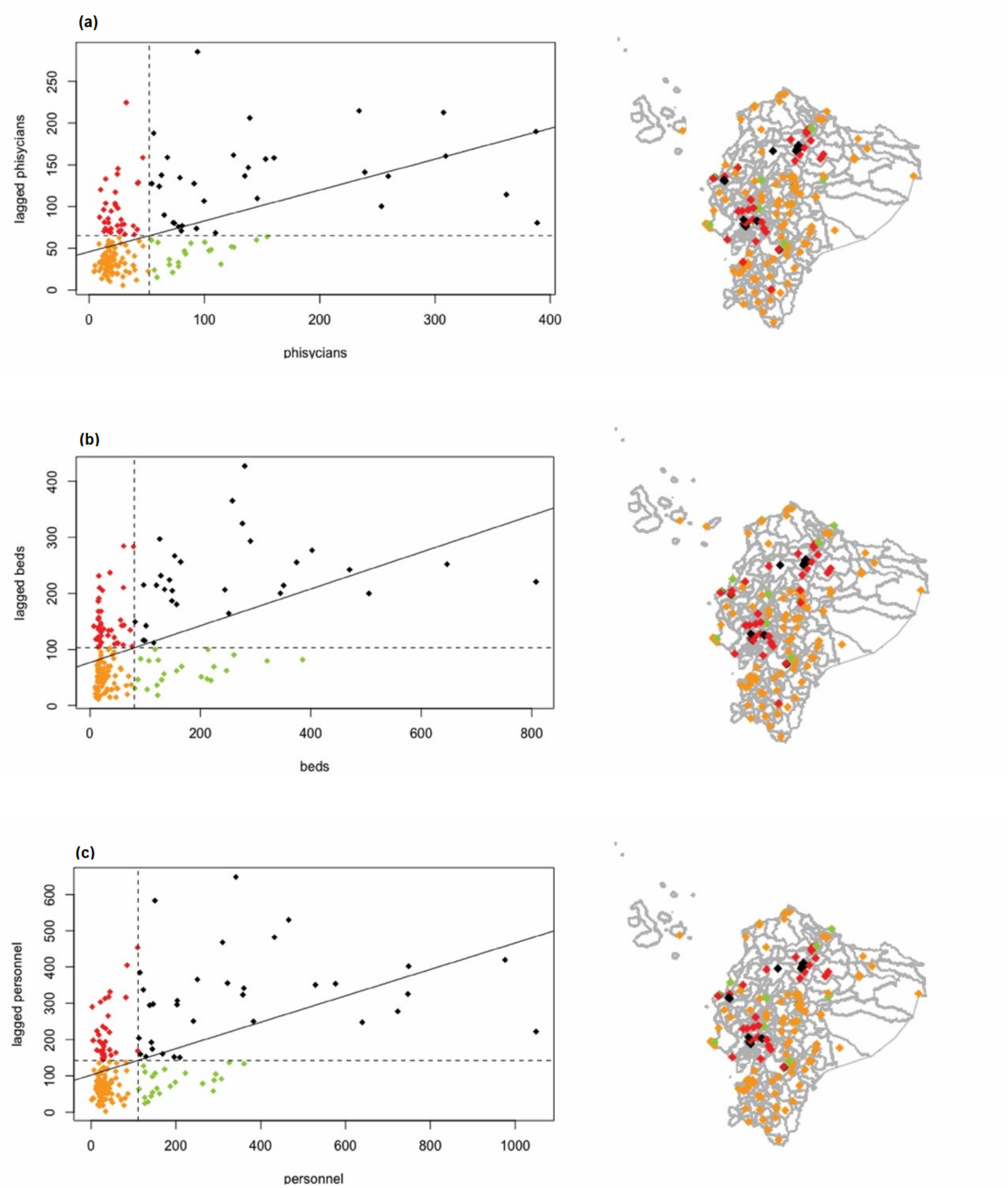


Figure 3.1: (continued)

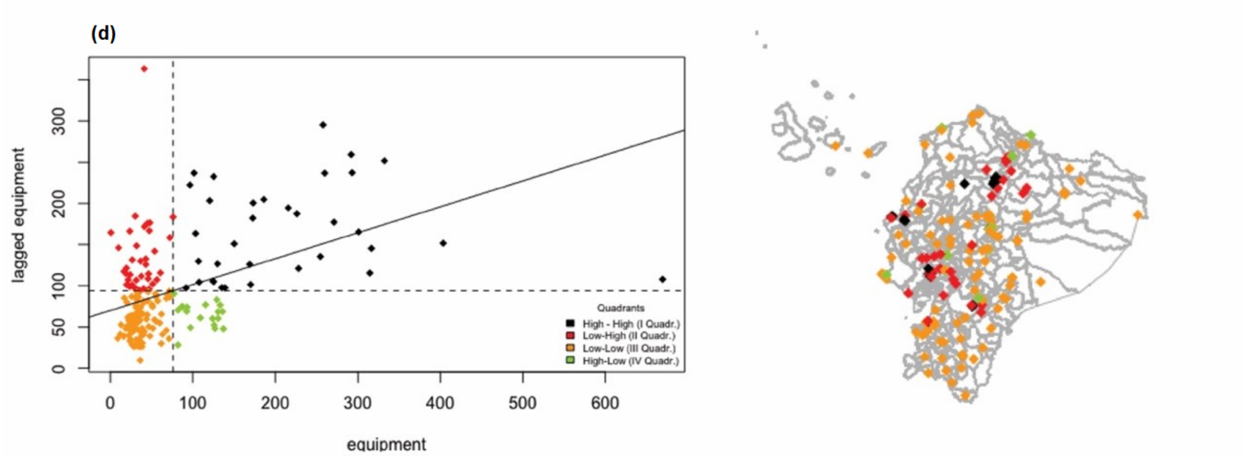


Table 3.2: Moran's I test of spatial dependence

Hospital Inputs	Moran's I	Prob
Physicians	0.3698	0.000
Beds	0.3279	0.000
Hospital Personnel	0.3638	0.000
Equipment and Infrastructure	0.3145	0.000

Source: The authors

The evidence shown in Table 3.2 shows an average positive spatial correlation for all the hospital features considered. Looking at the maps, the hospitals that present positive spatial autocorrelation (black points) are clustered around Quito and Guayaquil, which are the two biggest and most developed cantons in Ecuador (Mendieta Muñoz and Pontarollo, 2016). It is also worth noting that the spatial pattern changes as hospitals move further away from these cantons. Hospitals that surround them present dissimilar amounts of resources, represented by the reddish points (low-high), and present a negative correlation as they move farther away, as depicted by the orange points (low-low).

The corresponding scatterplots confirm the finding of positive autocorrelation. Most of the hospitals' resources cluster in quadrant III, whereas few are in quadrant I. This result assesses not only the high heterogeneity in terms of technological endowment for healthcare in Ecuador, but also the uneven distribution of these high-tech hospitals in the territory, which confirms the findings of the descriptive statistics mentioned above.

The evidence issued from this preliminary analysis implies that classic econometric approximations to study the Ecuadorian public healthcare system would fail to obtain unbiased results given the existence of spatial dependence. We need to consider a model that incorporates this dependence and which can disentangle the spillover effects that cause it (Anselin, 1988).

In order to perform our analysis, we need an appropriate measure that allows us to estimate to what extent healthcare resources are efficiently used in the production of a healthcare output. In this respect, many methods have been proposed in the literature (Cantor and Poh, 2018), but few of them have been applied in combination with spatial econometric techniques (Felder and Tauchmann, 2013).

Another novelty of this contribution is that it bridges these two strands of literature by proposing an empirical two-stage approach. In the first stage we estimate the efficiency scores (from Chapter 2) that are robust over time and have the advantage of considering the technological differences in the public healthcare Ecuadorian sector. Then, in the second stage, we select these measures of efficiency as dependent variables to perform spatial panel econometric estimates. In this way, we can determine the spatial dependence in efficiency across hospitals. This empirical framework also allows for disentangling the extent to which direct and spillover effects issuing from external factors – particularly hospitals' occupancy rates – affect the efficiency performance of hospitals over time.

3.7 Estimation results

Table 3.3 shows the regression results from the SAC spatial econometric model for equation (3.9). The first set of estimations refers to the model with the selected weight matrices and without incorporating the technological discrepancies. Hereinafter, we label this first type of setting as the baseline model.

Table 3.3: Spatial regression results. Direct, indirect, and total effects

Variables	$\boldsymbol{W}_d$			$\boldsymbol{W}_v$		
	Direct	Indirect	Total	Direct	Indirect	Total
log occupancy rate	0.140***	0.071***	0.211***	0.130***	0.0983***	0.228***
	(0.019)	(0.018)	(0.0311)	(0.0189)	(0.0211)	(0.0354)
log market share	-0.0372***	-0.0188***	-0.0560***	-0.0342***	-0.0259***	-0.0601***
	(0.011)	(0.0069)	(0.0166)	(0.0108)	(0.0094)	(0.0195)
log mortality rate	-0.0317***	-0.0160***	-0.0478***	-0.0289***	-0.0220***	-0.0509***
	(0.008)	(0.0057)	(0.0134)	(0.0083)	(0.0074)	(0.0151)
disease 1	-0.0076***	-0.0038**	-0.0114***	-0.0058**	-0.0044**	-0.0102**
	(0.003)	(0.0016)	(0.0041)	(0.0026)	(0.0022)	(0.0047)
disease 2	-0.0031	-0.0015	-0.0046	-0.0034	-0.0025	-0.0059
	(0.003)	(0.0014)	(0.0039)	(0.0026)	(0.002)	(0.0046)
disease 3	0.0784***	0.0395***	0.118***	0.0841***	0.0639***	0.148***
	(0.02)	(0.015)	(0.0366)	(0.0240)	(0.0219)	(0.0440)
disease 4	0.0232***	0.0117***	0.0350***	0.0212***	0.0161***	0.0373***
	(0.004)	(0.0034)	(0.0066)	(0.0039)	(0.0042)	(0.0075)
disease 5	-0.0255	-0.0129	-0.0384	-0.0289	-0.0221	-0.0510
	(0.018)	(0.0101)	(0.0283)	(0.0184)	(0.0149)	(0.0329)
disease 6	0.0066***	0.0034***	0.0100***	0.0067***	0.0051***	0.0118***
	(0.0008)	(0.0008)	(0.0015)	(0.0009)	(0.0011)	(0.0018)

Table 3.3 (continued)

Variables	$\boldsymbol{W}_d$			$\boldsymbol{W}_v$		
	Direct	Indirect	Total	Direct	Indirect	Total
disease 7	0.0056***	0.0028**	0.0084***	0.0045**	0.0034**	0.0079**
	(0.002)	(0.0012)	(0.0031)	(0.002)	(0.0017)	(0.0036)
disease 8	-0.0302***	-0.0152**	-0.0454***	-0.0294**	-0.0224**	-0.0518**
	(0.012)	(0.0067)	(0.0176)	(0.0118)	(0.009)	(0.0211)
disease 9	0.0060**	0.0030**	0.0090**	0.0059**	0.0045**	0.0104**
	(0.0025)	(0.0014)	(0.0038)	(0.0025)	(0.002)	(0.0045)
log GVA	0.0877**	0.0438**	0.131**	0.0650*	0.0490*	0.114*
	(0.036)	(0.0199)	(0.0545)	(0.0357)	(0.0278)	(0.0625)
log density	-0.610**	-0.300**	-0.910**	-0.663***	-0.499***	-1.162***
	(0.248)	(0.130)	(0.363)	(0.132)	(0.114)	(0.224)
log mortality (cantonal)	0.0678*	0.0342	0.102*	0.0515	0.0391	0.0906
	(0.038)	(0.0215)	(0.0584)	(0.0377)	(0.03)	(0.0669)
log pop > 65	-0.0126	-0.0075	-0.0201	-0.126	-0.0976	-0.224
	(0.120)	(0.0606)	(0.180)	(0.115)	(0.0933)	(0.207)
log inpatient migration	0.0045	0.0023	0.0068	0.0045	0.0035	0.0080
	(0.012)	(0.006)	(0.0177)	(0.0117)	(0.009)	(0.0206)
ρ	0.355***			0.453***		
	(0.053)			(0.0454)		
λ	-0.419***			-0.513***		
	(0.064)			(0.0627)		
N	1,674			1,674		
Number of hospitals	186			186		

Note: Dependent variable is the log of hospital efficiency. ML estimations were also run and are comparable. Direct, indirect, and spillover effects and related standard errors in parentheses computed using 2000 draws. *** p<0.01, ** p<0.05, * p<0.1.
Source: The authors.

The results confirm the existence of positive spatial dependence among hospital efficiency in the sample. These results are robust for both types of spatial matrices. Considering the weight matrix based on the shortest travel time distance,[23] the estimate of ρ indicates that a 1% increase in the efficiency of neighboring hospitals j is increasing the efficiency of the hospital i by 0.45%. Referring to our efficiency measurement setting, the results suggest significant strategic complementarity effects in hospitals' efficiency. These results contrast with those found in Longo et al. (2017), in which they use different efficiency ratios to proxy efficiency.

The statistical significance of the estimates for λ suggests the presence of a negative spatial error correlation. This result involves the existence of other sources of spatial correlation in our sample that were not properly captured in the model. The results are in line with previous findings in the literature. The existence of spatial error correlation is not new in spatial health econometrics (Baltagi et al., 2018). There are several risk factors

[23]Henceforth, this will be used for interpretation, as it is a more realistic matrix of hospital interactions than that of Euclidean distances ($\boldsymbol{W}_d$).

that are difficult to measure but they are so geographically concentrated that they affect health outcomes (Tosetti et al., 2018). These factors may not be necessarily linked to interactions among hospitals, but could rather be associated with interactions among spatial units observed at a different scale. For instance, Martini et al. (2014) discuss the importance of ward-level analysis in measuring efficiency rather than hospital aggregation, as similar behavior can occur among wards that provide homogeneous treatments. The spatial interaction in hospital efficiency can also originate in a more in depth disaggregation. For example, hospital efficiency can be affected by the physicians' productivity (Johannessen et al., 2017): the concentration of these physicians in large hospitals, mostly located in developed cantons, can generate interactions among them, giving rise to a spatial pattern that cannot be captured by the data. Conversely, the sources of spatial dependence can also come from macroeconomic phenomena such as immigration or unemployment which can cause inefficiency in the provision of healthcare services (Herwartz and Schley, 2018), and are very likely to be influencing hospital performance in Ecuador given its strong spatial dependence (Mendieta Muñoz and Pontarollo, 2016; Szeles and Mendieta Muñoz, 2016).

Another potential source of spatial correlation in errors could come from the omission of budgetary information, which has proved to be a relevant factor of influence in hospitals' efficiency and quality, especially when there are financial pressures due to budget constraints (Herr, 2008; Mas, 2015). In this respect, it is worth pointing out an important limitation of our dataset; this is the impossibility of retrieving the quality of hospitals' budgetary information or public investment to properly match our dataset.

Due to the scarce literature that exploits a similar approach, especially for Latin American countries, a comparative analysis becomes difficult. Nevertheless, the sign of the spatial correlation and the effect of both the spatially lagged efficiency score and the error term go in line with those of Felder and Tauchmann (2013). Although they perform a cross-sectional analysis at the district level in Germany, the average effect of spatial dependence for the hospitals' efficiency –measured by efficiency measurement nonparametric models – does not seem to be unrealistic in the Ecuadorian context.

Table 3.3 provides additional information about spillover effects. We present total effects disaggregated in direct and indirect (spillover) effects (LeSage and Pace, 2009). The logarithm of occupancy rate shows that an increase in 1% in a hospital's occupancy rate increases the efficiency of the same hospital by 0.13% and the efficiency of all neighboring hospitals by 0.09%. These findings would reject the hypothesis that higher demand for medical services (translating into higher occupancy rates) is the source of the decrease in hospitals' efficiency, and instead shows the opposite.[24] This finding is in line with the argument of the inefficient use of the spare resources in the public healthcare system as argued by Herwartz and Strumann (2014).

[24]We tested the direction of the causality between hospital efficiency and the demand by means of the Granger (1969) causality test for panel data models adapted by Dumitrescu and Hurlin (2012). The test rejects the null hypothesis of non-causality.

Instead, market share is associated with a negative estimated coefficient.[25] Its direct and indirect effects show that a 1% increase in this variable diminishes the efficiency performance by 0.03% for the selected hospital and 0.02% for neighboring hospitals. This implies that hospitals that host more patients tend to experience an inefficient use of resources. However, the magnitude of this effect could be different in accordance with the type of hospital under consideration.

In addition, it is interesting to review the negative effect of cantonal density, which means that hospitals located in denser areas tend to record lower performance. However, as we have previously mentioned, the higher level of efficiency in less populated cantons does not necessarily mean that these hospitals are outperforming those in denser territories, as it might be the result of patient migration outflow to the former.[26] Furthermore, the non-significative effect of cantonal inpatient migration might not necessarily mean that it has no effect on hospital efficiency and instead show that it is failing to capture the true effect of patient migration.[27]

The negative effect of hospital mortality provides evidence that a high performance rate is positively correlated with low mortality, which has been a common finding in recent literature (Ferreira and Marques, 2019; Herwartz and Strumann, 2012, 2014).

However, the previous results do not accommodate technological heterogeneities among hospitals. We go a step further than the applied literature by including technological differences as interactions with hospital-related variables, since these are the ones that tend to be relevant for the analysis.

Table 3.4 presents the estimated results including technological interactions. Model (1) presents the baseline model using $\boldsymbol{W}_v$. Models (2), (3), and (4) show the estimation results with the covariates at the hospital level interacted with two dummies of cluster 2 (intermediate-tech) and cluster 3 (high-tech).[28] For the sake of simplicity, we exclude from the table the morbidity estimations' parameters.[29]

[25] For the definition of market share, refer to the Appendix 3.9.1.
[26] We provide more evidence for this when we consider technological effects.
[27] Estimation results with the other weight matrix are comparable and available upon request.
[28] Estimation results with the other weight matrix are comparable and available upon request.
[29] Complete results tables are available upon request.

Table 3.4: Spatial panel regression results, including technological interactions.

Variables	(1)			(2)			(3)			(4)		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
log occupancy rate	0.130***	0.098***	0.228***	0.127***	0.078***	0.205***	0.132***	0.074***	0.206***	0.129***	0.073***	0.202***
	(0.019)	(0.0219)	(0.035)	(0.022)	(0.022)	(0.039)	(0.022)	(0.021)	(0.039)	(0.022)	(0.019)	(0.037)
log market share	-0.034***	-0.026***	-0.060***	-0.037***	-0.023***	-0.061***	-0.060***	-0.034***	-0.094***	-0.059***	-0.034***	-0.093***
	(0.011)	(0.009)	(0.019)	(0.011)	(0.008)	(0.018)	(0.012)	(0.010)	(0.020)	(0.012)	(0.011)	(0.020)
log mortality rate	-0.029***	-0.022***	-0.051***	-0.031***	-0.019***	-0.050***	-0.029***	-0.016***	-0.045***	-0.022**	-0.012**	-0.034**
	(0.008)	(0.007)	(0.015)	(0.008)	(0.007)	(0.0145)	(0.008)	(0.006)	(0.014)	(0.009)	(0.006)	(0.015)
log GVA	0.065*	0.049*	0.114*	0.092**	0.057**	0.148**	0.113***	0.063**	0.175***	0.109***	0.062**	0.171***
	(0.036)	(0.028)	(0.063)	(0.036)	(0.026)	(0.060)	(0.039)	(0.026)	(0.062)	(0.038)	(0.026)	(0.062)
log density	-0.663***	-0.499***	-1.162***	-0.640***	-0.392**	-1.032***	-0.730***	-0.400***	-1.130***	-0.714***	-0.395***	-1.109***
	(0.132)	(0.114)	(0.224)	(0.240)	(0.161)	(0.385)	(0.242)	(0.142)	(0.362)	(0.240)	(0.145)	(0.366)
log mortality (cantonal)	0.052	0.039	0.091	0.074*	0.045*	0.119*	0.071*	0.039	0.110*	0.072*	0.041	0.114*
	(0.038)	(0.030)	(0.067)	(0.041)	(0.026)	(0.066)	(0.041)	(0.026)	(0.065)	(0.039)	(0.026)	(0.064)
log pop > 65	-0.126	-0.098	-0.224	-0.052	-0.034	-0.086	-0.053	-0.032	-0.085	-0.041	-0.025	-0.065
	(0.115)	(0.093)	(0.207)	(0.122)	(0.079)	(0.200)	(0.122)	(0.071)	(0.192)	(0.114)	(0.067)	(0.179)
log inpatient migration	0.005	0.003	0.008	0.003	0.002	0.004	0.005	0.003	0.007	0.005	0.003	0.009
	(0.012)	(0.009)	(0.021)	(0.012)	(0.008)	(0.012)	(0.012)	(0.007)	(0.019)	(0.012)	(0.007)	(0.019)
log occupancy rate*cluster 2				0.031	0.019	0.050	0.008	0.004	0.0128	0.013	0.007	0.021
				(0.037)	(0.024)	(0.060)	(0.038)	(0.022)	(0.059)	(0.038)	(0.022)	(0.059)
log occupancy rate*cluster 3				-0.002	-0.002	-0.004	-0.048	-0.025	-0.072	-0.034	-0.019	-0.053
				(0.169)	(0.112)	(0.279)	(0.169)	(0.098)	(0.265)	(0.169)	(0.099)	(0.267)
log market share*cluster 2							0.106***	0.059***	0.165***	0.100***	0.056***	0.156***
							(0.028)	(0.019)	(0.044)	(0.028)	(0.019)	(0.045)
log market share*cluster 3							0.284***	0.159***	0.443***	0.250***	0.142**	0.392**
							(0.088)	(0.061)	(0.142)	(0.094)	(0.065)	(0.153)
log mortality rate*cluster 2										-0.047*	-0.027	-0.074*
										(0.028)	(0.016)	(0.043)

Table 3.4 (continued)

Variables	(1)			(2)			(3)			(4)		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
log mortality rate*cluster 3										-0.064	-0.035	-0.099
										(0.059)	(0.035)	(0.094)
ρ	0.453***			0.397***			0.373***			0.375***		
	(0.045)			(0.053)			(0.058)			(0.058)		
λ	-0.513***			-0.486***			-0.447***			-0.451***		
	(0.062)			(0.068)			(0.074)			(0.074)		
N	1,674	1,674	1,674	1,674	1,674	1,674	1,674	1,674	1,674	1,674	1,674	1,674
Number of hospitals	186	186	186	186	186	186	186	186	186	186	186	186

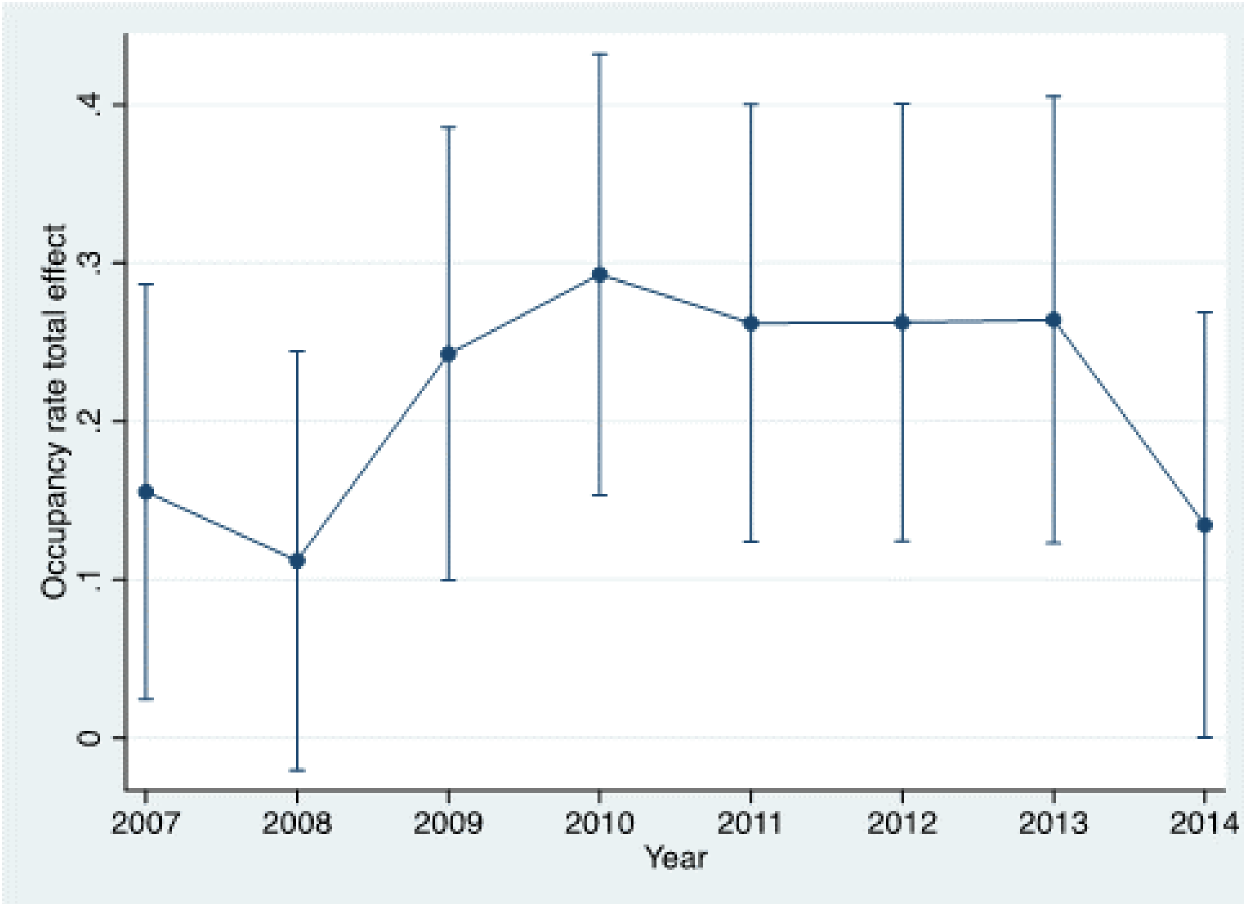
Note: Dependent variable is the log of hospital efficiency. ML estimations were also run and are comparable. Direct, indirect, and spillover effects and related standard errors in parentheses computed using 2000 draws. *** p<0.01, ** p<0.05, * p<0.1.
Source: The authors.

The most interesting finding refers to the market share. The estimated coefficient is significant and robust, positively associated with the technological endowment of public hospitals: the estimates are positive for high-and-intermediate-tech hospitals, something which is at odds with previous results. Indeed, the estimations provide evidence that, in case of more concentration, high-and-intermediate-tech hospitals' efficiency performance increases, enforcing spillover effects. These results are not far from recent findings in the literature. Pross et al. (2018) find that regional and hospital level concentration can improve quality and resource efficiency. Gobillon and Milcent (2013) identify that the higher local concentration of patients in a few large hospitals rather than many small ones improves the hospitals' performance. As these authors state, this can be the result of a learning-by-doing process. The hospitals with the best technology (better equipment, more specialized physicians, better infrastructure, etc.), having treated more patients and more severe cases over time, experience improvements in their treatment capacity through experience. These results might evidence policy recommendations for public investment in favor of hospital competition (which usually seek higher quality and efficiency of the healthcare system) that is well targeted in order to avoid a negative impact. The concentration of resources in developed areas (where most of the high-tech hospitals are located) can be beneficial for the hospital performance in those areas. It is desirable that public investment target less-developed areas where the low-tech hospitals concentrate without having many close competitors. Increasing the number of hospitals in less-developed areas would cause hospitals to compete by increasing their quality and performance in order to avoid patient outflow. Such a measure could also attract skilled and specialized physicians to these regions given the increased demand for qualified personnel . As a consequence, more patients could choose to receive treatment there if they perceive that these hospitals are increasing in quality and efficiency (Ippoliti and Falavigna, 2012), and thus the regional performance of the health sector would be enhanced.

Our estimations also stress that there are no significant changes in occupancy rate and mortality rate when referring to the technological endowment. Regardless of the technological differences, higher demand translates into higher efficiency. The reason for this might be that all hospitals, regardless of their technological level, show low levels of efficiency, implying an inefficient use of their spare inputs, which gives them room for improvement when there is a higher demand for medical treatment.

To find out how the occupancy rate has influenced the efficiency of public hospitals, we draw in Figure 3.2 the tendency of the total effect of $ocrate_t$ over time. There is a notable cut in the total effect after 2008, suggesting that the increase in demand after this year yields an increase in hospital performance due to a more efficient use of spare resources. This effect might also be the result of proper managerial planning that could have anticipated an increase in the bulk of patients, given that the Ecuadorian population had time to become informed about the potential changes that the constitution embraced.

Figure 3.2: Occupancy rate's total marginal effect with 95% CI



To verify whether this discontinuity in efficiency was statistically significant, Table 3.5 presents the correspondent hypotheses tests for both direct and indirect effects. The test rejects both hypotheses with 95% of confidence. This result implies that the period after the adoption of the new constitution enforced not only a significant upturn in the direct effect that an increase in demand generated in a specific hospital, but bigger spillover effects for neighboring hospitals as well.

The results presented so far highlight the importance that covariates (mainly higher demand and more competition) can bring to the efficiency performance of the public healthcare system, and the potential effect that policy implementation can have on it when it is well planned at the territorial level. As has been proved, these policies do not bring benefits exclusively for the selected hospital, but also affect neighboring hospitals due to spillover. Nevertheless, it is worth pointing out that there are still some explanatory variables that worsen the performance of the system. Some of these are still unknown and more research must be done in this direction.

Finally, Table 3.6 presents the different neighboring order coefficient estimates of the partitioning analysis. The direct partitioning effect in Table 3.6 shows a significant im-

Table 3.5: Occupancy rate effects and hypotheses tests

	Effect		
Occupancy rate	Direct	Indirect	Total
2007	0.114**	0.0147**	0.129**
	(0.0477)	(0.0064)	(0.0538)
2008	0.0662	0.00847	0.0747
	(0.0493)	(0.0064)	(0.0555)
2009	0.161***	0.0207***	0.181***
	(0.0534)	(0.0072)	(0.0600)
2010	0.187***	0.0241***	0.211***
	(0.0486)	(0.0066)	(0.0545)
2011	0.160***	0.0206***	0.180***
	(0.0487)	(0.0066)	(0.0549)
2012	0.175***	0.0226***	0.197***
	(0.0510)	(0.0071)	(0.0576)
2013	0.180***	0.0233***	0.203***
	(0.0512)	(0.0072)	(0.0578)
2014	0.0751	0.00962	0.0847
	(0.0479)	(0.0062)	(0.0539)
Test statistics			
$H_o : ocrate_1 = ocrate_2$	6.43**	6.02**	6.45**

Note: Dependent variable is the log of hospital efficiency. ML estimations were also run and are comparable. Direct, indirect, and spillover effects and related standard errors in parentheses computed using 2000 draws. *** p<0.01, ** p<0.05, * p<0.1.
Source: The authors.

pact beyond the so-called zero-order neighbor (W^0, see Appendix 3.9.3) that decreases significantly in size from W^1 on.[30] This implies that for direct impacts, those immediate neighbors play a strong role.[31]

Table 3.6: Spatial partitioning results of direct, indirect, and total effects of hospital demand

	Direct	
	log occupancy rate ($\boldsymbol{W}_d$)	log occupancy rate ($\boldsymbol{W}_v$)
W^1	0.12799***	0.12571***
	(6.022)	(5.824)
W^2	0	0
W^3	0.00272***	0.00265***
	(2.909)	(3.358)
W^4	0.00021**	0.00034**
	(2.056)	(2.423)
W^5	0.00018	0.00023*
	(1.566)	(1.858)
W^6	0.00002	0.00005
	(1.251)	(1.388)
W^7	0.00001	0.00003
	(1.030)	(1.129)
	Indirect	
W^1	0	0
W^2	0.04151***	0.04744***
	(4.562)	(4.908)
W^3	0.01074***	0.01525***
	(2.909)	(3.358)
W^4	0.00415**	0.00641**
	(2.056)	(2.423)
W^5	0.00123	0.00231*
	(1.566)	(1.858)
W^6	0.000004	0.00091
	(1.251)	(1.388)
W^7	0.000002	0.00033
	(1.030)	(1.129)
	Total	
W^1	0.12799***	0.12571***
	(6.022)	(5.824)

[30]The reader will appreciate that the coefficients for W^2 and for W^1 in Table 3.6 are zero for the direct and indirect partitioning effects, respectively. This is because the first term of the series expansion in (3.12) (see Appendix 3.9.3) contains zeros on the off-diagonal. Consequently, W^2 will always be equal to zero for the direct effect. Conversely, given that the spatial weight matrix $\boldsymbol{W}$ contains zero on the main diagonal, by definition, W^1 will always be zero for the indirect effect (Jensen and Lacombe, 2012).

[31]The impact of the marginal change of demand of hospital i on its own efficiency is the result of local effects plus feedback effects that pass mainly through its direct neighbor j.

Table 3.6 (continued)

	log occupancy rate ($\boldsymbol{W}_d$)	log occupancy rate ($\boldsymbol{W}_v$)
W^2	0.04151***	0.04744***
	(4.562)	(4.908)
W^3	0.01346***	0.01791***
	(2.909)	(3.358)
W^4	0.00436**	0.00675**
	(2.056)	(2.423)
W^5	0.00141	0.00255*
	(1.566)	(1.858)
W^6	0.00002	0.00096
	(1.251)	(1.388)
W^7	0.00001	0.00036
	(1.030)	(1.129)

Note: Dependent variable is log of hospital efficiency. Z-values in parenthesis computed using 2000 draws for the direct, indirect, and total effects. *** p<0.01, ** p<0.05, * p<0.1.
Source: The authors.

Regarding the indirect partitioning effects, these are significant for the second, third and fourth-order neighbors (and significant at 90% of confidence for the fifth-order neighbor, considering $\boldsymbol{W}_v$) and strongly decreasing in size after W^3. This effect suggests that, although significant, demand has a limited effect over space for hospital efficiency, with spillover effects being strong in small, concentrated areas and generating small feedback effects.

The aforementioned results provide a useful tool for policy decisions. We have demonstrated here not just the existence of positive spillover effects of demand on hospital efficiency, but that these spillover effects spread to a limited extent in small, concentrated areas. Policy reforms that enhance hospital demand will have positive effect on efficiency performance, but this will spread through spillover effects to a limited extent due to the concentrated spatial territory of the country only.

3.7.1 Robustness checks

To test for the robustness of the results of previous estimations, we run the same estimations for the model (4) by applying the Generalized Method of Moments (GMM) models for endogeneity to control for heteroscedasticity. Although this method displays some advantages over ML methods (Tosetti et al., 2018), GMM have been little exploited in spatial health economics, and for this reason their application has been recently encouraged (Baltagi et al., 2018). We also examine whether the results are sensitive when we consider the remoteness between hospitals by introducing the inverse of the squared distance to define the weight matrix $\boldsymbol{W}_{d^2}$ and time travel distance for the weight matrix $\boldsymbol{W}_{v^2}$, so those hospitals that are quite far apart are weighted less.

Table 3.7 presents GMM estimations as well as the results with the new squared matrices. The estimation is based on the Kelejian and Prucha (1999) model that was first extended to the panel case by Druska and Horrace (2004) and later by Kapoor et al. (2007) for the case of the random effects. The estimation in a fixed effects framework was later adapted by Mutl and Pfaffermayr (2011). One drawback of this method is that it does not provide an estimate of the dispersion of λ, so no significance test is possible (Croissant and Millo, 2018).

Table 3.7: Spatial regression results. GMM estimators.

Variables	$\boldsymbol{W}_d$			$\boldsymbol{W}_{d^2}$			$\boldsymbol{W}_v$			$\boldsymbol{W}_{v^2}$		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
log occupancy rate	0.134***	0.07***	0.204***	0.134***	0.065**	0.199***	0.135***	0.022*	0.157***	0.135***	0.019	0.154***
	(0.024)	(0.027)	(0.043)	(0.024)	(0.032)	(0.047)	(0.024)	(0.013)	(0.031)	(0.024)	(0.016)	(0.032)
log market share	-0.061***	-0.032**	-0.093***	-0.064***	-0.031**	-0.095***	-0.062***	-0.010*	-0.072***	-0.063***	-0.009	-0.073***
	(0.014)	(0.014)	(0.024)	(0.014)	(0.017)	(0.026)	(0.013)	(0.006)	(0.016)	(0.014)	(0.007)	(0.018)
log mortality rate	-0.024**	-0.013*	-0.037**	-0.024**	-0.012	-0.036**	-0.024**	-0.004	-0.027**	-0.024**	-0.003	-0.027**
	(0.011)	(0.008)	(0.017)	(0.011)	(0.009)	(0.019)	(0.011)	(0.003)	(0.013)	(0.011)	(0.003)	(0.013)
log GVA	0.095**	0.049*	0.145**	0.102**	0.049*	0.151**	0.102**	0.017	0.1183**	0.105**	0.015	0.121**
	(0.044)	(0.029)	(0.069)	(0.045)	(0.034)	(0.074)	(0.045)	(0.012)	(0.052)	(0.045)	(0.014)	(0.053)
log density	-0.565*	-0.296*	-0.861*	-0.638**	-0.309*	-0.947**	-0.749**	-0.122	-0.872***	-0.775***	-0.113	-0.888***
	(0.296)	(0.167)	(0.439)	(0.297)	(0.190)	(0.453)	(0.298)	(0.078)	(0.336)	(0.299)	(0.089)	(0.342)
log mortality (cantonal)	0.075	0.039	0.114	0.078*	0.038	0.116	0.072	0.012	0.083	0.074	0.011	0.085
	(0.047)	(0.031)	(0.076)	(0.047)	(0.034)	(0.078)	(0.047)	(0.011)	(0.056)	(0.047)	(0.013)	(0.057)
log pop > 65	-0.051	-0.026	-0.077	-0.039	-0.019	-0.058	-0.001	-0.0002	-0.001	-0.0012	-0.0002	-0.002
	(0.146)	(0.087)	(0.231)	(0.147)	(0.095)	(0.238)	(0.147)	(0.029)	(0.173)	(0.148)	(0.032)	(0.176)
log inpatient migration	0.004	0.002	0.006	0.003	0.002	0.005	0.006	0.0009	0.007	0.005	0.0007	0.006
	(0.014)	(0.008)	(0.022)	(0.014)	(0.009)	(0.022)	(0.014)	(0.003)	(0.016)	(0.014)	(0.003)	(0.017)
log occupancy rate*cluster 2	0.017	0.009	0.026	0.015	0.007	0.022	0.014	0.002	0.017	0.014	0.002	0.016
	(0.043)	(0.025)	(0.068)	(0.043)	(0.027)	(0.0695)	(0.043)	(0.009)	(0.051)	(0.043)	(0.009)	(0.052)
log occupancy rate*cluster 3	-0.079	-0.041	-0.119	-0.086	-0.042	-0.128	-0.079	-0.013	-0.092	-0.087	-0.013	-0.1
	(0.179)	(0.105)	(0.280)	(0.178)	(0.114)	(0.287)	(0.179)	(0.036)	(0.211)	(0.178)	(0.039)	(0.213)
log market share*cluster 2	0.121***	0.063**	0.184***	0.125***	0.060**	0.185***	0.124***	0.020	0.144***	0.125***	0.018	0.144***
	(0.032)	(0.029)	(0.057)	(0.032)	(0.036)	(0.062)	(0.031)	(0.013)	(0.039)	(0.031)	(0.015)	(0.039)
log market share*cluster 3	0.264**	0.138*	0.402**	0.257**	0.125*	0.382**	0.247**	0.0402	0.287**	0.247**	0.0362	0.284**
	(0.106)	(0.080)	(0.176)	(0.107)	(0.090)	(0.184)	(0.104)	(0.031)	(0.126)	(0.106)	(0.036)	(0.130)
log mortality rate*cluster 2	-0.042	-0.022	-0.064	-0.042	-0.020	-0.062	-0.043	-0.007	-0.050	-0.044	-0.006	-0.0502
	(0.031)	(0.019)	(0.048)	(0.031)	(0.021)	(0.050)	(0.031)	(0.007)	(0.036)	(0.031)	(0.008)	(0.037)

Table 3.7 (continued)

Variables	$\boldsymbol{W}_d$			$\boldsymbol{W}_{d^2}$			$\boldsymbol{W}_v$			$\boldsymbol{W}_{v^2}$		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
log mortality rate*cluster 3	-0.084	-0.044	-0.128	-0.081	-0.039	-0.121	-0.083	-0.013	-0.096	-0.084	-0.012	-0.096
	(0.064)	(0.042)	(0.103)	(0.063)	(0.045)	(0.104)	(0.065)	(0.016)	(0.079)	(0.065)	(0.018)	(0.079)
ρ	0.364***			0.339***			0.146**			0.132*		
	(0.079)			(0.089)			(0.072)			(0.081)		
λ	-0.18			-0.144			-0.078			-0.057		
N	1,674			1,677			1,674			1,674		
Number of hospitals	186			189			186			186		

Note: Dependent variable is log of hospital efficiency. Direct, indirect, and spillover effects and related standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Source: The authors.

Looking at the ρ coefficients of the first two models, the estimations show no substantial difference with respect to the previous estimates, even if we cannot determine whether the coefficients λ are statistically significant. Regarding the effects of the covariates, the results are robust, and the size of the estimation is comparable. Occupancy rates are still significant regardless of the technological disparities, while the opposite occurs for the market share. Although, the spillover effects of market share for intermediate and high-tech hospitals display no statistical significance taking into account the squared weight matrices. It may show that the indirect effect issued from the concentration of hospitals has an impact on the closest neighbors only, and this effect decays proportionally with their distance. The findings support the partitioning analysis carried out in section 3.7, demonstrating weaker spillover effects of demand in efficiency after the second-order neighbor.

In addition, we identify a weaker effect for ρ. Despite the magnitude of the estimates, once more, spatial dependence in efficiency is confirmed.

Furthermore, we test the robustness of the results of the competition detected among hospitals. There is the possibility that the heterogeneity that we recognize in terms of technology may also be visible in terms of the spectrum of diseases treated. Therefore, the spatial dependence found might not be due to competition for a greater demand for patients, but due to the existence of specialized hospitals versus other general hospitals that include more treatments, which do not compete. To test this statement, we run the equation (2.6), form Chapter 2, on three different subgroups provided by our dataset: acute, chronic, and basic hospitals. Hence, we analyze homogeneous hospitals in terms of functioning and treatment.

Table 3.8 shows the estimations of the baseline model on the three subgroups using the weight matrix based on the shortest travel distance.[32] For basic hospitals the results are robust and comparable with the previous ones, supporting the existence of spatial strategic interactions in hospital efficiency. Instead, acute, and chronic hospitals do not display spatial dependence in hospital efficiency, although chronic hospitals present a significant spatial dependence in the error term, comparable to previous results. This could suggest that basic hospitals (which constitute more than 50% of the sample) are those that compete in terms of efficiency with their neighbors. However, the estimations should be interpreted with caution given the loss of information in the regressions when we split the sample. Future research will focus on expanding the time span to reach conclusive results.

Finally, it is interesting to remark that inpatient migration turns out to be significant for chronic hospitals, especially when these are mainly oncological. The variable suggests a negative relationship with the efficiency performance of chronic hospitals and sets new insights for future research on hospital patient migration dynamics.

[32] As we consider three different subgroups of hospitals, the spatial weight matrix was calculated for each hospital type. Also, these regressions were run by QML to be comparable with the main results.

Table 3.8: Spatial regression results. Direct, indirect, and total effects by hospital type.

Variables		Acute			Chronic			Basic	
Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total	
log occupancy rate	0.104***	0.003	0.107***	0.456***	0.028	0.483***	0.121***	0.066***	0.187***
	(0.028)	(0.015)	(0.034)	(0.059)	(0.048)	(0.060)	(0.026)	(0.019)	(0.041)
log market share	0.082***	0.002	0.084***	-0.084***	-0.005	-0.088***	-0.151***	-0.083***	-0.233***
	(0.017)	(0.012)	(0.021)	(0.031)	(0.009)	(0.033)	(0.014)	(0.019)	(0.027)
log mortality rate	-0.041***	-0.001	-0.042***	-0.129***	-0.008	-0.137***	-0.014	-0.008	-0.021
	(0.014)	(0.006)	(0.016)	(0.024)	(0.014)	(0.028)	(0.010)	(0.006)	(0.016)
log GVA	-0.042	-0.001	-0.043	-0.088	-0.002	-0.09	0.069*	0.038	0.107*
	(0.076)	(0.013)	(0.078)	(0.249)	(0.031)	(0.264)	(0.042)	(0.024)	(0.065)
log density	-2.202***	-0.016	-2.217***	4.264**	0.231	4.495**	-0.532**	-0.286*	-0.819*
	(0.572)	(0.305)	(0.525)	(2.159)	(0.497)	(2.248)	(0.272)	(0.160)	(0.419)
log mortality (cantonal)	0.063	-0.0002	0.063	-0.006	-0.002	-0.009	-0.006	-0.003	-0.0098
	(0.097)	(0.017)	(0.099)	(0.367)	(0.047)	(0.392)	(0.042)	(0.024)	(0.066)
log pop > 65	0.21	-0.006	0.203	1.078	0.079	1.157	0.028	0.013	0.041
	(0.223)	(0.043)	(0.222)	(0.746)	(0.159)	(0.817)	(0.199)	(0.110)	(0.308)
log inpatient migration	0.019	-2.91E-05	0.0196	-0.184**	-0.01	-0.194**	0.002	0.001	0.0038
	(0.028)	(0.005)	(0.029)	(0.081)	(0.021)	(0.085)	(0.012)	(0.007)	(0.019)
ρ	0.014			0.06			0.378***		
	(0.133)			(0.094)			(0.057)		
λ	-0.085			-0.655***			-0.515***		
	(0.147)			(0.089)			(0.072)		
N	711			81			882		
Number of hospitals	79			9			98		

Note: Dependent variable is log of hospital efficiency. Direct, indirect, and spillover effects and related standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Source: The authors.

3.8 Conclusions

This study aims to analyze the spatial dependence of hospital efficiency in Ecuador. To address this question, we apply an innovative methodology proposed in Chapter 2 to obtain robust efficiency scores for a sample of public hospitals in Ecuador between 2006 and 2014, taking into account their technological differences to avoid biased results. We then use this efficiency score as a dependent variable in a spatial econometric SAC model to consider spatial autocorrelation in efficiency and disturbances. The results confirm that an increase in the efficiency of surrounding hospitals increases the efficiency of a selected hospital. The direction of these effects is robust to different specifications and estimation methods. Spatial autocorrelation and spillover effects seem to diminish as the hospitals' distance from the most developed areas increases. As Longo et al. (2017) state, the positive dependence between neighboring hospitals suggests that they are acting as strategic complements in efficiency.

We also address the question of whether the variations in demand for a given hospital, which we measure through occupancy rates, affect nearby hospitals' efficiency through spillover effects.The results confirm that increases in demand for medical services for public hospitals are causing neighboring hospitals to attract some of this demand, and that this is boosting their own efficiency, regardless of the technological endowment of the hospitals.[33] A large portion of this positive effect can be explained because the public healthcare hospitals show low levels of occupancy rates, which might imply the existence of spare resources that are inefficiently used to produce healthcare outputs. The increase in demand forces hospitals to make better use of these spare resources and so their efficiency performance rises. In addition, the estimates assess that after 2008, the direct and indirect impact of occupancy rates on the efficiency performance significantly increased. While waiting for the approval of the new constitution, which was expected to entail an increase in the number of patients seeking medical treatments, hospital managers could have planned strategies to adapt to these changes, and this could in part be reflected in this higher effect after 2008.

The technological disparities among hospitals also play a key role, especially when analyzed jointly with market share. We find evidence that high- and intermediate-tech hospitals have a differential effect. That is, the increase of concentration of patients in technologically better hospitals increases their efficiency and that of surrounding hospitals, whereas the opposite effect is found for low-tech hospitals. These results provide some evidence of a potential learning-by-doing process in high- and intermediate-tech hospitals.

These differences have important policy implications. Taking into consideration that high-tech hospitals are mostly concentrated in well-developed areas, policy decisions and

[33] Focusing on the types of hospitals, the results hold for basic hospitals, while acute and chronic hospitals do not show spatial dependence in efficiency. However, these results should be taken with caution due to the information loss when splitting the sample.

public funding should be allocated taking into consideration the territorial development within the country. The rationale is that policy reforms and public investment that imply more competition (by investing in the construction of more hospitals) can be counterproductive for the healthcare performance of well-developed areas but beneficial for less-developed ones.

In this line, policymakers could exploit spillover effects in well-developed areas to reinforce the hospital performance. However, they should be aware that these spillover effects will spread to a limited extent over space, emphasizing the importance of well targeted policy decisions. Clearer criteria for public funding allocation and stronger regulation of hospital resource consumption controlling (or limiting) for hospital costs' inflation can have a positive impact in these regions. With more control to prevent costs' inflation, hospitals would have incentives to increase their profits by improving their resource use and thus increase their efficiency. Due to spillover effects, this efficiency improvement would spread throughout the region, enhancing the performance of the public healthcare system without increasing the allocation of resources or public investment. Instead, public investment and resource allocation could focus on less-developed areas, where a higher supply of hospitals could motivate existing hospitals to compete for patient inflow by increasing their quality and efficiency. These improvements can be a potential solution that could reduce the existing regional gap in the Ecuadorian healthcare system.

The empirical application carried out in this study can also be extended to other Latin American countries that share many socio-economic, political, and cultural characteristics (Atun et al., 2015; Levy and Schady, 2013) and whose spatial disparities have been well documented (Cuadrado-Roura and Aroca, 2013).

However, this study leaves open some questions for future research. In a country with an important heterogeneity in the healthcare system, it would be useful to understand whether internal patient migration flows affect or are affected by the hospitals' performance. High-performing hospitals might be attracting patients from low-performing ones in neighboring regions. Understanding the mobilization patterns of patients is crucial to improving the healthcare system. Understanding interregional patient migration patterns can help central and local authorities as well as hospitals themselves to identify under-performing hospitals, which could benefit from an increase in healthcare budget and resource allocation in order to improve their performance and attract more demand. Also, patient mobility flows are likely to follow a spatial pattern, as patients will be willing to travel to the nearest high-performing hospital. In this sense, policymakers can identify spatial clusters of hospitals and promote policies that encourage efficiency gain.

Further methodological innovations can also be implemented. In this analysis, we assume that hospitals interact with each other within a bounded area, in the presence of local competition. However, hospitals can experience global forms of interactions that might not necessarily depend on their geographical distances but rather on long-range interdependencies (Lisi et al., 2017). By keeping w_{ij} unknown, and estimating it by graphical modeling (Moscone et al., 2017, 2018), future research could test the existence of these interdependencies in the case of developing countries such as Ecuador. Moreover, one-stage SFA panel models that account for hospital heterogeneity and address spatial dependence such as those recently proposed by Pross et al. (2018) can also be implemented to control for possible bias in the efficiency estimations in two-stage approaches (Simar and Wilson, 2007).[34] However, the main setback of SFA approximations is that they rely on a production function that has to be defined *a priori* (O'Neill et al., 2008), and that cannot be simply proposed in the context of a developing country. Future work defining the theoretical framework of a proper production function is therefore desirable to provide the background for empirical applications.

Finally, we need to point out some issues referring to data availability. It is recommended that future research take into account further information that has been proven to have a significant effect on hospitals' efficiency, such as the quality of treatments and budgetary information. Here, we proxy hospital quality with mortality variables, which have been widely used to approximate hospital quality and performance (Hafidz et al., 2018; Lisi et al., 2017; O'Neill et al., 2008). However, mortality can be influenced by other external factors, such as the severity of the disease that patients suffer when they enter the hospital or other complications that the hospitals cannot control for, and that does not reflect the quality of the treatment received. The same type of consideration applies to the readmission rates, the level of specialization (Gravelle et al., 2014; Longo et al., 2017) or nosocomial infections (Prior, 2006) that could bring more elements for better understanding the public healthcare quality-efficiency relationship in the healthcare system.

Another relevant missing piece of information refers to hospital budgets and public investments. Hospitals can adopt a different behavior when they face financial pressures (Mas, 2015). Such troubles are quite common in developing economies such as Ecuador, where hospitals might be forced to make efforts towards cost limitations that could affect their performance. The large public investment made by the government after 2008 is very likely to have an impact on hospital efficiency. It is expected to relax some financial pressures and could have been targeted from a territorial viewpoint and thus affect health outcomes directly. Future research should fill this gap to drive additional empirical research in order to bring relevant insights for policy decisions. In this line, a clear suggestion for policymakers is to implement strong monitoring systems that provide researchers and healthcare managers with reliable and robust data.

[34]The main setback of two-stage approaches relies on the impossibility of knowing the underlying Data Generating Process of DEA efficiency estimates, which raises some doubts over what is estimated in the second stage. In addition, DEA estimates are serially correlated and consequently lead to unreliable inference (Simar and Wilson, 2007). We control for this latter issue by accounting for panel-data robust efficiency estimations that take into consideration the panel structure of the data.

3.9 Appendix

3.9.1 Data description

Table 3.9: Description of the variables

Variable	Description	Variable construction
Output		
Number of discharges (weighted)	Patients treated in a given hospital	Number of discharges*Case-Mix index
Inputs		
Number of physicians	Physicians and general physicians in a given hospital	Total number of physicians
Number of beds	Total number of beds per hospital	Total number of beds
Number of hospital personnel	Medical staff not including physicians. (e.g. Nurses, technologists, administrative staff, dentists, etc.)	Total number of hospital personnel
Number of equipment and infrastructure	Physical infrastructure (surgery rooms, intensive care rooms, etc.) and medical equipment (imaging, diagnosis, sterilization, etc.)	Total number of equipment and infrastructure
Explanatory Variables		
Occupancy rate	Inpatient days of care per beds available in a given hospital	(Inpatient days of care/Bed days available) *100
Market share	Concentration of inpatients in a given hospital relative to the total amount of patients in the canton	(Total number ith hospital inpatients/Total number of cantonal patients)*100
Mortality rate	Percentage of deceased patients in a given hospital	Hospital mortality*100
Number of disease 1	Inpatients with certain infectious and parasitic diseases	Total inpatients with disease/100
Number of disease 2	Inpatients with neoplasms	Total inpatients with disease/100
Number of disease 3	Inpatients with diseases of the nervous system	Total inpatients with disease/100
Number of disease 4	Inpatients with diseases of the respiratory system	Total inpatients with disease/100
Number of disease 5	Inpatients with diseases of the skin and subcutaneous tissue	Total inpatients with disease/100
Number of disease 6	Inpatients with pregnancy, childbirth, and puerperium	Total inpatients with disease/100
Number of disease 7	Inpatients with certain conditions originating in the perinatal period	Total inpatients with disease/100

Table 3.9 (continued)

Variable	Description	Variable construction
Number of disease 8	Inpatients with congenital malformations, deformations and chromosomal abnormalities	Total inpatients with disease/100
Number of disease 9	Inpatients with injury, poisoning, and certain other consequences of external causes	Total inpatients with disease/100
GVA	Gross Value Added	Total, cantonal
Density (population per Km2)	Cantonal population per Km2	Population/km2
Mortality rate (% cantonal)	Percentage of deceased patients in a given canton relative to cantonal population	Cantonal mortality*100
Total population over 65	Cantonal population over 65 years old	Total, cantonal
Total patient migration	Patients treated in a given hospital residing in a different canton from the one they are treated in	Total, cantonal

Source: The authors.

3.9.2 Model specification

The following selection model strategy begins with a baseline model and develops around some tests to achieve an econometric specification that fits the data at hand. First, we present the panel LM and robust-LM tests to provide an initial idea of the potential sources of spatial autocorrelation in Table 3.10. To develop these tests and the following model specification in this appendix, we rely on the matrix of the inverse Euclidean distance $\boldsymbol{W}_d$.

The robust test fails to reject the null hypothesis of no spatial autocorrelation (at 90% and 95% of confidence) in both the dependent variable and the errors. The initial evidence leads to take into consideration both types of spatial autocorrelation.[35]

Table 3.10: LM and robust-LM tests

	Value	Prob
LM-Lag	0.2929	0.5884
LM-Err	1.386	0.2391
Robust LM-Lag	3.2491	0.0715
Robust LM-Err	4.3422	0.0372

Source: The authors.

[35]The reader must take these initial results with caution, given that the classical panel data tests (Anselin et al., 2006) and their robust counterpart (Elhorst, 2010) do not allow for any spatial or time-specific effects. The tests run here are controlled in an *ad hoc* way for individual effects by demeaning the data as in Croissant and Millo (2018).

We then compare the appropriateness of a scope of spatial models taking a fixed effects model as a benchmark in Table 3.11. The Hausman test rejects the null hypothesis of no systematic difference between fixed and random effects, so it is coherent to apply a fixed effect estimation. Following LeSage and Pace (2009) and Elhorst (2010), we explore the most suitable econometric estimation by starting with the general SDM model and then refining it towards a SAR or SEM model. Following the SDM model, we cannot find statistical evidence of spatial dependence in efficiency. SAR and SEM models produce the same results in efficiency and error spatial dependence, respectively. The SAC model, on the other hand, provides significant evidence of spatial dependence both in dependent variable and error term. Merging these outcomes with the results of the LM-tests, the SAC model seems to be the most advisable one to apply to our data. However, the SAC model is not nested within the SDM model (Elhorst, 2014), and so we can rely on alternative information criteria to select between them (Belotti et al., 2016). Akaike and Bayesian information criteria endorse the selection of the SAC model as the best specification. The recent literature in this regard supports this finding and SAC models usually account for spatial dependence in efficiency and potential unmeasurable variables that can affect the hospitals' efficient performance (Felder and Tauchmann, 2013; Herwartz and Strumann, 2012, 2014).

Table 3.11: Model specification

Variables	Panel	SDM	SAR	SEM	SAC
log occupancy rate	0.142***	0.153***	0.142***	0.143***	0.135***
	(0.046)	(0.019)	(0.019)	(0.019)	(0.018)
log market share	-0.037	-0.041***	-0.037***	-0.038***	-0.036***
	(0.038)	(0.011)	(0.011)	(0.011)	(0.011)
log mortality rate	-0.032*	-0.034***	-0.032***	-0.032***	-0.032***
	(0.017)	(0.009)	(0.009)	(0.009)	(0.009)
disease 1	-0.007	-0.008***	-0.007***	-0.007***	-0.007***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
disease 2	-0.005	-0.006**	-0.005*	-0.005*	-0.003
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
disease 3	0.076*	0.077***	0.076***	0.077***	0.075***
	(0.045)	(0.025)	(0.025)	(0.025)	(0.023)
disease 4	0.024***	0.024***	0.024***	0.024***	0.023***
	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
disease 5	-0.023	-0.025	-0.023	-0.023	-0.024
	(0.047)	(0.020)	(0.020)	(0.020)	(0.019)
disease 6	0.007***	0.007***	0.007***	0.007***	0.006***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
disease 7	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
disease 8	-0.032*	-0.033***	-0.032***	-0.032***	-0.029**
	(0.017)	(0.012)	(0.012)	(0.012)	(0.011)
disease 9	0.007*	0.007***	0.007***	0.007***	0.006**

Table 3.11 (continued)

	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)
log GVA	0.093	0.019	0.094**	0.096**	0.087**
	(0.073)	(0.051)	(0.043)	(0.043)	(0.037)
log density	-0.811*	-0.519	-0.823***	-0.836***	-0.606***
	(0.470)	(0.343)	(0.274)	(0.271)	(0.233)
log mortality (cantonal)	0.068	0.043	0.068	0.068	0.068*
	(0.047)	(0.050)	(0.045)	(0.045)	(0.039)
log pop > 65	0.033	0.051	0.036	0.037	-0.012
	(0.173)	(0.169)	(0.139)	(0.138)	(0.115)
log inpatient migration	0.003	-0.006	0.003	0.003	0.003
	(0.016)	(0.014)	(0.013)	(0.013)	(0.011)
Lagged Independent Variables					
log occupancy rate		0.019			
		(0.032)			
log market share		-0.02			
		(0.021)			
log mortality rate		0.005			
		(0.019)			
disease 1		-0.007			
		(0.008)			
disease 2		0.005			
		(0.005)			
disease 3		0.113**			
		(0.055)			
disease 4		-0.009			
		(0.009)			
disease 5		0.009			
		(0.039)			
disease 6		0.001			
		(0.002)			
disease 7		-0.005			
		(0.007)			
disease 8		0.002			
		(0.028)			
disease 9		0.003			
		(0.005)			
log GVA		0.268***			
		(0.096)			
log density		-0.867			
		(0.539)			
log mortality (cantonal)		0.135			
		(0.091)			
log pop > 65		-0.118			
		(0.227)			

Table 3.11 (continued)

log inpatient migration		0.046**			
		(0.023)			
Spatial					
ρ		-0.054	-0.011		0.355***
		(0.035)	(0.033)		(0.053)
λ				-0.026	-0.419***
				(0.034)	(0.064)
Hausman (p-value)	0.0000	0.0000	0.0002	0.0000	
AIC	91.93	85.05	79.82	79.35	66.84
BIC	227.5	280.3	182.9	182.4	175.3
Observations	1,674	1,675	1,674	1,674	1,674
Number of hospitals	186	187	186	186	186

Note: Dependent variable is log of hospital efficiency. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Source: The authors.

3.9.3 Modeling spatial effects

Following LeSage and Pace (2009), if a particular explanatory variable in an observed spatial unit changes (e.g. the change in the demand of hospital i), not only will the dependent variable in that unit itself change (efficiency of hospital i) but also the dependent variable in other units (efficiency of hospital j). The former are called *direct effects* and the latter *indirect (spillover) effects*. In the SAC model, direct effects are the result of local effects plus feedback effects mediated by spatial spillovers.[36]

In particular, taking the matrix of partial derivatives of the expected value of the logarithm of the efficiency $\boldsymbol{e}_t$ with respect to the z_{th} explanatory variable of $\boldsymbol{Z}_t$ in all hospitals (from 1 to 186) for the SAC model, we have:

$$\left[\frac{\partial E(\boldsymbol{e}_t)}{\partial z_{1,t}}, \ldots, \frac{\partial E(\boldsymbol{e}_t)}{\partial z_{186,t}}\right] = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\beta_z \tag{3.10}$$

Where β_z is the vector of coefficients. LeSage and Pace (2009) define the diagonal element of (3.10) as the direct effects, while the off-diagonal contains the indirect effects. The infinite series expansion of the spatial multiplier matrix $(\boldsymbol{I} - \rho \boldsymbol{W})^{-1}$ can be expressed as follows:

$$(\boldsymbol{I} - \rho \boldsymbol{W})^{-1} = \boldsymbol{I} + \rho \boldsymbol{W} + \rho^2 \boldsymbol{W}^2 + \rho^3 \boldsymbol{W}^3 + \cdots + \rho^q \boldsymbol{W}^q \tag{3.11}$$

[36]This feedback effect is derived from the impacts passing through neighboring hospitals and back to the hospital in which the change originated (from hospital i to j to k and back to hospital i).

Note that, since the off-diagonal elements of the identity matrix $\boldsymbol{I}$ are zero, the term represents a direct effect of a change in $\boldsymbol{Z}_t$. Furthermore, since the diagonal elements of $\rho\boldsymbol{W}$ are zero by assumption, the term represents the indirect effect of a change in $\boldsymbol{Z}_t$. The remaining terms in the right-hand side of (3.11) represent the second and higher order direct and spillover effects. Thus, the spatial multiplier, as shown in (3.11) can be expanded to determine the impacts that the explanatory variables have on the higher order of contiguity in the following manner:

$$(\boldsymbol{I}-\rho\boldsymbol{W})^{-1}\beta_z = \underbrace{\boldsymbol{I}\beta_z}_{W^0} + \underbrace{\rho\boldsymbol{W}\beta_z}_{W^1} + \underbrace{\rho^2\boldsymbol{W}^2\beta_z}_{W^2} + \underbrace{\rho^3\boldsymbol{W}^3\beta_z}_{W^3} + \cdots + \underbrace{\rho^q\boldsymbol{W}^q\beta_z}_{W^q} \tag{3.12}$$

The power of the autoregressive parameter, ρ, ensures that the marginal effect of a given variable decreases with a higher order of contiguity. In other words, the effect of a change of an explanatory variable declines as we move over space (LeSage and Pace, 2009).

However, the presentation of both direct and indirect effects can be challenging, since they vary from different units in the sample. Therefore, (LeSage and Pace, 2009) propose to report direct effects as the average of the diagonal elements, while one spillover effect can be measured by the average row sums of the off-diagonal elements. The sum of the average direct and spillover effects is the total effect.

In order to draw inferences regarding the statistical significance of the direct and spillover effects, (LeSage and Pace, 2009) propose simulating the distribution of the direct and spillover effects using the variance-covariance matrix estimated by the ML method. This is because it cannot be simply seen from the coefficient estimates and the corresponding standard errors or t-values of the variance-covariance matrix whether the indirect effects in models containing endogenous interaction effects are significant (Elhorst, 2014; LeSage and Pace, 2009).

Chapter 4

On the dynamics of patient migration flows: Is efficiency performance explaining inflows for neighboring hospitals? An application to the Ecuadorian healthcare system.*

Abstract

This study aims to analyze whether higher efficiency performance of Ecuadorian hospitals attracts larger inflows of interregional patients to a given hospital and the existence of spatial dependence in terms of larger inflows of patients for neighboring hospitals in the region. We develop a novel two-stage approach. In the first stage, we use conditional order-m estimations to obtain robust efficiency values for each hospital. In the second stage, we use a spatial Durbin interaction model to estimate the effect of hospital efficiency on patient migration flows and disentangle the spillover effects in the migration dyad. The results show a positive effect of specialized hospitals' efficiency in attracting patients from other regions. In addition, patient inflows present spillover effects not just on neighboring hospitals in the same region but also from hospitals in regions neighboring the origin. Policy implications mostly drive the attention to the importance to elaborate well planned healthcare strategies taking care of territorial externalities. Negative shocks affecting specialized hospitals could imply an adverse effect on the flow of patients to the whole region, affecting the regional public healthcare performance and

*We want to thank the participants of the seminars at UAB, the University of Brescia, the University of Gothenburg, the WRSA and RSAI international conferences, the scientific committee of the PhD on Applied Economics. To Riccardo Turati, Gabriel Facchini, José Luis Roig, Nicola Pontarollo, Rosella Levaggi, and Francesco Moscone for their comments to improve this chapter. To James LeSage and Christine Thomas-Agnan for facilitating the MATLAB and R codes. Any remaining errors are our own responsibility.

potential welfare gains. Conversely, more resources could be directed to less-developed regions to incentivize competition.

Keywords: hospital efficiency, patient migration, spatial dependence, spillover effects.
JEL: C18, C61, H75, I11, R23.

4.1 Introduction

In healthcare system analysis, patient choice of hospitals and the resulting patient mobility has been a topic that has occupied a vast body of the literature over the past two decades (Balia et al., 2014). Models that allow patient choice of hospitals have a wide spread of useful applications both for governments and the hospitals' own governance (Lowe and Sen, 1996). In this context, Balia et al. (2014) state that the importance to assess patient mobility can be twofold. First, the geography of patient mobility yields indications on the actual level of services provided. This can be particularly useful given that the preferences of the individuals are not perfectly observable. For example, patient outflows might reveal the possible inefficiency or low quality of public healthcare supply in a given region.[1] Second, the flow imbalances across regions may challenge the stability of their healthcare budgets. This kind of information can be useful for central planners and regional authorities interested in correcting inefficiencies in the system as well as improving the healthcare system performance. Understanding the mobility patterns of healthcare consumers may represent an important tool for the central government and regional planners to identify clusters of hospitals and take advantage of spillover effects to better allocate the resources and enhance the efficiency of the system.

Essentially, patients move because they want to get the best hospital treatment that the system can provide, or at least better services than those offered in their local region. They can be expected to move when possible inefficiencies translate into longer waiting lists but also when the perceived quality of the local healthcare services is low (Aiura, 2013; Balia et al., 2014). These movements might be permanent over time if the local regions in a country present a certain level of asymmetry in their systems (Balia et al., 2018).

In this sense, there is a strand running through the literature stating that eliminating barriers of access to healthcare, and thus giving patients the ability to choose between hospitals, creates a financial incentive for providers to compete among them, which leads to improvements in quality of care (Bloom et al., 2015; Gravelle et al., 2014; Propper, 2012).

[1] Throughout this paper, hospital efficiency reflects the ability of a hospital to properly make use of its resources or inputs (e.g., physicians, medical equipment, capacity, etc.) to provide medical attention derived from given outputs (e.g., patients treated, treatments carried out, etc.). In this sense, a fully efficient hospital can maximize its outputs with a given amount of inputs. This is commonly known in the healthcare efficiency measurement literature as *technical efficiency* (Hollingsworth, 2008).

This theory might hold in a country where the healthcare system is rather homogeneous across regions. However, when regional disparities are significant and persistent over time, high-income regions tend to offer a better quality of care. This motivates patients to move from low- to high-income regions seeking better treatment. In turn, the dynamic of such flows, closely relating with the spatial pattern, could be giving rise to network effects often detected in the data and translating into a structure correlation, known in the literature as spatial dependence (Anselin, 2010).

An interesting context of analysis is brought to this setting by Ecuador, whose marked regional disparities offer us a framework of study that can allow us to understand the interregional and intraregional dynamics of patient mobility that can be driven by the performance gaps of their heterogeneous hospitals.

Like other Latin American countries, Ecuador has suffered a continuous process of deterioration of its public healthcare system, which has been exacerbated by the neoliberal reforms of the 1990s and the 2000s, resulting in a widening of the existing territorial disparities in the country. These disparities derived in a concentration of healthcare resources in a few public hospitals (the high-performers), which at the same time were located in developed regions.[2] With the approval of the new constitution in 2008, new healthcare reforms were enhanced to promote free access of medical care and an increase of social security coverage. This gave patients the possibility of choosing the hospital where they wanted to receive treatment.[3] At the same time, this increased the demand for medical attention, promoting a behavior of mobilization to seek treatment in developed regions.

As the barriers of access vanished, patients were expected to seek better treatment in areas where they perceived would get the best possible treatment, leading to patient mobility. Mobility then caused an increase in patient demand, and this can result in two different outcomes. On the one hand, higher demand fuels competition among hospitals in the region, resulting in an increase in quality of care or more efficient use of resources in order to cope with the demand. On the other hand, when demand for hospital treatment increases, hospitals become crowded and additional resources are needed to reduce congestion, entailing eventually inefficiencies like longer waiting times and finally in an underprovision of public services such as healthcare (Aiura, 2013). Moreover, if developed regions are the receivers of a bigger share of patients, one can expect that other adjacent hospitals may receive patients driven by the demand at their neighbors.[4]

So far, the literature on patient mobility has focused on identifying and measuring the

[2] Refer to Appendix 4.10.1 for a description of the Ecuadorian healthcare institutional framework

[3] The new constitution approved in 2008 (which stated that health is a right guaranteed by the state who will ensure full exercise of the right and access to social insurance) provided reforms aiming at providing higher access to medical treatment, like the gratuity of medical services provided by the Public Ministry of Health (MSP) or laws that deprived the liberty to employers that do not affiliate workers (Orellana et al., 2017)

[4] For example, if a given hospital has a long waiting list, patients could try to receive attention in alternative hospitals in the region.

effects of the determinants on patient flows either between regions or between healthcare institutions, but there has not been an empirical study that assesses the dynamics of interregional patient mobility in the hospitals within a given region. Understanding these dynamics can help regional planners and hospital managers to understand the patterns of demand as not just interregional but also intraregional patient flows. High-performing hospitals can be prepared for potential boosts in demand generated by new reforms that widen the insured population or allow for the gratuity of medical services. They can account for these demand increases and plan to improve their capacity, medical staff, or technological endowment. Low-performing hospitals can also benefit from this, and enhance their medical resources as well, to increase their performance and avoid possible outflows of patients.

In Chapter 3, we emphasize the important influence that patient mobility can have on the performance of any given public hospital in Ecuador and that of surrounding hospitals as well, given the spillover effects in hospital efficiency.[5] Here we seek to understand the patterns of these patient flows and determine the extent to which these performance gaps are driving people to move from different regions to be treated in a (high-performing) hospital, and what the repercussions are for their surrounding hospitals.

Thus, this study aims to analyze whether higher hospitals' efficiency performance encourages larger inflows of interregional patients to a given hospital and whether these are accompanied by larger inflows of patients for neighboring hospitals in the region. So far, the literature on healthcare economics has focused on the measurement of the effects of hospital competition, patterns of access to hospital services, and the determinants of patient migration flows by just accounting for the spatial distance between hospitals or regions, using gravity models (e.g., Congdon, 2001; Varkevisser et al., 2012; Moscelli et al., 2016). A large part of the literature has concluded that the healthcare efficient performance of hospitals and regions is a strong driver of patient mobility. But there has not been an attempt—to our understanding—to consider the possible spillover effects that give rise to higher patient migration flows to neighboring hospitals. In this respect, our contribution to the literature is to provide a robust measure of hospital efficiency, consistent with economic theory, that allows us to identify its effect to attract patients. In addition, if spillover effects in the patient migration network are significant, this measure can serve as a reliable tool for decision-making to identify key hospitals that attract demand and foster competition.

To that end, we follow an innovative two-stage approach. In the first stage, we make use of the conditional order-*m* efficiency measurement proposed by Cazals et al. (2002), Daraio and Simar (2005), and Daraio and Simar (2007b) to obtain robust efficiency measures for Ecuadorian public hospitals in 2014. This method is based upon the economic concept of Pareto efficient allocation and takes into consideration the effect of other en-

[5]In Chapter 3, we provide evidence of the existence of positive spatial dependence in public hospital efficiency deriving from the existence of global and local spillover effects. In other words, the increase in the efficiency of neighboring hospitals is having a positive impact on the efficiency of an observed hospital as well.

vironmental variables (related to the region) in the hospital performance. In the second stage, we address patient mobility flows with spatial interaction models proposed by LeSage and Pace (2008) and LeSage and Pace (2009), which take into account traditional origin-destination (OD) models, but incorporate spatial lags of the dependent variable in order to account for spatial dependence, represented by flows from neighboring regions in these models and accommodating for endogenous interactions (i.e., global spillovers). In addition, we consider exogenous interaction arising from contextual effects, accommodating for spatial dependence of the explanatory variables, and representing characteristics of the neighboring regions and hospitals (i.e., local spillovers) (LeSage and Fischer, 2016). In the applied literature, these models have been used in cases where origins and destinations coincide (LeSage and Thomas-Agnan, 2015). However, this is not our case: the list of origins (regions/cantons) differs from the list of destinations (hospitals).[6] This calls for a modification in the econometric estimation which has been recently addressed by Laurent et al. (2019) that, to our understanding, has not yet been applied, and constitutes an additional contribution of our study.

In our context, the presence of endogenous interaction effects and, therefore, global spillovers mean that patient flows between an OD pair directly affect one another.[7] For example, a change in patient inflows traveling along a given OD pair, generated by variations in efficiency, potentially impact patient movements originating from a canton and going to alternative hospitals, originating from alternative cantons to a given hospital or originating from alternative cantons going to alternative hospitals. In contrast, exogenous interaction effects, hence, local spillovers imply that changes in the characteristics of neighboring cantons or regions affect the variations in patient flows across OD dyads. Taking once again efficiency as an example, the existence of local spillovers would be suggesting a competition effect among hospitals within the canton, as the increase in neighboring hospital efficiency would imply a higher inflow of patients for the region.

Our results show that efficiency is a strong determinant of interregional patients migration. However, this effect is significant just when we consider specialized hospitals (as opposed to basic hospitals). We observe significant global spillover effects in the form of patients traveling to neighboring hospitals within a region and coming from neighboring regions of the origin canton. These findings represent a useful tool for policy makers. Future healthcare reforms need to be well controlled and implemented since they need to consider territorial differences not just in terms of healthcare resources but in the level of specialization as well. In Ecuador, the specialized hospitals are concentrated in a few developed areas, and their performance is affecting the flow of patients coming from other cantons. Because spillover effects are present, other hospitals within the region seem to be benefiting from this inflow. Higher competition among hospitals could lead to higher quality of treatment (Gravelle et al., 2014; Longo et al., 2017), but it could be detrimental

[6]In Ecuador, cantons are the second-level administrative divisions. The Republic of Ecuador is divided into 24 provinces, which in turn are divided into 221 cantons. The cantons in turn are subdivided into parishes.

[7]Hereinafter, we will refer to cantons (or regions) as the origin observations of our OD dyad. Conversely, hospitals will be referred to as the destinations of the OD dyad.

if bigger inflows give rise to congestion effects. Furthermore, future public investment in healthcare services could target clusters of hospitals in low-income regions who are likely to be the origin of patient migration flows toward high-performing hospitals. A sustainable strategy could be to support the construction of more specialized hospitals –or the implementation of specialization wards in existing ones– that could serve more patients and focus on incentives to fuel local hospital competition so as to reduce the healthcare quality gap with respect to high-performing hospitals.

This study is structured in the following way. Section 4.2 reviews the literature on hospital patient migration. In Section 4.3, the theoretical model is described, as introduced by Brekke et al. (2016), which is followed throughout this study. Section 4.4 explains the methodology of the order-m efficiency measurement and the spatial interaction model, while Section 4.5 introduces the empirical approximation used. Section 4.6 describes our dataset and Sections 4.7 and 4.8 present the results and robustness analysis, respectively. Finally, the main conclusions are presented in Section 4.9.

4.2 Literature review

The aim of our study is to single out the effect that hospital efficiency has on interregional patient mobility. Moreover, we want to disentangle the potential spillover effects found in these mobilization flows between and within regions so, we can identify demand patterns of healthcare treatment that can be used as a tool for decision-making. In so doing, we combine two different strands of the literature: healthcare efficiency measurement and patient mobilization literature. There is a vast body of literature on healthcare efficiency measurement that focuses on obtaining a single value that measures the efficiency performance of an observed unit through parametric and non-parametric methods that combine multiple inputs and outputs. The idea of efficiency is linked to the concept of Pareto efficient allocation, where those efficient units are either minimizing inputs or maximizing outputs in the production of health (i.e. in providing medical attention). The main advantage of these approaches is that we can rely in a single estimated efficiency score, more consistent with economic theory, as it allocates technical or Pareto inefficiencies instead of measuring efficiency based on single averages (Cantor and Poh, 2018). A rich review of this literature can be found in O'Neill et al. (2008), Hollingsworth (2008) and Cantor and Poh (2018).

Furthermore, we rely on the hypothesis that the performance of a given set of hospitals is going to be determined –to a certain extent– by regional characteristics, and specially by the level of development or income level in the region (Brekke et al., 2016) due to the evident territorial disparities in Ecuador. In order to estimate efficiency scores that introduce environmental variables as a constraint of hospital performance, the applied literature indicates that they can be treated in one-step or two-step estimation models. The main setback of two-step approaches relies on a separability condition between the input-output space and the space of the contextual factors, assuming that these have no effect on the production process (Daraio and Simar, 2007b). To avoid the separability assumption and provide meaningful results, we implement a non-parametric method know as the conditional order-*m* efficiency estimation (Cazals et al., 2002; Daraio and Simar, 2005).[8] Recent applications of this technique include Halkos and Tzeremes (2011), who perform a conditional order-*m* efficiency analysis on Greek prefectures, and find a negative relationship between per-capita GVA and efficiency; whilst population density has a positive effect in hospital performance. Other micro-level approaches as Mastromarco et al. (2019) analyze the cost efficiency of Czech Republic hospitals during the period 2006-2010. They implement an order-*m* efficiency estimation controlling for non-profit status, teaching status, presence a specialized center (in the hospital) and occupancy rate, finding that non-profit hospitals, university hospitals and hospitals with specialized centers are generally less efficient. Another advantage of conditional order-*m* estimation is that we do not need to assume a production function in the estimation process. This is particularly important in our study, as the multidimensional nature of public hospitals and regional heterogeneity in the country posits a difficulty at the time of defending the assumption of a single production function for all hospitals in the sample.

However, despite the clear advantages of these methods to provide a robust estimation of efficiency, there has not been an attempt to combine them along with econometric models to study patient mobility patterns. The empirical literature directly focused on patient mobility has been developed in the past decade. Instead of focusing on specific determinants of patient flows, it centers on modeling hospital choices and flows across different jurisdictions (Balia et al., 2014). Some micro-level studies single out potential determinants of mobility. Victoor et al. (2012) offer a survey in which they put in evidence that some common determinants of patient mobility can refer to patient characteristics (e.g. education, income, and age) and provider characteristics. They classify the former in Structure indicators (which concern the organization of healthcare), Process indicators (which relate to the care delivery process), and Outcome indicators (which indicate the effect of the care delivered). In most of these studies, the performance of a hospital has been proxied by basic productivity indexes and capacity indicators.

In our setting, we need to take into account macro economics (regional) variables since they impact patient decision of seeking care across regional borders. In this respect (macro-level) applied economic studies have mainly been based on gravity models, commonly used to model flows that take many forms, like population migration, commodity flows and traffic flows (Thomas-Agnan and LeSage, 2014). These models embed movements of individuals between origin and destination regions. Levaggi and Zanola (2004) look for the determinants of net patient flows from regions of Italy to the rest of the country. They estimate gravity models for a sample of Italian regions from 1995-1997 and conclude that regions characterized by lower outflows are the ones that provide better or faster services. Cantarero (2006) develops the same analysis to patient flows across regions in Spain between 1996 and 1999 and identify that patients from the economically lagged regions move more than those regions that provide better health services. Fabbri and Robone (2010) explore the "trade" phenomenon in hospital care, exploring the role

[8] We explain this method on a deeper extent on Section 4.4

of the scale economies and the impact of North-South economic divide on the mobility of Italian Local Health Authorities (LHAs) controlling for push and pull factors of patients related to origin and destination. They find that richer LHAs have a higher probability of attracting more patients, who present the most severe cases.

However, the use of traditional gravity models to explain spatial interaction can be limited. These models rely on a function of the distance of the OD to clear spatial correlation and cross-section independence. As LeSage and Pace (2008, 2009) state, the notion that use of distance functions to effectively capture the spatial dependence of observations can be erroneous. Also, the idea that flows are independent since OD flows are fundamentally spatial in nature. In our framework of analysis, we expect to find a behavior pattern where high-income regions are the main receivers of patients, following a spatial pattern, that, if not controlled for in the econometric estimation, could lead to biased conclusions.

So far, no studies have tried to account for the spatial dependence in patient mobility. Moreover, even when a big part of the literature implicitly concludes that healthcare performance is a strong driver of patient flows, but there has not been an attempt to disentangle its sole effect. The closest paper to our approximation is Balia et al. (2018) who account for local spillover effects by incorporating the spatial lags of the exogenous variables in the gravity model. They use a spatial panel data framework of Italian hospital discharges between 2001 and 2010 to assess the effect of the main determinants of interregional patient flows, differentiating between the impacts of regional health policies and other exogenous factors. Their results show that neighboring regions´ supply factors, specialization and performance largely affect mobility by generating local externalities that explain OD patient flows; bringing some insights of the inherent spatial-dependent nature of hospital performance, and, its effects on patient mobility.

Our empirical estimation, hence, goes beyond the incorporation of local spillover effects as in Balia et al. (2018), and includes potential global spillover effects likely found in OD flows, as stated by LeSage and Pace (2008, 2009). In so doing, we use the extended gravity models developed by LeSage and Pace (2008, 2009) to allow for spatial dependence in the sample, represented by the flows from cantons (regions) to public hospitals in these models. Additionally, we consider exogenous interactions of the explanatory variables (LeSage and Fischer, 2016) to accommodate for the contextual effect of the neighboring regions and hospitals in the OD dyad, as in Balia et al. (2018). The introduction of endogenous and exogenous interactions in the econometric model allow us to take into consideration the spatial structure present in OD flow data that is not completely captured by the sole inclusion of the distance between origin and destination. If spillover effects are found statistically significant, then policy implications may be directed to identify key players within the flow network that have an indirect effect over other hospitals. Policy decisions can target those key players to improve healthcare performance of the region.

4.3 Theoretical framework

In our framework of study, the high-performing hospitals are mainly located in developed regions (see Section 4.6) that have historically concentrated the healthcare resources in the country. These asymmetries in hospital performance have derived in regional healthcare performance gaps that may incentive those patients residing in less developed regions (cantons) to seek treatment in high-performing hospitals. In this context, the backbone of our theoretical framework builds upon Brekke et al. (2014) and Brekke et al. (2016). They take a context of asymmetrical regions, where the regions differ in their ability to provide healthcare services, the higher the performance gap between providers, the higher the number of patients who will seek medical care in high-income regions. Here, Brekke et al. (2014) state that patient mobility can have significant participation in the improvement of welfare. Albeit, this welfare improvement comes with asymmetric effects. If competition promotes performance, then patients living in regions with high-performing hospitals are better-off than in a system without mobility. Conversely, in areas of low-performing hospitals, only patients who move to high-performing areas benefit from the quality improvement in healthcare. Additionally, Brekke et al. (2016) consider a framework with heterogeneous income across and within regions. They find that reducing barriers to free patient mobility represents an incentive to reduce quality for low-middle income regions while increasing income disparities between regions increase the interregional quality gap.

We take upon Brekke et al. (2016) cross-border patient mobility theory. The theoretical model relies on the idea that, in equilibrium, regions with higher income offer better quality, which creates an incentive for patient mobility from lower to higher income regions. This conception can be applied to our setting, as the best-performing hospitals are mainly located in high-income regions.[9] Following Brekke et al. (2016), let us define a uniformly distributed healthcare market where patients are distributed on a circle with circumference equal to 1 and the total patient mass normalized to 1. Consider three different neighboring regions of equal size ($i = L, M, H$) covering 1/3 of the circle. The index i denotes a Low, Middle or High average income regions. Healthcare is supplied by three hospitals, each in each region, where the hospital in region i is located at s_i. Assuming that each hospital is located at the center of its region, the residents of region i are located in the line segment $[s_i - 1/6, s_i + 1/6]$. Each patient consumes one unit of healthcare from the most preferred hospital. The model assumes public provision of healthcare with general income taxation funding and free consumption.

If the patient receives treatment in their local region, we define the net utility of a

[9] In fact, Brekke et al. (2014) develop an Hotelling model with two regions that differ in healthcare technology, where regions with more efficient technology supply higher healthcare quality, attracting patients from neighboring regions with less-efficient technology. However, the restriction of incorporating two regions prevents from considering a case where a region can be both importing and exporting patients as opposed to Brekke et al. (2016) (whom incorporate a three-region specification). In addition, the framework used in Brekke et al. (2016) allows for extra expenses when patients demand care outside their region, and allow for heterogeneity in income within regions (with wealthier patients more likely to move).

patient located in z_i receiving treatment from hospital in region i as:

$$U(z_i, s_i) = v + bq_i - t|z_i - s_i| + u(A_i^x) \tag{4.1}$$

where v in the patient's gross utility of being treated ($v > 0$), q_i is the quality offered by the hospital in region i, $q_i \geq \underline{q}$, with $\underline{q}$ representing the lowest possible quality the hospitals can provide without being charged with malpractice (for simplicity it is set to 0), $b > 0$ measures the marginal utility of quality, and t is the marginal disutility of traveling. The utility function, $u(.)$ of income is strictly concave, while A_i^x is the net income of type-x patient in region i. Assuming that patients are heterogeneous in income a^x with $x = P, R$, i. e., including high-income (Rich) and low-income (Poor) patients, which implies $a^R > a^P$. We include an income tax rate (social security contribution) $\tau > 0$, set by the central government.[10] Then, the net income of a type-x patient in region i is given by

$$A_i^x = a^x(1 - \tau) \tag{4.2}$$

Additionally, we assume heterogeneity of residents' income, with the proportion of high-income residents λ_i being $\lambda_H > \lambda_M > \lambda_L > 0$ (High, Middle and Low income residents). The average gross income in region i is set to:

$$\bar{a}_i = \lambda_i a^R + (1 - \lambda_i) a^P \quad \text{for } i = H, M, L \tag{4.3}$$

The net utility of a patient located at z_i, receiving treatment from hospital in a neighboring region j (different from the patient residence region), located at s_j is given by

$$U(z_i, s_j) = v + bq_j - t|z_i - s_j| + u(A_i^x) - F \tag{4.4}$$

Where F are the non-monetary costs of looking for care in a different region. The model also includes additional costs (π) that patients who get treatment in a different region must pay (like co-payments or other out-of-pocket expenses), such that the net income of type-x patient in region i who seeks care in neighboring region j is set by:

$$\widehat{A_i^x} = a^x(1 - \tau) - \pi \tag{4.5}$$

Assuming a patient utility-maximizing choice of hospital, type-x patients traveling from i to j for treatment are located on a line segment of length $max\{0, \phi_{ij}^x\}$, where:

[10] Note that we can also allow for an income tax rate set by the government of region i as τ_i

$$\phi_{ij}^{x} = \frac{1}{2t}(b(q_j - q_i) + u(\widehat{A_i^x}) - u(A_i^x) - F) \tag{4.6}$$

Notice that

$$\frac{\partial \phi_{ij}^{x}}{\partial a_x} = (\frac{1-\tau}{2t})(u'(\widehat{A_i^x}) - u(A_i^x)) > 0 \; if \; \pi > 0 \tag{4.7}$$

with $u'(\cdot)$ being the first derivative of utility function $u(\cdot)$.

As long as $\pi > 0$, richer patient have disutility of paying for extra costs and are more prone to choose cross-border healthcare. The total number of patients traveling from region i to region j is then given by $max\{o, \Phi_{ij}\}$ where

$$\Phi_{ij} = \lambda_i \phi_{ij}^{R} + (1-\lambda_i)\phi_{ij}^{P} \tag{4.8}$$

Notice that

$$\frac{\partial \Phi_{ij}}{\partial q_j} = -\frac{\partial \Phi_{ij}}{\partial q_i} = \frac{b}{2t} \tag{4.9}$$

Finally, Brekke et al. (2016) demonstrate that (in equilibrium) the optimal choice of healthcare quality will be higher in richer regions, in such a way that $q_H^* > q_M^* > q_L^*$; with q_i^* being the optimal quality choice in region i. This creates an incentive for patient migration from poorer to richer regions.[11]

Therefore, in order to analyze patient mobility across regions we rely on OD flows akin to what is done in international trade and migration models, which are heavily drawn on gravity model specifications. Thus, we define the following gravity function to be estimated

$$E(Y_{ij}) = f(X_i, X_j, G_{ij}) \tag{4.10}$$

Where $E(Y_{ij})$ is the expected flow from i to j; $G_{ij} = f(g_{ij})$, being g_{ij} a vector of separation (distance) measures. X_i and X_j are origin (canton) and destination (hospital) covariates, respectively. Cantonal environmental variables include measures that approximate the regional income level and healthcare quality such as per-capita gross value added

[11]Refer to Brekke et al. (2016) Section 3.2.

(GVApc), population density, cantonal mortality, insured population rate, and a multidimensional poverty index. In such a way, we can identify the poor regions that are more likely to push away patients to neighboring (wealthier) regions.

To proxy hospital quality and performance, the literature has usually relied on basic ratios, such as mortality rate or readmission rates, which in many cases can lead to mixed results. We go a step forward in this approach and use a value that measures the performance of a given hospital, comprising all their inputs and outputs and considering other environmental variables (respective to the region where each hospital locates) that captures the pull effect of those hospitals to attract patients. This way, we rely on a single (robust) measure that can facilitate policy decisions.

We also need to consider those potential spillover effects that may arrive from migration flows in our data. If hospitals located in high-income regions are expected to attract poor-income regions' patients, then we can assume that other regions (neighboring those that push away patients) may also present outflow of patients, attracted by high-performing hospitals. Conversely, spillover effects could appear in the receiving regions as patient flow movements in their hospitals. This question is the core of our empirical exercise.

4.4 Methodology

The method used in this study is developed in two stages. First, we need to obtain the efficiency measures for each hospital, conditional to the environmental variables they face and can constrain their performance. In the second stage, we develop a spatial interaction model (based on the conventional gravity specification) to estimate the impact that the efficiency value has on migration flows, accommodating for potential spillover effects.

4.4.1 Order-*m* efficiency estimation

The first stage of our strategy uses a nonparametric order-*m* efficiency estimation approach, introduced by Cazals et al. (2002), Daraio and Simar (2005) and Daraio and Simar (2007b) that relies upon the production theory (Debreu, 1951; Koopmans, 1951).[12] Introducing the notation used in this paper, we assume a set of $y \in \mathbb{R}_+^p$ outputs produced by a set of $x \in \mathbb{R}_+^q$ inputs, the production technology is the set of all feasible input-output combinations.

$$\Psi = \left\{(x, y) \in \mathbb{R}_+^{p+q} | \; x \; can \; produce \; y\right\} \tag{4.11}$$

[12] We take an output oriented approach, as we expect that patients can perceive –to a certain extent– the performance of an hospital based on the amount of patients treated.

The multidimensional nature of public hospitals, with different functions that are difficult to quantify, plus the impossibility to obtain input and output prices information makes Ψ impossible to observe. To account for this, we need to estimate Ψ from a random sample of production units denoted by $X = \left\{(x_i, y_i) \in \mathbb{R}_+^{p+q} | \; i = 1, \ldots, n\right\}$. Following this framework, an observed production unit (x_i, y_i) defines an individual production possibility set $\Psi(x_i, y_i)$, which under the free disposability of inputs and outputs can be expressed as:

$$\Psi(x_i, y_i) = \left\{(x, y) \in \mathbb{R}_+^{p+q} | \; (y \geq y_i) x \leq x_i\right\} \tag{4.12}$$

Nevertheless, there could be other environmental factors $Z \in \mathbb{R}^r$ exogenous to the production process that could be affecting the production and the distribution of efficiency scores. In this matter, Cazals et al. (2002), Daraio and Simar (2005) and Daraio and Simar (2007b) use a probabilistic formulation of the production process to develop a conditional efficiency approach to account for the environmental variables in the efficiency estimation, conditioning the production process to a given value of $Z = z$. This conditional function is given by:

$$S_Y\left(y \;\middle|\; x, z\right) = Prob(Y \geq y | X \leq x, Z = z) \tag{4.13}$$

representing the probability of a unit operating at level (x, y) being dominated by other units facing the same environmental conditions z. This way, the conditional output efficiency can be defined as the Farrell (1957) efficiency measure:

$$\theta\left(x, y \;\middle|\; z\right) = sup\{\theta | S_Y(\theta \cdot y | x, z) > 0\} \tag{4.14}$$

Those points where $\theta\left(x, y \;\middle|\; z\right) = 1$ are the technically efficient ones and correspond to the efficiency frontier, while those with $\theta\left(x, y \;\middle|\; z\right) > 1$ are technically inefficient. To obtain the nonparametric estimators of the conditional frontier $\theta\left(x, y \;\middle|\; z\right)$, mitigating the impact of outliers, we use the order-m frontier (Cazals et al., 2002). The order-m frontier considers as a benchmark the expectations of the best practice among m peers randomly drawn from the population of units from which $X \leq x$.[13] The procedure is repeated B times resulting in multiple efficiency measures $(\widehat{\theta}_m^1, \ldots, \widehat{\theta}_m^B)$, where the final order-$m$ efficiency value is the sample mean $(\widehat{\theta}_m)$. This way, the efficiency of a decision making unit (DMU)[14] can be compared with m potential DMUs that have a production larger or equal to y. The conditional order-m output efficiency score is defined as in Daraio and Simar (2007a):

[13]We fix the value of $m = 90$, following the approach of Daraio and Simar (2005) for which the decrease in *super-efficient* observations($\theta\left(x, y \;\middle|\; z\right) < 1$) stabilizes.

[14]We can call DMU to any unit of analysis, say, individuals, departments, firms, municipalities, or in the case of this study, hospitals.

$$\widehat{\theta}_m\left(x_0, y_0 \mid z_0\right) = E(max\{\theta_1, \theta_2, \ldots, \theta_m | X \leq x_0,\ Z = z_0\})$$

$$\int_0^{\infty} \left[1 - (1 - S_{Y|X,Z}(uy_0 | X \leq x_0,\ Z = z_0))^m\right] du \tag{4.15}$$

The efficient frontier corresponds to the DMUs where $\widehat{\theta}_m\left(x, y \mid z\right) = 1$. Notice that the efficiency values can take a score lower than one. In this case, the hospitals are labeled as *super-efficient*, meaning that they exhibit higher levels of outputs than the order-m frontier.

To estimate the conditional order-m model, we need to incorporate smoothing techniques such that in the reference samples of size m units with comparable z-values have a higher probability of being chosen. Hence, we rely on the estimation of nonparametric kernel functions to select the reference observations, and a bandwidth parameter h in the estimated probability function $S_Y(y|x, z)$, given by:

$$\hat{S}_{Y,n}\left(y \mid x, z\right) = \frac{\sum_{i=1}^{n} I(x_i \leq x, y_i \geq y) K_{\hat{h}}(z, z_i)}{\sum_{i=1}^{n} I(x_i \leq x) K_{\hat{h}}(z, z_i)} \tag{4.16}$$

Where $K_{\hat{h}}(\cdot)$ represents the kernel function, $I(\cdot)$ is an indicator function, n represent the number of observations and h is the appropriate bandwidth. Considering that our environmental variables Z are continuous, we estimate the appropriate bandwidth h following Daraio and Simar (2005) and use the k-Nearest Neighbor (k-NN) method. Finally, a non-parametric estimate of $\widehat{\theta}_m\left(x_0, y_0 \mid z_0\right)$ is obtained by plugging $\hat{S}_{Y,n}\left(y_0 \mid x_0, z_0\right)$ in equation (4.15).

4.4.2 Spatial interaction model specification

In the second stage of our strategy we make use of spatial interaction models, which rely on gravity models to explain OD migrations flows. In the empirical literature, gravity models have long been one of the most successful approaches, modeling remarkably well the observed variations in economic interactions across space (Anderson, 2011). Gravity models have commonly been used to explain OD flows that arise in trade, transportation, migration, among others. In the regional economics literature, these models are usually known as spatial interaction models (Sen and Smith, 1995), as the regional interaction is directly proportional to the product of regional size measures (e.g. regional income in the case of interregional commodity flows). One advantage of gravity models is that due to the nature of gravity itself, it does not apply to individuals but to spatial units as regions, cities or countries. (Lowe and Sen, 1996). This allow us to focus exclusively on inference about the determinants of patient migration, from the patterns of distribution of patients, without the need to involve what determines the total supply of medical care of all destinations or the total demand of patients from all origins.[15]

However, a potential drawback for gravity models is that they rely just on a function of OD distance to account for spatial correlation and ensure cross-section independence (Balia et al., 2018). These assumptions have been challenged by many authors. Porojan (2001) and Lee and Pace (2005) find evidence of spatial dependence in the residuals of international trade and retail sales flows, respectively; while LeSage and Pace (2008, 2009) point out that the assumption of independence among observations might be difficult to defend, as OD flows are fundamentally spatial in nature. The explicit consideration of flow data correlation due to the spatial configuration of the units involved has been drawing much attention in the literature as the so-called network autocorrelation (Patuelli and Arbia, 2016).

To embed spatial dependence in a spatial interaction setting, LeSage and Pace (2008) consider spatial spillovers at three dimensions: *origin-based*, *destination-based*, and *origin-destination based*. Using this definition of spatial dependence means that we need to model spatial dependence for flows of patients as a spatial autoregressive specification, accommodating endogenous interactions. This definition will allow us to define spatial spillover effects to hospitals neighboring the destination hospital in the flow of patients.

Additionally, we can accommodate the model for exogenous interactions in a Spatial Durbin Model (SDM) representing a situation where local spillovers arise from changes in the characteristics of neighboring hospitals and environmental features of neighboring regions (cantons). The exogenous interactions can be modeled by including the spatially lagged covariates in the econometric specification (along with the spatial lag of the endogenous variable). If statistically significant, the omission of these interactions can lead to problems of omitted variable bias (LeSage and Fischer, 2016). We control for this issue: we begin by defining the model with no spatial interactions (based on the conventional gravity model) and adjust it for exogenous interaction specifications as in LeSage and Fischer (2016),[16] and, then, we move to its SDM extension as illustrated in LeSage and Pace (2008) and Laurent et al. (2019).[17]

[15]This property is also known as "modularity" in trade models developed by Anderson and Van Wincoop (2003).

[16]This model specification is commonly referred in the literature as the spatial lag of X (SLX) model (Halleck Vega and Elhorst, 2015)

[17]We move in this direction to identify the sources of spatial autocorrelation and avoid model misspecification and omitted variable bias. Following this sequence, we can determine the significant effect of the exogenous interactions by means of an SLX, and, then, those of the endogenous interactions with the SDM model. In such a way, we can select the appropriate framework of analysis that provides the best fit to our data.

4.4.3 Spatial Lagged X interaction model

Form equation (4.10), we begin by setting a Y matrix of patients' migration flows, whose columns reflect origins (cantons), and rows destinations (hospitals). Let n_o be the number of geographical observations at the origin and n_d the number of geographical observations at the destination, then $N = n_o \times n_d$. The $n_o \times n_d$ flow matrix Y can be converted to an N vector by stacking columns. The flow matrix can be arranged so the $(i,j)th$ observation reflects a flow from j to i ($y^o = vec(Y)$), which is labeled origin-centric ordering. Then, the destination-centric ordering can be obtained by $y^d = vec(Y')$ reflecting a flow from i to j. We can use G to represent the $n_o \times n_d$ matrix of distances between origins and destinations. Then, $g = vec(G)$ is an N vector of these distances formed by staking the columns of the OD distance matrix. If we assume a destination-based order, the logged-transformed gravity regression model would be as follows:[18]

$$y = \propto l_N + X_o\beta_o + X_d\beta_d + \gamma g + \varepsilon \tag{4.17}$$

Where y is the N vector of patient migration (logged) flows, that has been obtained by stacking the columns of the matrix Y; X_o, X_d represent the $N \times k$ matrices of (logged) explanatory variables containing the origin and destination characteristics respectively, which are expected to reflect the regional and hospital factors that sustain patient choice for medical care; β_o, β_d are the associated $k \times 1$ parameter vectors. The scalar parameter γ is the effect of the (logged) distance g, and $\propto$ is the constant with l_N vector of ones. Finally, we have an $N \times 1$ vector of disturbances ($\varepsilon = vec(E)$).

From (4.17), we consider an SLX interaction model in the following specification:[19]

$$y = \propto l_N + X_o\beta_o + X_d\beta_d + W_oX_o\emptyset_o + W_dX_d\emptyset_d + \gamma g + \varepsilon \tag{4.18}$$

Where W_o and W_d are conventional (row-normalized) spatial weight matrices for the origin and destination observations, respectively. It is worth noting that here we do not account for spatial weights in W based on geographical distances, as other conventional spatial econometric models. This is because we are considering for OD distances in the matrix defined as G in the gravity model specification. We define the W matrix of spatial weights to be a contiguity (row-normalized) matrix to consider the spatial configuration of the hospitals and regions that leads to a flow of patient data correlation.[20]

[18] If we start with the standard gravity model and apply a log transformation, the resulting model would be as shown in (4.17) (Sen and Smith, 1995).

[19] We estimate an SLX model in order to test and identify the existence of local spillovers. Their omission from the econometric estimation could lead to potential problems of omitted variable bias (LeSage and Fischer, 2016).

[20] In section 4.5, we will explain the empirical strategy followed in this paper, along with the specifications for W_o and W_d.

The spatial lags of the exogenous variables W_oX_o and W_dX_d help explain variations in flows across dyads coming from changes in the characteristics of the regions neighboring the origin and hospitals neighboring the destination respectively. $\emptyset_o$, $\emptyset_d$ are the parameters associated to W_oX_o and W_dX_d. In our study, we enrich equation (4.17) and control for the spatial lags of distance g, in the following manner:

$$y = \propto l_N + X_o\beta_o + X_d\beta_d + W_oX_o\emptyset_o + W_dX_d\emptyset_d + \gamma g + W_og\gamma_o + W_dg\gamma_d + \varepsilon \tag{4.19}$$

Where W_dg and W_og explain the variations in flows arising from changes in the distance of neighboring hospitals in the same canton, and from neighboring cantons respectively. This aligns with the idea that patients will select the hospital to be treated depending not just on their proximity (to a given hospital), but to the that of their neighbors, as well. Finally, γ_d and γ_o are the parameters corresponding to W_dg and W_og.

4.4.4 Spatial Durbin interaction model

From equation (4.17), one can consider that a change at the characteristics of an observation i can impact inflows or outflows (or both) of other observations connected with element i which are not explained in (4.17) (Thomas-Agnan and LeSage, 2014). LeSage and Pace (2008) suggest that flows across networks can exhibit spatial dependence and propose a spatial autoregressive extension of the non-spatial model in (4.17), which can be viewed as filtering for spatial dependence related to origin and destination.

$$(I_{\mathrm{N}}-\rho_oW_o)(I_{\mathrm{N}}-\rho_dW_d)y=\propto l_{\mathrm{N}}+X_o\beta_o+X_d\beta_d+\gamma g+\varepsilon \tag{4.20}$$

Here, $(I_{\mathrm{N}}-p_oW_o)(I_{\mathrm{N}}-p_dW_d)$ is the filter that capture global spillover effects, translated into *origin-based*, *destination-based*, and *origin-destination-based dependence*.[21] As described by LeSage and Pace (2009), origin-based spatial dependence reflects the notion that forces leading to flows from any origin to a particular destination may create similar flows from neighboring origins. Destination-based spatial dependence is related to idea that forces leading to flows from the origin to a destination may generate similar flows to nearby destinations. Thus, Origin-destination-based spatial dependence reflect those forces that create flows from neighbors to the origin to neighbors to the destination. The model (4.20) can be further enriched considering the spatial lags of the explanatory variables into an SDM as follows:[22]

[21] $I_{\mathrm{N}}-\rho_dW_d-\rho_oW_o-\rho_wW_w$. Where W_w is the product of the two weight matrices ($W_o.W_d$). The reader can refer to LeSage and Pace (2008) for a better understanding.

[22] The selection of the SDM model allow us to test and identify the existence of global spillover effects in our dataset. It also incorporates local spillover effects modeled as the spatial lags of the covariates, as in the SLX model. Furthermore, the selection of an SDM model will produce unbiased coefficient estimates when the source of spatial correlation is unknown (LeSage and Pace, 2009).

$$y = \rho_d W_d y + \rho_o W_o y + \rho_w W_w y + \propto l_N + X_o\beta_o + X_d\beta_d + \gamma g + W_o X_o \emptyset_o + W_d X_d \emptyset_d + W_o g \gamma_o + W_d g \gamma_d + \varepsilon \tag{4.21}$$

Then, the spatial lag $\rho_d W_d y$ reflects flows from neighbors to each destination observation in the vector of origin-destination flow to form a linear combination of flows from neighboring destinations. While ρ_d captures the strength of destination-based dependence. Similarly, $\rho_o W_o y$ reflects a linear combination of flows from regions neighboring the origin; and ρ_o reflects the strength of origin-based dependence. Hence, $\rho_w W_w y$ forms a linear combination of flows from neighbors to the origin and flows from neighbors to the destination, and the parameter ρ_w represents the strength of this dependence.

Finally, the spatial autoregressive model can be estimated by Maximum Likelihood (see LeSage and Pace, 2008). LeSage and Pace (2009) also show how to produce Bayesian Markov Chain Monte Carlo (MCMC) estimates for the model.[23] In this study, we follow the Bayesian approach using the computational methods proposed in Laurent et al. (2019). Our decision is motivated by the flexibility that Bayesian methods offer to capture complex spatio-temporal relationships with heterogeneous data. The use of prior distributions allows for prior constraints in the parameters which reduces the risk of over-parametrization. In addition, it allow us to accommodate econometric specifications with more than one spatial weight matrix, adjusting to our model when origins differ from destinations.

4.5 Empirical application

Here, we will define our empirical strategy to understand whether the efficiency performance of high-performing hospitals is attracting more patients, and whether this is accompanied by higher patient inflows to neighboring hospitals.

In our empirical application, we define an OD patient flow matrix between regions (origin) and hospitals (destination), which are different units of analysis; this approach constitutes a different strategy to that used in other interregional studies. In most of the empirical literature, the OD flows have been measured by accounting for patient migration from one region to another, making it difficult to analyze the intraregional dynamics that occur among hospitals within a region; we contribute to the current literature on interregional patient mobility by accounting for patient migration flows, from which the origin represents the region and the destination a given hospital in a particular region. This way we will be able to consider the dynamic of the destination's neighboring hospitals in the same region.

[23] Refer to Appendix 4.10.2 for an explanation on LeSage and Pace (2009) MCMC estimation.

In this study, we expect to find spatial dependence, embedded in the size of patient flows from a region to a hospital as well as its neighboring hospitals. Following the methodology described above, this would mean the presence of destination-based-dependence spillovers, which in the econometric model means that $\rho_d \neq 0$. However, given the context of analysis were patients travel to just some certain high-performing hospitals, we can expect that not just the patients from one region travel to a given hospital to be treated, but also patients from neighboring regions. Thus, spillover effects, are embedded in the flow size from neighboring regions (origin-based-dependence), which would mean that $\rho_o \neq 0$. Finally, the spillovers can also come in the form of flows from regions neighboring the origin to hospitals neighboring the destination (origin-destination-based-dependence), thus $\rho_w \neq 0$. We define a contiguity matrix W_d where hospitals are neighbors if they are located in the same canton. Hence, W_o defines as neighbor those cantons that share a border line. The vector of distances g is composed by the euclidean distances between origins and destinations. On this basis, we can define the following model:

$$y = \rho_d W_d y + \rho_o W_o y + \rho_w W_w y + \propto l_N + e_d \beta_d + X_o \beta_o + \gamma g + W_d e_d \emptyset_d + W_o X_o \emptyset_o + W_o g \gamma_o + W_d g \gamma_d + \varepsilon \tag{4.22}$$

The vector e_d contains the robust logged efficiency scores obtained with (4.15) specific to every hospital. This measure is estimated by taking into consideration the environmental conditions that limit the hospital production, so an observed hospital is benchmarked with a sample of hospitals facing the same external conditions.[24] Therefore, it can properly be used as an indicator that measures the performance of the hospital as a pulling factor that attracts patients. The matrix X_o accounts for economic and demographic regional characteristics that proxy the regional income-level and health conditions, and impact patient choice to seek treatment in other (developed) regions. We use cantonal variables such as logged GVApc, logged population density, logged cantonal mortality, logged unsatisfied basic needs index (NBI),[25] and logged insured population rate.[26]

A problem that might arise in the application of the model (especially for the regions that present large inflow of patients) is the presence of large flows of patients in the matrix of OD flows, relative to smaller (or zero). This would produce the non-normality in flows and jeopardize the estimations (LeSage and Pace, 2008, 2009; Thomas-Agnan and LeSage, 2014). In our setting, this would be representing an intraregional flow of patients (e.g. residents of developed regions getting treatment in their local area). To deal with this problem, LeSage and Fischer (2010) propose to modify the independent variables, by replacing with zero the values of the independent variables for the intraregional flows.

[24] The environmental conditions considered are the Gross Value Added (GVA) per-capita and density of the canton where the hospital is located as well as the occupancy rate of the hospital in the canton.

[25] The NBI is a multidimensional poverty index, commonly used in Latin American countries (explained in Section 4.6).

[26] In order to avoid taking the log of zero, we have added the unity to the dependent and independent variables as in LeSage and Thomas-Agnan (2015).

The intraregional variations will be captured in a new set of explanatory variables X_i, W_iX_i, with non-zero observation for the intraregional observations as well as adding a new intercept term, α_i. If we allow for c=$vec(l_{\mathrm{N}})$, the new model would be as follows:

$$y = \rho_d W_d \mathrm{y} + \rho_o W_o \mathrm{y} + \rho_w W_w y + c\alpha_i + \propto l_N + e_d\beta_d + X_o\beta_o + X_i\beta_i + W_d e_d \emptyset_d + W_o X_o \emptyset_o + \\ W_i X_i \emptyset_i + \gamma g + W_o g \gamma_o + W_d g \gamma_d + \varepsilon \tag{4.23}$$

Note that we cannot interpret β_d (nor any other estimated parameter associated with origin-destination characteristics) as the partial derivative on flows arising from changes in the destination-efficiency. As pointed out by LeSage and Pace (2009), in the spatial econometric specification of the interaction model, changes in the *kth* characteristic of an observation i will produce changes in flows into the *ith* observation from other observations, as well as flows out of the observation i to other observations. Unlike conventional regression models where it leads to changes only in observation i of the dependent variable, y_i.

LeSage and Thomas-Agnan (2015) propose scalar summary measures of the impacts arising from changes in characteristics of the observations that involves averaging the *cumulative flow impact* associated with changes in all observations, resulting in the so called *origin effects*, *destination effects*, and *network effects*. Origin and destination effects express the mean impact on flows arising from changes in the origin and destination characteristics, respectively. In turn, network effects characterize the mean impact of a change in the characteristics of the origin i on all the flows coming from other origins, different from i to a destination j.[27]

In our setting we have $n_o \neq n_d$ and different covariates' matrices for origins (X_o) and destinations (X_d), which requires to follow the *computational inefficient* method to calculate the scalar marginal effects. This means that we need to calculate changes in each of the n_o and n_d elements of the vectors X_o and X_d, respectively, to obtain scalar summary measures of the impact of these changes on the patient flows.

4.6 Data and variables

To estimate (4.23) we collect data for the year of 2014. Hospital information comes from the Annual Survey of Hospital Beds and Discharges and the Survey of Health Activities and Resources provided by the National Institute of Statistics and Census (INEC, by the acronyms in Spanish). We excluded the psychiatric, dermatologic and geriatric hospi-

[27] Refer to Thomas-Agnan and LeSage (2014) and LeSage and Thomas-Agnan (2015) for a deeper understanding on the scalar summary measures.

tals, and took out from the sample those that presented irregularities in the data.[28] As described above, the migration flow matrix considers the rows to be the hospital destination, while the columns are the cantons (regions) of origin. We retrieved a sample of 176 destination hospitals and 106 cantons of origin. By vectorizing the flow matrix, using the destination-centric arrangement described in Section 4.4, we obtain a vector of 18.656 observations.

The cantonal economic and demographic variables were retrieved from the Ecuadorian Central Bank (BCE, by the acronyms in Spanish) and INEC's public statistics, respectively. While the poverty and population insurance data was collected from the 2010 national census. The description for all the variables is presented in Appendix 4.10.3.

4.6.1 Variables for the conditional order-*m* efficiency measurement

For the selection of input and output variables to estimate model (4.15) in the first stage of our strategy, we followed previous literature on efficiency measurement. A complete review of the literature is offered in Hollingsworth (2008); O'Neill et al. (2008) and Cantor and Poh (2018).

Regarding the input variables, we use the number of beds, the medical equipment, and infrastructure, widely used as a proxy for hospital size and capital investment (O'Neill et al., 2008). To proxy labor costs, clinical staff is usually included (Hollingsworth, 2003, 2008). To that end, we introduce the number of physicians and healthcare professionals beyond the number of physicians of the hospital.

As for the outputs we use the hospital's patient discharged to measure the final production of health. To control for the heterogeneity of the disease-case attended, we build a *case-mix*. This index is used in the healthcare efficiency measurement literature to control for the severity of the cases treated, as not all the patients can be treated with the same amount of resources, nor all of them have the means to treat the most severe cases (Cantor and Poh, 2018). As in Chapters 2 and 3, we use the case-mix index proposed by Herr (2008), which relies on the assumption of a positive correlation between length of stay and the severity of illness. The index is built according to the three-digit International Statistical Classification of Diseases and Related Health Problems (ICD-10).

In addition, the Survey of Health Activities and Resources in 2014 provides information on the total number of morbidity and emergency consults, and the total number emergencies treated, commonly used in the literature to measure the activity of hospitals (Cantor and Poh, 2018).

[28] We excluded psychiatric, dermatologic and geriatric hospitals as they focus on specific illness and patients that require different treatments that could bias the efficiency values. For example, psychiatric hospitals might require inpatients to stay for long periods of time, wherein our analysis would reflect it as a criteria for less efficiency.

Furthermore, we tried to account for a quality hospital related output with the hospital survival rate for patients after 48 hours of admission. The intuition is that the mortality rate after 48 hours has a stronger relationship with the resolutive capacity of the hospital employees. Therefore, it has a higher correlation with the quality of the treatment provided. Hospital mortality rates have been usually employed to proxy the quality of the hospital treatment (Hollingsworth, 2008), but in healthcare efficiency measurement we need to measure the outputs as health gains of the patients; which is why the survival rate (1-mortality rate) is usually employed.[29]

In this respect, other hospital indicators such as readmission rates, the level of specialization (Gravelle et al., 2014; Longo et al., 2017) or the nosocomial infections (Prior, 2006) have been usually employed to proxy hospital quality. However, we do not account for this information in our dataset, which is one of the limitations that we faced in our study.

Finally, we consider three environmental variables that can potentially affect the hospital performance: the cantonal GVApc, cantonal population density and the hospital occupancy rate. The former two explain the territorial inequalities in the country, which have a big influence in their regional development (Mendieta Muñoz and Pontarollo, 2016). Those developed regions present a high concentration of hospitals and health resources, that influence healthcare performance. In Chapter 2, we find empirical evidence that these developed regions do not just concentrate better-endowed hospitals, but these hospitals are also the best performers in terms of efficiency. The empirical evidence of the effect of GVApc and population density on healthcare efficiency is supported by Halkos and Tzeremes (2011).

The third environmental variable is commonly used to proxy the utilization of potential capacity in a hospital and determine whether it is adjusting their working staff to the increase of treated patients in the short-run (Herwartz and Strumann, 2012, 2014). The idea behind is that hospitals with low occupancy rate may be signaling an oversized staff and capacity, unlikely to meet the demand for medical treatment efficiently. The occupancy rate has been used as an environmental variable for conditional order-m approaches in recent work by Mastromarco et al. (2019). Furthermore, in Chapter 3, we provide empirical evidence of its positive direct and spillover effects on hospital's efficiency.

Table 4.1 presents the descriptive statistics of the variables for the conditional order-m efficiency estimation. Overall, we can distinguish a big gap of hospital's inputs and outputs (as well as in cantonal variables), observed in the difference of the minimum and maximum values that describes the marked discrepancies across hospitals and cantons. In fact, in Chapter 3 we emphasize that those hospitals that present a high amount of resources and treated population settle in regions densely populated and with high production (measured by the GVA). This initial evidence supports our hypothesis that patient

[29]One additional advantage of the conditional order-m estimation model is that we can use output variables expressed both in percentage and volume and obtain robust efficiency measures. If the same structure was applied to classical methods (such as DEA) the results would be inconsistent (Olesen et al., 2015).

Table 4.1: Conditional order-m variables' summary statistics

	Mean	Median	SD	Min	Max
Outputs					
Discharges (weighted)	2262.23	1059.03	2750.83	79.1	16262.77
Morbidity consults	41210.26	21923.5	64187.04	168	529420
Emergencies consults	43969.8	24942	58596.16	70	407485
Survival rate	0.98	0.99	0.03	0.79	1
Inputs					
Physicians	71.49	30	107.39	4	786
Beds	85.87	32	126.14	6	856
Medical personnel	137.57	52	212.58	3	1453
Equipment	96.02	50.5	112.31	3	776
Environmental variables					
Per-capita GVA	3081.89	2653.4	1699.52	646.32	6388.77
Density	288.02	129.32	502.56	0.4	4271.17
Occupancy rate	59.05	56.72	26.23	0.46	154.8

Source: The authors, based on information from INEC and BCE.

movement is likely to be directed to those developed regions, where more healthcare resources are concentrated.

4.6.2 Variables for the spatial Durbin interaction model

First, at the hospital level, we use the efficiency scores obtained in the first stage as a variable of hospital performance. The variable proxies the pulling effect for a hospital to attract patients (e_d in equation (4.23)). A negative sign of the efficiency variable *destination effect* means a good performance, attracting patients from other cantons.[30] The rationale could be twofold. On the one side, patients identify –to a certain extent– those best performing hospitals and prefer to travel to other canton (potentially the developed ones) to get treatment in what they perceive as the best facility. So, the efficiency performance of a hospital would also be explaining the quality perception of the patient. On the other side, this inflow of patients can also be explained by referrals from low-tech hospitals that might not have enough resources to treat a complex pathology. Unfortunately, we do not account with information regarding hospital referrals in our dataset to test this hypothesis. However, in both hypotheses, the significative effect of the efficiency performance is helping to explain the patient interregional mobility and the quality perception either by the patient or the hospital that is referring the patient (or both).

[30]Recall that hospitals with efficiency values higher than one are technically inefficient hospitals. Hence, a negative relationship with patient flows would mean that the best-performers are attracting more patients.

We proxy the cantonal level variables (X_o in (4.23)) that will impact on patients' decision to look for medical treatment with five variables. First, we use GVApc and population density to proxy the level of development of the region. As in the first stage, these variables can explain the regional heterogeneity that characterizes the country. Hence, it is very likely that the most important hospitals where population and economic activities are more concentrated are located in developed cantons, and, over time, this can foster quality differentials (Balia et al., 2020). In Ecuador, this statement has been empirically demonstrated in the previous chapters. The use of these variables to explain patient mobility has been extensively applied in the literature (see for example Cantarero, 2006; Fabbri and Robone, 2010). We use cantonal mortality (per 1000 individuals) and the insured population rate to proxy the healthcare conditions in the region, and control for the accessibility to medical treatment. The intuition is that higher mortality rates would be associated with poorer healthcare conditions in the canton.

Finally, we control for the poverty level in the canton by introducing the unsatisfied basic needs index (NBI). The index was developed by The Economic Commission for Latin America and the Caribbean (ECLAC / CEPAL by their Spanish acronyms), and has been widely applied in Latin American countries since the 1980s as a multidimensional measurement of poverty (CEPAL, 2007). Considering that poverty is a complex and multidimensional phenomenon, the NBI evaluates different dimensions of deprivation of goods and services required to the satisfaction of basic needs. In Ecuador, these dimensions comprehend economic capacity, basic education access, housing access, basic services access, and overcrowding. As stated in equation (4.7), richer patients are more prone to choose cross-border healthcare, we try to proxy this dimension of regional patient heterogeneity with the poverty index.

Table 4.2 presents the descriptive statistics of the variables used in the SDM model. Additionally, Figure 4.1 shows the distributions of hospitals by efficiency performance (e_d in (4.23)) at the top panel (a), and the migration flow dynamic of the sample (y in (4.23)) at the bottom panel (b). Panel a) of the Figure shows the most efficienct hospitals (that is, the hospitals with an efficiency value lower than 1) to be mainly concentrated in two of the most developed regions of Ecuador, where most of the healthcare resources are located.[31]

The panel b) of Figure 4.1 shows the patient flows from origin to destination, organized by intervals. We observe that there is a clear dynamic of patients traveling to the regions where the best performing hospitals concentrate. We can appreciate that most of the patient inflow is coming from neighboring cantons, which is a first signal of potential spatial autocorrelation in the migration flow. Hence, we use spatial interaction models that allow to disentangle the spillover effects of this migration dyad. Our empirical strategy begins by running the spatial interaction model, specified in (4.17), (4.18) and (4.19) to determine the econometric specification that better fits our data.

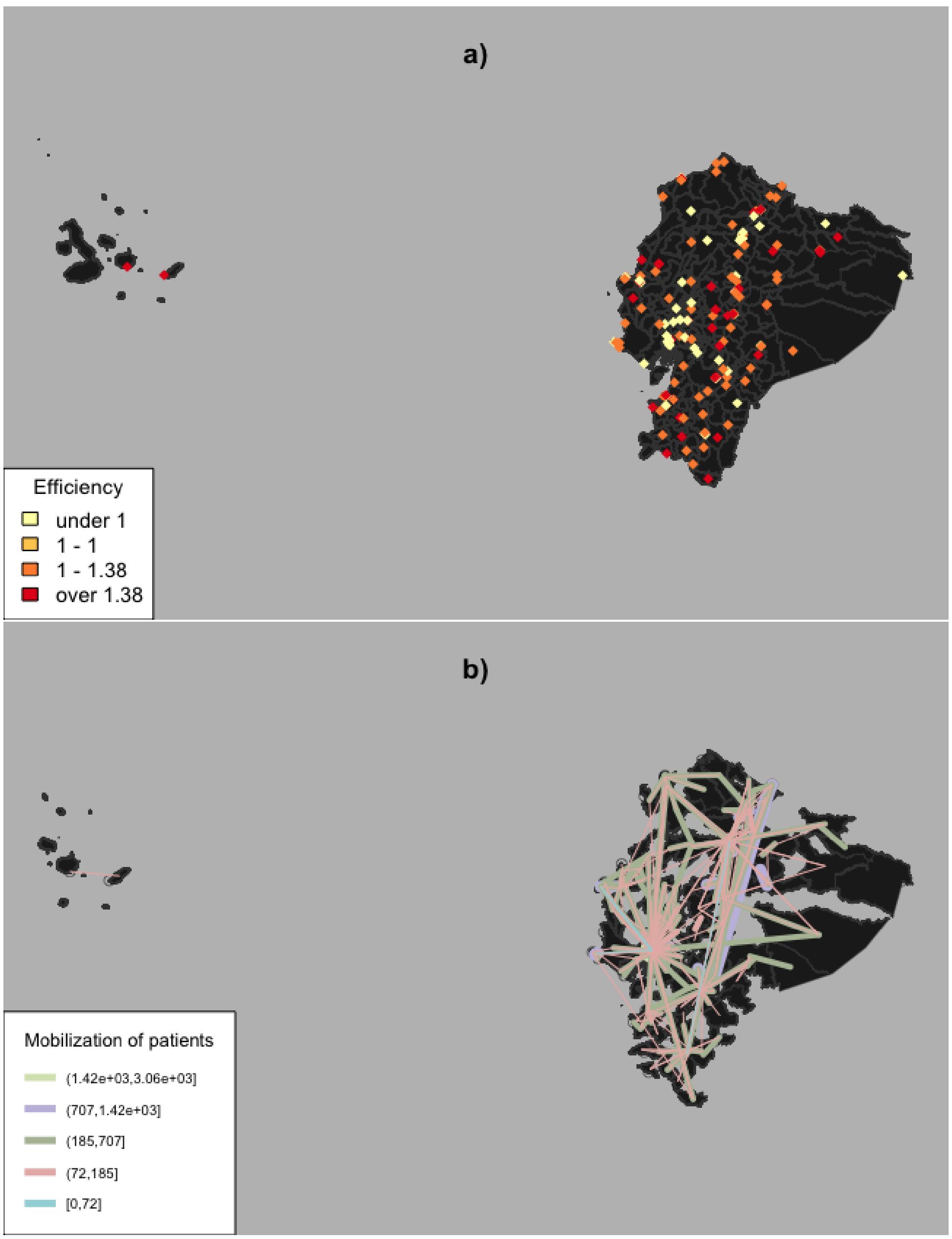
[31] These hospitals concentrate mainly in Quito and Guayaquil which are the two bigger and most developed cantons in Ecuador (Mendieta Muñoz and Pontarollo, 2016).

Table 4.2: Spatial interaction model variables' summary statistics

	Mean	SD	Min	Max
Order-m conditional efficiency	1.3207	0.7192	0.7264	5.7265
Per-capita GVA	3081.8877	1699.5163	646.315	6388.7741
Density	288.016	502.5621	0.3954	4271.174
Cantonal mortality	3.9611	1.206	0.5902	5.8391
NBI	0.6395	0.1832	0.297	0.987
Insured rate	0.2291	0.0878	0.0519	0.4844

Source: The authors, based on information from INEC and BCE.

Figure 4.1: Hospital efficiency and patient migration flows



4.7 Results and discussion

Table 4.3 presents the estimation results of the interaction model (4.17), adjusting for the intraregional patient flows in column (1). Column (2) incorporates the SLX interaction model (4.18); while column (3) includes the spatial lags of the distance variable g, of equation (4.19). As the traditional gravity model posits, the flows are inversely proportional to distances, as shown by the negative and statistically significant effect of distance (g). The estimated parameters are statistically different from zero. Although, as assessed in Section 4.4, they should not be interpreted as partial derivatives (LeSage and Fischer, 2016).

We can additionally use the estimates in Table 4.3 to emphasize that a non-spatial specification could suffer from omitted variable bias if the exogenous effects are not accounted for. This is endorsed by the fact that all the spatial lagged variables are significantly different from zero. The selection of the spatial specification in column (3) is endorsed by the Akaike and Bayes selection criteria –as well as the LR test and the R squared– as the best specification. Hereinafter, we will refer to this model as the baseline model.[32] [33]

Table 4.3: Spatial interaction model

	(1)	(2)	(3)
Constant	7.089***	6.462***	6.398***
	(0.27)	(0.43)	(0.43)
α_i	-7.148***	-6.535**	-5.51**
	(2.58)	(2.58)	(2.61)
log conditional efficiency	-0.256***	-0.235***	-0.234***
	(0.03)	(0.03)	(0.03)
log GVApc	0.309***	0.206***	0.207***
	(0.02)	(0.02)	(0.02)
log density	-0.015**	0.041***	0.045***
	(0.01)	(0.01)	(0.01)
log cantonal mortality	-0.099***	0.098***	0.11***
	(0.03)	(0.04)	(0.04)
log nbi	0.327**	0.043	0.1
	(0.14)	(0.17)	(0.17)
log insured	1.883***	0.886***	0.937***
	(0.19)	(0.23)	(0.23)
log conditional efficiency_i	-0.913***	-0.911***	-0.918***
	(0.33)	(0.33)	(0.33)
log GVApc_i	1.245***	1.271***	1.221***

[32] We test the the absence of spatial autocorrelation for the OD, patient migration flow, using the Moran test with both W_d and W_o spatial weight matrices. The tests reject the null of absence of spatial autocorrelation with Moran's I values of 0.5055 and 0.036, respectively.

[33] We tested the direction of the causality between migration flows and hospital efficiency by means of Granger (1969) causality test. The test rejects the null hypothesis of non-causality.

Table 4.3 (continued)

	(1)	(2)	(3)
	(0.24)	(0.24)	(0.24)
log density_i	-0.095	-0.103	-0.098
	(0.06)	(0.06)	(0.06)
log cantonal mortality_i	1.448***	1.475***	1.404***
	(0.34)	(0.33)	(0.34)
log nbi_i	5.137***	5.252***	4.76***
	(1.58)	(1.57)	(1.57)
log insured_i	5.909***	6.019***	5.995***
	(1.99)	(1.98)	(1.98)
W_d log conditional efficiency		-0.119***	-0.115***
		(0.04)	(0.04)
W_o log GVApc		0.217***	0.224***
		(0.03)	(0.03)
W_o log density		-0.089***	-0.093***
		(0.01)	(0.01)
W_o log cantonal mortality		-0.435***	-0.431***
		(0.05)	(0.05)
W_o log nbi		1.395***	1.352***
		(0.21)	(0.21)
W_o log insured		1.48***	1.433***
		(0.27)	(0.27)
log g	-0.741***	-0.771***	-0.659***
	(0.01)	(0.01)	(0.06)
W_d log g			0.085*
			(0.05)
W_o log g			-0.2***
			(0.05)
N	18656	18656	18656
Adj R-squared	0.4484	0.4572	0.4578
LogLik	-25844.97	-25692.14	-25681.66
AIC	51721.9358	51428.2796	51411.3173
BIC	51847.2786	51600.6259	51599.3315

Note: Dependent variable is the vector of (logged) migration flows. Estimations obtained by ML. Standard errors in parenthesis.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: The authors.

Once we have identified our baseline model, we estimate the SDM model as in equation (4.23). The Bayesian MCMC estimates based on 1000 draws are presented in Table 4.4. Lower and upper 0.05 and 0.95 credible intervals are reported, as well as the t-statistic.

The estimates show not just a high level of destination-based spatial dependence, but origin-based and origin-destination-based spatial dependence as well. The coefficients ρ_d and ρ_o are 0.31 and 0.53, respectively. The estimated parameter ρ_w is -0.11 and statistically different from zero. The 95 percent intervals suggest a small standard deviation and hence, big precision on the estimation.

These results provide evidence of the existence of spillover effects arriving from patient migration flows. Destination-based spatial dependence posits that flows coming from a given canton of origin to a destination hospital creates similar flows to neighboring hospitals (located in the same destination canton). In addition, origin-based spatial dependence shows that flows from any origin canton to a destination hospital creates similar flows from neighboring origins. Finally, origin-destination spatial dependence evidence that larger outflows from cantons neighboring the origin generate larger inflows to hospitals neighboring the destination. These findings point out the existence of spillovers steaming not just among cantons, but within cantons.

As noted by Thomas-Agnan and LeSage (2014) and LeSage and Thomas-Agnan (2015), the coefficients and *t*-statistics reported in Table 4.4 should not be interpreted as reflecting the partial derivative effects of changes in origin and destination characteristics. In turn, we need to calculate, origin, destination, and network summary measures to draw valid inferences on how changes in origin and destination characteristics impact the decision of patient migration flows.

In this respect, Table 4.5 reports the scalar summary effects for the model (4.23). In terms of hospital efficiency, the estimates show a significant expected negative effect. The increase in efficiency of an observed hospital leads to higher inflow of patients. Specifically, a 1 percent increase in efficiency on an average hospital would lead to a 0.3 percent increase in patient inflows.[34]As mentioned, these results are supporting the hypothesis that patients are selecting those hospitals that present a higher performance as more qualified. Higher efficiency performance seems to be working as a pull factor that attracts patients from neighboring regions. This effect can also be arising from patient referrals from other (low-performer) hospitals, which do not account with the necessary resources to treat complex pathologies. The information available in our dataset does not allow us to disentangle the size of these effects. We leave this question to be explored in future research.

Interestingly, hospital efficiency is also displaying a significative and negative network effect. This means that 1 percent increase in the efficiency of a given hospital is increasing the patient movements going to neighboring hospitals –different from their initially preferred hospital of destination– in 0.15 percent. These finding goes in line with our findings in Chapter 3, suggesting a competitive effect where higher efficiency in neighboring hospitals increase patient inflows.

Changes in the characteristics of the canton of origin provide additional information on the patient travel decision. For example, the positive and significant impact of GVApc

[34]Note that values bigger than 1 are inefficient.

Table 4.4: Spatial Durbin interaction model

	Mean	Lower 0.05	Upper 0.95	t-stat
Constant	0.5591	-0.0081	1.1639	1.6045
α_i	-1.2612	-4.7121	2.0721	-0.6144
log conditional efficiency	-0.1248	-0.1677	-0.0821	-4.7200
log GVApc	0.1351	0.1051	0.1643	7.5366
log density	0.0413	0.0296	0.0521	6.0521
log cantonal mortality	0.1149	0.0683	0.1603	4.0281
log nbi	-0.1674	-0.3791	0.0560	-1.2582
log insured	0.3172	0.0154	0.6082	1.7675
log conditional efficiency_i	-0.8186	-1.2245	-0.4078	-3.2243
log GVApc_i	0.4082	0.1022	0.7233	2.1771
log density_i	-0.0593	-0.1415	0.0225	-1.1901
log cantonal mortality_i	0.1978	-0.2413	0.6655	0.7218
log nbi_i	4.5273	2.3865	6.6296	3.5703
log insured_i	3.1162	0.4836	5.6679	1.9522
W_d log conditional efficiency	-0.0251	-0.0731	0.0267	-0.8293
W_o log GVApc	-0.0332	-0.0787	0.0096	-1.2351
W_o log density	-0.0565	-0.0727	-0.0397	-5.3821
W_o log cantonal mortality	-0.2131	-0.2728	-0.1491	-5.5530
W_o log nbi	0.7526	0.4872	1.0264	4.7012
W_o log insured	-0.3580	-0.6936	-0.0311	-1.6957
W_d log g	0.5162	0.4523	0.5792	13.3431
W_o log g	0.5477	0.4866	0.6114	14.4044
log g	-1.1637	-1.2472	-1.0820	-22.8236
ρ_d	0.3068	0.2967	0.3172	29.2126
ρ_o	0.5346	0.5283	0.5442	31.4697
ρ_w	-0.1085	-0.1245	-0.1004	-4.0614

Note: Dependent variable is the vector of (logged) migration flows. Bayesian MCMC estimates based on 1000 draws. N=18656 Source: The authors

Table 4.5: Scalar summary effects

	Mean	Lower 0.05	Upper 0.95	t-value
Destination Effects				
log conditional efficiency	-0.3015	-0.3987	-0.2026	-4.9594
Origin Effects				
log GVApc	0.2185	0.1653	0.2698	6.7937
log density	0.0473	0.0322	0.0620	5.1626
log cantonal mortality	0.1155	0.0414	0.1904	2.4933
log nbi	0.0558	-0.2490	0.3811	0.2817
log insured	0.4012	-0.0441	0.8513	1.4762
Network effects				
log conditional efficiency	-0.1529	-0.2525	-0.0590	-2.5719
log GVApc	0.1725	0.0214	0.3221	1.8814
log density	-0.1026	-0.1436	-0.0607	-3.9597
log cantonal mortality	-0.4405	-0.6302	-0.2635	-3.9308
log nbi	2.0840	1.2827	2.9013	4.4193
log insured	-0.5705	-1.5459	0.3676	-0.9851

Note: Dependent variable is the vector of (logged) migration flows. Bayesian MCMC estimates based on 1000 draws. N=18656

Source: The authors

could be measuring the (ceteris paribus) wealth effect of the origin. If the GVApc of the canton of origin increased, patients would have higher resources to devote to traveling costs to get medical treatment in other regions, such to create a pushing effect in that region. The positive and significant network effects of GVApc point to an increase in outflows from cantons neighboring the origin, when their wealth increases. This is supporting our assumption that regional income level is going to be a determinant of cross-border patient migration, as stated in Section 4.3.

Furthermore, densely populated cantons with high mortality rates are expected to push away patients, as expected. However, it is interesting to observe a negative network effect for both these variables. An explanation to the latter could be that high density and mortality in a neighboring canton reduces the incentives of patients to seek treatment in other regions different from their origin.

Before drawing any conclusions, we need to test the robustness of our results. Thus, we provide a robustness analysis in Section 4.8.

4.8 Robustness analysis

In order to check the robustness of our results, we carried out several tests. First, we want to test our efficiency estimator. In so doing, we perform a sensitivity analysis of the order-m estimation to different m values of peers randomly drawn from the population. We simulate different scenarios of estimated efficiencies, with $m = 1, \ldots, 150$, and test the difference in distributions between m and $m+1$ (H_0 : efficiency m = efficiency $m+1$) by means of Simar and Zelenyuk (2006) adaptation of the Li test for unknown distributions.[35] Cazals et al. (2002) show that, when m increases and converges to ∞, the order-m estimator converge to the full frontier. Hence, for a finite m the frontier will not embed all the data points and so is much more robust than other classic non-parametric approaches (like Data Envelopment Analysis of Free Disposal Hull) to outliers. The results show a convergence after $m = 30$ (depicted in Figure 4.2 Appendix 4.10.4) where H_0 cannot be rejected. Therefore, we can confirm that there are no significant differences within the range of the m value selected.[36]

A second concern is the validity of environmental variables included in the conditional order-m estimation. We rely on the fact that the level of development of a canton has an external effect on the efficiency performance of hospitals located within. So, best performers would be located in developed regions and would attract more patients. To find out whether environmental variables have a significative effect on the production of healthcare, we follow the procedure described in Daraio and Simar (2005) and Daraio and Simar (2007b) and regress the ratio $\widehat{R}$ of estimated conditional and unconditional efficiency scores ($\widehat{R} = \frac{\widehat{\theta}_m(x,y|z)}{\widehat{\theta}_m(x,y)}$) on the environmental variables Z, using a non-parametric smoothed regression. As stated by Daraio and Simar (2005), in an output oriented framework, an increasing regression means a favorable Z: the environmental variable acts as a sort of "extra input" favorable for the production process.[37] Conversely, an unfavorable Z would be observed with a decreasing regression, where the environmental variable is –in a certain sense– penalizing the production of the outputs of interest. Then, we test the significance of each variable.

The results show a significant and favorable impact of GVApc (p-value = 0.004) and occupancy rate (p-value = 2e-16) on $\widehat{R}$ (Figure 4.4 of Appendix 4.10.4) at the 99% confidence level.[38] This validates our hypothesis that hospital performance is being affected by

[35]The Li (1996) method relies on kernel smoothing to non-parametrically estimate two density functions. Simar and Zelenyuk (2006) modify this method in order to test efficiency values estimated by non-parametric approaches and provide consistent bootstrap estimates of the p values of the Li test.

[36]Recall that we have fixed $m = 90$.

[37]The value of $\widehat{\theta}_m(x, y \mid z)$ would be smaller (more efficient) than $\widehat{\theta}_m(x,y)$ for small values of Z than for large values. Hence, $\widehat{R}$ will increase with Z, on average.

[38]Although density does not seem to have a significant effect on $\widehat{R}$, the results do not vary when we take it off the efficiency estimation.

the regional income levels, and this effect is being captured with the conditional model.[39]

Another point to test is the endogeneity of the efficiency value. As a random variable, there is the possibility that it is correlated with the error term. To test the hypothesis of no endogeneity (H_0 : true correlation equal to 0) we perform a t test between the efficiency score and the error terms after running equation (4.19). The test confirm that we do not suffer from endogeneity in the efficiency term (p-value = 1).[40]

There is also the possibility that the effect of hospital efficiency that we find may be biased given that our dataset si constructed using information from one single year. In this respect, patients may be driving their decisions based upon the perceived hospital performance of previous years. To test this hypothesis, we retrieved the hospital information coming from the Annual Survey of Hospital Beds and Discharges and the Survey of Health Activities and Resources of 2013. We repeated our approximation and calculated the hospital efficiency score with equation (4.15) and used it in a second stage to estimate equation (4.23). We find very little variation in the size of the scalar summary values with an efficiency destination and network effects of -3.059 and -0.017, respectively.

Regarding the spatial econometric specification, we test the robustness of the estimations from equation (4.23) with a new efficiency value. In so doing, we calculate the efficiency value of equation (4.15) taking out the emergency consults from the outputs. We consider this alternative estimation of hospital efficiency given that the patients do not usually have a decision over the hospital where they get treatment in these cases. The destination and network marginal effects of the hospital efficiency are significant and comparable (-0.303 and -0.147, respectively).

In addition, we examine whether the results are sensitive to alternative specifications of the spatial weight matrix W_d. Rather than considering the neighboring dimension for hospitals that are located in the same canton, we chose to consider those hospitals located within time travel distance radius. Thus, we define W_{dt} to be the inverse of the shortest time travel distance by car between any pair of hospitals.[41]In addition, we consider remoteness between hospitals by introducing the inverse of the squared travel time distance for the weight matrix, W_{dt}^2, so closer hospitals receive a higher weight.

Table 4.6 show the destination and network effects estimated for our variable of interest, as well as the parameters ρ_d, ρ_o and ρ_w corresponding to each weight matrix after running equation (4.23). The results for the destination effects are robust and comparable. The network effects are not statistically significant with W_{dt}^2, which suggest that the competition effect (in efficiency) among hospitals is diminishing for those that locate further apart. Regarding the spillover effects on migration flows, the results are robust and

[39]Note that occupancy rate also presents a significant favorable effect on efficiency, which is signaling that hospitals are making a better use of their resources and capacity to treat incoming patients as found in Chapter 3

[40]We performed the same test for all explanatory variables, with comparable results to those of efficiency.

[41]Defining the spatial weight matrix using a measure of distance between spatial units has commonly been used in the literature when the data covers healthcare providers (Tosetti et al., 2018)

comparable in size for origin and destination spillovers (ρ_o, ρ_d, respectively), but loose significance for origin-destination spillovers. The intuition behind could be associated with the proximity of hospitals within the region. By using W_{dt} and W_{dt}^2, we consider a given hospital as neighbor if it is located within a radius, so origin-destination spatial spillover effects do not seem to be happening to those immediate neighbor hospitals but rather on those located throughout the region (when we use W_d).

Table 4.6: Scalar summary effects, using W_{dt} and W_{dt}^2

	W_{dt}				W_{dt}^2			
	Mean	Lower 0.05	Upper 0.95	t-value	Mean	Lower 0.05	Upper 0.95	t-value
log conditional efficiency (Destination Eff.)	-0.2139	-0.3248	-0.1348	-3.4104	-0.1798	-0.2653	-0.0600	-2.9476
log conditional efficiency (Network Eff.)	-0.4854	-0.8970	-0.2000	-2.3507	-0.4017	-0.7996	-0.0368	-1.5881
ρ_d	0.4757	0.4144	0.4964	15.7229	0.4308	0.2275	0.4695	5.9839
ρ_o	0.5159	0.3334	0.5428	3.9444	0.5160	0.4362	0.5339	17.8886
ρ_w	-0.1695	-0.2249	1.6087	0.0107	-0.1078	-0.1532	0.0357	-1.1939

Note: Dependent variable is the vector of (logged) migration flows. Bayesian MCMC estimates based on 1000 draws. N=18656
Source: The authors

Another dimension to check the robustness of the results is by considering the spectrum of treated diseases. There is the possibility that the pulling effect could be mainly driven by the presence of specialized hospitals versus other basic hospitals that provide another scope of treatments. Thus, we split the sample in two different subgroups by distinguishing between basic and specialized hospitals (this latter include chronic and acute hospitals).[42]

Table 4.7 present the scalar summary effects for efficiency and the parameters ρ_d, ρ_o and ρ_w for each hospital type. It is not surprising to note that the destination effect for basic hospitals disappears, suggesting that the pulling effect of hospital efficiency performance is mainly being captured by specialized hospitals, because the magnitude of the estimation is larger. As basic hospitals spread across the country, what seems to be driving people to travel to high-income regions is the performance of specialized medical institutions, which are more concentrated in those cantons (see Figure 4.4 of Appendix 4.10.5). However, high performance of an average basic hospital is not enough to attract interregional patients as they are prone to receive medical attention in their local hospital to treat a common disease. Instead, in the case of specific or severe illnesses, patients will select a particular hospital on the basis of the quality of the treatment they perceive they will attain over there, which is being captured by our efficiency variable. Nevertheless, spillover effects are still statistically robust and comparable in both the cases, which means that both arrangements are valid to guarantee patient mobility across the territory. One explanation endorsing these results (particularly for basic hospitals) is that even though the increase in efficiency of a given hospital is not enough to attract intraregional patients,

[42]In our database, acute hospitals embed infectious hospitals, obstetric-gynecological hospitals, pediatric hospitals, general hospitals that treat acute diseases and other hospitals of specialization. Whilst chronic hospitals embed oncology and pneumology hospitals

those hospitals are taking advantage of patient inflows, initially attracted by other hospitals (most likely neighbor specialized hospitals).[43]

Table 4.7: Scalar summary effects by hospital type

	Basic (N=7743)				Specialized (N=2610)			
	Mean	Lower 0.05	Upper 0.95	t-value	Mean	Lower 0.05	Upper 0.95	t-value
log conditional efficiency (Destination Eff.)	-0.0083	-0.0828	0.0664	-0.1822	-0.8088	-1.1888	-0.4502	-3.7294
log conditional efficiency (Network Eff.)	0.3624	0.2718	0.4556	6.3162	-0.7771	-1.0282	-0.5211	-5.0242
ρ_d	0.1521	0.1383	0.1634	12.1592	0.3325	0.2874	0.3690	15.6936
ρ_o	0.2709	0.2586	0.2834	31.9435	0.4225	0.4039	0.4464	12.1845
ρ_w	-0.0324	0.0458	-0.0196	-3.7970	-0.1222	-0.1138	-0.1501	2.1341

Note: Dependent variable is the vector of (logged) migration flows. Bayesian MCMC estimates based on 1000 draws.
Source: The authors

Considering that the efficiency performance of specialized hospitals seems to be the main determinant to attract interregional patients, one could doubt that patients choose to travel because they want to receive a better treatment than what they could obtain in their local hospital, but because there are no other alternatives to treat their disease. Hence, the decision to travel may be forced by the complexity of the treatment, which is not available in the hospital of their respective region. To corroborate this, we provide a frequency table of the top five morbidity causes of interregional patients (i.e., patients that get medical attention in a hospital located in a canton different from where they reside) treated in specialized hospitals, in Appendix 4.10.6 (Table 4.9). We observe that the main causes of (interregional) patient migration are mainly related to pregnancy (with more than 6% of treated patients). The intuition behind lead us to think that, being pregnancy-related treatments something that is usually planned and monitored, and could be carried out in any hospital, patients are choosing to incur in travel expenses to receive the best treatment possible in their closest best-performing hospitals (located in developed cantons).[44] This is backed up in Table 4.10 of Appendix 4.10.6, where we present the amount of patients treated in hospitals located at the three high-income cantons in Ecuador (Quito, Guayaquil and Cuenca), divided by the patient's province of residence.[45] The table shows that, for example, in Cuenca the majority of interregional patients belong to neighboring cantons located in the same province (and that holds for Quito and Guayaquil).

Furthermore, we plot two figures in Appendix 4.10.6. Figure 4.5 describes the demographics (available in our dataset) of the patients with the top five morbidity causes, while Figure 4.6 describes the public entity embedding the public hospital (MSP, Social Security

[43]For example, patients traveling to get specialized medical attention could incur in additional costs that are not covered by their insurance, but similar treatments could be offered in alternative public hospitals. Other scenario could imply that patients would seek medical attention in adjacent hospitals if the waiting time for specialized ones is long enough.

[44]Note that other morbidity causes relate to appendicitis or calculus of the gallbladder, which are not as complex as cancer, for example.

[45]Remind that, in Ecuador, the provinces are the first-level administrative division. The cantons of Quito, Guayaquil and Cuenca belong to the provinces of Pichincha, Guayas and Azuay, respectively.

hospitals, and other public hospitals patronized by their respective municipality). Unfortunately, our database does not account with information about the patient's income level, but it includes their self-perceived ethnicity which can be used to proxy this variable.[46] Figure 4.5 shows that more than 60% of the interregional patients describe themselves as mixed-race. In addition, Figure 4.6 shows that mixed-race and white patients are the ones that make use of the social security institutions. These former are the ones that account with a formal job, hence, having access to social security services. More than 90% of indigenous and afro-ecuadorian (interregional) patients get medical attention in MSP hospitals (which offer free healthcare). The descriptives support our theory that wealthier patients are more prone to seek medical attention outside their area of residence. Also, they seem to be choosing to go beyond the regional borders to treat their pathologies, rather than being forced by the complexity of their disease.

Our results open up an important discussion in terms of policy implications. Hospital efficiency performance seems to be capturing a deal of quality perception by public hospital patients that cannot be neglected. In this respect, policy makers need to take into consideration that the effect of an unexpected healthcare reform could entail a broader spectrum of consequences beyond the ones addressed to those healthcare institutions initially targeted. For example, new reforms that decrease the barriers to access to more specialized and sophisticated treatments (only available in specialized hospitals) need to be well planned and allocated. If the increase of the demand driven by these reforms is not controlled, they could lead to congestion effects that can impact the performance of specialized hospitals. Due to spillover effects, neighboring hospitals (including the basic ones) could experience detrimental consequences,[47] leading to a deterioration of the regional healthcare performance.

So far, Ecuadorian healthcare reforms have been accompanied by an increase of hospital efficiency, and hospitals adapted the spare resources to treat the higher inflow of patients, but those reforms have been mainly focused on offering general treatment in public hospitals that are abundant and spread around quite homogeneously across the country. However, there is a lower supply of specialized hospitals which are much more territorially concentrated. These findings highlight the importance to implement tailored regional healthcare policies.

As Brekke et al. (2014) suggest, high-income regions could be benefiting from welfare improvements, as we found a competition effect in efficiency among hospitals within the same regions that leads to higher regional performance and quality. However, the welfare effects could generate asymmetric effects as low-income regions are not accounting with high-performing specialized hospitals, and only the patients that move to other regions benefit from these welfare improvements. Future public investment could be focused on

[46]Mixed-race population is more likely to belong to the middle-income class, while indigenous, afro-ecuadorian, and other indigenous ethnicities (apart from white) are more likely to belong the low-income class.

[47]For example, they could be obliged to attend bigger amount of complex pathologies for which they do not have the medical resources to treat.

increasing specialized services for hospital clusters of low-income areas. More supply of specialized hospitals could attract patients and motivate competition among hospitals to provide welfare improvements and reduce the quality gap between regions.

4.9 Conclusions

This study aims to analyze whether the higher efficiency performance of Ecuadorian public hospitals is resulting in a higher inflow of interregional patients to a destination hospital, and whether this is also leading to a higher inflow of patients to neighboring hospitals within the same region. To determine the effect of efficiency on the patient migration network, we follow an innovative two-stage strategy where the first step is to estimate robust conditional order-*m* efficiency values, based on the economic concept of Pareto efficient allocation and the second step makes use of a spatial Durbin interaction model to estimate the effect of hospital efficiency in patient migration flows, and separates the spillover effects in the form of larger inflows of patients for neighboring hospitals. We contribute to the empirical applied literature by estimating a model that considers different origins and destinations in the OD dyad, that—to the best of our knowledge—has not yet been applied.

We are referring to a structure in which regional disparities are modeled by means of healthcare asymmetries over time, producing a healthcare performance gap across regions and motivating a patient mobilization pattern since the majority of the influx of patients was concentrated in developed regions. Our results support the hypothesis that hospital efficiency performance is a strong pulling factor for this inflow, and the direction of this effect is robust according to different specifications and estimation methods. However, when we split the sample separating basic and specialized hospitals, this effect disappears for the former, but gets even stronger for the latter. In addition, we identify spillover effects in the mobilization flows, not just in the form of patients arriving at neighboring destination hospitals from an origin canton, but from patients arriving at a given hospital from cantons close to that origin, and arriving at adjacent hospitals as well.

This evidence has two implications. First, the efficiency effect suggests that patients are perceiving—to some extent—hospital performance as a proxy for hospital quality that is encouraging cross-border migration to receive a better medical treatment than what they can get in their local area. However, this decision is based on the availability of specialized hospitals in the destination region, which are mostly concentrated in highly developed areas. The possibility also exists that other hospitals are referring patients for complex diseases, as they do not possess the resources to treat them. Second, spillover effects present in the data are suggesting that other hospitals neighboring the specialized ones are also capturing some of those inflows of patients. According to Brekke et al. (2014), if there were competition among hospitals (which we find with the statistical significance of the network effects), this could entail a beneficial effect on the welfare of the population, as more competition encourages higher quality of care. However, hospitals from less-developed regions might not be benefiting from that welfare increase, as there is no incentive to provide better medical attention and hence just those that travel beyond regional borders may enjoy it.

Our results deliver useful suggestions for policy makers. On the one hand, new reforms need to be well-planned not just in terms of territorial discrepancies but also in terms of hospital specialization. For example, decreasing the limitations to specialized care could incur an increase of healthcare demand, that, if not controlled, could lead to negative consequences like congestion effects. Negative shocks to specialized hospitals induce a negative impact on their performance as well as the demand for the hospitals that surround them and as consequence, affects the efficiency of the hospitals of the whole region and the welfare of the population. Public authorities could identify those key players in the healthcare network to target strengthening reforms that could encourage better performance within the public healthcare system of the region due to spillover effects.

Public healthcare policy can devote a larger share of their resources to targeting investment in those less-developed regions. The significant origin-based spatial dependence suggests the existence of clusters of less-developed cantons that are recording an outflow of patients. If there were not enough demand for local hospitals to compete, there would be no incentive to increase the quality of care over there. Therefore, public investment could be focused on the creation of specialized hospitals –or specialized wards in existing hospitals– in these regions to attract more demand. Once the inflow of patients is stablished, new spillover effects could arise, benefiting adjacent hospitals and bringing improvements both for the regional healthcare performance and welfare so as to benefit the low-income patients of that place, who cannot afford to receive treatment in other cantons.

Finally, future research implications can be derived from this study. As pointed out, the effect of efficiency performance on migration flows could be driven by the perception of patients selecting a given hospital (where they perceive they could receive better medical treatment) or by other hospitals referring highly complex cases to those best-performers (or both). Unfortunately, our dataset does not account for information on patient referrals to disentangle the size of these effects, but it opens up interesting methodological research strategies to be investigated in future studies.

Further research can also aim to explore the determinants of maternal mobility. The exploratory analysis performed in our study points to an outflow of patients looking for obstetric services in (high-performing) specialized hospitals. These preliminary results suggest a bad quality of public obstetric healthcare in hospitals locating in less-developed areas. In this respect, future studies can focus on identifying the hospital's (or regional) features that produce this pushing effect. The results may be used to address important issues such as reducing child or maternal mortality in low-income areas of Ecuador.

4.10 Appendix

4.10.1 Institutional setting

The Ecuadorian healthcare system accounts for public and private service sectors. The public sector accounts for the majority of the insured population, with a 66% covered by the year 2014, according to the Survey of Life Conditions. Private insurance accounts for 6% only. The institutions belonging to the public healthcare sector are:

1. The Public Ministry of Health (MSP) and the Ministry of Social and Economic Inclusion (MIES), which provide health services to the whole population, including those that do not account with any type of insurance.

2. The social security institutions which embed the Ecuadorian Social Security Institute (IESS), the Social Security Institute of the Armed Forces (ISSFA) and the Social Security Institute of the National Police (ISSPOL). The former provides medical services to all social security contributors; while the latter two grant medical attention to the army and national police corps, respectively.

Ecuador is a country that has suffered from a continuous process of healthcare deterioration that began in the 1990s, with a period of democratic instability that hinged the performance of healthcare with a reduction of budget for healthcare provision, worsening infrastructure due to lack of investment, low quality of healthcare services and a deficient institutional structure (Granda and Jimenez, 2019).

In 2008 the new constitution came into force and many reforms have been carried out to promote access to medical treatment and reduce financial and social barriers to healthcare. For instance, the gratuity of medical services provided by the MSP in 2008 or mandatory enrollment of employees to social security in 2011. After the implementation of these policies, there was in increase in the annual growth rate of active beneficiaries (Orellana et al., 2017), and a rise of 40% of patients attended in public hospitals between 2006 and 2014. These reforms were supported by an increasing public investment for the core system, mostly involving the endowment of medical infrastructure and training.

4.10.2 Bayesian Markov Chain Monte Carlo estimation

In this Appendix, we describe the Bayesian MCMC robust estimation proposed in LeSage and Pace (2009). We depart from the spatial econometric interaction model specified in equation (4.20) and introduce a set of latent variance scalars for each observation, so we have:

$$\varepsilon \sim N(0, \sigma^2 \tilde{V})$$

$$\tilde{V}_{ii} = V_i, i = 1, ..., N$$
$$V = vec(R)$$

$$R = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n_d} \\ v_{21} & v_{22} & & v_{2n_d} \\ \vdots & & \ddots & \vdots \\ v_{n_o1} & & & v_{n_on_d} \end{pmatrix} \tag{4.24}$$

Estimates of the N variance scalars are obtained using an *iid* $\chi^2(\lambda)$ prior on each v_{ij} contained in matrix R, with mean of unity and a mode and variance that depend on the hyperparameter λ of the prior.

In order to obtain the MCMC estimations, we need to sample sequentially from the set of full conditional distributions for all the parameter of the model: $\delta, \sigma, \rho_d, \rho_o, \rho_w$ and $\tilde{V}_{ii}$, where $\delta = [\alpha, \beta_d, \beta_o, \gamma]'$.

The conditional distributions for δ and σ^2 are stablished by assigning uninformative priors to the parameters δ, and independent *inverse gamma* distribution ($IG(a,b)$, with $a = b = 0$) prior to σ^2. We rely on a uniform prior over the range $-1 < \rho_d, \rho_o, \rho_w < 1$ and impose stability restrictions such that $\sum_i \rho_i > -1, \sum_i \rho_i < 1, i = d, o, w$ using rejection sampling. The prior for the variance scalars v_{ij} are based on Geweke's *iid* chi-squared with λ degrees of freedom. The prior distributions, indicated with π are expressed as:

$$\pi(\delta) \propto N(c, T), c = 0, T \to \infty \tag{4.25}$$
$$\pi(\lambda/v_{ij}) \sim iid\chi^2(\lambda) \tag{4.26}$$
$$\pi(\sigma^2) \sim IG(a,b) \tag{4.27}$$
$$\pi(\rho_i) \sim U(-1,1), i = d, o, w \tag{4.28}$$

The full conditional distribution for the parameters δ, σ^2, and each variance scalar v_{ij} can be taken from LeSage and Pace (2009).[48] In addition, we need to sample each of the three parameters ρ_d, ρ_o, ρ_w conditional on the two other dependence parameters and the remaining parameters (δ, σ^2, V), which is carried out using a Metropolis-Hastings algorithm based on a tuned normal random walk.[49]

[48] Refer to LeSage and Pace (2009) Chapter 8
[49] Refer to LeSage and Pace (2009) Chapter 5

4.10.3 Variable description

Table 4.8: Variable description

Variable	Description	Variable construction
Output		
Number of discharges (weighted)	Treated patients in a given hospital	Number of discharges*Case-Mix index
Morbidity consults	Morbidity consults in a given hospital	Total number of morbidity consults
Emergency consults	Emergency consults in a given hospital	Total number of emergency consults
Survival rate	Rate of non-deceased discharged patients in a given hospital	1-hospital mortality rate
Inputs		
Number of physicians	Physicians and general physicians in a given hospital	Total number of physicians
Number of beds	Total amount of beds per hospital	Total number of beds
Number of hospital personnel	Medical staff not including physicians. E.g. Nurses, technologists, administrative staff, dentist, etc.	Total number of hospital personnel
Number of equipment and infrastructure	Physical infrastructure (surgery rooms, intensive care rooms, etc.) and medical equipment (imaging, diagnosis, sterilization, etc.)	Total number of equipment and infrastructure
Environmental Variables		
Per-capita GVA	Cantonal per-capita Gross Value Added	GVA/cantonal population
Density	Cantonal population density	Cantonal population/Canton's area in Km^2
Occupancy rate	Incoming patients days of care per beds available in a given hospital	(Inpatient days of care/Bed days available) *100
Cantonal Mortality	Percentage of deceased patients in a given canton (per 1000 population) relative to cantonal population	Cantonal mortality*1000
NBI	Percentage of households that present at least one unsatisfied basic necessity, relative to the total households in a respective canton	NBI_household/Total households
Insurance Rate	Percentage of insured population relative to the cantonal population	Population insured/cantonal population

Source: The authors.

4.10.4 Order-*m* robustness analysis

Figure 4.2: Order-*m* p-values

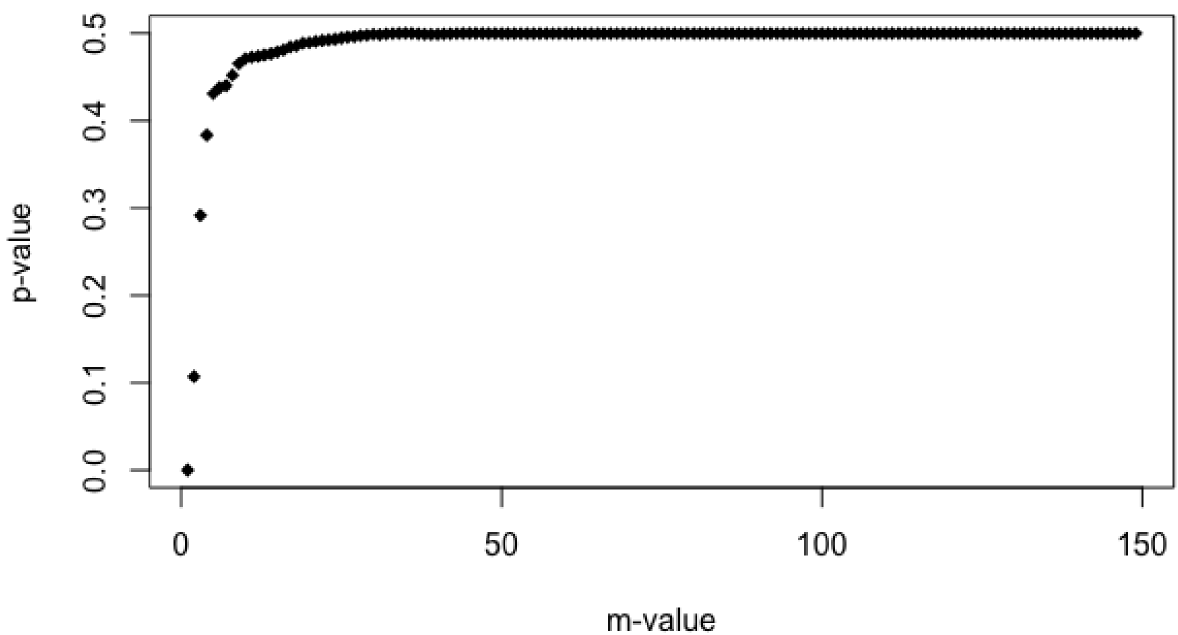
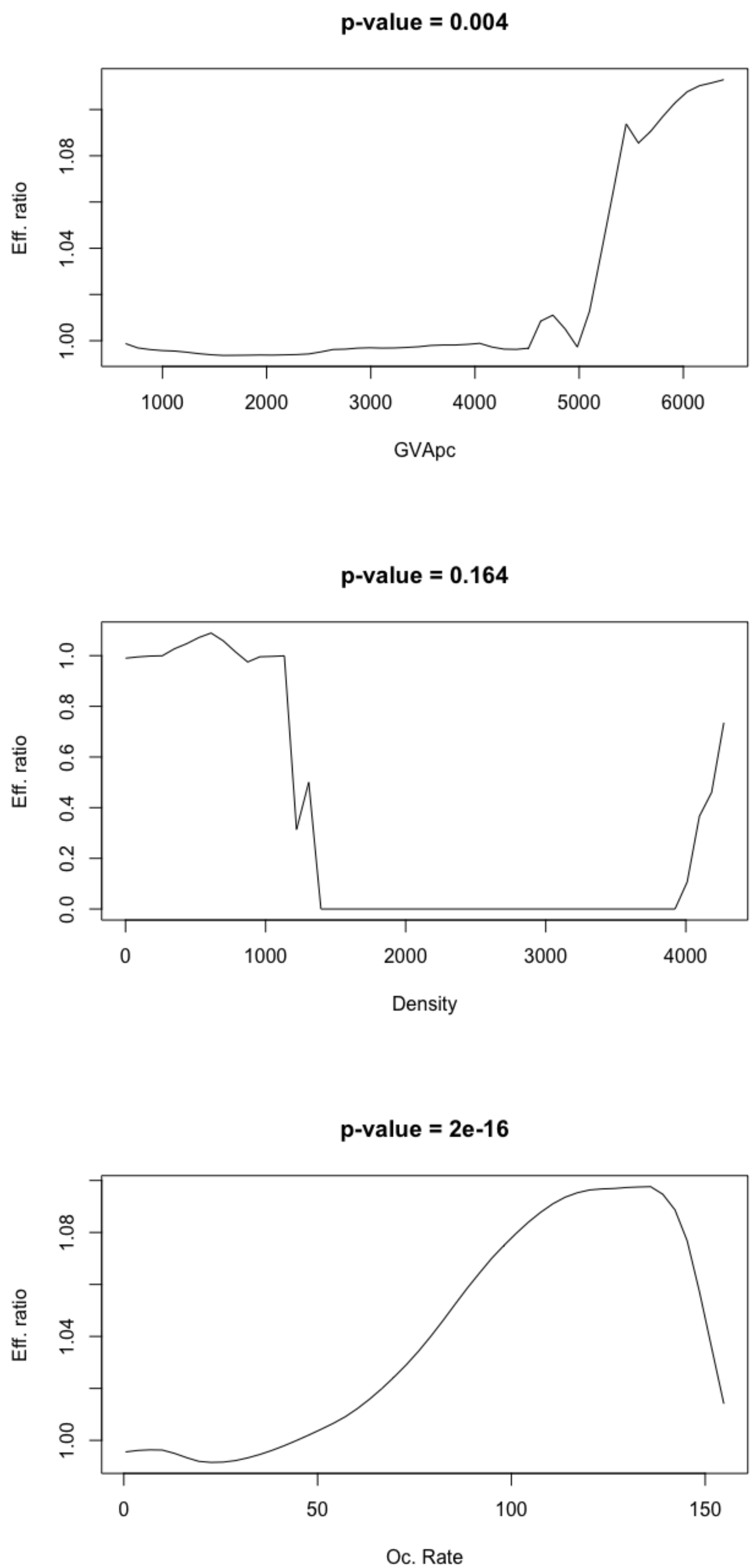
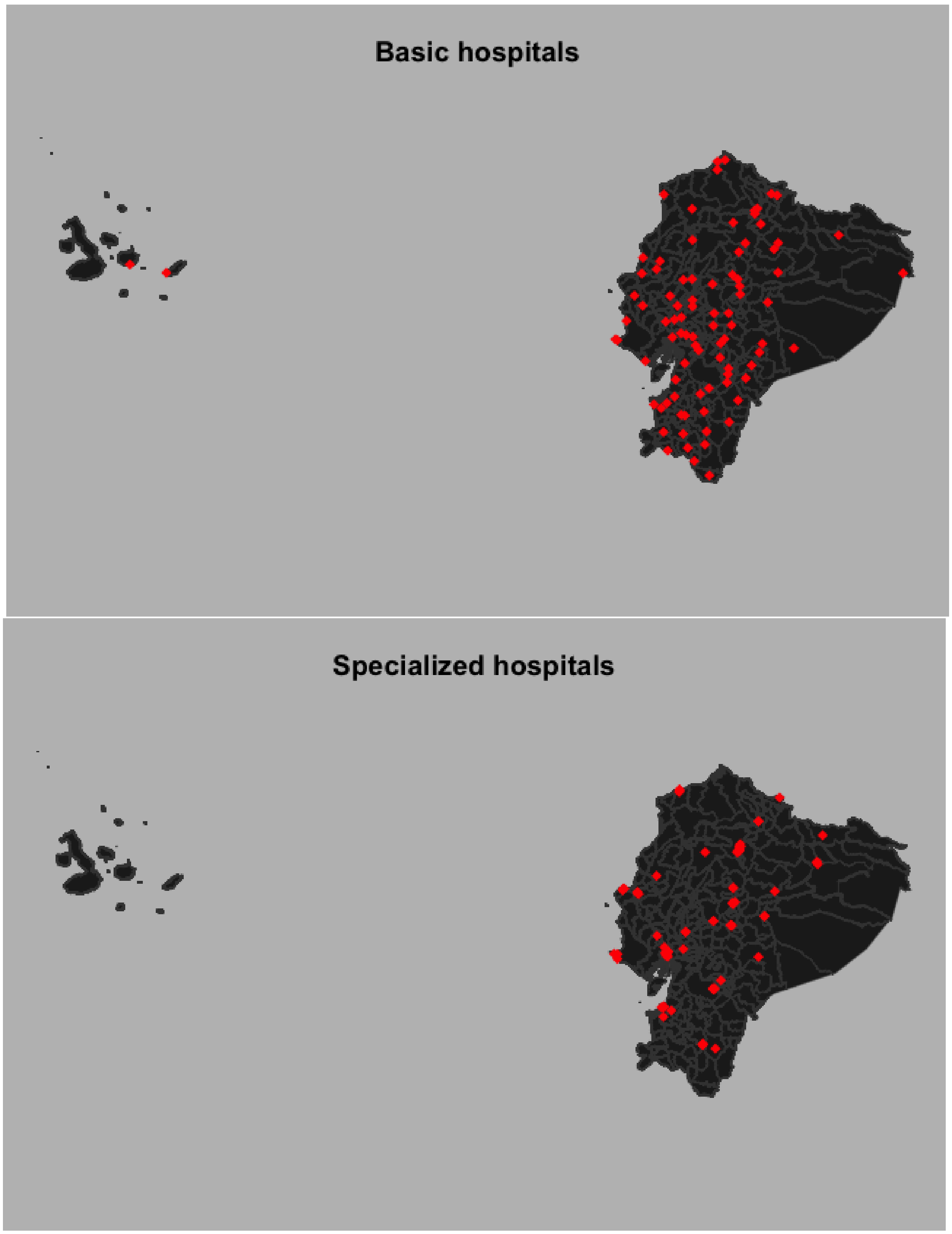


Figure 4.3: Conditional order-m partial regression plots



4.10.5 Hospital distribution

Figure 4.4: Territorial distribution of basic and specialized hospitals



4.10.6 Interregional patients' demographics

The information provided in this Appendix is collected for the interregional patients treated for the top five morbidity causes in specialized hospitals.

Table 4.9: Top five morbidity causes of interregional patients in specialized hospitals

Morbidity cause	Total patients	Percentage
Pregnancy (single spontanous delivery)	8307	4.66
Acute appendicitis	5068	2.84
Pregnancy (caesarean section)	3821	2.14
Calculus of the gallbladder (without cholecystitis)	3143	1.76
Pneumonia	3012	1.69

Note: Percentages calculated relative to the total amount of patients treated in specialized hospitals
Source: The authors.

Table 4.10: Total interregional patients by canton and province of residence

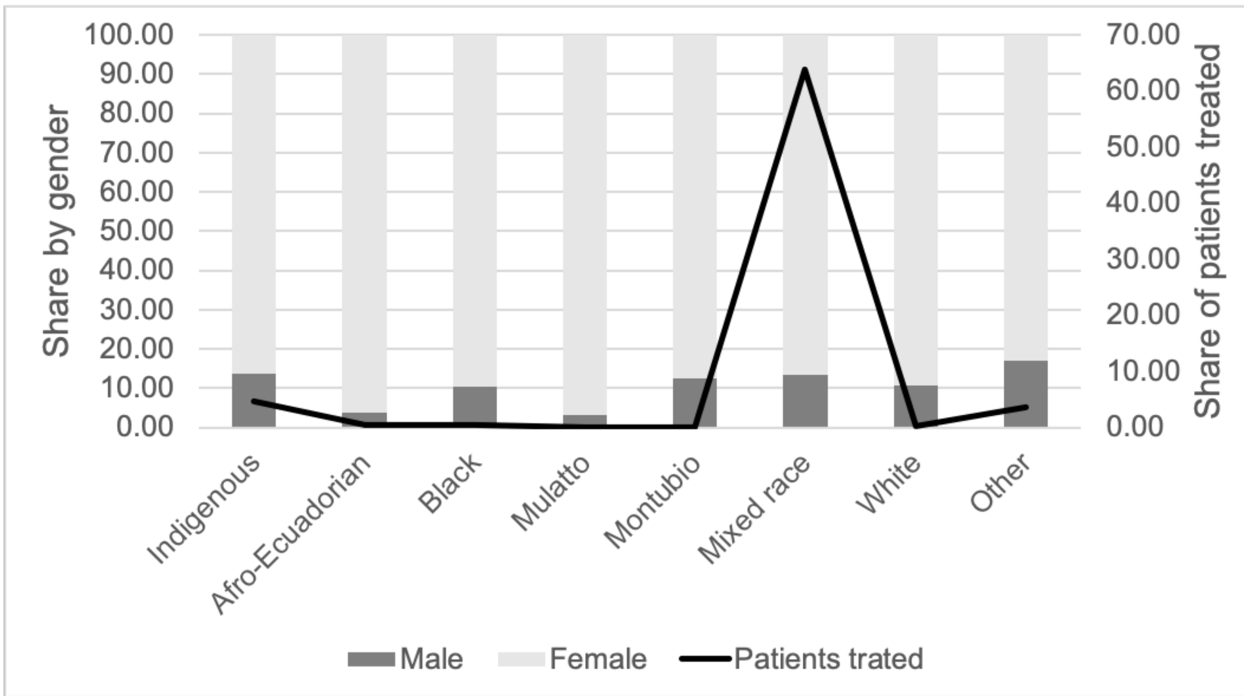
Province of residence/Canton of the hospital	Cuenca	Guayaquil	Quito
Azuay	284	18	1
Bolívar	1	23	40
Cañar	119	41	2
Carchi	0	1	35
Cotopaxi	1	12	76
Chimborazo	7	21	35
El Oro	49	50	15
Esmeraldas	0	52	62
Guayas	23	1709	9
Imbabura	1	3	89
Loja	28	8	14
Los Ríos	2	260	18
Manabí	7	152	43
Morona Santiago	62	1	7
Napo	0	1	20
Pastaza	0	15	10
Pichincha	4	15	588
Tungurahua	2	4	36
Zamora Chinchipe	5	2	1
Galápagos	0	5	3
Sucumbíos	1	3	37
Orellana	0	5	35
Santo Domingo de los Tsáchilas	0	16	63
Santa Elena	0	69	2

Table 4.10 (continued)

Province of residence/Canton of the hospital	Cuenca	Guayaquil	Quito
Exterior	1	2	1

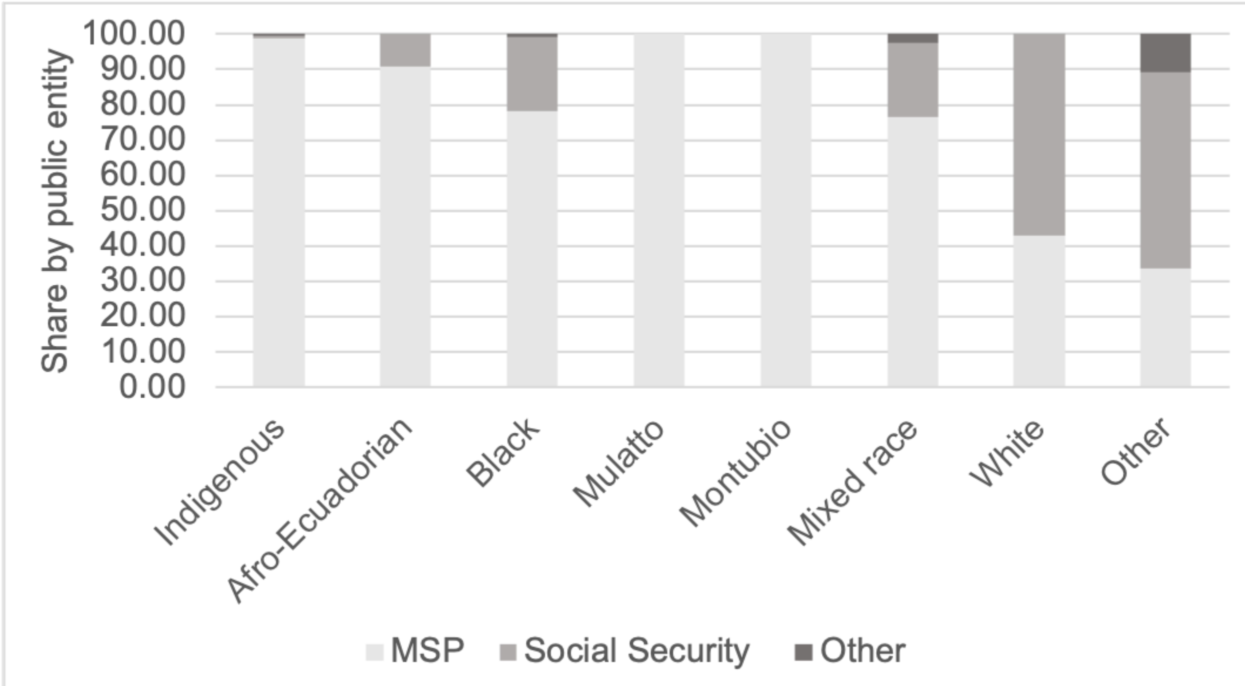
Source: The author.

Figure 4.5: Share of interregional patients by gender and ethnic group.



Note: Montubio is the name given to the peasant of the Ecuadorian coast

Figure 4.6: Share of interregional patients by hospital's public entity and ethnic group.



Note: Montubio is the name given to the peasant of the Ecuadorian coast

Chapter 5

Conclusions

The objective of this dissertation is to assess the efficiency performance of the public healthcare system of a developing country such as Ecuador. This country experienced a series of reforms aimed at guaranteeing higher equity in healthcare access and universal coverage to overcome a deficient healthcare system that was deteriorated by erroneous political reforms, economics crises and profound regional heterogeneities. With this in mind, our scope is to contribute to the academic discussion of healthcare efficiency in developing countries. This topic has been left aside in the literature and has received little attention in Ecuador.

The focus of our research hover at the implementation of the new constitution in 2008. This event represents a milestone in the Ecuadorian history, bringing upon new healthcare and social security reforms that decreased the barriers of access to medical attention to the population. These reforms were accompanied by a vast deployment of public investment, mainly directed towards new medical infrastructure and training to improve the healthcare system's performance. However, in the short-run, this performance could have been affected by the sudden increase in demand (derived from the healthcare reforms).

This dissertation has addressed selected and relevant issues associated with this effect in hospital efficiency performance. In Chapter 2, we take a first look to the evolution of hospital efficiency over the period 2006-2014 (considering two years before the new Constitution came into force), and analyze statistically significant changes on its behavior after 2008. In so doing, we consider the technological asymmetries among public hospitals in the country, as a consequence of a marked spatial concentration of healthcare resources. Our findings emphasize a significant increasing trend in hospital performance after 2008, especially for low and intermediate-tech hospitals who seem to have been able to adapt their spare medical resources and capacity –inefficiently utilized- to the patient inflow in the short-term.

Another important aspect we took into consideration in our analysis is the evident territorial concentration of hospitals and healthcare resources, which may be resulting in agglomeration economies that lead to interactions among hospitals limited to the spatial pattern, generating spillover effects that shape their strategies. In Chapter 3, we provide evidence that these spillover effects are translating into strategic interactions among hospitals in terms of efficiency. That is, public hospitals are reacting to changes in the efficiency performance of their closest neighbors. Specifically, the increase in the efficiency performance of a given hospital is producing increases in efficiency of neighboring hospitals, entailing strategic complementarities due to global spillover effects.

We go deeper into this analysis and explore the question of whether demand increases have direct and indirect (spillover) effects on hospital efficiency and whether these effects change after 2008 (approval of the new constitution). We find that –independently to the hospitals' technological endowment– the increase in demand has a positive effect on the hospitals' efficiency and to those closest neighbors (through spillover effects), and these effects have been reinforced after 2008. These findings support the results of Chapter 1 and confirm that hospitals have been able to adapt to demand changes stemming from new reforms by making a more efficient use of their spare capacity and medical resources.

The importance that demand changes have on the improvement of hospital performance raise the interest to understand the determinants of patient mobilization. In particular, in a country where those high-performing hospitals are mainly located in developed regions, encouraging interregional patient mobilization to receive the best medical treatment. In Chapter 4, we conclude that hospitals' efficiency performance has a pulling effect to attract patients from neighboring (less-developed) regions. However, this effect is significant for specialized hospitals –which are more concentrated in developed areas– as opposed to basic hospitals, which are more homogeneously distributed across the country.

In sum, contrary to what should have been expected, the increase of medical treatment demand after the healthcare reform promoted by the new Constitution of 2008 had a positive effect on hospital performance, both through direct and indirect (spillover) effects. The inefficient use of spare resources and capacity of the hospitals before the reform, jointly with the delay that hospitals had to adapt to the upcoming inflow of patients and the public investment deployed in public health could be strong drivers of this effect. An important determinant of this demand is the performance shown by specialized hospitals, concentrated mainly in developed regions. These results suggest that specialized hospitals are the key players. They are attracting patients form neighboring regions pushing them to neighboring local hospitals, enhancing competition effects that may be translating into higher hospital quality. However, as emphasized in Chapter 2, the average efficiency after 2008 decreased. Nevertheless, the negative effects cannot be captured in our model, as shown in Chapter 3, and thus opens new questions for future research.

5.1 Methodological matters

In this thesis, we exploited a novel database, constructed with public information coming from different official institutions. We managed to retrieve detailed information of hospital and regional features over the period of 2006-2014. We have implemented methodological innovations over this dissertation. However, the results and conclusions are often conditioned by data availability. Overall, the lack of information that allows to properly proxy hospital quality constitutes one of the main limitations throughout the thesis. Quality indicators, commonly used in the applied literature, such as readmission rates or nosocomial infections were not available in the databases.

Another relevant piece of information refers to hospital budgets and public investment. As stated in Chapter 3, hospitals may adapt their behavior if they face financial pressures, specially in developing countries with limited resources and healthcare budget such as Ecuador. The investment deployed in public healthcare could have likely affected hospital efficiency by reducing financial pressures and cost limitations. In this respect, a significant extension of this research would involve exploiting information on healthcare investment and new datasources that allow us to proxy hospital quality and bring more elements for better understanding the public healthcare's quality-efficiency relationship.

Another improvement would be to enlarge the time period with information stemming from hospital surveys available after 2014. The main inconvenient is the lack of a common identifier, not published after 2014, but is available upon request. In this spirit, a potential extension of our research question is thus to enlarge the time period of analysis to achieve results referring to the medium-term.

Another potential issue may come form the potentials setback of two-stage approaches in Chapter 3. Future research can innovate with one-stage SFA panel models that account for hospital heterogeneity, addressing spatial dependence to control for possible bias in the efficiency estimation of two-stage approaches.

5.2 Policy implications

Throughout this dissertation we have observed the low efficiency performance that public hospitals present in the period of analysis and the challenges that policy and decision-makers are forced to face. All in all, the increase of demand derived from the healthcare reforms after the new constitution was implemented seems to have benefited from the performance of public hospitals. Nevertheless, the welfare improvements translated from this may differ on the basis of the regional development of the country. If a single policy recommendation had to be outlined, it would consist on designing clear and tailored policy decisions for the distribution of public funding across the country. To design these policies, we suggest to take into account three key features: regional development, hospital technological endowment and specialization.

Our findings suggest that efficient specialized hospitals are those that are attracting patients to receive medical treatments. More demand is triggering a better performance for those hospitals, but also for neighboring ones, who react by increasing their own efficiency to capture some of this patient inflow, by means of strategic complementarities among public hospitals located within a given area (due to the spatial dependence). However, these positive effects on efficiency are different depending on the technological endowment of the hospital. High and intermediate-tech hospitals –mainly concentrated in developed regions– show a positive association between efficiency and concentration of patients, as opposed to low-tech hospitals. In this respect, decision-makers could focus on reforms aimed at straightening hospital performance, without the need to deploying vast quantities of public investment in the construction of more hospitals, for example. But rather, they could exploit spillover effects among hospitals of developed regions, encouraging competition with reforms that promote quality and efficiency improvement of existing hospitals.

In this context, clear criteria for public funding allocation and stronger regulation of resource consumption is essential to control for cost inflation and *cream skimming*. In addition, public authorities need to be aware that, even if changes in demand –in the short-term– have had an overall positive effect on public hospitals' performance, potential decisions that translate into additional demand (specially for specialized hospitals) may have a counter productive effect on the performance of public hospitals –located in developed areas– if this increase translates into congestion effects.

Conversely, our results evidence the poor performance of public hospitals, mainly located in less-developed areas, with a continuous regional outflow of patients that seek specialized treatment in high-performing hospitals. This outflow of patients in these clusters of less-developed regions suggest a clear need of specialized treatment that could be the target of public investment. The construction of specialized hospitals or the implementation of specialization wards in the existing ones may be a sustainable strategy to fuel demand for medical treatment in these areas and stimulate competition that may translate into higher performance and quality.

As a final remark, this dissertation constitutes one of the first attempts to analyze hospital efficiency performance over a period characterized by one of the biggest milestones of Ecuadorian reforms in the last twenty years. We have assessed the efficiency of the Ecuadorian public healthcare system and addressed it under a framework of analysis where the spatial structure plays a key role. However, our research is a first step to explore questions that has been barely touched in the applied literature for Latin American countries. In this line, there is still much ground to cover that could be exploited in the future. For example, in this thesis, we focused on the public healthcare system of Ecuador. Any interaction between the public and private sector (the so-called private and public partnerships) are topics of interest that can be addresses in future work and provide useful insights for policy recommendations. As assessed, this thesis is intended as a first contribution to the academic debate and a tool for decision makers. But also to plant seed for future research in a fascinating field of work that can contribute to the welfare improvement of developing economies.

Bibliography

Abayomi, K., Gelman, A., and Levy, M. (2008). Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(3):273–291.

Aiura, H. (2013). Inter-regional competition and quality in hospital care. *The European Journal of Health Economics*, 14(3):515–526.

Allin, S., Grignon, M., and Wang, L. (2016). The determinants of efficiency in the Canadian health care system. *Health Economics, Policy and Law*, 11(01):39–65.

Anderson, J. E. (2011). The gravity model. *Annual Review of Economics*, 3(1):133–160.

Anderson, J. E. and Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review*, 93(1):170–192.

Andrews, D. F. and Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society, Series B (Methodological)*, 40(1):85–93.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, volume 4 of *Studies in Operational Regional Science*. Springer Netherlands, Dordrecht.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1):3–25.

Anselin, L., Le Gallo, J., and Hubert, J. (2006). Spatial Panel Econometrics. In *The econometrics of panel data, fundamentals and recent developments in theory and practice*, pages 901–969. Kluwer, Dordrecht.

Arocena, P. and García-Prado, A. (2007). Accounting for quality in the measurement of hospital performance: evidence from Costa Rica. *Health Economics*, 16(7):667–685.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5):941–973.

Atun, R., de Andrade, L. O. M., Almeida, G., Cotlear, D., Dmytraczenko, T., Frenz, P., Garcia, P., Gómez-Dantés, O., Knaul, F. M., Muntaner, C., de Paula, J. B., Rígoli, F., Serrate, P. C.-F., and Wagstaff, A. (2015). Health-system reform and universal health coverage in Latin America. *The Lancet*, 385:1230–1247.

Au, N., Hollingsworth, B., and Spinks, J. (2014). Measuring the efficiency of health services in lower-income countries: The case of papua new guinea. *Development Policy Review*, 32(2):259–272.

Baicker, K., Chernew, M. E., and Robbins, J. A. (2013). The spillover effects of Medicare managed care: Medicare Advantage and hospital utilization. *Journal of Health Economics*, 32(6):1289–1300.

Balaguer-Coll, M. T., Prior, D., and Tortosa-Ausina, E. (2013). Output complexity, environmental conditions, and the efficiency of municipalities. *Journal of Productivity Analysis*, 39(3):303–324.

Balia, S., Brau, R., and Marrocu, E. (2014). What drives patient mobility across Italian regions? Evidence from hospital discharge data. In *Developments in health economics and public policy*, volume 12, pages 133–154.

Balia, S., Brau, R., and Marrocu, E. (2018). Interregional patient mobility in a decentralized healthcare system. *Regional Studies*, 52(3):388–402.

Balia, S., Brau, R., and Moro, D. (2020). Choice of hospital and long-distances: Evidence from Italy. *Regional Science and Urban Economics*, 81(103502).

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data*. Wiley, Chichester, 5th edition.

Baltagi, B. H., Moscone, F., and Santos, R. (2018). Spatial Health Econometrics. In *Contributions to Economic Analysis: Health Econometrics*, volume 294, pages 305–326. Emerald Publishing Limited.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for Eestimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9):1078–1092.

Banker, R. D. and Morey, R. C. (1986). The use of categorical variables in data envelopment analysis. *Management Science*, 32(12):1613–1627.

Banker, R. D. and Morey, R. C. (1996). Estimating production frontier shifts: An application of DEA to technology assessment. *Annals of Operations Research*, 66(3):181–196.

Battese, G. E. and Rao, D. S. P. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business and Economics*, 1(2):87–93.

Battese, G. E., Rao, D. S. P., and O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, 21(1):91–103.

Bech, M. and Lauridsen, J. (2008). Exploring the spatial pattern in hospital admissions. *Health policy*, 87:50–62.

Behrens, K. and Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. In *Handbook of Regional and Urban Economics*, volume 5, pages 171–245.

Belotti, F., Hughes, G., and Piano Mortari, A. (2016). Spatial panel data models using Stata. *Stata Journal*, 16(1):139–180.

Bhattacharjee, A., Maiti, T., and Petrie, D. (2014). General equilibrium effects of spatial structure: Health outcomes and health behaviours in Scotland. *Regional Science and Urban Economics*, 49:286–297.

Bloom, N., Propper, C., Seiler, S., and Van Reenen, J. (2015). The impact of competition on management quality: Evidence from public hospitals. *The Review of Economic Studies*, 82(2):457–489.

Brekke, K. R., Levaggi, R., Siciliani, L., and Straume, O. R. (2016). Patient mobility and health care quality when regions and patients differ in income. *Journal of Health Economics*, 50:372–387.

Brekke, K. R., Levaggi, R., Siciliani, L., and Straume, R. (2014). Patient mobility, health care quality and welfare. *Journal of Economic Behavior and Organization*, 105:140–157.

Brueckner, J. K. (2003). Strategic interaction among governments: An overview of empirical studies. *International Regional Science Review*, 26(2):175–188.

Caliñski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27.

Cantarero, D. (2006). Health care and patients' migration across Spanish regions. *The European Journal of Health Economics*, 7(2):114–116.

Cantor, V. J. M. and Poh, K. L. (2018). Integrated analysis of healthcare efficiency: A systematic review. *Journal of Medical Systems*, 42(1):8.

Carr-Hill, R. A. (1994). Efficiency and equity implications of the health care reforms. *Social Science & Medicine*, 39(9):1189–1201.

Carrillo, M. and Jorge, J. M. (2017). DEA-like efficiency ranking of regional health systems in Spain. *Social Indicators Research*, 133(3):1133–1149.

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

CEPAL (2007). La medida de necesidades básicas insatisfechas (NBI) como instrumento de medición de la pobreza y focalización de programas. Technical report, Bogotá.

Chandra, A. and Staiger, D. O. (2007). Productivity spillovers in health care: Evidence from the treatment of heart attacks. *Journal of Political Economy*, 115(1):103–140.

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.

Chen, K.-C., Chien, L.-N., Hsu, Y.-H., and Yu, M.-M. (2016). Metafrontier frameworks for studying hospital productivity growth and quality changes. *International Journal for Quality in Health Care*, 28(6):650–656.

Cheng, T. C., Haisken-DeNew, J. P., and Yong, J. (2015). Cream skimming and hospital transfers in a mixed public-private system. *Social Science & Medicine*, 132:156–164.

Choi, H. and Park, M. J. (2019). Evaluating the efficiency of governmental excellence for social progress: Focusing on low- and lower-middle-income countries. *Social Indicators Research*, 141(1):111–130.

Chowdhury, H., Zelenyuk, V., Laporte, A., and Wodchis, W. P. (2014). Analysis of productivity, efficiency and technological changes in hospital services in Ontario: How does case-mix matter? *International Journal of Production Economics*, 150:74–82.

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

Cohen, J. P. and Morrison Paul, C. (2008). Agglomeration and cost economies for Washington State hospital services. *Regional Science and Urban Economics*, 38(6):553–564.

Congdon, P. (2001). The development of gravity models for hospital patient flows under system change: a Bayesian modelling approach. *Health care management science*, 4(4):289–304.

Cooper, W., Seiford, L., and Tone, K. (2006). *Introduction to data envelopment analysis and its uses*. Kluwer Academic Publishers, Boston, 1 edition.

Cozad, M. and Wichmann, B. (2013). Efficiency of health care delivery systems: effects of health insurance coverage. *Applied Economics*, 45(29):4082–4094.

Croissant, Y. and Millo, G. (2018). *Panel Data Econometrics with R*. Wiley.

Cuadrado-Roura, J. R. and Aroca, P. (2013). *Regional Problems and Policies in Latin America*. Advances in Spatial Science. Springer Berlin Heidelberg, Berlin, Heidelberg.

Culyer, A. (1971). The nature of the commodity 'health care' and its efficent allocation. *Oxford Economic Papers*, 23(2):189–211.

Culyer, A. (2006). The bogus conflict between efficiency and vertical equity. *Health Economics*, 15(11):1155–1158.

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. (2007a). *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*, volume 4 of *Studies in Productivity and Efficiency*. Springer US, Boston, MA.

Daraio, C. and Simar, L. (2007b). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *Journal of Productivity Analysis*, 28(1-2):13–32.

De Castro Lobo, M., Lins, M., da Silva, A., and Fiszman, R. (2010a). Assessment of teaching-health care integration and performance in university hospitals. *Revista de Saude Publica*, 44(4):581–590.

De Castro Lobo, M., Ozcan, Y., da Silva, A., Lins, M., and Fiszman, R. (2010b). Financing reform and productivity change in Brazilian teaching hospitals: Malmquist approach. *Central European Journal of Operations Research*, 18(2):141–152.

De Paepe, P., Tapia, R. E., Santacruz, E. A., and Unger, J. P. (2012). Ecuador's silent health reform. *International Journal of Health Services*, 42(2):219–233.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19(3):273–292.

Druska, V. and Horrace, W. C. (2004). Generalized moments estimation for spatial panel data: Indonesian rice farming. *American Journal of Agricultural Economics*, 86(1):185–198.

Dumitrescu, E.-I. and Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling*, 29(4):1450–1460.

El-Mahgary, S. and Lahdelma, R. (1995). Data envelopment analysis: Visualizing the results. *European Journal of Operational Research*, 83(3):700–710.

Elhorst, J. P. (2010). Spatial Panel Data Models. In *Handbook of Applied Spatial Analysis*, pages 377–407. Springer Berlin Heidelberg, Berlin, Heidelberg.

Elhorst, J. P. (2014). Linear spatial dependence models for corss-section data. In *Spatial Econometrics*, volume 1 of *SpringerBriefs in Regional Science*, pages 59–63. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ellis, R. P. (1998). Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics*, 17(5):537–555.

Espinosa, V., de la Torre, D., Acuña, C., and Cadena, C. (2017). Los recursos humanos en salud según el nuevo modelo de atención en Ecuador. *Revista Panamericana de Salud Pública*, 41(52).

Fabbri, D. and Robone, S. (2010). The geography of hospital admission in a national health service with patient choice. *Health Economics*, 19(9):1029–1047.

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

Felder, S. and Tauchmann, H. (2013). Federal state differentials in the efficiency of health production in Germany: an artifact of spatial dependence? *The European Journal of Health Economics*, 14(1):21–39.

Ferreira, D. and Marques, R. (2019). Do quality and access to hospital services impact on their technical efficiency? *Omega*, 86:218–236.

Giannakas, K., Tran, K. C., and Tzouvelekas, V. (2003). On the choice of functional form in stochastic frontier modeling. *Empirical Economics*, 28(1):75–100.

Giménez, V., Prieto, W., Prior, D., and Tortosa-Ausina, E. (2019). Evaluation of efficiency in Colombian hospitals: An analysis for the post-reform period. *Socio-Economic Planning Sciences*, 65:20–35.

Gobillon, L. and Milcent, C. (2013). Spatial disparities in hospital performance. *Journal of Economic Geography*, 13(6):1013–1040.

Goodman, A. C. and Smith, B. C. (2018). Location of health professionals: The supply side. *Regional Science and Urban Economics*, 68:148–159.

Granda, M. L. and Jimenez, W. G. (2019). The evolution of socioeconomic health inequalities in Ecuador during a public health system reform (2006-2014). *International Journal for Equity in Health*, 18(1):1–12.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424.

Gravelle, H., Santos, R., and Siciliani, L. (2014). Does a hospital's quality depend on the quality of other hospitals? A spatial econometrics approach. *Regional Science and Urban Economics*, 49:203–216.

Hafidz, F., Ensor, T., and Tubeuf, S. (2018). Efficiency measurement in health facilities: A systematic review in low- and middle-income countries. *Applied Health Economics and Health Policy*, 16(4):465–480.

Halkos, G. E. and Tzeremes, N. G. (2011). A conditional nonparametric analysis for measuring the efficiency of regional public healthcare delivery: An application to Greek prefectures. *Health Policy*, 103(1):73–82.

Halleck Vega, S. and Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science*, 55(3):339–363.

Hartmann, C. (2016). Postneoliberal public health care reforms: Neoliberalism, social medicine, and persistent health inequalities in Latin America. *American Journal of Public Health*, 106(12):2145–2151.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251.

Herr, A. (2008). Cost and technical efficiency of German hospitals: does ownership matter? *Health Economics*, 17(9):1057–1071.

Herr, A., Schmitz, H., and Augurzky, B. (2011). Profit efficiency and ownership of German hospitals. *Health Economics*, 20(6):660–674.

Herwartz, H. and Schley, K. (2018). Improving health care service provision by adapting to regional diversity: An efficiency analysis for the case of Germany. *Health Policy*, 122:293–300.

Herwartz, H. and Strumann, C. (2012). On the effect of prospective payment on local hospital competition in Germany. *Health Care Management Science*, 15(1):48–62.

Herwartz, H. and Strumann, C. (2014). Hospital efficiency under prospective reimbursement schemes: an empirical assessment for the case of Germany. *The European Journal of Health Economics*, 15(2):175–186.

Hollingsworth, B. (2003). Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science*, 6(4):203–218.

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

Hollingsworth, B. (2012). Revolution, evolution, or status quo? Guidelines for efficiency measurement in health care. *Journal of Productivity Analysis*, 37(1):1–5.

Homedes, N. and Ugalde, A. (2005). Why neoliberal health reforms have failed in Latin America. *Health Policy*, 71(1):83–96.

Husson, F., Lê, S., and Pagès, J. (2010). *Exploratory multivariate analysis by example using R*. Chapman & Hall/CRC Computer Science & Data Analysis. CRC Press.

Ippoliti, R. and Falavigna, G. (2012). Efficiency of the medical care industry: Evidence from the Italian regional system. *European Journal of Operational Research*, 217(3):643–652.

Jensen, C. D. and Lacombe, D. J. (2012). A note on partitioning effects estimates over space. *Letters in Spatial and Resource Sciences*, 5(1):47–53.

Johannessen, K. A., Kittelsen, S. A., and Hagen, T. P. (2017). Assessing physician productivity following Norwegian hospital reform: A panel and data envelopment analysis. *Social Science and Medicine*, 175:117–126.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1):97–130.

Keith, J. and Prior, D. (2014). Scale and scope economies in Mexican private medical units. *Salud Publica de Mexico*, 56(4):348–354.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2):509–533.

Kinfu, Y. and Sawhney, M. (2015). Inefficiency, heterogeneity and spillover effects in maternal care in India: a spatial stochastic frontier analysis. *BMC Health Services Research*, 15(1):118.

Koopmans, T. C. (1951). An analysis of production as an efficient combination of activities. In *Activity analysis of production and allocation, cowles commission for research in economics*. John Willey and Sons, Inc., New York.

Kumbhakar, S. C. (2010). Efficiency and productivity of world health systems: where does your country stand? *Applied Economics*, 42(13):1641–1659.

Laurent, T., Margaretic, P., and Thomas-Agnan, C. (2019). Modelling spatial flows with R. *Mimeo*.

Leach, J. (2010). Ex post welfare under alternative health care systems. *Journal of Public Economic Theory*, 12(6):1027–1057.

Lee, M.-L. and Pace, R. K. (2005). Spatial distribution of retail sales. *The Journal of Real Estate Finance and Economics*, 31(1):53–69.

LeSage, J. and Fischer, M. M. (2010). Spatial Econometric Methods for Modeling Origin-Destination Flows. In *Handbook of Applied Spatial Analysis*, pages 409–433. Springer Berlin Heidelberg, Berlin, Heidelberg.

LeSage, J. and Fischer, M. M. (2016). Spatial Regression-Based Model Specifications for Exogenous and Endogenous Spatial Interaction. In Patuelli, R. and Arbia, G., editors, *Spatial Econometric Interaction Modelling. Advances in Spatial Science (The Regional Science Series)*, pages 15–36. Springer, Cham.

LeSage, J. and Pace, R. K. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48(5):941–967.

LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.

LeSage, J. and Thomas-Agnan, C. (2015). Interpreting spatial econometric origin-destination flow models. *Journal of Regional Science*, 55(2):188–208.

Levaggi, R. and Zanola, R. (2004). Patients' migration across regions: the case of Italy. *Applied Economics*, 36(16):1751–1757.

Levy, S. and Schady, N. (2013). Latin America's social policy challenge: Education, social insurance, redistribution. *Journal of Economic Perspectives*, 27(2):193–218.

Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews*, 15(3):261–274.

Li, Q., Maasoumi, E., and Racine, J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics*, 148(2):186–200.

Linna, M. (1998). Measuring hospital cost efficiency with panel data models. *Health economics*, 7(5):415–427.

Lisi, D., Moscone, F., Tosetti, E., and Vinciotti, V. (2017). Hospital interdependence in a competitive institutional environment : Evidence from Italy. *Health, Econometrics and Data Group (HEDG) Working Papers*.

Longo, F., Siciliani, L., Gravelle, H., and Santos, R. (2017). Do hospitals respond to rivals' quality and efficiency? A spatial panel econometric analysis. *Health Economics*, 26(S2):38–62.

Longo, F., Siciliani, L., Moscelli, G., and Gravelle, H. (2019). Does hospital competition improve efficiency? The effect of the patient choice reform in England. *Health Economics*, 53:2324–2345.

López-Cevallos, D., Chi, C., and Ortega, F. (2014). Consideraciones para la transformación del sistema de salud del Ecuador desde una perspectiva de equidad. *Revista de Salud Pública*, 16(3):346–359.

López-Cevallos, D. F. and Chi, C. (2010). Assessing the context of health care utilization in Ecuador: A spatial and multilevel analysis. *BMC Health Services Research*, 10(1):64.

Lowe, J. M. and Sen, A. (1996). Gravity model applications in health planning: analysis of an urban hospital market. *Journal of Regional Science*, 36(3):437–461.

Lucio, R., Villacrés, N., and Henríquez, R. (2011). Sistema de salud de Ecuador. *Salud Publica de Mexico*, 53(2):177–187.

Malo-Serrano, M. and Malo-Corral, N. (2014). Reforma de salud en Ecuador: Nunca más el derecho a la salud como un privilegio. *Revista Peruana de Medicina Experimental y Salud Publica*, 31(4):754–761.

Maniadakis, N., Hollingsworth, B., and Thanassoulis, E. (1999). The impact of the internal market on hospital efficiency, productivity and service quality. *Health Care Management Science*, 2(2):75–85.

Martini, G., Berta, P., Mullahy, J., and Vittadini, G. (2014). The effectiveness-efficiency trade-off in health care: The case of hospitals in Lombardy, Italy. *Regional Science and Urban Economics*, 49:217–231.

Mas, N. (2015). Hospital financial pressures and the health of the uninsured: Who gets hurt? *International Journal of Health Services*, 45(1):53–71.

Mastromarco, C., Stastna, L., and Votapkova, J. (2019). Efficiency of hospitals in the Czech Republic: Conditional efficiency approach. *Journal of Productivity Analysis*, 51(1):73–89.

Mendieta Muñoz, R. and Pontarollo, N. (2016). Cantonal convergence in Ecuador: A spatial econometric perspective. *Journal of Applied Economic Sciences*, 11(39):107–110.

Mendieta Muñoz, R., Raileanu Szeles, M., Beltrán Romero, P., and Piedra Peña, J. A. (2015). Explaining the regional economic heterogeneity in Ecuador. *Bulletin of the Transilvania University of Brasov*, 8(2):399–406.

Ministerio de Salud Pública (2012). Manual del Modelo de Atención Integral del Sistema Nacional de Salud Familiar Comunitario e Intelectual. page 210.

Mitropoulos, P., Talias, M. A., and Mitropoulos, I. (2015). Combining stochastic DEA with Bayesian analysis to obtain statistical properties of the efficiency scores: An application to Greek public hospitals. *European Journal of Operational Research*, 243(1):302–311.

Mobley, L. R. (2003). Estimating hospital market pricing: An equilibrium approach using spatial econometrics. *Regional Science and Urban Economics*, 33(4):489–516.

Mobley, L. R., Frech, H. E., and Anselin, L. (2009). Spatial interaction, spatial multipliers and hospital competition. *International Journal of the Economics of Business*, 16(1):1–17.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society: Series B (Methodological)*, 10(2):243–251.

Moscelli, G., Siciliani, L., Gutacker, N., and Gravelle, H. (2016). Location, quality and choice of hospital: Evidence from England 2002–2013. *Regional Science and Urban Economics*, 60:112–124.

Moscone, F. and Tosetti, E. (2014). Spatial Econometrics: Theory and Applications in Health Economics. In Culyer, A., editor, *Encyclopedia of Health Economics*, pages 329–334. Elsevier.

Moscone, F., Tosetti, E., and Vinciotti, V. (2017). Sparse estimation of huge networks with a block-wise structure. *The Econometrics Journal*, 20(3):S61–S85.

Moscone, F., Vinciotti, V., and Tosetti, E. (2018). Large Network Inference: New Insights in Health Economics. In *Contributions to Economic Analysis: Health Econometrics*, volume 294, pages 359–378.

Mutl, J. and Pfaffermayr, M. (2011). The Hausman test in a Cliff and Ord panel model. *The Econometrics Journal*, 14(1):48–76.

Nord, E., Richardson, J., Street, A., Kuhse, H., and Singer, P. (1995). Maximizing health benefits vs egalitarianism: An Australian survey of health issues. *Social Science & Medicine*, 41(10):1429–1437.

O'Donnell, C. J., Rao, D. S. P., and Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics*, 34(2):231–255.

Olesen, O. B., Petersen, N. C., and Podinovski, V. V. (2015). Efficiency analysis with ratio measures. *European Journal of Operational Research*, 245(2):446–462.

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.

Orellana, M. R., Piedra Peña, J. A., and Sarmiento Moscoso, L. S. (2017). Evidence about moral hazard in the Ecuadorian health system. *Journal of Smart Economic Growth*, 1(2):109–132.

Organización Panamericana de la Salud (2008). Perfil de Sistema de Salud: Ecuador, monitoreo y análisis de los procesos de cambio y reforma. Technical report, Organización Panamericana de la Salud, Washington, D.C.

Papanicolas, I. and Smith, P. (2013). *Health system performance comparison. An agenda for policy, information and research*. Open University Press.

Pastor, J. M. and Tortosa-Ausina, E. (2008). Social capital and bank performance: an international comparison for OECD countries. *The Manchester School*, 76(2):223–265.

Patuelli, R. and Arbia, G. (2016). Spatial Econometric Interaction Modelling: Where Spatial Econometrics and Spatial Interaction Modelling Meet. In *Spatial Econometric Interaction Modelling*, pages 1–12.

Pérez-López, G., Prior, D., and Zafra-Gómez, J. L. (2018). Temporal scale efficiency in DEA panel data estimations. An application to the solid waste disposal service in Spain. *Omega*, 76:18–27.

Piedra-Peña, J. and Prior, D. (2020). Analyzing the effect of health reforms on the efficiency of Ecuadorian public hospitals. Working paper, Graduate Program in Applied Economic Research, 2020-01, Barcelona.

Piedra-Peña, J. (2020). Spatial dependence in hospitals efficiency: A spatial econometric approach for Ecuadorian public hospitals. Working paper, Graduate Program in Applied Economic Research, 2020-05, Barcelona.

Podinovski, V. V. (2005). Selective convexity in DEA models. *European Journal of Operational Research*, 161(2):552–563.

Porojan, A. (2001). Trade flows and spatial effects: The gravity model revisited. *Open Economies Review*, 12(3):265–280.

Prior, D. (2006). Efficiency and total quality management in health care organizations: A dynamic frontier approach. *Annals of Operations Research*, 145(1):281–299.

Prior, D. and Surroca, J. (2010). Performance measurement and achievable targets for public hospitals. *Journal of Accounting, Auditing and Finance*, 25(4):749–766.

Propper, C. (2012). Competition, incentives and the English NHS. *Health Economics*, 21(1):33–40.

Pross, C., Strumann, C., Geissler, A., Herwartz, H., and Klein, N. (2018). Quality and resource efficiency in hospital service provision: A geoadditive stochastic frontier analysis of stroke quality of care in Germany. *PLoS ONE*, 13(9).

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics*, 4(1):87.

Ruiz-Rodriguez, M., Rodriguez-Villamizar, L. A., and Heredia-Pi, I. (2016). Technical efficiency of women's health prevention programs in Bucaramanga, Colombia: a four-stage analysis. *BMC Health Services Research*, 16(1):576.

Sen, A. and Smith, T. E. (1995). *Gravity Models of Spatial Interaction Behavior*. Advances in Spatial and Network Economics. Springer Berlin Heidelberg, Berlin, Heidelberg.

Simar, L. and Wilson, P. W. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research*, 139(1):115–132.

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). Statistical inference in nonparametric frontier models: recent developments and perspectives. In *The Measurement of Productive Efficiency and Productivity Change*, pages 421–521. Oxford University Press.

Simar, L. and Wilson, P. W. (2011). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36(1):33–53.

Simar, L. and Zelenyuk, V. (2006). On testing equality of distributions of technical efficiency scores. *Econometric Reviews*, 25(4):497–522.

Smith, P. C. and Yip, W. (2016). The economics of health system design. *Oxford Review of Economic Policy*, 32(1):21–40.

Sommersguter-Reichmann, M. (2000). The impact of the Austrian hospital financing reform on hospital productivity: empirical evidence on efficiency and technology changes using a non-parametric input-based Malmquist approach. *Health Care Management Science*, 3(4):309–321.

Staat, M. (2011). Estimating the efficiency of general practitioners controlling for case mix and outlier effects. *Empirical Economics*, 40(2):321–342.

Surroca, J., Prior, D., and Tribó Giné, J. A. (2016). Using panel data dea to measure CEOs' focus of attention: An application to the study of cognitive group membership and performance. *Strategic Management Journal*, 37(2):370–388.

Szeles, M. and Mendieta Muñoz, R. (2016). Analyzing the regional economic convergence in Ecuador. Insights from parametric and nonparametric models. *Romanian Journal of Economic Forecasting*, 19(2):43–65.

Tay, A. (2003). Assessing competition in hospital care markets: The importance of accounting for quality differentiation. *The RAND Journal of Economics*, 34(4):786.

Thomas-Agnan, C. and LeSage, J. P. (2014). Spatial Econometric OD-Flow Models. In *Handbook of Regional Science*, pages 1653–1673. Springer Berlin Heidelberg, Berlin, Heidelberg.

Tosetti, E., Santos, R., Moscone, F., and Arbia, G. (2018). The Spatial Dimension of Health Systems. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Town, R. and Vistnes, G. (2001). Hospital competition in HMO networks. *Journal of Health Economics*, 20(5):733–753.

Tulkens, H. (1986). La performance productive d'un service public. Définitions, méthodes de mesure et application à la Régie des Postes en Belgique. *L'Actualité économique*, 62(2):306.

Valdmanis, V., Rosko, M., Mancuso, P., Tavakoli, M., and Farrar, S. (2017). Measuring performance change in Scottish hospitals: a Malmquist and times-series approach. *Health Services and Outcomes Research Methodology*, 17(2):113–126.

Van Ineveld, M., van Oostrum, J., Vermeulen, R., Steenhoek, A., and van de Klundert, J. (2016). Productivity and quality of Dutch hospitals during system reform. *Health Care Management Science*, 19(3):279–290.

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.

Varkevisser, M., van der Geest, S. A., and Schut, F. T. (2012). Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands. *Journal of Health Economics*, 31(2):371–378.

Victoor, A., Delnoij, D. M., Friele, R. D., and Rademakers, J. J. (2012). Determinants of patient choice of healthcare providers: a scoping review. *BMC Health Services Research*, 12(1):272.

Villacrés, T. and Mena, A. C. (2017). Mecanismos de pago y gestión de recursos financieros para la consolidación del Sistema de Salud de Ecuador. *Revista Panamericana de Salud Pública*, 41:1–6.

Villalobos-Cid, M., Chacón, M., Zitko, P., and Inostroza-Ponta, M. (2016). A new strategy to evaluate technical efficiency in hospitals using homogeneous groups of casemix. *Journal of Medical Systems*, 40(4):103.

Wagstaff, A. (1989). Estimating eficiency in the hospital sector: a comparison of three statistical cost frontier models. *Applied Economics*, 21:659–672.

Williams, H., Gentzkow, M., and Finkelstein, A. (2016). Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics*, 131(4):1681–1726.

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

World Health Organization (2000). *The World Health Report 2000. Health Systems: Improving Perfomance*. Geneva.

Worthington, A. C. (2004). Frontier efficiency measurement in health care: A review of empirical techniques and selected applications. *Medical Care Research and Review*, 61(2):135–170.

Xenos, P., Yfantopoulos, J., Nektarios, M., Polyzos, N., Tinios, P., and Constantopoulos, A. (2017). Efficiency and productivity assessment of public hospitals in Greece during the crisis period 2009–2012. *Cost Effectiveness and Resource Allocation*, 15(1):6.

Xingzhu, L. (2003). *Policy Tools for Allocative Efficiency*. World Health Organization, Geneva.