working paper

GEAR

GRADUATE PROGRAM IN APPLIED ECONOMIC RESEARCH

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

Juan Piedra-Peña

Working paper:
2021-01

Universitat Autònoma de Barcelona

This collection belongs to:

Departament
d'Economia Aplicada

UAB

Avinguda de l'Eix Central
Edifici B2
Campus de la UAB
08193 Bellaterra
(Cerdanyola del Vallès)
Barcelona · Spain
Tel. +34 93 581 16 80
Fax +34 93 581 22 92
d.econ.aplicada@uab.cat
www.uab.cat/departament/
economia-aplicada/

Coordinator: Rosella Nicolini (rosella.nicolini@uab.cat)

This collection includes a selection of research by students of the PhD Program in Applied Economics (UAB) and the Master of Applied Research in Economics and Business (MAREB) - specialization in Applied Economics. Research contributions can be published in English or Spanish.

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

Juan Piedra-Peña
Department of Applied Economics
Universitat Autònoma de Barcelona, Barcelona, Spain

February 24, 2021

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 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

*I want to acknowledge the support of the Pre-doctoral Trainee Research Scholarship (PIF) from the Autonomous University of Barcelona. I want to thank my supervisors Rosella Nicolini and Diego Prior. The participants of the seminars at UAB, the University of Brescia, the University of Gothenburg, the WRSA international conference, 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 paper. To James LeSage and Christine Thomas-Agnan for facilitating the MATLAB and R codes. Any remaining errors are the author's own responsibility.

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). 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

[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).

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 (Piedra-Peña, 2020).[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 (Piedra-Peña, 2020).

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 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 en-

[2]Refer to Appendix A 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; Piedra-Peña, 2020)

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

hance their medical resources as well, to increase their performance and avoid possible outflows of patients.

Piedra-Peña (2020) emphasizes 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 ifs 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 environmental 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).

[5] Piedra-Peña (2020) provides 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.

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 (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 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 2 reviews the literature on hos-

[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.

pital patient migration. In Section 3, the theoretical model is described, as introduced by Brekke et al. (2016), which is followed throughout this study. Section 4 explains the methodology of the order-m efficiency measurement and the spatial interaction model, while Section 5 introduces the empirical approximation used. Section 6 describes our dataset and Sections 7 and 8 present the results and robustness analysis, respectively. Finally, the main conclusions are presented in Section 9.

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 (Piedra-Peña, 2020). 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

[8]We explain this method on a deeper extent on Section 3

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 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 inter-regional 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.

3 Theoretical framework

In our framework of study, the high-performing hospitals are mainly located in developed regions (see Section 6) that have historically concentrated the healthcare resources in the country (Piedra-Peña, 2020; Piedra-Peña and Prior, 2020). 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) corss-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 (Piedra-Peña, 2020; Piedra-Peña and Prior, 2020).[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 patient located in z receiving treatment from hospital in region i as:

$$U(z, s_i) = v + bq_i - t|z - s_i| + u(A_i^x) \tag{1}$$

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

[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 form 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).

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

$$A_i^x = a^x(1-\tau) \tag{2}$$

Additionally, we assume that 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:

$$\bar{a}_i = \lambda_i a^R + (1-\lambda_i)a^P \tag{3}$$

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

$$U(z, s_j) = v + bq_j - t|z - s_j| + u(A_i^x) - F \tag{4}$$

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:

$$\widehat{A_i^x} = a^x(1-\tau) - \pi \tag{5}$$

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:

$$\phi_{ij}^x = \frac{1}{2t}(b(q_j - q_i) + u(\widehat{A_i^x}) - u(A_i^x) - F) \tag{6}$$

Notice that

$$\frac{\delta \phi_{ij}^x}{\delta a_x} = (\frac{1-\tau}{2t})(u'(\hat{A_i^x}) - u(A_i^x)) > 0 \; if \; \pi > 0 \tag{7}$$

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

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

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

$$\Phi_{ij} = \lambda_i \phi_{ij}^R + (1 - \lambda_i)\phi_{ij}^P \tag{8}$$

Notice that

$$\frac{\delta \Phi_{ij}}{\delta q_j} = -\frac{\delta \Phi_{ij}}{\delta q_i} = \frac{b}{2t} \tag{9}$$

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

$$E(Y_{ij}) = f(X_i, X_j, G_{ij}) \tag{10}$$

Where $E(Y_{ij})$ are the expectations of observed flows 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 (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.

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

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 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.

4.1 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.

$$\Psi = \left\{(x, y) \in \mathbb{R}_+^{p+q} | \; x \; can \; produce \; y\right\} \tag{11}$$

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 $X = \left\{(x, y) \in \mathbb{R}_+^{p+q} | \; i = 1, \ldots, n\right\}$. 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:

$$\Psi\left(x_i, y_i\right) = \left\{(x, y) \in \mathbb{R}_+^{p+q} | \; (y \geq y_i) x \leq x_i\right\} \tag{12}$$

[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.

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:

$$S_Y\left(y \mid x, z\right) = Prob(Y \geq y|X \leq x, Z = z) \tag{13}$$

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:

$$\theta\left(x, y \mid z\right) = sup\left\{\theta|S_Y(\theta_y|x,z) > 0\right\} \tag{14}$$

Those points where $\theta\left(x,y \mid z\right) = 1$ are the technically efficient ones and correspond to the efficiency frontier, while those with $\theta\left(x,y \mid z\right) > 1$ are technically inefficient. To obtain the nonparametric estimators of the conditional frontier $\theta\left(x,y \mid z\right)$, 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 estimator can be obtained by the computation of the one-dimensional numerical integral defined as in Daraio and Simar (2007a):

$$\widehat{\theta}_m\left(x_0, y_0 \mid z_0\right) = \int_0^\infty \left[1 - (1 - S_{Y|X,Z}(uy_0|X \leq x_0,\ Z = Z_0))^m\right] du \tag{15}$$

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.

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

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

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:

$$\hat{S}_{Y,n}\left(y \mid x,z\right) = \frac{\sum_{i=1}^{n} I(x_i \leq x, y_i \geq y) K_{\hat{h}}(z,z_i)}{\sum_{i=1}^{n} I(x_i \leq x) K_{\hat{h}}(z,z_i)} \tag{16}$$

Where $K_{\hat{h}}(.)$ represents the kernel function, $I(.)$ 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.

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

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

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]

4.3 Spatial interaction model

Form equation (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 $N = n_o n_d$. The n_o x n_d flow matrix Y can be converted to an N vector by stacking columns. The flow matrix can be arranged so the i, jth 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 x 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]

[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.

[18]If we start with the standard gravity model and apply a log transformation, the resulting model would be as shown in (17) (Sen and Smith, 1995).

$$y = \propto l_N + X_o\beta_o + X_d\beta_d + \gamma g + \varepsilon \tag{17}$$

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 Nxk 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 $k\ x\ 1$ 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 $N\ x\ 1$ vector of disturbances ($\varepsilon = vec(E)$).

From (17), we consider an SLX interaction model in the following specification:[19]

$$y = \propto l_N + X_o\beta_o + X_d\beta_d + W_oX_o\emptyset_o + W_dX_d\emptyset_d + \gamma g + \varepsilon \tag{18}$$

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. However, the sole inclusion of a distance function in the estimation will not allow incorporating the spatial structure into the model explicitly. 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]

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 (17) and control for the spatial lags of distance g, in the following manner:

$$y = \propto l_N + X_o\beta_o + X_d\beta_d + W_oX_o\emptyset_o + W_dX_d\emptyset_d + \gamma g + W_og\gamma_o + W_dg\gamma_d + \varepsilon \tag{19}$$

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.

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

4.4 Spatial Durbin Interaction Model

From equation (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 (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 (17), which can be viewed as filtering for spatial dependence related to origin and destination.

$$(I_{\mathrm{N}}-\rho_o W_o)(I_{\mathrm{N}}-\rho_d W_d)y=\propto l_{\mathrm{N}}+X_o\beta_o+X_d\beta_d+\gamma g+\varepsilon \tag{20}$$

Here, $(I_{\mathrm{N}}-p_o W_o)(I_{\mathrm{N}}-p_d W_d)$ 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 (20) can be further decomposed considering the spatial lags of the explanatory variables into an SDM as follows:[22]

$$\begin{aligned} y=&\rho_d W_d \mathrm{y}+\rho_o W_o \mathrm{y}+\rho_w W_w y+ \propto l_{n^2} + X_o\beta_o + X_d\beta_d + \gamma g + W_o X_o \emptyset_o + \\ &W_d X_d \emptyset_d+W_o g\gamma_o + W_d g\gamma_d + \varepsilon \end{aligned} \tag{21}$$

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, $\rho_o W_o \mathrm{y}$ 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

[21] The expansion of the filter product ends in the expression $I_{\mathrm{N}}-\rho_d W_d-\rho_o W_o-\rho_w W_w$. Where W_w is the product of the two weight matrices $(W_o.W_d)$ and reflects the origin-destination-based dependence. 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).

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.

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.

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 baisis, we can define the following model:

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

$$y = \rho_d W_d \mathrm{y} + \rho_o W_o \mathrm{y} + \rho_w W_w y + \propto l_N + e_d \beta_d + X_o \beta_o + \gamma g + W_d e_d \emptyset_d + W_o X_o \emptyset_o + W_o g \gamma_o + W_d g \gamma_d + \varepsilon \tag{22}$$

The vector e_d contains the robust logged efficiency scores obtained with (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. The intraregional variations will be captured in a new set of explanatory variables X_i, $W_i X_i$, 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:

$$y = \rho_d W_d \mathrm{y} + \rho_o W_o \mathrm{y} + \rho_w W_w y + c\alpha_i + \propto l_N + e_d \beta_d + X_o \beta_o + X_i \beta_i + W_d e_d \emptyset_d + W_o X_o \emptyset_o + W_i X_i \emptyset_i + \gamma g + W_o g \gamma_o + W_d g \gamma_d + \varepsilon \tag{23}$$

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

[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 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)

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 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.

6 Data and Variables

To estimate (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 hospitals, 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, 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 C.

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

[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.

6.1 Variables for the conditional order-m efficiency measurement

For the selection of input and output variables to estimate model (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). We follow Piedra-Peña (2020) and Piedra-Peña and Prior (2020), and 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).

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.

[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).

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. Piedra-Peña and Prior (2020) 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 (Piedra-Peña, 2020). The occupancy rate has been used as an environmental variable for conditional order-m approaches in recent work by Mastromarco et al. (2019). Furthermore, Piedra-Peña (2020) provide empirical evidence of its positive direct and spillover effects on hospital's efficiency.

Table 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, Piedra-Peña (2020) 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 movement is likely to be directed to those developed regions, where more healthcare resources are concentrated.

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 (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

[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.

Table 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.

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).

We proxy the cantonal level variables (X_o in (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 by Piedra-Peña (2020) and Piedra-Peña and Prior (2020). 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

Table 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.

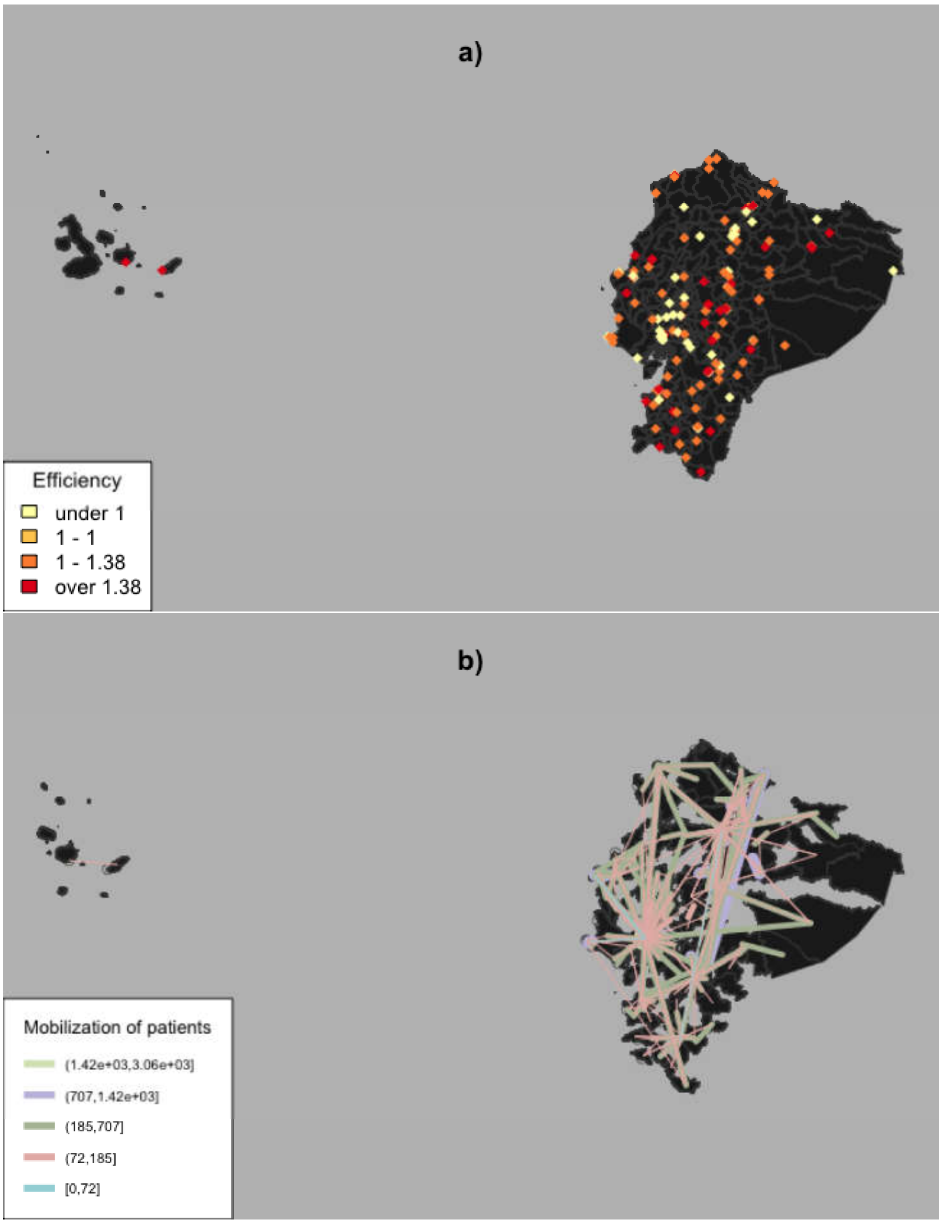
widely applied in Latin American countries since the 1980s as a multidimensional measurement of poverty (CEPAL, 2007). Considering that poverty is a complex and multidimensional phenomenon, the NBI evaluates different dimensions of deprivation of goods and services required to the satisfaction of basic needs. In Ecuador, these dimensions comprehend economic capacity, basic education access, housing access, basic services access, and overcrowding. As stated in equation (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 2 presents the descriptive statistics of the variables used in the SDM model. Additionally, Figure 1 shows the distributions of hospitals by efficiency performance (e_d in (23)) at the top panel (a), and the migration flow dynamic of the sample (y in (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 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 (17), (18) and (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).

Figure 1: Hospital efficiency and patient migration flows



7 Results and discussion

Table 3 presents the estimation results of the interaction model (17), adjusting for the intraregional patient flows in column (1). Column (2) incorporates the SLX interaction model (18); while column (3) includes the spatial lags of the distance variable g, of equation (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, they should not be interpreted as partial derivatives (LeSage and Fischer, 2016).

We can additionally use the estimates in Table 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 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***
	(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***

[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 3 (continued)

	(1)	(2)	(3)
	(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 (23). The Bayesian MCMC estimates based on 1000 draws are presented in Table 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 $\rho_w = -\rho_d \rho_o$ 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 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 5 reports the scalar summary effects for the model (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 Piedra-Peña (2020), 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 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 3.

Furthermore, densely populated cantons with high mortality rates are expected to

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

Table 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 5: Scalar summary effects

	Mean	t-value
Destination Effects		
log conditional efficiency	-0.3015	-4.9594
Origin Effects		
log GVApc	0.2185	6.7937
log density	0.0473	5.1626
log cantonal mortality	0.1155	2.4933
log nbi	0.0558	0.2817
log insured	0.4012	1.4762
Network effects		
log conditional efficiency	-0.1529	-2.5719
log GVApc	0.1725	1.8814
log density	-0.1026	-3.9597
log cantonal mortality	-0.4405	-3.9308
log nbi	2.0840	4.4193
log insured	-0.5705	-0.9851

Note: Dependent variable is the vector of (logged) migration flows.

Bayesian MCMC estimates based on 1000 draws. N=18656

Source: The authors

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 8.

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 2 appendix D) 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 of appendix D) at the 99% confidence level.[38] This validates our hypothesis that hospital performance is being affected by 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 (19). The test confirm that we do not

[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.

[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 by Piedra-Peña (2020)

suffer from endogeneity in the efficiency term (p-value = 1).[40]

Regarding the spatial econometric specification, we tested the robustness of the estimations from equation (23) with a new efficiency value. In so doing, we calculate the efficiency value of equation (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 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 (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 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).

[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)

Table 6: Scalar summary effects, using W_{dt} and W_{dt}^2

	W_{dt}		W_{dt}^2	
	Mean	t stat	Mean	t stat
log conditional efficiency (Destination Eff.)	-0.2139	-3.4104	-0.1798	-2.9476
log conditional efficiency (Network Eff.)	-0.4854	-2.3507	-0.4017	-1.5881
ρ_d	0.4757	15.7229	0.4308	5.9839
ρ_o	0.3057	3.9444	0.5159	17.8886
ρ_w	0.0010	0.0107	-0.1078	-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 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 of appendix E). 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, those hospitals are taking advantage of patient inflows, initially attracted by other hospitals (most

[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

likely neighbor specialized hospitals).[43]

Table 7: Scalar summary effects by hospital type

	Basic (N=7743)		Specialized (N=2610)	
	Mean	t stat	Mean	t stat
log conditional efficiency (Destination Eff.)	-0.0077	-0.1669	-0.8085	-3.9508
log conditional efficiency (Network Eff.)	0.3621	6.2390	-0.7748	-5.2670
ρ_d	0.1530	19.2537	0.3453	15.6936
ρ_o	0.2692	31.9435	0.4202	12.1845
ρ_w	-0.0318	-3.8361	-0.1222	-5.5952

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 F (Table 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 10 of Appendix F, 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).

[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.

Furthermore, we plot two figures in Appendix F. Figure 5 describes the demographics (available in our dataset) of the patients with the top five morbidity causes, while Figure 6 describes the public entity embedding the public hospital (MSP, Social Security 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 5 shows that more than 60% of the interregional patients describe themselves as mixed-race. In addition, Figure 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 (Piedra-Peña, 2020), 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

[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.

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 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.

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.

References

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

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.

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

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).

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.

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.

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.

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.

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.

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

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.

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.

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.

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

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.

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.

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

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

Propper, C. (2012). Competition, incentives and the English NHS. *Health Economics*, 21(1):33–40.

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 Zelenyuk, V. (2006). On testing equality of distributions of technical efficiency scores. *Econometric Reviews*, 25(4):497–522.

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.

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.

Appendices

A 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 (Piedra-Peña and Prior, 2020). These reforms were supported by an increasing public investment for the core system, mostly involving the endowment of medical infrastructure and training.

B 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 (20) and introduce a set of latent variance scalars for each observation, so we have:

$$\varepsilon \sim N(0, \sigma^2 \tilde{V})$$
$$\tilde{V}_{ii} = V_i, i = 1, ..., N$$
$$V = vec(R)$$

$$R = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n_d} \\ v_{21} & v_{22} & & v_{2n_d} \\ \vdots & & \ddots & \vdots \\ v_{n_o 1} & & & v_{n_o n_d} \end{pmatrix} \tag{24}$$

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 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:

$$\pi(\delta) \propto N(c, T), c = 0, T \to \infty \tag{25}$$
$$\pi(\lambda / v_{ij}) \sim iid\chi^2(\lambda) \tag{26}$$
$$\pi(\sigma^2) \sim IG(a, b) \tag{27}$$
$$\pi(\rho_i) \sim U(-1, 1), i = d, o, w \tag{28}$$

The conditional posterior 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]

C Variable description

Table 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.

[48] Refer to LeSage and Pace (2009) chapter 8

[49] Refer to LeSage and Pace (2009) chapter 5

D Order-*m* robustness analysis

Figure 2: Order-*m* p-values

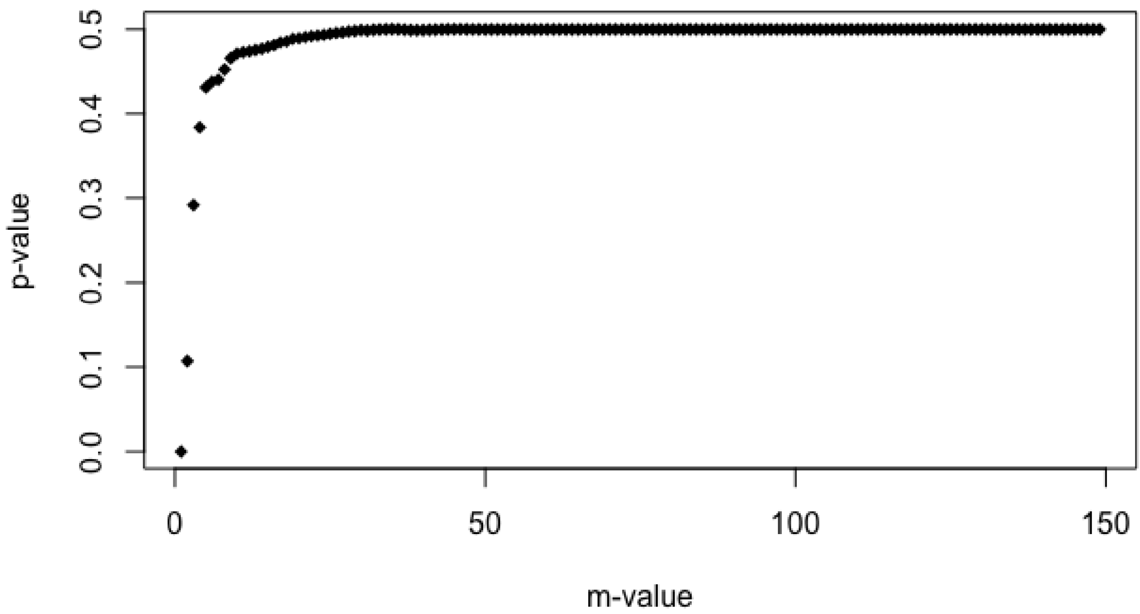
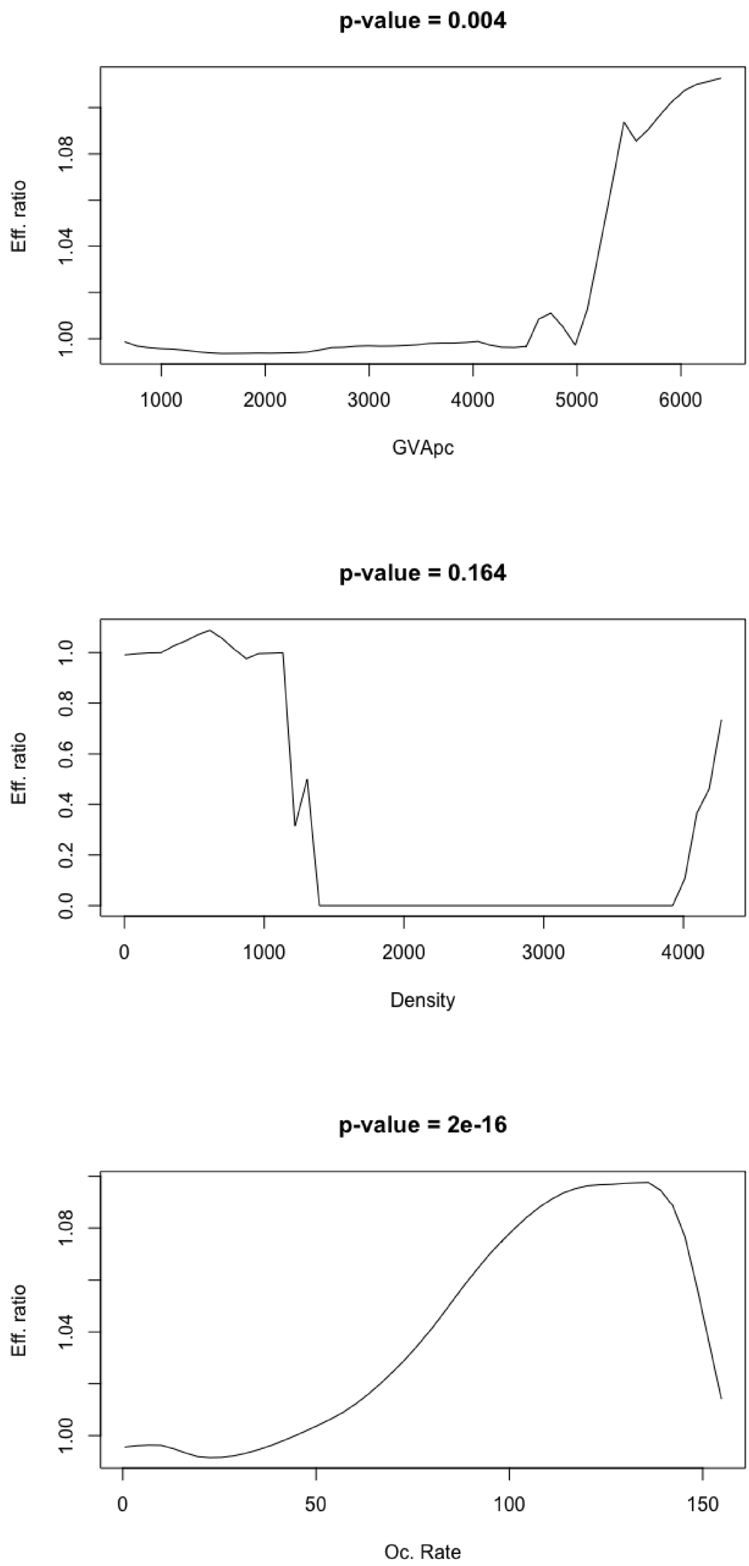
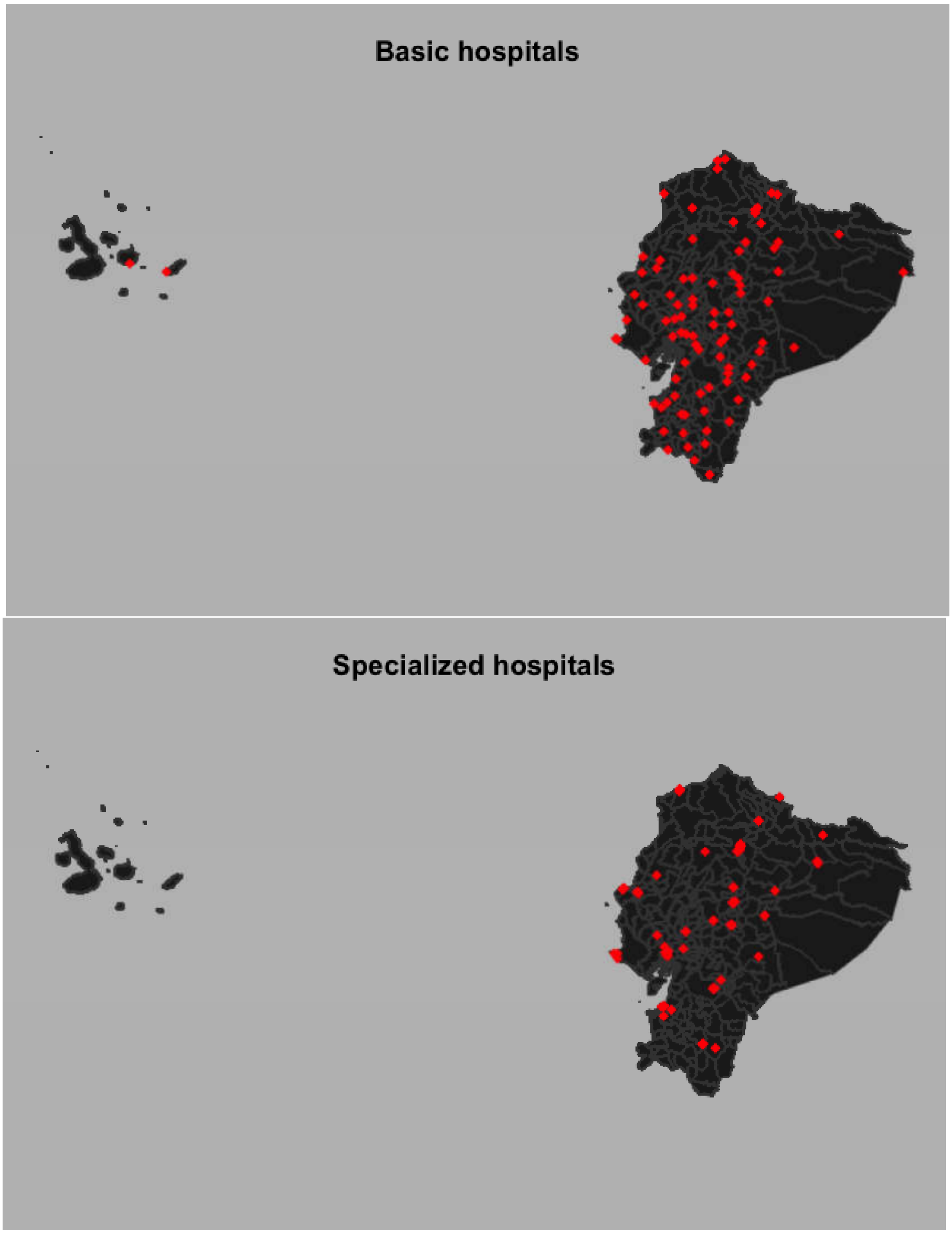


Figure 3: Conditional order-m partial regression plots



E Hospital distribution

Figure 4: Territorial distribution of basic and specialized hospitals



F 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 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 10: Total interregional patients by canton and province of residence

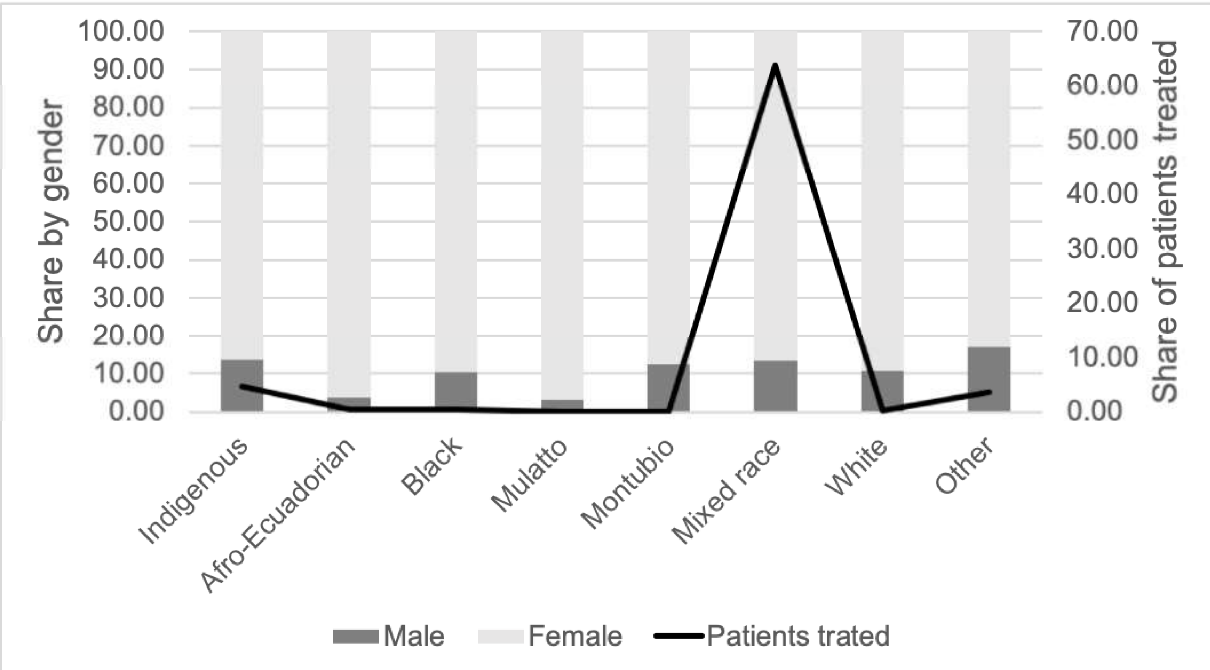
Province of residence/Canton of the hospital	Cuenca	Guayaquil	Quito
Azuay	284	18	1
Bolívar	1	23	40
Cañar	119	41	2
Carchi	0	1	35
Cotopaxi	1	12	76
Chimborazo	7	21	35
El Oro	49	50	15
Esmeraldas	0	52	62
Guayas	23	1709	9
Imbabura	1	3	89
Loja	28	8	14
Los Ríos	2	260	18
Manabí	7	152	43
Morona Santiago	62	1	7
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 10 (continued)

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

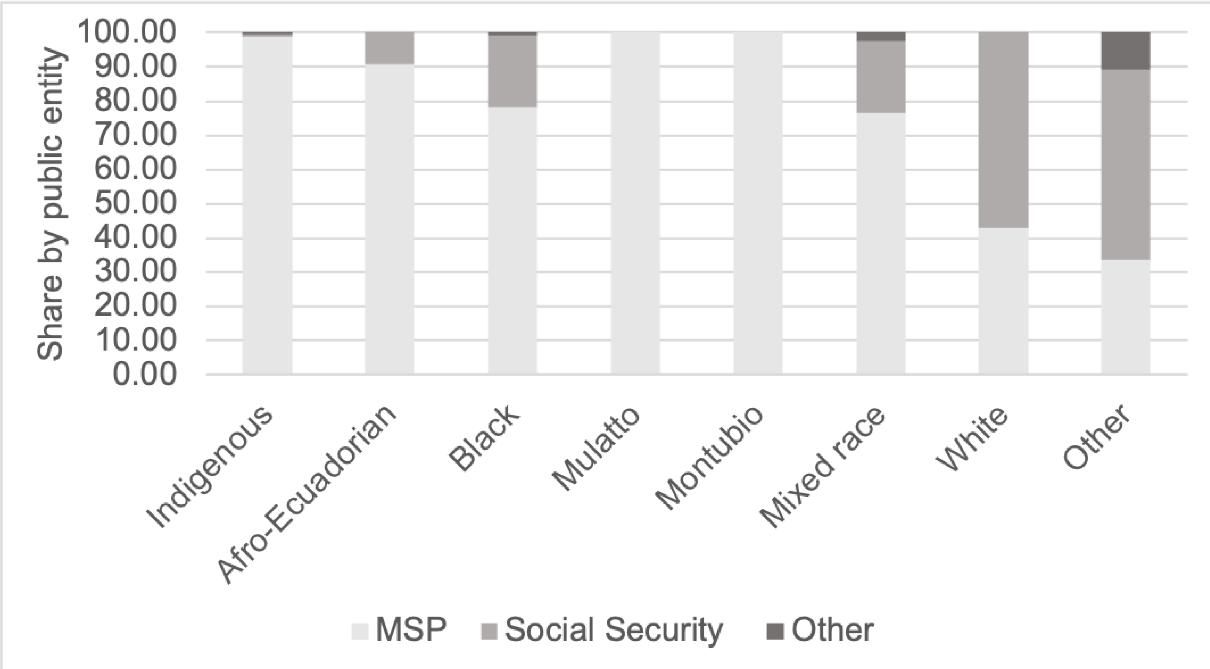
Source: The author.

Figure 5: Share of interregional patients by gender and ethnic group.



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

Figure 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

GEAR

GRADUATE
PROGRAM
IN APPLIED
ECONOMIC
RESEARCH

UAB Universitat Autònoma de Barcelona