TAMPERE UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Master's Degree Programme in Public Choice

MASTER'S THESIS

Evaluating the incidence of regional patient migration in Italy through the formulation of a theoretical model and the conduction of a spatial econometric analysis

Candidate

Alessio Moro

Supervisor

Katri K. Sieberg

May 2019

Abstract

This research thesis investigates the phenomenon of regional patient migration in Italy, which relates to individuals moving from a region to another to receive planned health treatments in specific local health authorities due to various possible motives. The main intent of the research is to provide an innovative contribution to fill a research gap that has been recognised in the existing literature. For this purpose, the thesis outlines the topic and reviews relevant information from the literature that upholds the conduction of the enquiry, defines the research methods and the analysis framework, which support the advancement of a theoretical model and the execution of an empirical analysis, and discusses the results to provide sound conclusions. In particular, the theoretical model illustrates how regional patient migration can emerge even from a situation of perfect equilibrium, while the empirical analysis of collected data, based on methods from the field of spatial econometrics, demonstrates how certain factors can be associated with its occurrence over time. The theoretical and empirical outcomes, combined with other concepts from the literature, are employed to deliver wisdom on the need for rational public policies and a distinctive solution for the issue, with the achievement of repairing a fractured equilibrium and sustaining it in the future to protect the public health care system and the welfare of the population. The final chapter concludes the thesis with a rundown of the research, the recognition of its limitations and suggestions for further enquiries on the matter.

Acknowledgements

I would like to communicate my sincere appreciation for the teachings and guidance of my supervisor Katri K. Sieberg, which had significantly contributed to the advancement of this research thesis. I also intend to acknowledge the roles of other lecturers, who had decided to share their meaningful expertise during the entire continuance of the studies. I wish to convey my deepest gratitude to the members of my family, whose fundamental encouragement and support have allowed me to realise my achievements and to become the person I am today.

Contents

Introduction

Leading a life in rather good health should be a fair achievement to often pursue over the course of a lifetime. The realisation of this objective can mainly occur through two paths that are connected to each other: on one hand, individuals should care for their own health in accordance with their capabilities, for instance by engaging in healthy behaviours over the course of their lives; on the other hand, the government of a state should introduce sound policies in areas of interest where certain factors may involve potential consequences on the health of its population, such as health care, employment or social support, so that the probability that negative health outcomes happen can be reduced. In democratic states, the design of rational policies is normally performed by representatives that are elected by the population. In the health care sector, the actual implementation of health policies takes place through the establishment of health care systems, which can be portrayed as complex organisations of people, institutions and resources that allow for the provision of health care services in a country. In general, they can be entirely privatised, publicly managed by the state or organised with a mixed type of offer; however, even though they can be defined by common basic principles, their specific institution is influenced by economic, political, social and cultural aspects that are peculiar for each different country in the world. An acceptable establishment of a fairly functioning health care system can provide each individual of a population with support in case of either illness or indigence. In combination with a society that is primarily composed by people who engage in rather healthy manners and retain overall sentiments of care for their health, such an establishment permits to attain positive social welfare, which may bring positive elements that interact with one another over time, such as a higher efficiency of labour, economic growth of the country, shares of income to devote towards savings or consumptions instead of direct health expenditures and a diminished pressure on the health care system, especially if its sustainment relies on taxes that are collected from the population.

In Italy, the health of the population is theoretically considered and defended with profoundly high regards. Indeed, the article 32 of the Constitution of the Italian Republic declares that health must be safeguarded as a fundamental right of the individual and a collective interest, with guarantees of free medical care to the indigents; this statement is certainly sound and aligned with the conceptual objective of the state being a positive contributor for the health of its population, which should occur through the resources of a functioning health care system and the design of sound policies in all the related areas of interest, as previously underlined. In general, Italy counts on an overall satisfactory provision of health care services and rather healthy behaviours of its population, which are highlighted by several positive end results. Nonetheless, some significant problems have conceived various insecurities that threaten the sustainability of the Italian health care system and the supply of excellent health care services to the society; for instance, a known issue concerns the persistence of regional variation in the quality of the supplied health care services, which is presumed to exist due to substantial regional differences in structuring health care offerings on a local level, within the general framework that is established nationally, as well as in terms of funding and management of the resources related to each regional health care system.

The argument of interest for this research thesis regards a reality that results from the existence of some of these problems of the health care system. The phenomenon is that of regional patient migration, which concerns individuals moving from a region to another to exercise their rights to health in local health authorities situated in different areas, for reasons of resource availability or higher quality of care; for sure, the uneven provision of health care services across the territory represents one of the issues that can influence the occurrence of patient migration to a certain degree, since it can cause the manifestation of various health outcomes among the regions and therefore individuals may be willing to move elsewhere to seek for the most appropriate health treatments in relation to their needs. Still, even when these regional movements could theoretically take place, the ability to relocate for health motivations depends on the characteristics, necessities and resources of a person and thus the phenomenon may take effect or not when looking at separate cases, allowing for the potential development of situations in which an individual has to renounce the obtainment of health treatments and to endure unintentionally neglected needs. The importance of regional patient migration and its related outcomes can be discerned when considering that if public health care services, funded by taxes collected from the entire population to provide those who are in need with free access to basic services and to ensure the fulfilment of the objectives outlined in the Constitution of the Italian Republic, cannot be evenly guaranteed to any entitled individual in every region of the country without divergences, then health and economic disparities are created among the population and thus the central state fails to guarantee what should otherwise be defended at a constitutional level.

First of all, a theoretical model is formulated to illustrate how patient migration can come into existence at the beginning. Then, the phenomenon in the country is examined with a data analysis to understand its occurrence rate and to accomplish three research purposes. The first research objective is to analyse whether certain factors, concerning the uneven delivery of health care services or other potentially complementary matters, have a significant relationship with the occurrence of patient migration; for this intent, the examination will account for appropriate quantitative data that relate to a series of factors which are depicted as important by the information gathered from the upcoming review of the literature. The second research scope deals with analysing whether some sort of spatial dependence in the happening of the phenomenon exists when taking an observation of interest and those that are close to it into account, since patient migration happens across various portions of the country and hence can be observed as a national phenomenon, which involves interactions between people in the whole territory and is not only related to a single individual who seeks for appropriate medical care in isolation from other people. The third research aim is similar to the second and is concerned with examining whether the occurrence of patient migration in an area is affected by one or more of the identified factors existing in nearby places. Considering this geographical nature of the matter, location must be considered as a fundamental factor of influence on the observations in the data, since things more likely influence each other the closer they are, so that the factors contributing to the occurrence of patient migration in the country can be further highlighted. The fulfilment of the described research purposes of the thesis will happen with the delivery of answers to three related research questions, which have been formulated as follows:

- To which extent do the identified factors influence the occurrence rate of regional patient migration?
- Is regional patient migration of a certain location influenced by its occurrence in neighbouring areas?
- Is regional patient migration of a certain location influenced by one or more of the identified factors existing in neighbouring areas?

The provision of answers to the research questions depends upon the formulation and testing of related hypotheses, which will take place in the analysis framework and the data analysis. For this purpose, spatial econometric analysis has been identified as the most appropriate methodology for the examination of collected data that associate with the subject, which will occur with the employment of dedicated statistical spatial models and statistical tests; to be more specific, spatial analysis can be described as a type of geographical examination which intends to detect the existence of patterns of human behaviour and to explain their characteristics using both global and local area analysis. The primary instruments used to perform this type of analysis will be the Moran's I test for spatial autocorrelation, which examines whether a phenomenon is clustered or not, and the implementation of spatial regression through statistical models, which considers eventual spatial dependency in the analysed data, that will be quantitative by nature and will not be accompanied by any sort of supplementary qualitative measure. The rationale that supports the choice of this research and analysis methodology mainly comes from the geographical nature of the matter and the absence of this particular employment of the method in other studies, as illustrated afterwards in the review of the literature. The insights of the theoretical model and the results of the data analysis will be used to portray valuable policy advice concerning the issue, that legislators should take into account to reduce the occurrence rate of regional patient migration by targeting the aspects that need to be considered with more urgency, while avoiding potential risks of overlooking details that may lead to the happening of unintended consequences, such as the stagnation of the phenomenon or an escalation of regional differences. Moreover, considering the potential occurrence of negative health outcomes among the population resulting from the features of the issue, especially in the long term, a timely employment of innovative policies on the subject is deemed to be fundamental.

The thesis inspects and reviews the outlined theme through numerous sections. The first section, "Background", gives a background on the topic of regional patient migration in Italy, while the second one, "Literature review", illustrates the relevant information on the matter of the existing scientific research to support the methodological choices of the research. The third section, "Research methods", highlights the details concerning the primary theoretical foundations that surround the research, such as the description of the statistical models to employ and the rationale concerning the selection of the most appropriate regression model for the specific data, the analysis framework that contains the theoretical model and definitions for the data analysis, as well as details on the process of data set preparation, which involves the collection of data and the selection and transformation of the variables to build the definitive data set. The fourth section, "Data analysis", is dedicated to examining the data and shows the results for each subtopic of the matter, which are further portrayed with other thoughts in the fifth section, "Discussion". The last section, "Conclusion", ends the thesis with final words, a few reflections on the limitations of this research and possible ideas for supplementary studies on the topic. Some appendices with further information follow the last section, together with a bibliography of the references, a list of figures and a list of tables.

Chapter 1

Background

The Italian health care system is constructed upon the ideology of universal health as an individual right and a collective interest, established by the article 32 of the Constitution of the Italian Republic since 1948. The outset of its history can be traced back even to the foundation of the Kingdom of Italy in 1861, from which many changes occurred and lead to the development of the health care system as it is known today in the Republic of Italy. The most significant events, happened after the enactment of its Constitution, can be considered the following: the institution of an independent Ministry of Health on 14 August 1958; the conversion of hospitals to public entities, whose functions and financing were regulated under the aims of national and regional plannings to offer health treatments to Italian and foreign indigents, in 1968 (Law No 132 of 12 February 1968); the creation of ordinary administrative regions, which were given administrative functions on health and hospital care in 1972 (Presidential Decree No 4 of 1972) and for which a national health fund, to be divided among them based on population density, was established in 1974 (Law No 386 of 17 August 1974); the constitution of the Italian National Health Service in 1978 (Law No 833 of 23 December 1978), as a result of cultural, political and social processes that had occurred in the previous years, which was gradually implemented by all regions and autonomous provinces between 1979 and 1981. Further recent policies have introduced changes on certain matters, such as more clearly divided responsibilities among government levels and promoted cooperation among health providers (Legislative Decree No 229 of 19 June 1999), deeper fiscal decentralisation and the abolishment of the national health fund (Legislative Decree No 56 of 18 February 2000) in favour of regional taxes and funding coming from a national solidarity fund in case of financial difficulties with the provision of the basic package of health care services.

The establishment of the Italian National Health Service effectively implemented the protection of individual and collective health as described by the article 32 of the Constitution of the Italian Republic, especially by abiding to a set of specific fundamental principles: universalism of access to uniform levels of health care; equality of treatment; respect of individual dignity and liberty; development of prevention schemes; public democratic control of the health care system. The current Italian health care system is based upon a mixed model in which the public offer, regulated by the parliament and effectively implemented through the National Health Service, prevails to fulfil the scope of the state supporting the health of its population, even with the accompanying offerings of private health care services in the market. The National Health Service can be defined as a series of national and regional entities and institutions that are organised according to different levels of governance and responsibility; while the state has the duty of ensuring all its citizens the right to health through a strong system of guarantees based on a series of essential levels of care, which form a statutory benefits package that must be offered equally to all the residents in the entire country, every region has direct responsibilities for the expenditures needed to achieve the national health aims and for the implementation of the government directives, which occurs through different local entities, such as local health authorities and general hospitals, that provide health care in their territory trough public or private accredited health facilities while being held accountable towards their respective region. The National Health Service is financed by a mix of general taxation and statutory health insurance contributions. The sources of funding include revenues collected by local health authorities in each region, whose amounts are defined by agreements made between the state and the regions, regional taxes, contributions from special administrative regions and autonomous provinces, as well as contributions from the state for the missing portion of the needed monetary resources. The public health spending has been following an increasing trend over time, as the government financing for the National Health Service had gradually risen from 71,3 billion € in 2001 to 111 billion € in 2016.

A series of positive results reflect the general effectiveness of the Italian health care system. For instance, life expectancy at birth reached 82,7 years in 2015 from 79,9 years in 2000, the second highest in the EU after Spain, and was paired with two-thirds of the Italian population reporting being in a good state of health, while the level of health spending of 2.502 € per capita in 2015 was 10% lower than the EU average of 2.797 €. Moreover, research has also underlined how favourable individual behaviour towards health has been preserved compared to other countries, which is shown by measures such as low rates of smoking and alcohol consumption [36]. Nevertheless, even though these positive results have been achieved and maintained, some problems, such as the mentioned regional variation of health care quality, have produced various uncertain circumstances that threaten the long-term sustainment of the health care system and the delivery of fairly distinguished health care services in the country, which represent essential components of exceptional population health; as previously said, the interest of this research surrounds the circumstance of patient migration that takes place among regions in the decentralised health care system of Italy. In particular, regional patient migration regards people moving from a region to another to gather health treatments elsewhere for various possible reasons, such as those of resource availability or higher quality of care. The phenomenon can happen because free patient choice is considered to be one fundamental feature of the decentralised Italian health care system; even though this freedom should also be an instrument for implementing competition mechanisms among health providers, containing the health expenditures and raising the quality of health treatments, advancements of health providers have seemed to be unbalanced among areas of the country. Internal movement of people that occurs from a region to another to obtain higher-quality treatments may cause several problems, such as cost and time issues for patients, inabilities to manage excessive amounts of individuals for receiving health authorities in relation to their available resources and development of inefficiencies for hospitals located in regions with a negative net migration balance, due to failures in reaching economies of scale and reimbursement obligations towards other regions for the costs sustained to treat their escaping patients. This phenomenon seems to be intertwined with the issue of regional variation in the quality and efficiency of the supply of health care services, with disparities that have appeared to create a clear divide mainly between the macro areas of Northern and Southern Italy.

In this framework, providing the definition of common-pool resources is deemed to be fundamental, since it can facilitate the comprehension of the underlying importance of the phenomenon. Researchers in the literature defined a common-pool resource as "a natural or man-made resource system that is sufficiently large as to make it costly (but not impossible) to exclude potential beneficiaries from obtaining benefits from its use" [38]; differently from a public good, for which its use by an individual does not subtract from its availability to others (e.g. individual consumption of public security does not reduce the general level of security that is available for a population), a common-pool resource can be influenced by effects of crowding and overexploitation by its users that lead it to approach the limit of the number of resource units it can produce. In the context of having a number of similar common-pool resources throughout a territory, organised by the size of local populations, the exploitation of certain resources in excess of their capabilities may be accompanied by the underuse of others, since local individuals gather fewer resource units than what can be produced, leading to provision inefficiencies and unbalances between the various resources over the entire territory. The definition and considerations are important for the following reasons. As previously mentioned, the health care system of Italy is composed by a series of regional health care system, which are organised and funded in accordance with the size of the population that resides in a region. Each regional health care system can be considered as a common-pool resource, because it is publicly funded and produces resource units, forming the local public health care supply, which can only be used in a limited manner given the finite availability of resources (e.g. medical equipment, personnel, finances); although the regional systems and the respective supply production capabilities are formed upon the needs of the local populations, the resource units can be obtained by everyone living in Italy, as patients have the right to gather health treatments in any area of the country independently from their region of residence; if an individual obtains certain health treatments in another region, the receiving region is reimbursed of costs and acquires the potential benefits (e.g. more efficient employment of resources, potentially higher attraction rates, further possibilities for personnel training), while the region of origin receives no benefit from the health treatments but bears the costs and eventual negative outcomes that may occur (e.g. failure to reach economies of scale, underuse of resources, demand for cost-cutting measures, potentially lower attraction rates). Taking this evidence into account permits to contemplate how the happening of regional patient migration may potentially create many imbalances: on one hand, the overuse of resources that exist in a region, caused by treating an excessive amount of patients, may withdraw usage opportunities that could be necessary for another person in the area, considering that such resources should be proportioned to the local population size; on the other hand, regions with high escape rates may become less capable to offer sufficient health treatments due to an unceasing incidence of negative outcomes that are intertwined with one another (e.g. costs are cut by lowering the rates of personnel, which also reduces the ability to retain patients), while regions with high attraction rates may continuously attain benefits thanks to the occurrence of positive outcomes (e.g. attraction of qualified personnel or stakeholders) that can lead to enhancements of the supply capacity, which could even counteract the potential concern of resource overuse. In addition to causing disparities among regions, the issue may also induce the creation of inequalities among the population (e.g. diverse health treatment costs when accounting for every sort of expenditure), depending upon individual preferences and opportunities, which can result in totally different outcomes when comparing individuals (e.g. one moves to another region without cost concerns, while another has to refrain from gathering some health treatments). These ideas will be further developed in a forthcoming theoretical model, under the sphere of a concept connected to common-pool resources that is known as “the tragedy of the commons”, while the analysis of certain data will provide additional insights on the matter.

Some negative consequences resulting from these problems have been reflected into the need for recent government interventions that differentiated virtuous regions from those in difficulty because of significant deficits, creating necessities for agreements that included objectives, limits, incentives and sanctions specific for each region that must have been respected for the obtainment of monetary financing from the central state. In recent years, regions in very aggravated financial and economic situations were placed under the scope of special plans; even though deficit issues have been somehow reduced, positive results have been achieved only through administrative and financial measures (e.g. shortage of employment, increase of out-of-pocket payments), without touching important structural issues that concerned the efficiency of resource usage, wasteful spending, disparities between the technical and perceived quality of health care services and their related outcomes. Furthermore, problems of supply imbalances may also be at risk of being aggravated given that the Italian National Health Service covers all citizens and legal foreign residents in a universal manner, with the opportunity for temporary visitors to receive health care services, albeit by paying for the costs of treatment, and for undocumented immigrants to access urgent and essential services. As a consequence, the phenomenon of health migration may also originate effects of crowding and overuse of resources in certain locations of the country. Together with economic and financial imbalances, these impacts may make it difficult for all local health authorities to deliver the health care basket benefits homogeneously within the national territory, as planned by the government and as an instrument to guarantee the constitutionally defended health rights, due to the development of concerns in the overall efficiency and quality of health care supply. Recent information on the matter indicates that regional patient migration is still a commonly occurring phenomenon, with regional differences in the provision of health care services being of significant relevance. Therefore, the present research thesis intends to analyse data on internal movement of patients across Italian regions in relation to various factors, with the purpose of responding to the research questions presented in the introduction, so that it will be possible to comprehend the scale of the issue, to detect eventual spatial patterns across the country that relate to it and to find out which factors may be significantly correlated with the occurrence rate of regional patient migration in various areas. The examination will take place through the implementation of certain spatial econometric analysis methods at a provincial level, which will deliver a unique contribution to the literature; in fact, as illustrated in the next chapter, many researchers have already examined the phenomenon of regional patient migration across Italy with various analysis techniques, but their studies did not include any sort of spatial approach that is similar to that of the present research and focused on a regional level. To be more specific, this research will concern the portion of planned health care treatments in public and accredited private health care facilities, related to the most common treatments for diseases that are not urgent and can be easily defined over time, which can follow either outpatient visits and recovery instructions through patient placement in waiting lists or preceding treatments in the context of pursuance of treatments for the same condition; the research focuses on the provision of care in the short to medium term, excluding other forms of treatment which are less general and for which a different analysis approach should be employed, such as long-term care. Established upon the research results, the research will portray some policy suggestions that can target the factors of interest with precision, focusing especially on long-term outcomes despite possible short-term pitfalls, so that it can be possible to reduce issues of economic differences among local health providers and occurrences of complications for patients (e.g. difficulties with receiving the necessary health treatments in timely manners, bearing of significant costs because of long waiting lists or insufficient service provision in a certain region). Even though the objective may be challenging to reach due to the complexity of the examined problem, the research is accomplished with the highest efforts, since advancements in aspects of the health care system will facilitate the respect of the fundamental principles surrounding the National Health Service and the concept of universal health of the population as regarded by the Constitution of the Italian Republic.

Chapter 2

Literature review

This chapter presents the relevant studies on regional patient migration in Italy, using a chronological order, and evaluates them to illustrate the current state of the research and to delineate the contribution of the thesis that intends to fill the mentioned void that was found in the literature. For these purposes, the review focuses on recognising the current significance of regional patient migration in the country, identifying the main factors that may influence free patient choice of treatment and hence the occurrence of patient movement among regions and investigating the research methods that other researchers have already employed to examine the phenomenon.

Levaggi and Zanola (2004) were concerned with the persistence of regional patient migration in Italy in the early 2000s, especially after certain legislative changes on the regionalisation process of its health care system had come into force in the 90s. Indeed, the introduction of regional funding schemes and free patient choice created potential for significant variance in the quality of health care services, despite the promotion of competition among health providers through elements of an internal market, and the consequent increase in the rate of patients escaping into other regions. Their research aim was to examine the determinants of patient migration to disclose useful insights, especially for the poorer regions that were affected by high escape rates and payment obligations to the others for the services bought by their emigrating patients. To identify the effects of certain factors, they used a modified gravity model of patient migration and estimated it with panel observations on regional migration and quality indicators for the period 1994-1997; even though they recognised a high degree of aggregation due to constraints in the available data as a limitation of the study, they found out that regions with lower patient outflows also had incomes that were greater than average [26].

Messina, Vigiani, Lispi and Nante (2008) analysed the occurrence of the phenomenon in the year 2003 to determine the hospital supply of health care services, with the goal of delivering suggestions about the perception of patients on their quality and organisation. To conduct the analysis, they evaluated the usage of hospital centres in 2003 through an instrument called Gandy Nomogram, which consists of a squared Cartesian area with the percentage of resident patients admitted to a local health unit in a certain district on the x axis and the total demand percentage satisfied in that district on the y axis, by dividing it into four areas to determine the condition of an observation with respect to numerical data on ordinary and day hospital patient discharges. Their findings showed that patient movements seemed to be prevalent towards nearby regions and decreasing as distance increased, with short-range emigration taking place in regions of Central and Northern Italy and long-range emigration prevailing in those of Southern Italy [30]. As Montefiori (2005) noted, these differences could have existed because patients may decide to endure both monetary and non-monetary distance costs if they expect to receive positive returns in terms of better quality from a health unit that is located the furthest away from their district of residence [31].

Messina et al. (2013) instead stratified specific portions of data on regional patient migration depending on disease severity in cardiac surgery units of three health areas in the single region of Tuscany, for the period 2001-2008, to study the influence of severity of patient condition on the occurrence of the phenomenon, examining it under a diverse light compared to what other studies had previously done and therefore filling a gap they had identified within the research literature. The analysis, which was conducted with the already mentioned Gandy Nomogram, showed that, with an increase in condition severity, more resident escapes than admissions occurred in one health area compared to the other two locations. As a consequence, the results clearly highlighted how patient migration can be affected by the specific aspect pertaining to the degree of severity of a condition, a finding which could be of certain interest when designing policies targeting the phenomenon [29].

Toth (2014) accounted for the migration of patients across regions of Italy as one of three specific indicators to analyse whether the gap between health care systems of Northern and Southern Italy widened or compressed over time, especially under the influence of the mentioned regionalisation processes. To conduct the analysis, he studied the defined indicator of regional patient migration by using a “synthetic mobility index”, which was calculated as a ratio of the attraction index to the escape index for each region, for the period 1999-2009, so that results could have been produced in combination with the other two indicators; as mentioned by the author, his analysis solely accounted for ordinary admissions of acute patients, with exclusion of treatments related to patients admitted without overnight stays and those following procedures of long-term care. His research findings described that, for the period 1999-2009, the overall influx of Southern treatment-seeking residents into the regions of Northern Italy had increased, while the opposite influx had decreased, depicting an increment of the gap between regions in the macro areas as well as the continuous significance of the patient migration phenomenon, albeit without any further enquiry on its specific causes [44].

Fattore, Petrarca and Torbica (2014) focused on migration of patients for aortic valve substitution, a specific health treatment procedure, for reasons related to the importance of patient migration for cardio-vascular diseases in the country and certainty in tracing the procedures from the data. In their analysis, they employed t-tests and chi-square tests to assess the differences of means and proportions, as well as logit and multi-level logit models to discover the factors related to patient migration. The authors found that this specific facet of patient migration, which had taken place primarily from Southern to Northern Italy, was characterised by three important aspects: age of patients, as those admitted in their regions of residence were more than 3 years older than those admitted in other regions; length of hospital stay, since patients admitted in their regions stayed in hospitals approximately 0,7 days longer than those admitted in other regions; presence of private accredited providers, which were more likely to admit patients incoming from other regions compared to public hospitals [14].

Brenna and Spandonaro (2015) also studied patient migration due to an interest about equity and financial reasons, resulted from the process of decentralisation of the health care system and the possibility for patients to exercise their rights to free choice of treatment in any region of the country. In particular, they examined cross-border regional patient migration using data on five sample regions for the year 2010, by calculating an index, measuring the ability of a sample region to attract patients from another, which was used to select further six regions with the highest percentage of their residents exported to each sample region and to compute attraction indices for hospital categories, separating boundary and distance cross-border migration. Their results portrayed generally higher attraction indices for private providers compared to public ones, for both types of patient migration, which appeared to drive flows of patients from Southern Italy to Northern and Central Italy; the reason for this connection may come from the gradual improvements of northern regional health care systems which had happened through accreditation processes with private providers and contrasted with a substantial lack of developments of health care systems of southern regions [9].

Pierini et al. (2015) assessed patient migration of individuals admitted for bone marrow transplant in the Hospital of Perugia, as it was the second most important structure in the country for the treatment. With the Gandy Nomogram, they analysed data about ordinary hospital discharge records of patients admitted for bone marrow transplant, of the period 2000-2013, to detect movements of patients over time; with a total number of incoming patients that was almost split in half between residents of the region Umbria and individuals coming from other areas, the results showed a high attraction strength of the structure, which had increased and remained stable for distant regions but has recently decreased for bordering regions; moreover, a portion of residents seeking for health treatments elsewhere highlighted a recent increase of escape rates, despite the initial ability to satisfy the needs of the local population. In addition to gathering useful insights on patient migration in a specific context, the researchers also illustrated the possible implications of the location aspect on the phenomenon [40].

Balia, Brau and Marrocu (2017) evaluated the causes of interregional patient migration by studying regional bilateral patient flows, using hospital discharge data concerning an extended time period between 2001-2010. Their findings underlined a significant role of the technological endowment and performance of regional health systems, while discovering that these characteristics of neighbouring regions produced exogenous spatial effects that influenced the phenomenon in other nearby areas [6].

The economic organisation OECD (2017) recently published its latest report that reviewed the state of the health care system of Italy and the condition of its population health, providing an overview from an international perspective. Among other matters, the organisation highlighted the actual relevance of patient migration in the country, stating that movements to gather health treatments appeared to occur towards regions in Northern and Central Italy, since those in Southern Italy had shown high escape rates and low attraction rates. Furthermore, it also warned that a significant portion of the population reported unmet needs for various reasons, including geographic barriers and long waiting lists, with individuals in the lowest income group being affected more than those in the highest income group (e.g. 15% compared to 1,5% in 2015). With regards to the causes surrounding patient migration, the OECD underlined how seeking for higher quality medical care in other areas seemed to be a widely accepted circumstance due to the existence of regional variations in the actual availability of resources and the perceived quality of care; these variations appeared to happen because of differences among regions in their abilities to deliver the services of the benefit package, that resulted from discrepancies between the allocated resources and those required, therefore creating the need for certain regions to provide additional monetary resources towards complete funding of the services. Regarding the availability of resources, this situation could be seen as more concerning when also considering that, in terms of resources for the entire country, the overall number of hospital beds for acute care had declined from an average of 4,2 beds per 1.000 population in 2000 to 2,8 beds in 2013, while the ratio of nurses per doctor had continued to be quite low (e.g. at 1,5 compared to an EU average of 2,3 in

2015), despite an increase in the total number of health personnel [36]. Concerning the perceived quality of care, useful information can be found into a report of the OECD on the quality of the Italian health care system for the year 2014, in which the organisation illustrated that the regions and autonomous provinces had been implementing the national guidelines through independent decisions, without a consistent framework of robust standardised means of implementation and monitoring, and applying the results of national frameworks on quality monitoring and improvement (e.g. Essential Levels of Care) in an inconsistent manner, while following only a minimum set of standards; as a consequence, the organisation called for a stronger role of the central state in defining and enforcing a more standardised realisation of the national guidelines in all regions and autonomous provinces [35].

The presented research studies clearly outline the continuous significance and happening of regional patient migration in Italy, therefore suggesting the phenomenon can be considered as persistently existent in the country and hinting at necessities for changes concerning how its decentralised health care system operates. As a matter of fact, as Tiebout (1956) suggested with his model and hypothesis, local provision of public goods can lead individuals disclose their inclination for them through their eventual decisions to move to another jurisdiction where the local expenditure more closely matches their preferences and maximises their personal utility, which is a mechanism that has been renamed as "foot voting" [42]; therefore, movements of individuals from a region to another may be taken into account as acts of preference disclosure and implicit voting that signal the need for modifications of certain components of specific regional health care systems, so that they can more closely match the quality of treatments to that obtained elsewhere and meet the needs of the local population in a region. The evaluation of these studies in the literature depicts that various analysis methodologies have been used to examine the phenomenon of regional patient migration over time, ranging from the development of indices and models to review the phenomenon as a whole, to looking at it under the light of specific health treatment procedures; as already hinted previously, this information is utilised to support the employment of spatial econometric analysis on a provincial level as an appropriate methodology to examine the data on the matter, which will permit the thesis to deliver a unique contribution to the literature. Moreover, apart from recognising the continuous relevance of the phenomenon and enquiring about the research methods that have already been employed, the factors that may influence the occurrence of patient migration need to be identified, so that it will be possible to define the statistical models for the data analysis. Certainly, as suggested by the mentioned resources in the literature, a series of factors relates to the quality of health care services offered by local health authorities in each region; this case can be supported by the evidence from the Italian Ministry of Health on quality monitoring of the services that form the essential levels of care, which takes effect using weighted indicators that evaluate them in terms of appropriateness, quality and efficiency to find out whether regions provide either an adequate or a compromised offering level, which have been showing that regions of Northern Italy have always been able to comply with the national objectives, while other regions, especially those of Southern Italy, have been more inconsistent and sometimes unable to compete on the same levels in terms of alignment with the national guidelines. On health care quality, an appropriate definition of its components can be based upon the model proposed by Donabedian (1966), which considers three indicators to be relevant: the outcome of medical care, which is a concrete measure and whose validity is rarely questioned; the process of medical care, which concerns the proper application of the medicine practice; the structural nature of the location of medical care, which enables good practice depending upon the availability of adequate conditions and equipments [11]. The reviewed studies have also illustrated the possible significance of other factors, such as income of regions, presence of accredited private health care providers and performance of regional health care systems, including potential spatial spillovers from externalities in nearby areas. In a dedicated subsection of the research methods, together with further evidence, this awareness will contribute to the definition of the set of variables to include into the data analysis.

Chapter 3

Research methods

The present chapter outlines various essential aspects that form the research methods of the thesis. First of all, the reader is introduced to certain theoretical foundations that are employed to establish the main features of the research found in the analysis framework, such as those on the tragedy of the commons and spatial econometric analysis. Secondly, the analysis framework illustrates a theoretical model and how the aspects of the theory are implemented for the analysis. Finally, the last section portrays the process of data set preparation, which involves the data collection, the selection of information from the data and the transformation of the defined variables for the analysis.

3.1 Theoretical foundations

3.1.1 The tragedy of the commons

As outlined in the background section, the various health care systems of each Italian region, which are organised and funded depending upon the size of the local populations, can be considered as common-pool resources that each regional resident can utilise to fulfil his or her health needs without problems. However, in the presence of differences between regional health systems, individual rights to free treatment choice can cause the occurrence of regional patient migration, which may create imbalances for local health authorities, especially when accounting for obligations on cost reimbursements in favour of receiving regions. In fact, regions with only high levels of incoming patients could be affected by overuse of resources if maximum capacity is reached, while regions with only high levels of outgoing patients could possess underused resources, without improvement opportunities and with cost obligations towards other regions.

First of all, the work of Olson (1965) on the logic of collective action can be associated with patient migration in a partially tangent manner. In his book, he described how members of an interest group, when driven by self-interest, can decide to free-ride on the action of others to receive the benefits of a collective cause without contributing to it; this event tends to be absent in small groups but to become relevant as they enlarge, as the significance of individual contributions for group performance and the per-capita share of benefits reduce as the total number of people in the group increases [37]. Bendor and Mookherjee (1987) contributed to these ideas by confirming that the problem cannot be solved neither through cooperation, which is unsustainable in large groups, nor through centralised solutions, which can become afflicted by a number of problems (e.g. difficult development of effective monitoring systems); instead, they proposed the organisation of interest groups within federal structures, which can enforce cooperation and eliminate free-riding through small groups that together form a larger group [8].

More importantly, the previous thoughts can relate the phenomenon to the theory about a concept known as the tragedy of the commons, primarily portrayed by Hardin (1968) when discussing the issue of overpopulation in a world with finite resources. In particular, it concerns the shared usage of a resource that is open to anyone; at first, the common use can continue to happen as various circumstances, such as those caused by nature, maintain an equilibrium over time; however, independent decisions made by rational individuals, who seek to maximise their own utilities by taking as much as possible from the resource while sharing the downsides with the others, will eventually lead to the collapse of the shared resource, which will not be able to sustain any sort of production of resource units for them anymore. A clear example the author made regarded the shared usage of an open pasture by various herdsmen; to maximise the individual benefit, each herdsman would add as much cattle as possible to the common pasture, so that eventual negative consequences caused by an overall overgrazing would be endured by all the herdsmen; over time, such a system will collapse, leading to ruin for every individual relying on it and causing a tragedy of that common [18].

The contributions provided by Ostrom (1990), that enquire about the presented concepts, provide further explanations on how individuals can collaborate in the presence of common resources. In her book, using examples of real communities where people cooperate to govern a shared resource, she suggested that neither a centralised solution overseen by a state nor a privatisation of the resource is able to sustain a common productive usage over the long term; for the former, the state would be prone to making errors on the organisation of the resource usage, while having issues with costs for monitoring individual behaviours and potential imperfect information; for the latter, dividing a resource between individuals through private rights could be prone to the occurrence of unfavourable random events against only some of its portions (e.g. rain not falling in certain areas of a privately divided soil, in which grass is supposed to grow for the nourishment of animals pasturing there) or to the need for sustaining additional costs that can be avoided when the resource is instead commonly used (e.g. insurance costs against these sorts of unfavourable random events). As an alternative solution to either the control of a central authority or complete privatisation, the author suggested that individuals in a community should make preliminary agreements before using a shared resource through a self-made binding contract, which balances the share of benefits with the costs of enforcing them, while ensuring that the resource exploitation will not take place outside of the commonly agreed terms; being constructed by the users sharing a resource, the enforcement mechanisms and the conditions of the contract can be optimally shaped upon the needs of the community, with opportunities for changes if the users demand the agreements to be updated. Many were the empirical examples that the author gave to provide evidence on the existence of self-organising communities over the world: commonly utilised lands in the village Törbel of Switzerland; shared terrains in three villages of Japan; collective exploitation of irrigation systems in some cities of Spain. Her distinction between different kinds of individuals in a community may also be helpful when discussing the design of policies to resolve the issue; appropriators are those that withdraw resource units from a common-pool resource; providers arrange its provision; producers ensure the sustainability of the resource system in the long term. Among these individuals, she recognised that some of them may act opportunistically when having the chance or if the benefits largely exceed the costs, therefore delivering potential issues to consider when enquiring about the problem [38].

A theoretical model will overview the implications resulting from the development of regional patient migration over time, which are considered of significance importance since, as Malthus (1798) underlined, "a great emigration necessarily implies unhappiness of some kind or other in the country that is deserted" [27, 9]. In the discussion section, these concepts on the logic of collective action and the exploitation of shared resources will be applied to the entire aspects of the issue in manners that cohere with its nature, that is deemed to be rather unique, to provide rational and realistic policy suggestions.

3.1.2 Spatial econometric analysis

In addition to the considerations that connected the research topic with concepts from the tragedy of the commons, further perceptions that associate it with other theoretical concepts need to be recognised. Regional patient migration can be seen as a matter that inherently retains a geographical nature, bringing the feature of location into light as a very important aspect, since individual movement instances do not happen in isolation, but rather globally across the country, and involve potential for interactions between individuals as well as the presence of externalities that produce significant spatial spillover effects that could influence the occurrence of a related event in a certain area from another location. The nature of the matter seems to be aligned with a statement of Tobler (1970), known as his First Law of Geography, in which he declared that "everything is related to everything else, but near things are more related than distant things" [43, 236]. Therefore, considering the apparent importance of space and location for patient migration, spatial econometric analysis is deemed to be the most appropriate analysis method to explore the topic and to answer the related research questions, which will happen with the employment of various statistical models.

The presented literature review has demonstrated how previous studies have enriched the literature on the topic of patient migration among Italian regions under different lights, for various purposes and through the employment of a variety of inspection methodologies. The scope of this thesis is to contribute to the literature in a different manner, which will occur through the investigation of regional patient migration as the main matter of interest, the development of a unique theoretical model, an application of spatial econometric analysis methods that has not been found in the literature yet and the consideration of a provincial level rather than a regional one. Therefore, the methodological approach presented in this thesis is also regarded to be appropriate from the point of view of representing an innovative contribution to the literature, further sustaining the underlying motivations concerning its strict usage for the research scope. To be precise, the analysis of regional patient migration is looked upon from two opposite but also strictly intertwined aspects. The first one relates to regional patient immigration, which regards individuals that emigrated to the region of a certain province from the provinces in other regions of Italy to obtain planned health care treatments in public or accredited private facilities during a certain year. The second one regards regional patient emigration, which concerns individuals residing in a certain province of Italy that emigrated from their region to another to gather planned health care treatments in public or accredited private facilities during a certain year. Furthermore, these aspects are examined with additional distinction between ordinary admissions, which require overnight stays of patients, and day admissions, which involve short hospitalisations occurring during the day without the need for overnight stays, but with potential returning requirements on one or more following days if more assessments or interventions need to be made. In particular, the transformation of the data and their analysis through the statistical models will be conducted with the R programming language, using the open source RStudio front end. Furthermore, the GeoDa [5] programme will be employed as a secondary tool to support the analysis, to highlight potential procedural errors and to provide further information whenever necessary.

3.1.3 Spatial weights

A few notions should be introduced to understand the foundations of spatial regression, before delving into an overview of the various statistical models and the ideas behind the procedures of statistical model selection. In particular, the concepts of spatial weights, neighbours and weights matrix are outlined here, based upon a comprehensive overview provided by Anselin and Rey (2014) [4].

Spatial weights are w_{ij} components (for $i = 1, ..., n$ and $j = 1, ..., n$) that permit to create spatially explicit variables and are used for the calculation of various spatial statistics. Together with one another, they form a $n \cdot n$ spatial weights matrix W representing the neighbouring structure between all the observations, which is defined by the following matrix structure:

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \tag{3.1}$$

When the observations i and j are neighbours, $w_{ij} \neq 0$; when the observations i and j are not neighbours, $w_{ij} = 0$; when $i = j$, $w_{ij} = 0$, since an observation is not normally considered as a neighbour of itself. The spatial weights matrix expresses the existence of neighbouring relationships by representing spatial units in a row i, with their potential neighbours in a column j, for $i \neq j$. For each row and column combination, $w_{ij} = 1$ if the relationship exists and $w_{ij} = 0$ otherwise.

In order to ensure that proportional weights are created when the observations have an unequal number of neighbours, each non-zero w_{ij} is row-standardised through the division of its value by the row sum:

$$w_{ij(s)} = \frac{w_{ij}}{\sum_j w_{ij}} \tag{3.2}$$

Furthermore, resulting from the row-standardisation process, it can be observed that the sum of all the row-standardised non-zero weights w_{ij} in the entire matrix is equal to the total number of observations n:

$$S_0 = \sum_i \sum_j w_{ij} = n \tag{3.3}$$

A spatial weights matrix can be specified according to various types, which establish the neighbouring structure using diverse methods. For instance, a contiguity matrix defines two spatial units as neighbours if they share a common border of non-zero length, while a distance-based matrix defines two spatial units as neighbours if specific conditions are satisfied given a certain distance between points. Furthermore, different criteria specify the characteristics of the weights matrix of the chosen type; for example, for a contiguity matrix, the queen criterion considers a common edge or vertex, while the rook criterion only accounts for a common edge; instead, for a distance-based matrix, the k-nearest neighbour criterion assigns the same number of closest neighbours to all spatial units, while the inverse distance criterion is based upon a step function that provides neighbours with decreasing weights as distance increases towards a cut-off point, from which units are not considered to be neighbours anymore. Nonetheless, as Elhorst (2010) correctly underlined, the spatial weights matrix W cannot be estimated and needs to be specified in advance [12, 17], hence its specification should be based upon judgements considering the nature of the observations to be studied.

3.1.4 Statistical models

The methodological approach to spatial analysis involves the examination of data and testing of various hypotheses through the employment of different statistical models, whose results are evaluated with a process of model selection that suggests which model better fits the data. The features of the various non-spatial and spatial models taken into account for this research are outlined here.

Multiple linear regression model (MLR)

$$Y = \alpha\iota_n + \beta X + \epsilon \tag{3.4}$$

The multiple linear regression model defines the dependent variable as a linear relationship of explanatory variables and an error term. In the equation, Y is a $n \cdot 1$ vector of the dependent variable, ι_n is a $n \cdot 1$ vector of ones related to the constant parameter α, X is a $n \cdot k$ vector of the independent variables, β is a $k \cdot 1$ vector of their parameters and ϵ is a $n \cdot 1$ vector of the error term. The relationship of the dependent variable with each explanatory variable is often estimated with the ordinary least squares method and the validity of the estimations depend on the following fundamental assumptions:

1. **Linearity** – The dependent variable can be calculated as a linear function of a specific set of explanatory variables plus an error term, as its relationship with each explanatory variable is linear in parameters and the error term enters additively;

2. **Independence** – The observations are independent and identically distributed: $\{x_i, y_i\}_{i=1}^{N}$ *i.i.d.* (*independent and identically distributed*);

3. **Exogeneity**:

   (a) The error term is normally distributed conditionally upon the explanatory variables: $\epsilon_i | x_i \sim N(0, \sigma_i^2)$;

   (b) The error term is independent from the explanatory variables: $\epsilon_i \perp x_i$;

   (c) The mean of the error term is independent from the explanatory variables: $E(\epsilon_i | x_i) = 0$;

   (d) The error term and explanatory variables are uncorrelated: $Cov(\epsilon_i, x_i) = 0$;

4. **Homoscedasticity** – The error term has the same variance at each set of values of the explanatory variables: $Var(\epsilon_i | x_i) = \sigma^2$;

5. **Multicollinearity** – No explanatory variable is an exact linear combination of the others.

The OLS estimators $\hat{\beta}_j$, for $j = 1, ..., k$, are the best linear unbiased estimators (BLUE) for the true parameters β_j in the multiple linear regression model when these conditions are satisfied, otherwise the validity of the estimations can be questioned.

Spatial cross-regressive model (SLX)

$$Y = \alpha\iota_n + \beta X + \theta WX + \epsilon \tag{3.5}$$

The spatial cross-regressive model includes spatial effects of the explanatory variables, defined as the spatial average of neighbouring characteristics [25]. The equation includes the term WX, a $n \cdot k$ vector of spatially lagged predictors, and the related coefficient θ. When $\theta = 0$, spatial effects of the explanatory variables are absent and the model can be reduced to a linear regression model.

Spatial autoregressive model (SAR)

$$Y = \rho WY + \alpha\iota_n + \beta X + \epsilon \tag{3.6}$$

The spatial autoregressive model involves spatial effects of the dependent variable, hence it adds a spatial autoregressive structure to the linear regression model [25]. The equation includes the term WY, a $n \cdot 1$ vector of the spatially lagged dependent variable, and the related coefficient ρ. When $\rho = 0$, spatial effects of the dependent variable are absent and the model can be reduced to a linear regression model.

Spatial error model (SEM)

$$\begin{gathered} Y = \alpha\iota_n + \beta X + \epsilon, \\ \epsilon = \lambda W\epsilon + \mu \end{gathered} \tag{3.7}$$

The spatial error model involves spatial effects of the error term, referred to as disturbances of the model [25]. The equation includes the term $W\epsilon$, a $n \cdot 1$ vector of the spatially lagged error term, and the related coefficient λ. When $\lambda = 0$, spatial effects of the error term are absent and the model can be reduced to a linear regression model.

Spatial Durbin model (SDM)

$$Y = \rho WY + \alpha \iota_n + \beta X + \theta WX + \epsilon \tag{3.8}$$

The spatial Durbin model involves spatial effects of the dependent variable and the independent variables. The equation includes the terms WY and WX, with the related coefficients ρ and θ. When $\rho = 0$, spatial effects of the dependent variable are absent and the model can be reduced to a SLX model. When $\theta = 0$ for all predictors, spatial effects of the explanatory variables are absent and the model can be reduced to a SAR model. For this case, if $\theta = -\rho\beta$, then $\lambda = \rho$ and the model can also be reduced to a SEM.

Spatial Durbin error model (SDEM)

$$\begin{gathered} Y = \alpha \iota_n + \beta X + \theta WX + \epsilon, \\ \epsilon = \lambda W \epsilon + \mu \end{gathered} \tag{3.9}$$

The spatial Durbin error model involves spatial effects of the independent variables and the error term. The equation includes the terms WX and $W\epsilon$, with the related coefficients θ and λ. When $\theta = 0$ for each predictor, spatial effects of the independent variables are absent and the model can be reduced to a SEM. When $\lambda = 0$, spatial effects of the error term are absent and the model can be reduced to a SLX model.

Spatial autoregressive model with autoregressive disturbances (SARAR)

$$\begin{gathered} Y = \rho WY + \alpha \iota_n + \beta X + \epsilon, \\ \epsilon = \lambda W \epsilon + \mu \end{gathered} \tag{3.10}$$

The spatial autoregressive model with autoregressive disturbances, originally introduced by Kelejian and Prucha (1998) [23], involves spatial effects of the dependent variable and the error term. The equation includes the terms WY and $W\epsilon$, with the related coefficients ρ and λ. When $\rho = 0$, spatial effects of the dependent variable are absent and the model can be reduced to a SEM. When $\lambda = 0$, spatial effects of the error term are absent and the model can be reduced to a SAR model.

Manski model

$$Y = \rho WY + \alpha\iota_n + \beta X + \theta WX + \epsilon, \\ \epsilon = \lambda W\epsilon + \mu \tag{3.11}$$

The Manski model, introduced upon the work of Manski (1993), accounts for every possible spatial effect: endogenous interactions, when individual decisions are affected by those of the neighbours; exogenous interactions, when individual decisions are influenced by observable features of the neighbours; correlated effects of unobservable features [28]. The equation includes the terms WY, WX and $W\epsilon$, with the related coefficients ρ, θ and λ. Various researchers suggest to begin from a simpler model [12], whose choice can occur through certain methods of model selection, as this model is complete and the separate coefficients ρ, θ and λ cannot be really estimated at the same time.

3.1.5 Statistical model selection

Two methods of statistical model selection, based on the same assumptions of having a known neighbourhood matrix and exogenous independent variables, can delineate the choice of the appropriate model from which to gather the results. They similarly employ specification tests and statistical measures to give advice on forward stages of analysis.

The first method is a bottom-up approach, which consists in beginning from a non-spatial regression model and eventually testing for a spatial autoregressive process [15]. As outlined by some authors in the literature, such as Anselin (1988), the choice between keeping the results of a non-spatial model or moving towards a spatial model can be driven by some regular and robust Lagrange multiplier tests, respectively related to the spatial lag of either the dependent variable (LMlag and RLMlag) or the error term (LMerr and RLMerr) [2]. The SAR model should be considered when the LMLag test is significant, while the SEM should be looked upon when the LMErr test is significant. If both are significant, the robust tests are compared; if both of these are significant, the suggestion is to consider the model related to the most significant test results [4, 110].

The second method is top-down approach, supported by other authors in the literature such as LeSage and Pace (2009), which concerns beginning from a more comprehensive spatial model, choosing between the SDEM, SARAR and SDM by setting one of the possible constraints (i.e. respectively $\rho = 0$, $\theta = 0$ or $\lambda = 0$), depending upon a choice that accounts for the context under examination, and to eventually reduce it to simpler models based on a set of statistical measures and tests, such as the likelihood ratio test for the goodness of fit [25]. The SDM is often suggested as the first model of choice. For instance, LeSage and Pace (2009) considered it to be a good starting point for discussion of spatial regression model estimation [25, 46]. Elhorst (2010) also described some considerable strengths of the model, such as the production of unbiased coefficient estimates and their correct standard errors or t-values, if the true data-generation process is a spatial lag or a spatial error model, or the absence of prior restrictions on the magnitude of potential spatial spillover effects; moreover, the SDM can be reduced to all the models with a single spatial effect, including the SEM, even though it does not include a spatial coefficient for the error term itself, since if $\theta = -\rho\beta$ then $\lambda = \rho$. By contrast, he highlighted that taking the SARAR model as the starting one can lead to omitted variables bias if the true data-generating process is represented by a SDM or SDEM. In addition, Elhorst also suggested that ignoring spatial dependence in the dependent variable or the predictors has a high cost that can lead to biased and inconsistent coefficient estimators for the remaining independent variables, thus excluding the spatially autocorrelated error term is the best choice among all possibilities [12, 10, 14].

The approach to data analysis chosen for this research uses both methodologies to identify the most appropriate model from which the results should be gathered to explain patterns in the data; the first method is chosen as the primary, thus the analysis will begin from a non-spatial regression model. As Elhorst (2010) suggested, providing a case for the SDM, the results from the various likelihood ratio tests can be combined with those of the robust LM specification tests to be directed towards the most appropriate model with sufficient certainty, with regards to the spatially lagged effect of the dependent variable and the error term; when the results from these tests differ, with regards to these effects, the more comprehensive model should be taken as valid, since it generalises the effects of the nested spatial models and hence should provide better coefficient estimates [12]. The following diagram summarises the various statistical models and the reduction possibilities from a model into a nested one, starting from the most comprehensive spatial model:

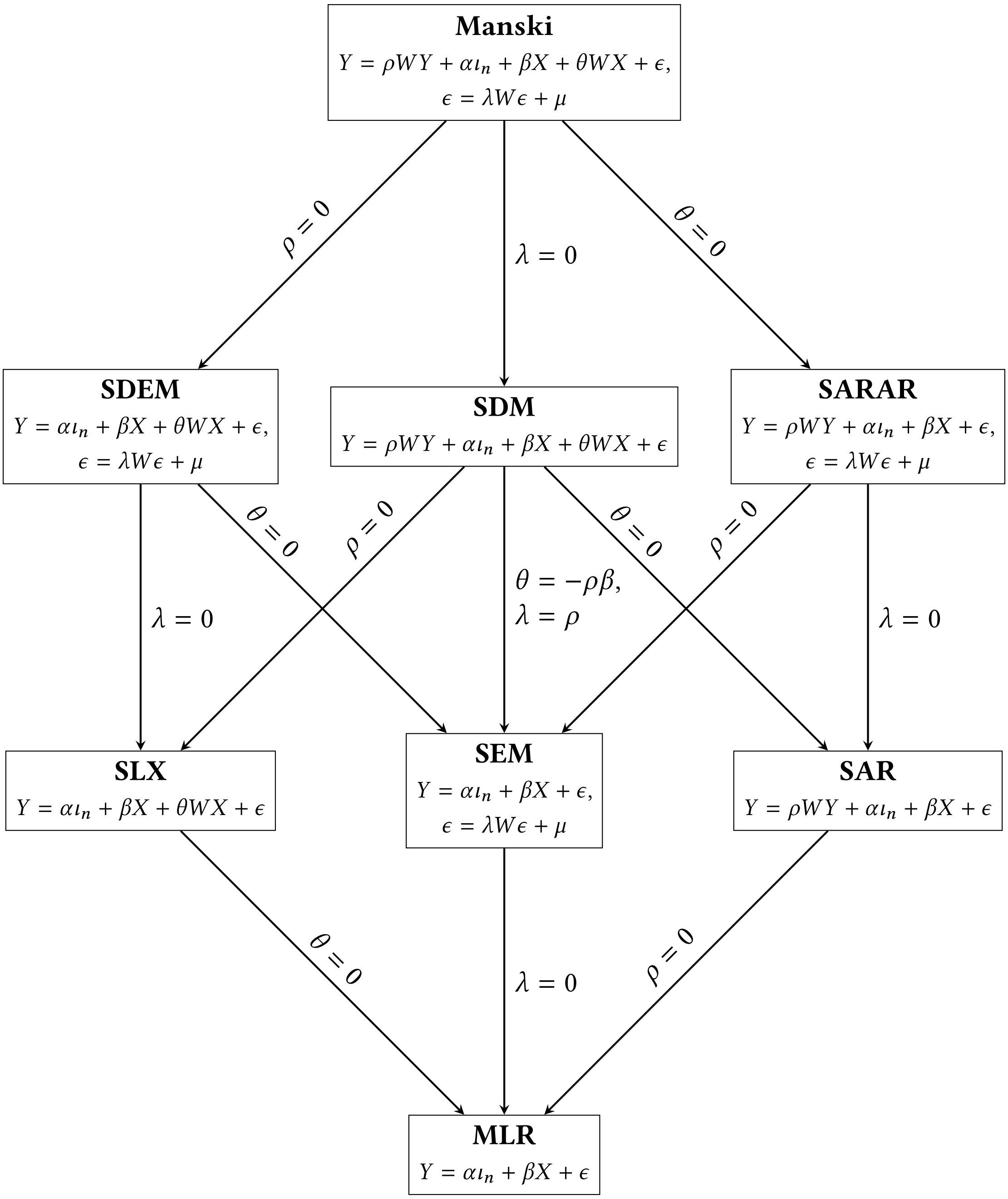


Figure 3.1: Nested structure of statistical models for spatial econometric analysis

3.1.6 Hypothesis testing

The statistical tests to execute throughout the spatial analysis rely upon testing a null hypothesis H_0, which has a specific formulation for each test and is accompanied by an alternative hypothesis H_1 representing the contrary. Every hypothesis test calculates its own resulting value and a related p-value statistic, defined as the probability of obtaining a result that is equal to or more extreme than what was actually observed in the data, which lets the researcher evaluate whether to reject or fail to reject the respective null hypothesis. Therefore, before running any test, it is fundamental for the research to establish a significance level α, which is defined as the probability of rejecting the null hypothesis when it is true, since it will be compared to each resulting p-value to decide upon the eventual rejection of the various null hypotheses; to be specific, a null hypothesis is rejected if the p-value is less than α, while it cannot be rejected if the p-value is greater than α. For this research thesis, a significance level $\alpha = 0{,}05$ is defined for comparisons with the p-values of the various statistical tests.

3.1.7 Statistical instruments

Various statistical instruments will be used for different purposes that relate to the data analysis, such as to control the validity of the assumptions that uphold the regression estimations. At first, when inspecting the collected data to prepare the definitive data set for the analysis, the Jarque-Bera test will be executed to determine whether the sample has the same skewness and kurtosis as the normal distribution under the null hypothesis of the residuals being normally distributed.

Another used test is the spatial Hausman test, as defined by LeSage and Pace (2009) upon the specifications of Hausman (1978), which tests the equality of the coefficient estimates produced by the linear model and SEM to investigate upon potentially omitted variables that correlate with variables in the SEM; failure to reject the null hypothesis of equality indicates that specification problems, such as omission of predictors, are absent from the SEM and, in this situation, if the SEM has significantly higher likelihood values compared to the linear model, then the spatial error term in the SEM captures the effects of omitted variables that are not correlated with those that the model includes [25, 62, 63] [19]. As LeSage and Pace (2006) underlined, a significant difference between the coefficient estimates of the linear model and SEM suggests how neither the former nor the latter produces correct estimates that counterpart the underlying parameters in the data generating process, for a given set of variables, warning against their usage [39].

A set of statistical measures will be used to test the possible existence of multicollinearity in the multiple regression equation. Multicollinearity is a statistical phenomenon in which two or more independent variables are highly correlated with one another; if present in a model, it should be taken into account and often removed before continuing with any kind of statistical analysis, since it can lead to unreliable and unstable estimates of regression coefficients, making it difficult to distinguish between the effect of a single predictor on the dependent variable when it is itself correlated with one or more other independent variables. Two types of indicators of multicollinearity will be employed to control for its existence in the linear regression: the variance inflation factor (VIF) and the collinearity condition indices. The VIF is an indicator that can be obtained for each predictor by executing a linear regression for a single predictor on all the others, gathering its related R^2 and calculating it with the following formula:

$$VIF = \frac{1}{(1 - R^2)} \tag{3.12}$$

The examined predictor is not linearly related to the others when the indicator is equal to 1, its lower bound, and thus $R^2 = 0$. Suggestion of intercorrelation with other variables comes when the indicator is higher than 1. Since no upper bound exists, several recommendations among the literature, such as those of Kutner et al. (2005), suggest a maximum cutoff of 10, with the largest VIF among all predictors being utilised as an indicator of the overall severity of multicollinearity [24].

The condition indices are calculated by the eigenvalues of the crossproduct matrix of the scaled but uncentered explanatory variables. Recommendations from the literature, such as those of Belsley (1991), suggest that the regression may be affected by severe multicollinearity if the condition number is above 30 [7], which is generally deemed to be a cutoff value, although weak dependencies might be starting to affect the regression estimates when the indicator is around 10.

Another series of statistical measures related to spatial analysis will be used to examine the eventual presence of global and local spatial autocorrelation: the Moran's I, the Geary's C, the Local Moran's I and the Local Geary's C. The Moran's I is a statistic that measures global spatial autocorrelation, ideated by Moran (1948, 1950) and later on popularised by the contributions of Cliff and Ord (1973) [32] [33] [10]. It is calculated as a cross-product statistic between a variable and its spatial lag, with the variable expressed in deviations from its mean, using the following equation:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (\text{for } j \neq i) \tag{3.13}$$

The Geary's C, developed by Geary (1952), is another used indicator of global spatial autocorrelation that focuses on the squared differences between pairs of data values [16]. It is calculated with the following equation:

$$C = \frac{n-1}{2S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (\text{for } j \neq i) \tag{3.14}$$

In both the equations, w_{ij} are the elements of the spatial weights matrix, $S_0 = \sum_i \sum_j w_{ij}$ is the sum of all the weights and n is the number of observations. The Moran and Geary tests assess whether the null hypothesis of absence of spatial autocorrelation can be rejected, but they do so in different manners; in fact, as can be seen from the differences between the two equations, the two indicators are inversely related to each other. The Moran's I has an expected value of $E(I) = \frac{-1}{n-1}$, while the Geary's C has an expected value of $E(C) = 1$. Positive spatial autocorrelation, which relates to a pattern of similar values at neighbouring locations, exists if the observed value of I is significantly greater than its expected value and if the observed value of C is significantly lower than 1. Negative spatial autocorrelation, which relates to a pattern of dissimilar values at neighbouring locations, exists if the observed value of I is significantly lower than its expected value and if the observed value of C is significantly greater than 1. In addition to the Moran's I and the Geary's C, which act as global indicators of spatial autocorrelation, two related Local Indicators of Spatial Autocorrelation (LISA) will help to identify local clusters of neighbours with similar values and local spatial outliers surrounded by neighbours with different values, as proposed by Anselin (1995): the Local Moran and the Local Geary statistics [3]. For scopes of simplification, only the Moran's I and the Local Moran's I will be shown in the data analysis, as the alternative statistics convey the same information.

Finally, various statistical indicators will be employed to compare statistical models and to select the most relevant for the data under examination. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) will be taken into account as relative measures for statistical model selection, as the model with the lowest AIC and BIC has the best goodness of fit for the data among all; since adding more parameters can improve the goodness of fit, the measures include a penalty, defined as an increasing function of the number of estimated parameters, which discourages the practice of overfitting. Likelihood ratio tests will instead be utilised for nested statistical models, to evaluate whether restricting a more comprehensive model to a simpler one can improve the goodness of fit and thus to select the appropriate model between the two. As Elhorst (2010) illustrated, a reduced model is nested into a more complex one if it contains at least one of the spatially lagged variables and coefficients [12, 13]. The R^2 and pseudo-R^2 will also be considered as comparative measures, with higher values indicating greater goodness of fit of a model for the analysed data compared to that of the others.

3.2 Analysis framework

3.2.1 Theoretical model

The theoretical model is constructed to fulfil various purposes, such as establishing and developing a new concept on how the phenomenon of regional patient migration occurs and progresses over time, discussing the predictions that can be induced and connecting the theory to the forthcoming empirical analysis of selected data; to be precise, the model examines the role of provinces in an aggregate manner to determine the implications for the entire regions, while the empirical analysis will consider each province separately, for motivations that will be explained later on. For the outlined purposes, the theoretical model begins from a situation of equilibrium and equality between regions, for scopes of simplification of the idea, while its advancement will account for the short, medium and long terms to identify the potential effects that may result over time. Its constitution and development are based upon the following assumptions:

1. **Assumption 1**: "At the beginning, certain characteristics are the same for every region: local population size, financial resources (as a sum of state contributions, local taxes, revenues, variable costs and fixed costs) and overall capacity of the regional health care supply" – this assumption permits to start from a situation of equilibrium, in which every region is equal to the others without differences and the phenomenon does not exist;

2. **Assumption 2**: "When conditions between regions are equal, patients prefer to gather health treatments in their region of residence, since its the closest in terms of distance and allows them to reduce their individual costs and maximise their personal benefits (i.e. for $PT_{R_1 n} \rightarrow R_1 > R_2 \sim R_3$; for $PT_{R_2 n} \rightarrow R_2 > R_1 \sim R_3$; for $PT_{R_3 n} \rightarrow R_3 > R_1 \sim R_2$)" – this assumption permits to discard patient preferences as the factors that start the phenomenon and therefore to retain the initial equilibrium between regions;

3. **Assumption 3**: "Each region is a rational actor which seeks to maximise its own utility, in terms of patient retention and attraction, by performing investments to enhance the quality of health care supply whenever it obtains further revenues" – this assumption permits to consider the potential occurrence of changes between regions for the demonstration of various possible scenarios;

4. **Assumption 4**: "Each patient is a rational actor who seeks to maximise his or her own utility in terms of gathering the highest quality of treatment while enduring the lowest individual costs" – this assumption supports the presence of an initial equilibrium between regions and permits to consider the potential occurrence of changes of patient preferences for the illustration of various possible scenarios;

5. **Assumption 5**: "Each region can cover the treatment costs of a patient and gain monetary or immaterial profits without losses, given the health care supply is not constrained at a given point in time" – this assumption supports the presence of an initial equilibrium between regions and an equal development of regional health care systems if conditions are held the same; the equilibrium outcome is certain as each region treats its internal patients and has perfect information on treatment costs, hence it can align the resources with the expected costs in advance;

6. **Assumption 6**: "Each region, when providing another region with repayments for the costs of treating one of its patients, incurs into cost reimbursements that may be equal or different from the expected treatment costs it would have endured had the patient been treated in his or her region of residence" – this assumption permits to consider the potential occurrence of changes between regions for the portrayal of various possible scenarios in absence of equilibrium; when a resident of a region is treated outside, imperfect information on treatment costs does not permit the region to align the resources with the expected costs in advance and the resultant outcome will depend on the monetary amount that the receiving region requests as a cost reimbursement, which is not known by the region of origin;

7. **Assumption 7**: "Regions make decisions that are independent from those of the others" – this assumption eliminates other circumstances that may occur between regions to isolate the effects of specific events, considered during the theoretical advancement of the model, that influence the establishment and development of the phenomenon over time;

8. **Assumption 8**: "Patients make decisions that are independent from those of the others" – this assumption eliminates other circumstances that may occur between patients to isolate the effects of specific events, considered during the theoretical advancement of the model, that influence the establishment and development of the phenomenon over time.

The theoretical model considers three example regions (R_1, R_2 and R_3), in a timeline with time periods t_k, each representing one year. Initially, the following characteristics are assumed to be equal for every region:

- POP_{R_i}: population of region i (for $i = 1, 2, 3$), which composes the total population TP when all local populations are summed together;

- FR_{R_i}: financial resources of region i (for $i = 1, 2, 3$), which are employed towards the sustainment and development of its regional health care system. These financial resources result from the sum of five monetary components:

  - SR_{R_i}: share of financial resources for the health care system that the state gives to region i (for $i = 1, 2, 3$) based upon the local population size. For simplification, they are defined as $SR_{R_i} = \frac{1 \cdot POP_{R_i}}{TP}$ (for $i = 1, 2, 3$);

  - TX_{R_i}: taxes for the health care system of region i (for $i = 1, 2, 3$), which are collected from the local population and therefore depend upon its size. For simplification, they are defined as $TX_{R_i} = 1 \cdot POP_{R_i}$ (for $i = 1, 2, 3$);

  - TR_{R_i}: revenues from treating patients of any region in region i (for $i = 1, 2, 3$). For simplification, they are defined as $TR_{R_i} = \sum_{in} \gamma_{PT_{R_i n}}$ (for $i = 1, 2, 3$);

- TC_{R_i}: costs for treating patients of any region in region i (for $i = 1, 2, 3$). For simplification, they are defined as $TC_{R_i} = \sum_{in} \kappa_{PT_{R_i n}}$ (for $i = 1, 2, 3$);

- FC_{R_i}: fixed costs for the health care system in region i (for $i = 1, 2, 3$). For simplification, they are defined with a fixed value and do not change over the time periods;

• H_{R_i}: capacity of the health care supply of region i (for $i = 1, 2, 3$) in terms of quantity and quality. For simplification, it is defined as a natural logarithmic function of financial resources, $H_{R_i} = ln(FR_{R_i})$, where their increases induce improvements in the capacity towards a theoretical maximum potential, while their decreases cause disinvestments towards eventual failures of the system. Since the capacity is proportioned to the local population size, a per-capita share of health care supply is defined as $PCH_{R_i} = \frac{H_{R_i}}{POP_{R_i}}$.

Considering that H_{R_i} depends upon the components of FR_{R_i}, which in turn depend upon the local population size POP_{R_i} and the number of treated patients $PT_{R_i n}$, equality between regions in terms of population size and absence of treatments at the beginning signify that the first two elements are equal everywhere. The following figures show the conditions of the three regions in a situation of equilibrium at time t_1, accounting for the previous assumptions and presuming that $POP_{R_i} = 100$:

$\mathbf{R_1}$	$\mathbf{R_2}$	$\mathbf{R_3}$
$POP_{R_1} = 100$	$POP_{R_2} = 100$	$POP_{R_3} = 100$
$PT_{R_1} = 0$	$PT_{R_2} = 0$	$PT_{R_3} = 0$
$SR_{R_1} = 100$	$SR_{R_2} = 100$	$SR_{R_3} = 100$
$TX_{R_1} = 100$	$TX_{R_2} = 100$	$TX_{R_3} = 100$
$TR_{R_1} = 0$	$TR_{R_2} = 0$	$TR_{R_3} = 0$
$TC_{R_1} = 0$	$TC_{R_2} = 0$	$TC_{R_3} = 0$
$FC_{R_1} = 50$	$FC_{R_2} = 50$	$FC_{R_3} = 50$
$H_{R_1} = 5{,}01$	$H_{R_2} = 5{,}01$	$H_{R_3} = 5{,}01$
$PCH_{R_1} = 0{,}0501$	$PCH_{R_2} = 0{,}0501$	$PCH_{R_3} = 0{,}0501$

Figure 3.2: Theoretical framework at time t_1

In the following time period t_2, it is assumed that one patient in each region needs to obtain a certain health treatment. According to previously defined preferences, even though the conditions among regional health care systems are equal, each patient should prefer to be treated in his or her own region to maximise the benefits and minimise the costs related to time and distance. However, a number of random occurrences that can relate to either the supply side or the demand side may lead to the replacement of these preferences with the selection of an alternative solution that involves being treated in another region; for instance, a temporary reduction of capacity H_{R_i} in the supply side of one region due to transitory conditions (e.g. closure of a department for renovations) could make another region more capable of offering a certain treatment at a point in time, while a person in need may momentarily reside in another region due to certain reasons (e.g. work transfer, holidays). As a consequence of the occurrences of random events of these types, it is assumed that patient $PT_{R_1 1}$ decides to seek for health treatments in region R_2 and patient $PT_{R_2 1}$ chooses to obtain them in region R_3; on the contrary, because the situation in region R_3 is deemed to be normal, patient $PT_{R_3 1}$ decides to gather them in his or her own region of residence without moving to other locations. This theoretical context permits to view the outcomes for three potential cases: a region with only patient emigration (R_1); a region with only patient immigration (R_3); a region with patient emigration and immigration (R_2). The following figures illustrates the results of these movements at time t_2:

$\mathbf{R_1}$		$\mathbf{R_2}$		$\mathbf{R_3}$
$POP_{R_1} = 100$		$POP_{R_2} = 100$		$POP_{R_3} = 100$
$PT_{R_1} = 0$		$PT_{R_2} = 2 - 1 = 1$		$PT_{R_3} = 1 + 1 = 2$
$SR_{R_1} = 100$		$SR_{R_2} = 100$		$SR_{R_3} = 100$
$TX_{R_1} = 100$	Emigration $\rightarrow$	$TX_{R_2} = 100$	Emigration $\rightarrow$	$TX_{R_3} = 100$
$TR_{R_1} = 0$		$TR_{R_2} = \gamma_{PT_{R_1 1}}$		$TR_{R_3} = \gamma_{PT_{R_3 1}} + \gamma_{PT_{R_2 1}}$
$TC_{R_1} = 0$		$TC_{R_2} = \kappa_{PT_{R_1 1}}$		$TC_{R_3} = \kappa_{PT_{R_3 1}} + \kappa_{PT_{R_2 1}}$
$FC_{R_1} = 50$		$FC_{R_2} = 50$		$FC_{R_3} = 50$
$H_{R_1} = 5{,}01$		$H_{R_2} = 4{,}9599$		$H_{R_3} = 4{,}9098$
$PCH_{R_1} = 0{,}0506$		$PCH_{R_2} = 0{,}0501$		$PCH_{R_3} = 0{,}04959$

Figure 3.3: Theoretical framework at time t_2

According to one of the previous theoretical assumptions, each region has perfect information concerning the residents of its local population, hence it is able to acquire either monetary or immaterial profits from treating an internal patient, while covering the related costs without issues, using the available financial resources. Therefore, in the theoretical model, it can be stated that the treatment revenues $\gamma_{PT_{R_1 1}}$ are always higher than the related costs $\kappa_{PT_{R_1 1}}$ for an example patient $PT_{R_1 1}$. As a result of the assumption and the differential circumstances created by the movements of patients at time t_2, each region confronts a particular situation that involves different outcomes, as described in the following overview:

- Region R_1 – 1 patient emigrates to region R_2, leading to a total of 0 patients that are treated in the region. Revenues are not obtained and costs are not sustained. The total available health care supply does not alter and the per-capita available health care supply increases for the other 99 individuals in the local population thanks to the emigration of 1 patient to region R_2. The potential of the health care supply increases but remains untouched, leading to the existence of underused resources that do not deliver any sort of benefit and a less efficient sustainment of fixed costs during the time period, without revenues to invest for improvements in the total capacity of the health care supply;

- Region R_2 – 1 patient emigrates to region R_3 and 1 patient immigrates from region R_1, leading to a total of 1 patient that is treated in the region. Revenues are obtained and costs are covered for 1 incoming patient. The total available health care supply reduces by the amount used by 1 patient and the per-capita available health care supply stays the same for the other 99 individuals in the local population because of the immigration of 1 patient from region R_1 and the emigration of 1 patient from the local population. The potential of the health care supply remains the same and is used as normal, leading to a regular sustainment of fixed costs during the time period, with revenues from the treatment of 1 patient to invest for improvements in the total capacity of the health care supply;

- Region R_3 – 1 patient immigrates from region R_2 and 1 patient does not relocate, leading to a total of 2 patients that are treated in the region. Revenues are obtained and costs are covered for 1 incoming patient and 1 local patient. The total available health care supply reduces by the amount used by 2 patients and the per-capita available health care supply diminishes for the other 99 individuals in the local population due to the immigration of 1 patient from region R_2 and the presence of 1 patient from the local population. The potential of the health care supply remains the same and is used more than normal, leading to a more efficient sustainment of fixed costs during the time period, with revenues from the treatment of 2 patients to invest for improvements in the total capacity of the health care supply.

The outcomes that occurred at time t_2, as a consequence of individual decisions that replaced the default treatment preferences, caused the movements of some revenues and costs from region R_1 to region R_3, higher efficiency in the usage of resources through economies of scale from region R_1 to region R_3 and a higher capacity of the health care supply from region R_3 to region R_1. In the short term, the following results take place:

- Region R_1 suffers from lower efficiency and absence of potential for advancement in the health care supply, but acquires higher temporary capacity to treat incoming patients; an increase in the rates of emigration would widen these outcomes;
- Region R_2 keeps the same characteristics on average; however, differences between the rates of immigration and emigration can skew the results towards a situation that could be similar to that of either region R_1 or region R_3;
- Region R_3 acquires higher efficiency and potential for improvements in the health care supply, but suffers from lower temporary capacity to treat incoming patients; an increase in the rates of immigration would widen these outcomes; if maximum capacity was reached at a specific point in time, local patients would have to seek for health treatments in another region, causing the situation to be similar to that of region R_2, or to delay their treatment needs to the future.

At time t_3, regions give monetary reimbursements to other regions depending upon the origin and destination of emigrating patients. In this case, region R_1 gives region R_2 the costs it covered for its patient $PT_{R_1 1}$ and region R_2 transfers to region R_3 the costs it covered for its patient $PT_{R_2 1}$. Therefore, the following situation develops:

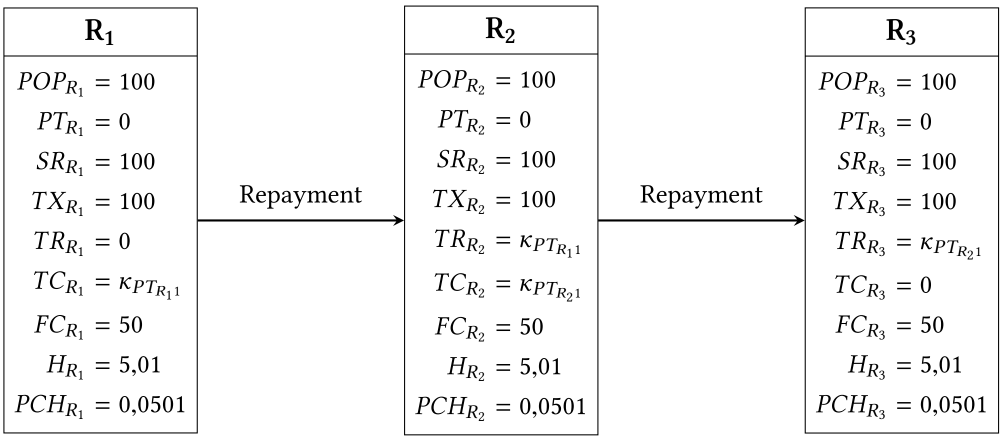


Figure 3.4: Theoretical framework at time t_3

As can be seen, the situation returns to the equilibrium present at time t_1, but with additional monetary reimbursements that will influence its state in the forthcoming time periods. In fact, in the condition of equilibrium, it was assumed that each region treating its own local patients can obtain at least a positive profit from a health treatment, with revenues being able to completely cover the costs. However, in the new situation, when considering the revenues obtained at time t_2 and the reimbursement transfers at time t_3, the outcomes for each region depend on the differentials between what was obtained from treating immigrating patients and the amount to be paid for emigrating ones, which may not be aligned with the same assumption. Hence, the following results occur:

- Region R_1 has to deal with costs for 1 emigrating patient that are not balanced by any sort of profit. The following scenarios can occur:

  1. Profit $= TR_{R_1} - TC_{R_1} = 0 - \kappa_{PT_{R_1 1}} = - \longrightarrow$ Disinvestments from the health care supply are made;

  2. Profit $= TR_{R_1} - TC_{R_1} = 0 - \kappa_{PT_{R_1 1}} = - \longrightarrow$ Taxes are increased;

- Region R_2 has to deal with costs for 1 emigrating patient which are balanced by profits obtained from the treatment of 1 immigrating patient. The following scenarios can occur:

  1. Profit $= TR_{R_2} - TC_{R_2} = (\gamma_{PT_{R_1}1} + \kappa_{PT_{R_1}1}) - (\kappa_{PT_{R_1}1} + \kappa_{PT_{R_2}1}) = \gamma_{PT_{R_1}1} - \kappa_{PT_{R_2}1} = -$ $\longrightarrow$ Disinvestments from the health care supply are made;
  2. Profit $= TR_{R_2} - TC_{R_2} = (\gamma_{PT_{R_1}1} + \kappa_{PT_{R_1}1}) - (\kappa_{PT_{R_1}1} + \kappa_{PT_{R_2}1}) = \gamma_{PT_{R_1}1} - \kappa_{PT_{R_2}1} = -$ $\longrightarrow$ Taxes are increased;
  3. Profit $= TR_{R_2} - TC_{R_2} = (\gamma_{PT_{R_1}1} + \kappa_{PT_{R_1}1}) - (\kappa_{PT_{R_1}1} + \kappa_{PT_{R_2}1}) = \gamma_{PT_{R_1}1} - \kappa_{PT_{R_2}1} = +$ $\longrightarrow$ Investments in the health care supply are made;
  4. Profit $= TR_{R_2} - TC_{R_2} = (\gamma_{PT_{R_1}1} + \kappa_{PT_{R_1}1}) - (\kappa_{PT_{R_1}1} + \kappa_{PT_{R_2}1}) = \gamma_{PT_{R_1}1} - \kappa_{PT_{R_2}1} = +$ $\longrightarrow$ Taxes are decreased;

- Region R_3 has to deal with costs for 1 local patient which are balanced by profits obtained from the treatments of 1 immigrating patient and 1 local patient. The following scenarios can occur:

  1. Profit $= TR_{R_3} - TC_{R_3} = (\gamma_{PT_{R_3}1} + \gamma_{PT_{R_2}1} + \kappa_{PT_{R_2}1}) - (\kappa_{PT_{R_3}1} + \kappa_{PT_{R_2}1}) = (\gamma_{PT_{R_3}1} + \gamma_{PT_{R_2}1}) - \kappa_{PT_{R_3}1} = + \longrightarrow$ Investments in the health care supply are made;
  2. Profit $= TR_{R_3} - TC_{R_3} = (\gamma_{PT_{R_3}1} + \gamma_{PT_{R_2}1} + \kappa_{PT_{R_2}1}) - (\kappa_{PT_{R_3}1} + \kappa_{PT_{R_2}1}) = (\gamma_{PT_{R_3}1} + \gamma_{PT_{R_2}1}) - \kappa_{PT_{R_3}1} = + \longrightarrow$ Taxes are decreased.

At time t_4, each region falls into one scenario depending upon the decisions it made according to these profits or losses, which influence its conditions and the equilibrium. To illustrate the development of each scenario, the following assumptions are made:

- Investments and disinvestments towards the health care supply derive from 10% increases or reductions in financial resources that equal the profits or losses;
- Tax increases and decreases consist of 10% increments or reductions of total taxes that correspond to the profits or losses.

The following figures show the conditions for regions R_1, R_2 and R_3 in each potential scenario at time t_4, which permit to discuss the related consequences in the long term:

R_1 – Scenario 1	**R_1 – Scenario 2**
$POP_{R_1} = 100$	$POP_{R_1} = 100$
$PT_{R_1} = 0$	$PT_{R_1} = 0$
$SR_{R_1} = 100 \cdot 0{,}9$	$SR_{R_1} = 100$
$TX_{R_1} = 100 \cdot 0{,}9$	$TX_{R_1} = 100 \cdot 1{,}1$
$TR_{R_1} = 0$	$TR_{R_1} = 0$
$TC_{R_1} = 0$	$TC_{R_1} = 0$
$FC_{R_1} = 50$	$FC_{R_1} = 50$
$H_{R_1} = 4{,}87$	$H_{R_1} = 5{,}01$
$PCH_{R_1} = 0{,}0487$	$PCH_{R_1} = 0{,}0501$

(a) Scenarios for region R_1 at time t_4

R_2 – Scenario 1	**R_2 – Scenario 2**	**R_2 – Scenario 3**	**R_2 – Scenario 4**
$POP_{R_2} = 100$	$POP_{R_2} = 100$	$POP_{R_2} = 100$	$POP_{R_2} = 100$
$PT_{R_2} = 0$	$PT_{R_2} = 0$	$PT_{R_2} = 0$	$PT_{R_2} = 0$
$SR_{R_2} = 100 \cdot 0{,}9$	$SR_{R_2} = 100$	$SR_{R_2} = 100 \cdot 1{,}1$	$SR_{R_2} = 100$
$TX_{R_2} = 100 \cdot 0{,}9$	$TX_{R_2} = 100 \cdot 1{,}1$	$TX_{R_2} = 100 \cdot 1{,}1$	$TX_{R_2} = 100 \cdot 0{,}9$
$TR_{R_2} = 0$	$TR_{R_2} = 0$	$TR_{R_2} = 0$	$TR_{R_2} = 0$
$TC_{R_2} = 0$	$TC_{R_2} = 0$	$TC_{R_2} = 0$	$TC_{R_2} = 0$
$FC_{R_2} = 50$	$FC_{R_2} = 50$	$FC_{R_2} = 50$	$FC_{R_2} = 50$
$H_{R_2} = 4{,}87$	$H_{R_2} = 5{,}01$	$H_{R_2} = 5{,}14$	$H_{R_2} = 5{,}01$
$PCH_{R_2} = 0{,}0487$	$PCH_{R_2} = 0{,}0501$	$PCH_{R_2} = 0{,}0514$	$PCH_{R_2} = 0{,}0501$

(b) Scenarios for region R_2 at time t_4

R_3 – Scenario 1	**R_3 – Scenario 2**
$POP_{R_3} = 100$	$POP_{R_3} = 100$
$PT_{R_3} = 0$	$PT_{R_3} = 0$
$SR_{R_3} = 100 \cdot 1{,}1$	$SR_{R_3} = 100$
$TX_{R_3} = 100 \cdot 1{,}1$	$TX_{R_3} = 100 \cdot 0{,}9$
$TR_{R_3} = 0$	$TR_{R_3} = 0$
$TC_{R_3} = 0$	$TC_{R_3} = 0$
$FC_{R_3} = 50$	$FC_{R_3} = 50$
$H_{R_3} = 5{,}14$	$H_{R_3} = 5{,}01$
$PCH_{R_3} = 0{,}0514$	$PCH_{R_3} = 0{,}0501$

(c) Scenarios for region R_3 at time t_4

Figure 3.5: Theoretical framework at time t_4

As can be seen, every region falls into scenarios that can be either similar or different from those of the others depending upon the events that occurred. For region R_1, each scenario can develop in the following ways:

1. Scenario 1 – Disinvestments diminish the total and per-capita capacity of supply. At time t_5, due to changes of individual preferences resulting from the differences in capacity with other regions, locals will more likely seek for health treatments elsewhere and external patients will less likely consider the region to get them;

2. Scenario 2 – Taxes increase to retain the total and per-capita capacity of supply at the initial levels. At time t_5, the situation can remain stable or degrade if even one resident moves to another region, as there will be a higher pressure on the others or a lower amount of taxes, depending on holding the same total taxes or pressure, which can lead to supply disinvestments or other tax increases at time t_6.

For region R_2, each scenario can develop in the following ways:

1. Scenario 1 – Disinvestments diminish the total and per-capita capacity of supply. At time t_5, due to changes of individual preferences resulting from the differences in capacity with other regions, locals will more likely seek for health treatments elsewhere and external patients will less likely consider the region to get them;

2. Scenario 2 – Taxes increase to retain the total and per-capita capacity of supply at the initial levels. At time t_5, the situation can remain stable or degrade if even one resident moves to another region, as there will be a higher pressure on the others or a lower amount of taxes, depending on holding the same total taxes or pressure, which can lead to supply disinvestments or other tax increases at time t_6;

3. Scenario 3 – Investments increase the total and per-capita capacity of supply. At time t_5, due to changes of individual preferences resulting from the differences in capacity with other regions, locals will less likely seek for health treatments elsewhere and external patients will more likely consider the region to get them;

4. Scenario 4 – Taxes decrease as profits are used to retain the total and per-capita capacity of supply at the initial levels. At time t_5, the situation can remain stable or improve if even one resident moves from another region, as there will be a lower pressure on the others or a higher amount of taxes, depending on holding the same total taxes or pressure, which can lead to supply investments or other tax decreases at time t_6.

For region R_3, each scenario can develop in the following ways:

1. Scenario 1 – Investments increase the total and per-capita capacity of supply. At time t_5, due to changes of individual preferences resulting from the differences in capacity with other regions, locals will less likely seek for health treatments elsewhere and external patients will more likely consider the region to get them;

2. Scenario 2 – Taxes decrease as the profits are used to retain the total and per-capita capacity of supply at the initial levels. At time t_5, the situation can remain stable or improve if even one resident moves from another region, as there will be a lower pressure on the others or a higher amount of taxes, depending on holding the same total taxes or pressure, which can lead to supply investments or other tax decreases at time t_6.

The initial outline of the theoretical model depicted a situation of equilibrium among regions and local populations that allows for the maximisation of the utility for every involved actor. However, as can be comprehended from the consecutive advancement of the theoretical model, the occurrence of random events in unplanned manners could temporarily influence individual preferences in the immediate term, so that a patient can continue to seek for health treatments in the most optimal way to maximise his or her own benefits, when free patient choice of treatment is available; nonetheless, the downside of this opportunity consists in the production of different outcomes among entities and people that remain over time due to cascade effects happening in a vicious cycle and therefore fracture a situation of equilibrium in the long term.

Considering the three example regions, the following statements can be made when deducting potential effects from the theory:

- Region R_1 will continue to encounter negative consequences, in absence of special policies (e.g. additional monetary assistance from the central government);
- Region R_2 will endure a situation that may either remain balanced or sway towards positive or negative ends;
- Region R_3 will be sustained with positive results, conditionally upon not reaching the point of maximum capacity, which however increases over time as a result of continuous improvements.

The formulation of the initial assumptions has been fundamental for the illustration of various potential scenarios and the deduction of theoretical outcomes. To be precise, assumptions 1 and 2 permitted to retain an initial situation of equilibrium when holding the same conditions, assumptions 3 and 4 conceded changes to happen as the conditions varied, assumptions 7 and 8 allowed for isolating the effects of specific events from those of other circumstances occurring among regions or patients, while assumptions 5 and 6 have been important to induce the theoretical development of the model.

In particular, assumption 5 stated that regions always profit from treating a patient, when free resources are available, because revenues are assumed to be higher than costs, a notion that questions the common assumptions about costs of providing health care services being higher than revenues; excluding fixed costs that can be efficiently covered through economies of scale, the reason is that variable treatment costs can be adequately offset not only with monetary revenues (e.g. copayments of patients), but also through intangible benefits which advance beyond visible short-term results and create positive circumstances that enhance the features of a regional health care system (e.g. training of personnel, investment appeal for providers, increased competition between public and accredited private providers, attraction of patients); first of all, as noted by Nyman

(2007) [34, 781] as well as Sieberg and Shevtsova (2012) [41, 136], the state involvement for a publicly supervised health care competition can reduce the high costs of a market-driven system that had been previously highlighted by various authors, such as Hsiao (1995) [21, 139] and Anderson et al. (2003) [1, 97-98]; secondly, the significance of the underlined immaterial benefits should be recognised more frequently, because they can produce valuable development opportunities for a health care system when coupled with monetary revenues, as proved by steady improvements made by regional health systems with high patient immigration rates over time; removing this assumption would render the advancement, probable in the case of region R_2 and certain in that of region R_3, more uncertain for receiving regions or even detrimental in absence of cost reimbursements during the short term, if an incoming patient causes the sustainment of excessively high costs without benefits. In addition, assumption 6 considered the cost repayments to be unknown for regions with escaping patients and thus has permitted to retain constant uncertainties in the outcomes, especially in the case of region R_2, which may then appear as irrelevant or conduce to severe issues; removing these uncertainties through perfect information would permit a region to align the repayments with the expected costs in advance, minimising losses in case of escaping patients treated somewhere else.

To conclude, the main outcome of these theoretical reflections can be the recognition of how the issue can originate from unplanned events and, if not consciously regulated, hold on or widen over time through influences on individual preferences when patients are provided with free choice of treatment. Furthermore, considering the case of region R_2 as the most probable and the other two as distant theoretical extremes can be a realistic conclusion, given the existence of many regions and an extended population in the country that may induce the occurrence of both aspects in the same region. In addition to this contribution, an empirical analysis will be conducted to identify the aspects that can be changed to counteract the presented effects towards a more balanced condition, while a conclusive discussion will present thoughts that combine its results with the theory and further information from other portions of the literature.

3.2.2 Spatial weights matrix

Before conducting spatial analysis on the data, a spatial weights matrix has to be created to represent the neighbouring structure between all the provinces of Italy and to calculate the weight of influence that a province has on another, when it is part of a group of neighbours, which varies depending on the size of the specific group of provinces and is defined in the spatial weights matrix. A first-order queen contiguity matrix has been chosen and created for this purpose, because:

- Understanding how it functions is fairly straightforward (i.e. two provinces are neighbours when they share a non-zero border);
- It is based on an objective definition compared to other types of weights matrices (e.g. distance-based matrix);
- It is recommended to use in most cases, in order to deal with potential inaccuracies in the polygon files (e.g. rounding errors) that define the spatial units;
- It gives at least one neighbour to all provinces, allowing to refrain from the issue of provinces without neighbours ("isolates"), which may generate problems when calculating spatially lagged variables and measures of local spatial autocorrelation;
- It gives a fairly balanced distribution of neighbours for the provinces, ranging from a minimum of 1 to a maximum of 9;
- The quality of the spatial data is appropriate for constructing a functioning contiguity weights matrix.

Certain instruments visually represent various characteristics of the created weights matrix when implemented for this context. For instance, the following histogram illustrates the distribution of neighbouring relationships between Italian provinces with the specified weights matrix:

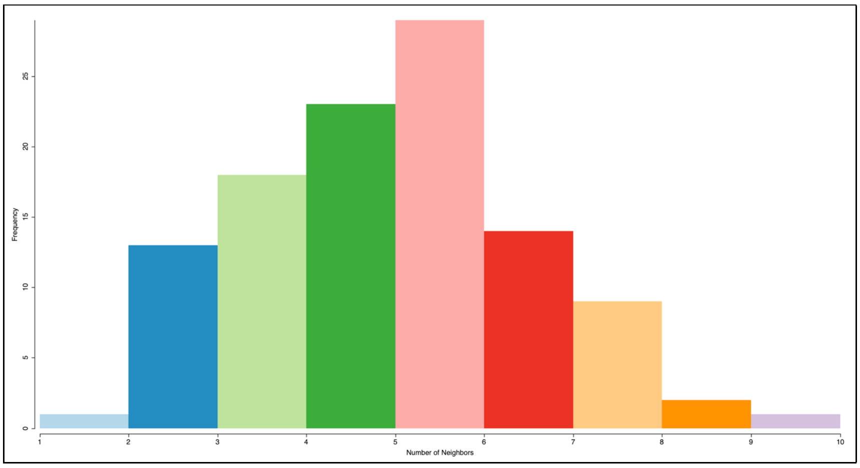


Figure 3.6: Histogram of the weights matrix

As can be seen from the histogram, it highlights that the number of neighbours for the provinces is quite balanced overall, ranging from a minimum of 1 to a maximum of 9 and with an average of 5. In this case, most of the provinces have a number of immediate neighbours that aligns with the average or is close to it, while just a few observations are provided with an amount that is near one of the extreme ends of the spectrum. The following examples can be helpful to illustrate the differences in number of neighbours assigned to each province based upon the defined weights matrix:

- The province of Trieste has 1 neighbouring province: Gorizia;
- The province of Como has 4 neighbouring provinces: Monza e Brianza, Varese, Lecco and Sondrio;
- The province of Caserta has 6 neighbouring provinces: Latina, Frosinone, Isernia, Campobasso, Benevento and Napoli;
- The province of Firenze has 9 neighbouring provinces: Lucca, Pisa, Siena, Arezzo, Forlí-Cesena, Ravenna, Bologna, Prato and Pistoia.

In addition, the following map illustrates the provinces of Italy and, using the province of Firenze as an example, it depicts how neighbours are considered within the defined weights matrix:

Figure 3.7: Map of the weights matrix

In the map, the province of Firenze is highlighted with a dark green colour and its 9 neighbours are depicted with a white colour. As can be seen, the neighbours share a border of any length with the province, while other provinces that do not share one are not taken into consideration. In this context, a province may be neighbour of Firenze independently from whether it is located into the same region or another; in fact, while Firenze is the capital of the region Toscana, 6 provinces belong to the same region (e.g. Siena), while the other 3 are found in the region of Emilia-Romagna (e.g. Bologna). This fact makes the analysis of regional patient migration interesting from a spatial point of view, since it can account for influences occurring between individuals in provinces located in the same region and surrounding ones.

To conclude, the following table outlines some results of the row-standardisation procedure that depict how weights assigned to various provinces have different values depending upon the number of their neighbours:

Observation	Neighbours	Weight $w_{ij(s)}$
1	7	0,1428571
2	5	0,20
3	4	0,25
⋮	⋮	⋮
54	6	0,1666667
55	5	0,20
56	5	0,20
⋮	⋮	⋮
108	2	0,50
109	3	0,3333333
110	2	0,50

Table 3.1: Row-standardised weights for some observations in the weights matrix

Each weight $w_{ij(s)}$ has a fundamental role in the execution of various tests for spatial autocorrelation, whose measures consist of compromises between attribute and locational similarity, with the latter being expressed through the spatial weights, and the definition of spatially explicit variable in the statistical spatial models (WY, WX or $W\epsilon$), which contribute to the regression results by taking into account the values observed at neighbouring locations, weighted by their degree of influence that is numerically expressed by the spatial weights. Taking the province of Firenze as an example again, its specific row in the spatial weights matrix is initially defined as follows, with a total of 9 neighbours among the total of 110 provinces:

$$W = \begin{bmatrix} 0 & \dots & 1 & 1 & \dots \\ 0 & \dots & 1 & 1 & \dots \\ 0 & \dots & 1 & 1 & 1 \\ \dots & 0 & \dots & 1 & \dots \\ 0 & \dots & 1 & \dots & 0 \end{bmatrix} \tag{3.15}$$

Using the row-standardisation procedure, the single spatial weights in the row are divided by 9, the total amount of neighbours for the province of Firenze, defining the following specific row in the standardised weights matrix:

$$W = \begin{bmatrix} 0 & \dots & 0,1111111 & 0,1111111 & \dots \\ 0 & \dots & 0,1111111 & 0,1111111 & \dots \\ 0 & \dots & 0,1111111 & 0,1111111 & 0,1111111 \\ \dots\ 0 & \dots & & 0,1111111 & \dots \\ 0 & \dots & 0,1111111 & \dots & 0 \end{bmatrix} \tag{3.16}$$

In this case, each nearby province has a spatial weight of around 0,11. As a counter-example, a province with only 1 neighbour is given a single spatial weight of 1.

3.2.3 Regression equation

A multiple linear regression will be used for analysis with the various statistical models and has been defined using the following equation:

$$y_i = \alpha + \sum_{j=1}^{k} \beta_j x_{ji} + \epsilon_i, \tag{3.17}$$

where, for $i = 1, ..., n$ and $j = 1, ..., k$, y_i is the dependent variable, x_{ji} is one of the k independent variables and ϵ_i is the error term. Considering that $i = 1, ..., n$, there exist the following n equations for each observation in the data:

$$y_1 = \alpha + \sum_{j=1}^{k} \beta_j x_{j1} + \epsilon_1, \tag{3.18}$$

$$y_2 = \alpha + \sum_{j=1}^{k} \beta_j x_{j2} + \epsilon_2, \tag{3.19}$$

$$\vdots$$

$$y_n = \alpha + \sum_{j=1}^{k} \beta_j x_{jn} + \epsilon_n, \tag{3.20}$$

which can be simplified and merged together using matrix notation. In particular, the multiple linear regression equations can be shown with the following matrix expression:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & \dots & x_{k1} \\ x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \tag{3.21}$$

which can further become simplified into the final matrix form of the multiple linear regression equation:

$$Y_i = \alpha \iota_n + \sum_{j=1}^{k} \beta_j X_{ji} + \epsilon_i, \tag{3.22}$$

for $i = 1, \ldots, n$. For each observation i, Y is an $n \cdot 1$ column vector of the dependent variable, ι_n is a $n \cdot 1$ column vector of ones related to the constant term α, X is an $n \cdot k$ matrix of the independent variables, β is a $k \cdot 1$ column vector of the predictor coefficients, which describe their related relationships with the dependent variable, and ϵ is an $n \cdot 1$ column vector of the error term. In this case, the matrix X and vector β are multiplied together with the methods of matrix multiplication, then the resulting vector βX is added to the vectors α and ϵ using the methods of matrix addition.

The linear regression model, as the equation illustrates, will be provided with systematic components α, the intercept that measures the value where the regression line crosses the y axis, and βX, which represent the independent variables and their respective coefficients, as well as a stochastic component ϵ. Moreover, the model may be further enhanced with the inclusion of one or more spatially lagged variables WY, WX and $W\epsilon$, with their respective coefficients ρ, θ and λ, to account for the presence of one or more spatial effects in the data, which will depend upon the selection of a certain spatial model for the data that results from evaluations that account for the outcomes of the various specifications tests and statistical instruments.

3.2.4 Variable identification

The collection of data requires the identification of the dependent and independent variables to be examined using the presented statistical models. As mentioned at the beginning of the theoretical portion, the selection of the dependent variables is influenced by the intention of looking upon the matter of regional patient migration from two points of view, patient immigration, which takes place into a region from another, and patient emigration, which happens from a region into another, with additional divisions into ordinary and day admissions. Therefore, a total of four dependent variables will be considered and analysed separately from one another to deliver a comprehensive overview on the phenomenon as a whole; in particular, the dependent variables will relate to the following specific subtopics: regional patient immigration for ordinary admissions; regional patient immigration for day admissions; regional patient emigration for ordinary admissions; regional patient emigration for day admissions.

With regards to the independent variables, their identification revolves around the importance of certain factors for each subtopic of regional patient migration, which can be supposed from the suggestions deriving from the previous literature review. First of all, factors related to the quality of health care services offered by local health authorities can be considered to be fundamental for every aspect of patient migration, in terms of influencing both the attraction and escape rates of a region; accounting for the reviewed components of health care quality, explanatory variables considered for inclusion could be related to the availability of various resources, the efficiency of the processes of medical care and the associated outcomes. Furthermore, other types of explanatory variables seem to be of relevance and worth taking into account, such as the income and the age of individuals in a province or the number of public and private providers that are located there. Nonetheless, the availability and quality of the data will influence the selection of potential variables of interest for the analysis, which will be discussed in the forthcoming section on the data set preparation.

3.2.5 Research hypotheses

As mentioned in the introduction, the ability to answer the research questions requires the definition of specific hypotheses to be tested in the data analysis.

The hypotheses related to the first research question, on the existence of a linear relationship between the dependent variable and one or more independent variables, are defined as follows:

$$H_0 : \beta_1 = \cdots = \beta_j = \cdots = \beta_k = 0 \tag{3.23a}$$

$$H_1 : \beta_j \neq 0 \text{ for at least one } j \tag{3.23b}$$

The hypotheses related to the second research question, on the absence or presence of spatial autocorrelation for the dependent variable, are defined as follows:

$$H_0 : \rho = 0 \tag{3.24a}$$

$$H_1 : \rho \neq 0 \tag{3.24b}$$

The hypotheses related to the third research question, on the absence or presence of spatial effects for the independent variables, are defined as follows:

$$H_0 : \theta = 0 \tag{3.25a}$$

$$H_1 : \theta \neq 0 \tag{3.25b}$$

Furthermore, to also account for additional consequences of unobserved factors, the hypotheses related to the absence or presence of spatial effects for the error term are defined as follows:

$$H_0 : \lambda = 0 \tag{3.26a}$$

$$H_1 : \lambda \neq 0 \tag{3.26b}$$

3.3 Data set preparation

3.3.1 Data collection

For the purpose of executing spatial analysis on the data for regional patient migration, digital information on the administrative boundaries of Italy was required. The Italian National Institute of Statistics (ISTAT) provides official data sets on the matter for statistical purposes on a yearly basis, using 1 January as a constant reference from 2002, on four hierarchical levels: geographical divisions, regions, provinces and municipalities. The geographical information on the administrative boundaries was defined with specific GIS programs, respecting the WGS84 standard, and coded into a shapefile, which is a geospatial vector data format, developed and regulated by the company Esri, that is used by geographic information system (GIS) softwares to capture, store, manipulate, analyse, manage and present spatial or geographic data. The vector is composed by at least three main mandatory files: .shp (the main file that stores the feature geometry), .shx (the index file that stores the index of the feature geometry) and .dbf (the database table that stores the attribute information of features).

The subject of regional patient migration is hereby examined at a provincial level, with the employment of the appropriate data set of geographical files provided by ISTAT, because of a series of reasonable motivations. First of all, this choice provides the researcher with a more robust sample size of 110 provinces, against a smaller one composed by solely 20 regions, which can strengthen the extrapolation of eventual results. Secondly, the examination of spatial data through a suitable model can optimally occur when each observed unit has at least one neighbour, when constructing a spatial weight matrix, so that weighing spatial effects on the variables of interest can occur appropriately; the usage of regions for spatial analysis would have created an important problem concerning the presence of "isolates", observational units that do not have any neighbour, which would have been the isles of Sicilia and Sardegna; on the other hand, a data analysis on provinces guarantees that each unit in the sample has at least one neighbour. Moreover, given that the topic of interest is patient migration among regions, examining spatial effects occurring between provinces that may be located in either the same area or different regions, depending on their geographical positions, could potentially capture interesting information on the matter. Finally, it seemed to be more reasonable to examine smaller units for the scope of avoiding the risk of neglecting effects due to mismatches between the extent of their occurrence and the scale of observation, which might have been too broad had regions been instead chosen.

The selection of the final geographical information files to use as the base for spatial analysis was preceded by preliminary steps of selection and resolution of technical issues. First of all, the files that relate to the year 2016 were chosen as the most appropriate, since the most recent ones reflect some administrative changes occurred in Italy, such as the creation of a new province and the suppression of others in the region of Sardegna, for which the structure of the database where the data was gathered from has not been updated yet. With regards to the geographical files, technical issues related to the geometry information of some Italian provinces were found and repaired with a dedicated programme, since their presence could have hindered the processes of analysis of spatial data; several tables that reveal the details about the correction of the problem can be found in the section "Repair of shapefile geometries" of the Appendix A named "Data set preparation".

The process of data collection to be examined through spatial analysis has been fairly straightforward, since potential sources of information had been flawlessly identified and inspected with ease and in a relatively short amount of time. The scope of the thesis and the analysis methodology used to find out the presence of relevant results leverage on the availability of data that is secondary by nature, since it was already collected, archived and made officially available by the Italian National Institute of Statistics; as a consequence, primary data collection has not been executed in this particular context. As Hox and Boeije (2005) suggested, the collection of secondary data presents researchers with three main problems: location of data sources useful for the research issue; retrieval of relevant data from said sources; evaluation of the data quality with regards to the current research requirements and the technical criteria of good scientific practice [20]; these problematics will be addressed shortly after the introduction of the secondary data source that has been used for the research.

The main source of data for the execution of spatial analysis is the "Health For All – Italia" database, provided by the Italian National Institute of Statistics, which contains 4.000 indicators about the health care system and population health in Italy in a format that is compatible with a dedicated HFA interrogation software, which was developed by the World Health Organization specifically for national requirements. The indicators are updated periodically to add new ones, to fill missing pieces of data or to strengthen their presence on the more detailed provincial level. The original data was exported from the HFA software into a .csv file, then it was appropriately transformed with programming code, written with the R language, into a new .csv file, for reasons that will be explained later, and finally merged into the main database through join procedures executed in the GeoDa software by specifying a common variable for the correct union of the data.

Returning to the issues pertaining to the usage of secondary data that were outlined previously for a brief moment, all three of them have been correctly addressed before conclusively gathering the needed information from the aforementioned data source. In particular, it is possible to make the following statements with regards to all the three points of issue underlined by the cited researchers:

- *Location of data sources useful for the research issue* – Since the topic relates to the population and the health care system of Italy, the search for data had begun from several Italian sources, which led the researcher identify the Italian National Institute of Statistics and the Italian Ministry of Health as potential information suppliers for such an important national matter. Eventually, the finding of the aforementioned database from ISTAT concluded the search for data, as it was found to include all the relevant information for the research;

- *Retrieval of relevant data from the sources* – The Health For All database contains all the currently available information on the Italian population health and the Italian health care system, especially since it was recently updated. Due to the magnitude of the database, the search for information might have seemed to be hindered at a first sight. Nonetheless, after an extensive review of the database, the retrieval of the indicators that are deemed to be the most appropriate for the research occurred with ease and sufficient confidence, especially considering their availability for both the dependent and independent variables in the statistical models and their completeness for the selected time frame between the years 2012 and 2014;
- *Evaluation of the data quality with regards to the current research requirements and the technical criteria of good scientific practice* – The available data was reviewed and the sole statistics needed for the research were extrapolated from the database after a quality assessment, which led to the retainment of the most complete portions of evidence for the variables of interest. The collection of such an exhaustive and high-quality data set meets the criteria of good scientific practice, such as conducing fair scientific research, undertaking responsibilities for the validity of the research results and respecting ethical standards when interpreting them, without conflicts of interest that may bias the statements made by the researcher.

3.3.2 Data selection

With regards to the gathered secondary data, the most appropriate indicators for the analysis had been selected among all and divided into two categories of dependent and independent variables to proceed with the analysis. In the database, each indicator was named with a concise abbreviation, since the .dbf file format associated with the accompanying .shp file has a limit of 10 characters for each variable name. The absence of data for some or even all observations in certain years led the researcher build a data set of the selected indicators that takes three years into account, from 2012 to 2014.

The following lists provide an overview of the dependent and independent variables, named by their abbreviations and accompanied by a description of what they refer to and measure. As already mentioned, the dependent variable choice was influenced by the existence of topic subdivisions, since the matter of patient migration is looked upon from the viewpoints of patient immigration and emigration, with further separation between ordinary and day admissions. Therefore, the following four dependent variables have been identified:

- **RHIOAPxx** – Percentage of acute patients that emigrated to the region of a certain province from the provinces in other regions of Italy to gather planned health care treatments, through ordinary admissions in public or accredited private facilities, in the year 20xx ($\frac{\text{Discharges of non-residents in region “L”}}{\text{Total discharges in region “L”}} \cdot 100$);
- **RHIDAPxx** – Percentage of acute patients that emigrated to the region of a certain province from the provinces in other regions of Italy to gather planned health care treatments, through day admissions in public or accredited private facilities, in the year 20xx ($\frac{\text{Discharges of non-residents in region “L”}}{\text{Total discharges in region “L”}} \cdot 100$);
- **RHEOAPxx** – Percentage of acute patients residing in a certain province of Italy that emigrated from their region to another to gather planned health care treatments, through ordinary admissions in public or accredited private facilities, in the year 20xx ($\frac{\text{Discharges of residents of region “L” in region “J”}}{\text{Total discharges of residents of region “L”}} \cdot 100$);
- **RHEDAPxx** – Percentage of acute patients residing in a certain province of Italy that emigrated from their region to another to gather planned health care treatments, through day admissions in public or accredited private facilities, in the year 20xx ($\frac{\text{Discharges of residents of region “L” in region “J”}}{\text{Total discharges of residents of region “L”}} \cdot 100$).

With regards to the independent variables, their selection was primarily influenced by the information from the literature and the availability of statistics from the data set, as stated in the section on variable identification, regardless of the topic subdivisions.

A comprehensive assessment resulted in their confinement into the specific quantitative area of health care resources. As a consequence, the following seven independent variables have been identified:

- **BedOARxx** – Rate per 10.000 population of beds for ordinary admissions of acute patients in public and accredited private health care facilities ($\frac{\text{Hospital beds}}{\text{Population}} \cdot 10.000$);
- **AvgOHDxx** – Average length of hospitalisation of acute patients through ordinary admissions ($\frac{\text{Days of hospitalisation}}{\text{Ordinary admissions of acute patients}}$);
- **BedDARxx** – Rate per 10.000 population of beds for day admissions of acute patients in public and accredited private health care facilities ($\frac{\text{Hospital beds}}{\text{Population}} \cdot 10.000$);
- **AvgDHCLxx** – Average length of "day hospital" cycle for day admissions of acute patients. A cycle starts with the opening of a medical record and ends with its closure; its duration refers to the number of days in which the patient visited a public or accredited private facility for health treatments ($\frac{\text{"Day hospital" cycle duration}}{\text{Discharges of day admitted patients}}$);
- **MedEqRxx** – Rate per 10.000 population of medical equipment in public and accredited private health care facilities ($\frac{\text{Medical equipment}}{\text{Population}} \cdot 10.000$);
- **DocDenRxx** – Rate per 10.000 population of doctors and dentists in public and accredited private health care facilities ($\frac{\text{Doctors and dentists}}{\text{Population}} \cdot 10.000$);
- **NursesRxx** – Rate per 10.000 population of nurses in public and accredited private health care facilities ($\frac{\text{Nurses}}{\text{Population}} \cdot 10.000$).

Throughout the data analysis, the independent variables BedOARxx and AvgOHDxx will be included in the statistical models that relate to the dependent variables RHIOAPxx and RHEOAPxx, while the independent variables BedDARxx and AvgDHCLxx will be comprised in those that relate to the dependent variables RHIDAPxx and RHEDAPxx. Furthermore, the independent variables MedEqRxx, DocDenRxx and NursesRxx will be considered in every statistical model for all four dependent variables.

3.3.3 Data transformation

For various motivations that will be outlined soon, all the data of both the categories of dependent and independent variables had to be transformed before proceeding with the data analysis. With regards to the dependent variables, their residuals followed strong positively skewed distributions, which are characterised by a long right tail, and their spread changed systematically with the values of the dependent variable, a statistical condition of the data named heteroscedasticity. The Jarque-Bera test, established by Jarque and Bera (1980), was employed to test whether the original data sample retained the same skewness and kurtosis as the normal distribution, which has respective values that are equal to 0 and 3, based on the null hypothesis of the residuals being normally distributed [22]; to be specific, the Jarque-Bera test statistic is defined as:

$$JB = \frac{N}{6}\left(W^2 + \frac{(K-3)^2}{4}\right) \tag{3.27}$$

The execution of the test on the original dependent variables provided the results shown in the following table, using the variables for the year 2014 as an example:

Variable	χ^2	**DF**	**p-value**
RHIOAP14	384.58	2	< 0.00000000000000022
RHIDAP14	52.132	2	0.000000000004783
RHEOAP14	29.68	2	0.0000003591
RHEDAP14	39.349	2	0.000000002854

Table 3.2: Jarque-Bera test for the original dependent variables (2014)

As can be seen, the results undoubtedly confirmed the presence of heteroscedasticity of residuals for every dependent variable, which always represents a problem for linear regression analysis with the ordinary least squares methods, because it violates one of the assumptions on the homoscedasticity of residuals, and therefore needs to be solved before continuing with the data analysis.

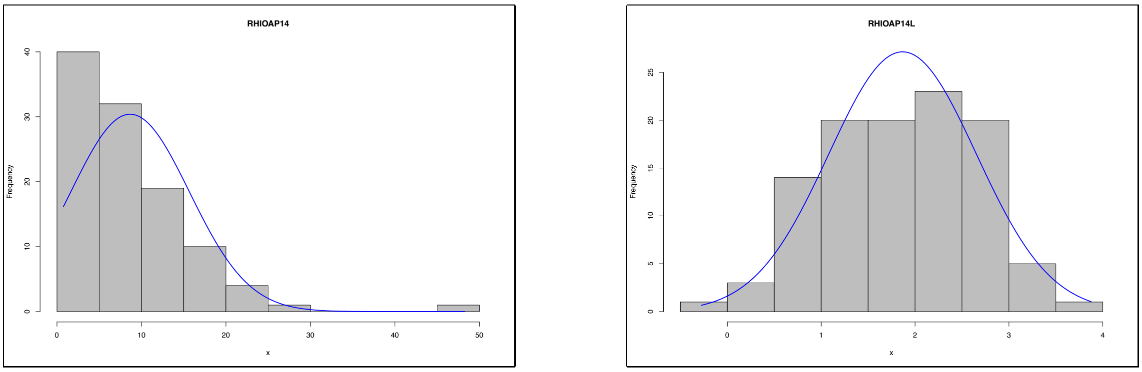
In this situation, the dependent variables were log-transformed by taking their natural logarithms, to obtain residuals that were approximately symmetrically distributed and to remove their systematic change in spread, roughly achieving the opposite statistical assumption of homoscedasticity, so that it could have been possible to correctly conduct the analysis. In fact, this type of transformation will permit the execution of all the statistical tests, which depend upon the assumption of normality of the residuals. The natural logarithmic transformation is often used in the fields of statistical analysis and social sciences since it is a simple process and, as Gelman and Hill (2006) suggested, "coefficients on the natural-log scale are directly interpretable as approximate proportional differences: with a coefficient of 0.06, a difference of 1 in x corresponds to an approximate 6% difference in y, and so forth" [17, 60-61]. For the data analysis in this research, the letter "L" at the end of the name of a variable indicates the execution of this logarithmic transformation procedure. The following table illustrates the results of the Jarque-Bera test for the new set of dependent variables resulted from the execution of the logarithmic transformation, using the variables for the year 2014 as an example:

Variable	χ^2	**DF**	**p-value**
RHIOAP14L	1.4939	2	0.4738
RHIDAP14L	4.1254	2	0.1271
RHEOAP14L	2.0297	2	0.3625
RHEDAP14L	2.2686	2	0.3216

Table 3.3: Jarque-Bera test for the log-transformed dependent variables (2014)

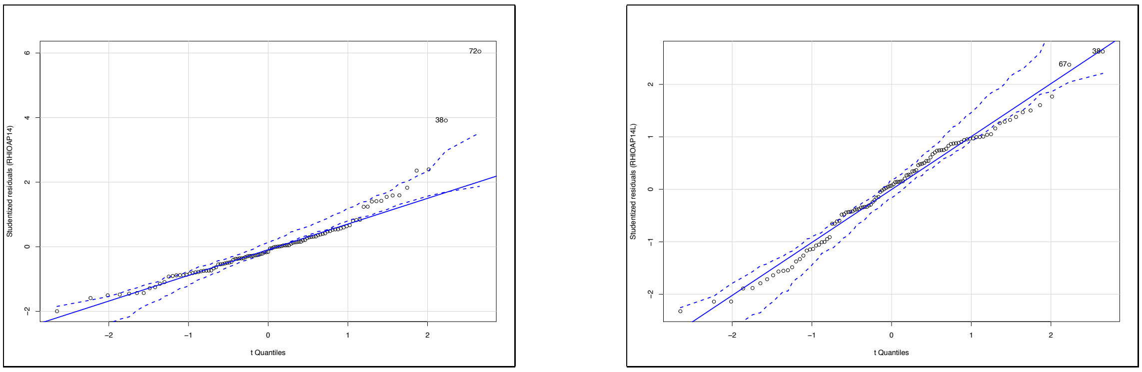
These outcomes indicate that the assumption of homoscedasticity of residuals has been satisfied, enabling to count on the results of the data analysis. This achievement is also confirmed with histograms, that show the distribution of a continuous variable and are used to determine if the values of each dependent variable are normally distributed, and probability plots, which represent the residuals of the data against the expected order statistics of the standard normal distribution and indicate negative or positive skewness depending upon showing curvatures with downward or upward concavity.

The following histograms and probability plots (Q-Q plots) illustrate the effects of the logarithmic transformation on the distribution of the dependent variables and their residuals, using the variable RHIOAP14 as an example:



(a) Distribution of RHIOAP14 **(b)** Distribution of RHIOAP14L

Figure 3.8: Logarithmic transformation of the dependent variable RHIOAP14

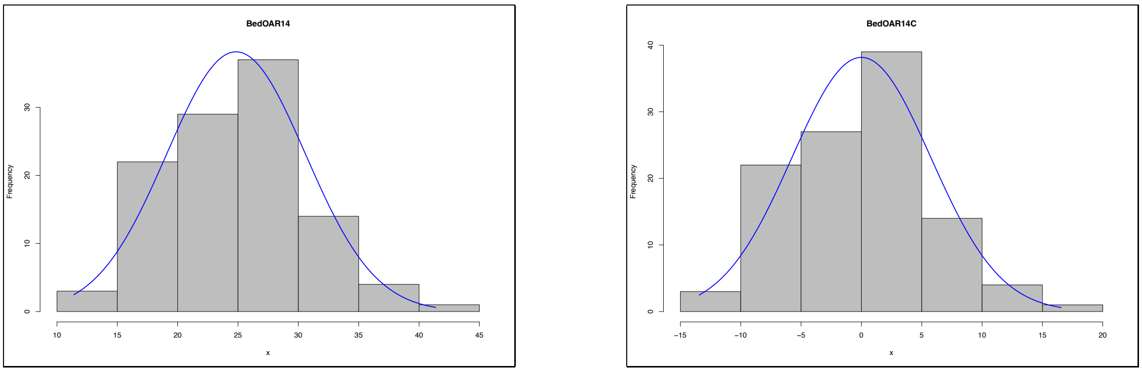


(a) Q-Q plot for RHIOAP14 **(b)** Q-Q plot for RHIOAP14L

Figure 3.9: Q-Q plots of residuals for RHIOAP14 and RHIOAP14L

The histograms illustrate that the original dependent variable followed a positively skewed distribution, while the log-transformed one is normally distributed. Moreover, the Q-Q plots show that the residuals of the original dependent variable followed a positively skewed distribution, while those of the log-transformed one are normally distributed, as indicated by the upward concavity in the first plot and the loose adherence to a straighter line at a 45° upward angle in the second plot. The logarithmic transformation has not altered the values of the data and the interpretation of the analysis results will just need to follow the guideline outlined for the natural logarithmic transformation.

Concerning the independent variables, a mean-centring procedure was executed to diminish the collinearity between them, avoiding problems of inflated multicollinearity indicators that could have wrongly questioned the selection of independent variables for the analysis with the statistical models. The procedure involved the subtraction of the mean from the values of each respective independent variable, which resulted in their centring around zero. In this case, the procedure has not affected neither the inherent meanings of the data nor any characteristic of the independent variables, such as the standard deviation and skewness. For the data analysis in this research, the letter "C" at the end of the name of an independent variable indicates the execution of this mean-centring procedure. The following histograms depict the results of the mean-centring procedure on the distribution of the independent variables, using the variable BedOAR14 as an example:



(a) Distribution of BedOAR14

(b) Distribution of BedOAR14C

Figure 3.10: Mean centring of the independent variable BedOAR14

The section "Data transformation" of the Appendix A named "Data set preparation" contains additional histograms and Q-Q plots that illustrate the effects of the logarithmic transformation on the distribution of all the dependent variables and their residuals, as well as the outcomes of the mean-centring procedure on the distribution of all the independent variables, for all the years taken into account for the data analysis.

Chapter 4

Data analysis

As outlined in the section on data selection, the research inspects the matter of regional patient migration from the two viewpoints of patient immigration and emigration, with further separation between ordinary and day admissions. These topic subsections define the organisation of this chapter on the analysis of data, separating it into two sections with two further subsections as per the following structure:

- **Regional patient immigration**, which regards individuals that emigrated to the region of a certain province from the provinces in other regions of Italy to obtain planned health care treatments in a particular year. The analysis is further divided into the inspection of data on ordinary and day admissions of acute patients.
- **Regional patient emigration**, which concerns individuals residing in a specific province of Italy that emigrated from their region to another to gather planned health care treatments in a particular year. The analysis is further divided into the examination of data on ordinary and day admissions of acute patients.

Some reasons support the separation of the data analysis into various portions. First of all, as already portrayed when discussing the preparation of the data set, the selection of independent variables changes when conducting an examination of various aspects of regional patient migration. Moreover, the data are separately available for ordinary and day admissions of acute patients, thus differentiating between two further subsections is considered as appropriate. Finally, using distinct subsections for each subset of the data permits to conduct the model selection procedures with manners that are appropriate for each case. In accordance with the previous outline on the gathering of data, the entire analysis accounts for a period of three years, between 2012 and 2014.

4.1 Regional patient immigration

4.1.1 Ordinary admissions

Overview

The first part of the analysis involves gathering information from the data to assimilate how the phenomenon had been happening in the country. First of all, the following table summarises the main information on the data regarding regional patient immigration for ordinary admissions, for each year during the period 2012-2014:

Variable	Minimum	Mean	Maximum
RHIOAP12	1,210	8,480	47,050
RHIOAP13	1,350	8,578	49,130
RHIOAP14	0,760	8,654	48,250

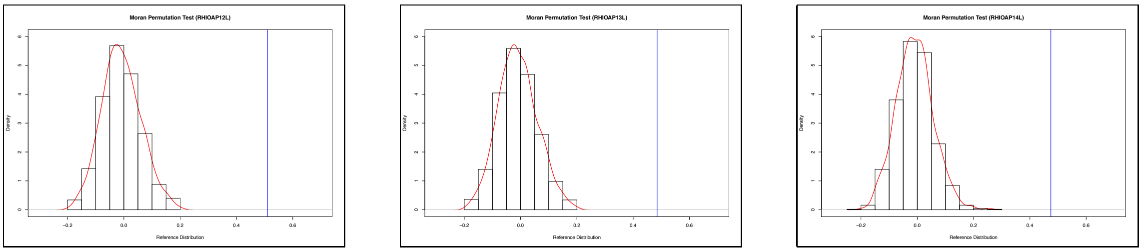
Table 4.1: Summary of regional patient immigration (ordinary admissions) (2012-2014)

The table illustrates that the percentage of patients gathering health treatments for ordinary admissions from a province in a particular region, coming from another region, had decreased in certain areas and increased in others over time, with an overall raising average percentage. Therefore, it can be declared that widening divergences had existed in the occurrence of regional patient immigration for ordinary admissions, making the phenomenon of interest for more research. Employing the log-transformed dependent variables, the Moran's I tests for RHIOAPxxL calculated the following Moran's I values for each year, excluding 3 observations without information in the data:

Variable	Moran's I	p-value
RHIOAP12L	0,509690975	7,384e^{-15}
RHIOAP13L	0,484920085	1,207e^{-13}
RHIOAP14L	0,474442953	3,595e^{-13}

Table 4.2: Moran's I values for RHIOAPxxL (2012-2014)

The following images display various density plots on the reference distribution for the Moran's I values related to each year, which illustrate how every observed value is statistically significant and quite distant from the expected value $E(I) = \frac{-1}{1-N} = \frac{-1}{1-107} = -0{,}009433962$:



(a) Moran permutation test for RHIOAP12L

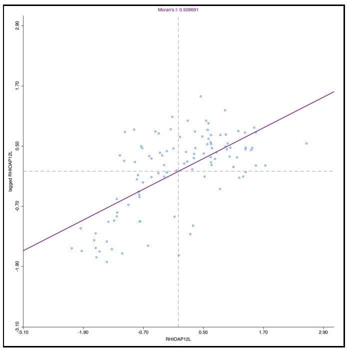
(b) Moran permutation test for RHIOAP13L

(c) Moran permutation test for RHIOAP14L

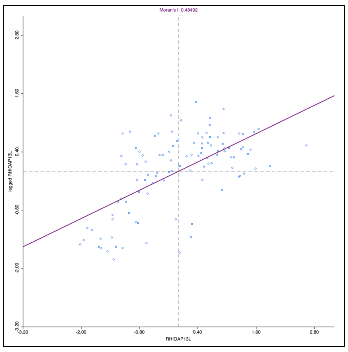
Figure 4.1: Moran permutation tests for RHIOAPxxL (2012-2014)

Taking the low p-values and the significant differences with the expected value into account, it is possible to reject the null hypothesis of absence of spatial autocorrelation and to declare that positive spatial autocorrelation in the data is observed for each year in the period 2012-2014. The underlying meaning is that the phenomenon of patient immigration for ordinary admissions had not been occurring in a random fashion across the country, but rather had tended to be clustered among its various areas, with provinces having high patient immigration percentages being closer to one another and provinces with low patient immigration percentages displaying the same disposition. This result is significant, since it illustrates that the behaviour of patients towards the treatment offers in a province was not independent from that of other patients found in close provinces, violating the assumption of independence of observations in a linear regression model and suggesting the need to conduct some sort of spatial analysis.

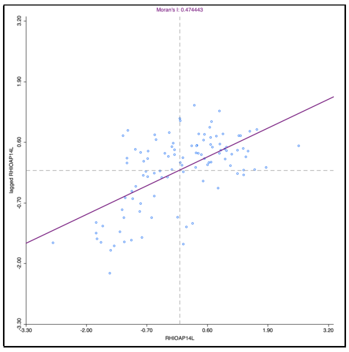
This situation can be more thoroughly discerned with the support of supplementary instruments that communicate further information. For instance, the following Moran scatter plots, obtained from the programme GeoDa, can assist with the identification of the presence and direction of spatial autocorrelation related to the dependent variables of patient immigration for ordinary admissions, for each year in the period 2012-2014:



(a) Moran scatter plot for RHIOAP12L



(b) Moran scatter plot for RHIOAP13L

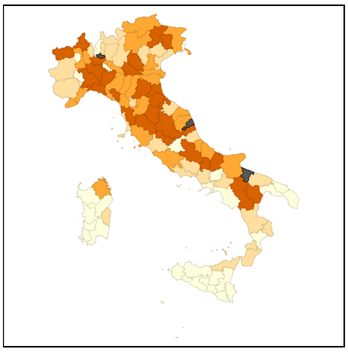


(c) Moran scatter plot for RHIOAP14L

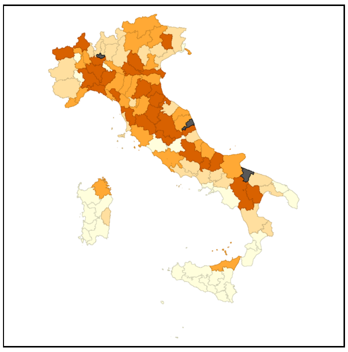
Figure 4.2: Moran scatter plots for RHIOAPxxL (2012-2014)

The Moran scatter plots portray the presence of a positive spatial autocorrelation of the phenomenon in each year between 2012 and 2014, driven by the observations in the lower-left and upper-right quadrants: some provinces with high patient immigration rates had tended to be close to others with high patient immigration rates as well (upper-right quadrant), while some provinces with low patient immigration rates had tended to be near others with low patient immigration rates too (lower-left quadrant). Considering the information from the data, it is possible to assert that the phenomenon had become slightly less clustered from 2012 to 2014, although while retaining a significant number of clusters of provinces with similar patient behaviour.

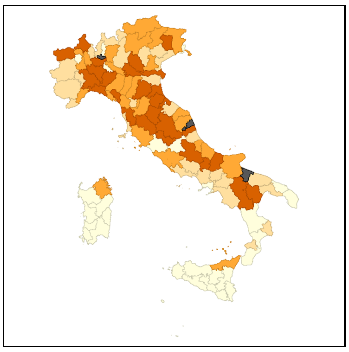
In addition, the following quartile maps depict how the percentage values of patient immigration for ordinary admissions are distributed when grouped into four classes:



(a) Quartile map for RHIOAP12L



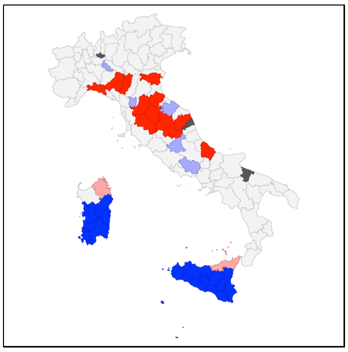
(b) Quartile map for RHIOAP13L



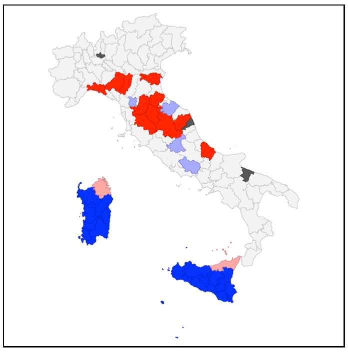
(c) Quartile map for RHIOAP14L

Figure 4.3: Quartile maps for RHIOAPxxL (2012-2014)

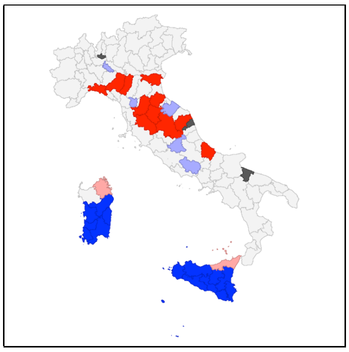
The phenomenon of regional patient immigration for ordinary admissions seemed to happen primarily in provinces of Northern and Central Italy, with some outliers in Southern and Insular Italy. The following LISA cluster maps and LISA significance maps are also employed to further discern the aspects of its occurrence in the country:



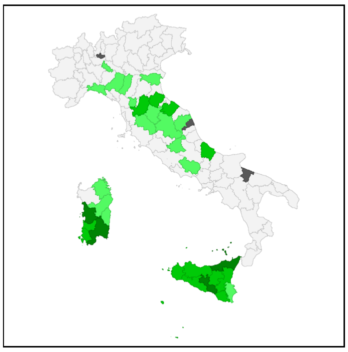
(a) LISA cluster map for RHIOAP12L



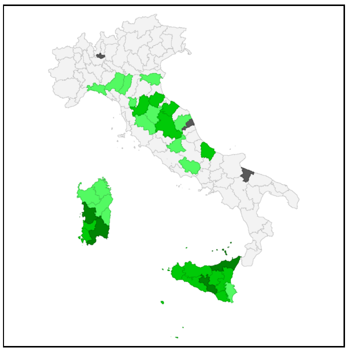
(b) LISA cluster map for RHIOAP13L



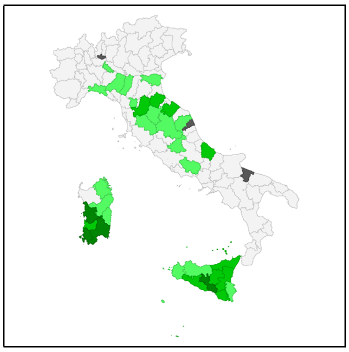
(c) LISA cluster map for RHIOAP14L



(d) LISA significance map for RHIOAP12L



(e) LISA significance map for RHIOAP13L



(f) LISA significance map for RHIOAP14L

Figure 4.4: LISA cluster and significance maps for RHIOAPxxL (2012-2014)

In the LISA cluster maps, a province that is marked with a colour represents the core of a cluster of neighbouring provinces, as defined by the specified weights matrix, which has percentages of patient immigration that are either similar or dissimilar to those of nearby provinces. A province is marked in red if it has a high percentage of patient immigration and is surrounded by neighbouring provinces with a high percentage, while it is marked in blue if it has a low percentage of patient immigration and is surrounded by neighbouring provinces with a low percentage. A light-red province consists of an outlier with a high percentage of patient immigration that is surrounded by neighbouring provinces with a low percentage, while a light-blue province consists of an outlier with a low percentage of patient immigration that is surrounded by neighbouring provinces with a high percentage. All the marked provinces reached statistical significance and their significance levels are mirrored in the LISA significance maps with various degrees below $\alpha = 0{,}05$. For this subtopic, values for three observations are missing as shown by the provinces marked in grey. In this situation, the cluster maps show a concentration of clusters with high patient immigration percentages around Northern and Central Italy and low patient immigration percentages in Insular Italy, with a few outliers present around these clusters as well.

Analysis framework

The second part of the analysis involves the definition of a specific analysis framework and the illustration of the diverse analysis procedures that depend upon it. In particular, the framework features a multiple linear regression equation and a set of variables that, to allow the data to be examined through various statistical models, are defined for the subtopic in question according to the following specifications (where "xx" corresponds to a specific year in the period 2012-2014):

$$Y_i = \alpha \iota_n + \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 X_{3_i} + \beta_4 X_{4_i} + \beta_5 X_{5_i} + \epsilon_i \qquad \text{for } i = 1, \ldots, n \tag{4.1}$$

Equation variable	**Specific variable**
Y	RHIOAPxxL
X_1	BedOARxxC
X_2	AvgOHDxxC
X_3	MedEqRxxC
X_4	DocDenRxxC
X_5	NursesRxxC

Table 4.3: Specific variables in equation 4.1 for regional patient immigration (ordinary admissions) (2012-2014)

Analysis procedure (2012)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedOAR12C	4,111104	4,753
AvgOHD12C	1,197256	
MedEqR12C	2,371673	
DocDenR12C	3,476604	
NursesR12C	4,454663	

Table 4.4: VIFs and condition number of the predictors in equation 4.1 (2012)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 17{,}19$ and $p\text{-}value = 2{,}815e^{-12}$) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}363150113$ is significantly diverse from the expected value $E(I) = -0{,}018782526$ ($p\text{-}value = 3{,}69e^{-9}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	34,588	4,074e^{-9}
LMerr	27,987	1,222e^{-7}
RLMlag	7,2108	0,007246
RLMerr	0,60948	0,435
SARMA	35,197	2,275e^{-8}

Table 4.5: Results of the specification tests for equation 4.1 (2012)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	AIC	BIC	Log Likelihood	R^2	LR Test
LM	197,9158	216,6256	-91,95792	0,4331	–
SLX	197,0614	229,1353	-86,53070	0,4611	–
SAR	171,5069	192,8896	-77,75346	0,6103669	–
SEM	174,224	195,6066	-79,11199	0,6117717	–
SDM	176,9998	211,7465	-75,91207	0,6277262	SAR / SEM
SDEM	177,4534	212,2002	-75,72672	0,6266914	SEM
SARAR	173,2808	197,3362	-77,64039	0,6211201	SAR / SEM

Table 4.6: Measures of goodness of fit for equation 4.1 (2012)

The SAR model has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SEM), a result that aligns with the outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SAR model or SEM, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. The information from the two approaches indicates that the SAR model has the best goodness of fit and should be taken as the source for the results.

Analysis procedure (2013)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedOAR13C	3,870235	4,546
AvgOHD13C	1,222987	
MedEqR13C	2,287293	
DocDenR13C	4,192351	
NursesR13C	4,235132	

Table 4.7: VIFs and condition number of the predictors in equation 4.1 (2013)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 18{,}08$ and $p\text{-}value = 8{,}982e^{-13}$) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}377416684$ is significantly diverse from the expected value $E(I) = -0{,}020872070$ ($p\text{-}value = 7{,}539e^{-10}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	32,136	1,437e^{-08}
LMerr	30,229	3,84e^{-08}
RLMlag	4,0564	0,044
RLMerr	2,1488	0,1427
SARMA	34,285	3,59e^{-08}

Table 4.8: Results of the specification tests for equation 4.1 (2013)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	**AIC**	**BIC**	**Log Likelihood**	$\mathbf{R^2}$	**LR Test**
LM	192,8445	211,5543	-89,42223	0,4462	–
SLX	193,0041	225,078	-84,50203	0,4685	–
SAR	168,9381	190,3207	-76,46903	0,608084	–
SEM	166,9987	188,3813	-75,49935	0,6299011	–
SDM	171,5716	206,3184	-72,78581	0,6395404	SEM / SAR
SDEM	172,498	207,2448	-73,24901	0,637293	SEM
SARAR	167,8799	191,9354	-74,93997	0,6740627	SEM / SAR

Table 4.9: Measures of goodness of fit for equation 4.1 (2013)

The SEM has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SAR), contrasting the outcome of the specification tests. However, the SEM is excluded because a spatial Hausman test suggests the model may not be correctly specified ($p\text{-}value$ = 0,001164). Among the other more encompassing models, an overall view of the measures suggests the SDM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SEM or SAR model, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. The information from the two approaches indicates that the SAR model has the best goodness of fit and should be taken as the source for the results.

Analysis procedure (2014)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedOAR14C	2,945655	4,379
AvgOHD14C	1,201309	
MedEqR14C	2,664688	
DocDenR14C	3,134417	
NursesR14C	4,396955	

Table 4.10: VIFs and condition number of the predictors in equation 4.1 (2014)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 17{,}52$ and $p\text{-}value = 1{,}845e^{-12}$) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}391571909$ is significantly diverse from the expected value $E(I) = -0{,}019749760$ ($p\text{-}value = 2{,}353e^{-10}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	34,18	$5{,}023e^{-09}$
LMerr	32,539	$1{,}168e^{-08}$
RLMlag	4,0731	0,04357
RLMerr	2,4315	0,1189
SARMA	36,612	$1{,}122e^{-08}$

Table 4.11: Results of the specification tests for equation 4.1 (2014)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	**AIC**	**BIC**	**Log Likelihood**	$\mathbf{R^2}$	**LR Test**
LM	198,3582	217,068	-92,17910	0,438	–
SLX	197,377	229,451	-86,68851	0,4664	–
SAR	172,9541	194,3367	-78,47704	0,6088274	–
SEM	171,0467	192,4293	-77,52333	0,6309987	–
SDM	173,7903	208,5371	-73,89514	0,6466963	SEM / SAR
SDEM	174,7299	209,4766	-74,36493	0,6456166	SEM
SARAR	171,7895	195,845	-76,89477	0,6761182	SEM / SAR

Table 4.12: Measures of goodness of fit for equation 4.1 (2014)

The SEM has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SAR), contrasting the outcome of the specification tests. However, the SEM is excluded because a spatial Hausman test suggests the model may not be correctly specified ($p\text{-}value$ = 0,007361). Among the other more encompassing models, an overall view of the measures suggests the SDM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SEM or SAR model, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. The information from the two approaches indicates that the SAR model has the best goodness of fit and should be taken as the source for the results.

Results

The third part of the analysis involves the presentation and explanation of the outcomes resulting from the outlined procedures of data analysis. First of all, to provide them in a clear manner, the following three tables illustrate the results for each considered year in the period 2012-2014, with p-values in parentheses and asterisks indicating which of them are statistically significant:

Variable	Direct impact	Indirect impact	Total impact
BedOAR12C	-0,001823463 (0,9036858)	-0,001381677 (0,9285650)	-0,003205140 (0,914095)
AvgOHD12C	-0,166209981* (0,0098083)	-0,125940830 (0,0559127)	-0,292150811* (0,017591)
MedEqR12C	0,042047037* (0,000002895)	0,031859933* (0,0047243)	0,073906970* (0,000027795)
DocDenR12C	-0,003516616 (0,8102198)	-0,002664615 (0,8695471)	-0,006181231 (0,835195)
NursesR12C	0,012167210 (0,2189679)	0,009219353 (0,2782462)	0,021386562 (0,233272)

Table 4.13: Impacts in the SAR model for RHIOAP12L (2012)

Variable	Direct impact	Indirect impact	Total impact
BedOAR13C	-0,008889287 (0,5880626)	-0,006258144 (0,631768)	-0,015147431 (0,6024123)
AvgOHD13C	-0,201232413* (0,0034188)	-0,141669566* (0,034546)	-0,342901979* (0,0064801)
MedEqR13C	0,039880293* (0,0000052561)	0,028076112* (0,003227)	0,067956405* (0,000012041)
DocDenR13C	0,002172306 (0,9219533)	0,001529325 (0,896012)	0,003701631 (0,9095934)
NursesR13C	0,015631866 (0,0654518)	0,011004985 (0,110800)	0,026636852 (0,0711463)

Table 4.14: Impacts in the SAR model for RHIOAP13L (2013)

Variable	Direct impact	Indirect impact	Total impact
BedOAR14C	-0,0161621271 (0,30895408)	-0,01176962497 (0,3679959)	-0,0279317521 (0,32424065)
AvgOHD14C	-0,2377150259* (0,00034169)	-0,17310943568* (0,0190202)	-0,4108244616* (0,00132946)
MedEqR14C	0,0385518985* (0,000078956)	0,02807436072* (0,0088738)	0,0666262592* (0,00024281)
DocDenR14C	-0,0001190408 (0,96751420)	-0,00008668816 (0,9774304)	-0,0002057289 (0,99218918)
NursesR14C	0,0194538713* (0,04076430)	0,01416674721 (0,0863517)	0,0336206185* (0,04687483)

Table 4.15: Impacts in the SAR model for RHIOAP14L (2014)

Since the outcomes have been retrieved from spatial models, the procedures of data analysis generated various types of effect concerning the independent variables that are represented by three types of impact. With regards to this particular subtopic of patient immigration for ordinary admissions, the impacts can be defined as follows:

- **Direct impact**: it measures the average effect that a factor in a province has on patient immigration for ordinary admissions in the same province;
- **Indirect impact**: it measures the average effect that a factor in a province has on patient immigration for ordinary admissions in the other provinces, in a direct manner or through its influence on the phenomenon in the same province;
- **Total impact**: it measures the average effect that a factor in a province has on patient immigration for ordinary admissions in all provinces in a global fashion, by merging the direct and indirect impacts.

Establishing a distinction between these effects permits to see whether the various impacts differ in terms of statistical significance (e.g. the direct or indirect impact may be statistically significant, while the total may not) and to evaluate the strengths of the direct and indirect impacts, which may be hidden if solely looking at the total impact.

In addition to the results for the independent variables, the analysis outcomes for each year also involve the following spatial coefficients:

- **RHIOAP12L** (SAR model): $\rho = 0{,}46773$ (with $p\text{-}value = 9{,}8212e^{-8}$);
- **RHIOAP13L** (SAR model): $\rho = 0{,}44726$ (with $p\text{-}value = 3{,}5838e^{-7}$);
- **RHIOAP14L** (SAR model): $\rho = 0{,}45664$ (with $p\text{-}value = 1{,}6508e^{-7}$).

The results for every year are gathered from the SAR model, which provides a spatial coefficient ρ of significant importance. In fact, ρ denotes the average influence that factors in a province have on patient immigration for ordinary admissions in all the other provinces in a global manner, through endogenous interactions occurring in the phenomenon itself that affect neighbouring and non-neighbouring provinces through spatial spillovers (e.g. one factor in a province influences the phenomenon there, which influences it in a neighbouring province, which in turn affects it in a province that is close only to the latter); furthermore, these spatial spillovers can return back and influence the phenomenon in the province of origin. As the results show, the coefficient had remained significantly high during that period, apart from slight fluctuations, indicating the continuous occurrence of indirect effects of factors that from a province had globally spilled over the other neighbouring and non-neighbouring provinces in the entire country, in addition to direct influences over the phenomenon in the province of origin.

Returning to the three main tables with the outcomes for the independent variables and considering just the statistically significant results, highlighted by an asterisk, the following statements on their relation to the phenomenon of patient immigration for ordinary admissions can be made:

- **Average duration of an ordinary admission** – In 2012, the direct effect indicates that an increase of 1 day could have reduced the phenomenon by 16,62% in the province of origin and the total effect indicates that an increase of 1 day could have reduced it by 29,22% overall. In 2013, the direct effect indicates that an increase of 1 day could have reduced the phenomenon by 20,12% in the province of origin, the indirect effect indicates that an increase of 1 day could have reduced it by 14,17% in the other provinces and the total effect indicates that an increase of 1 day could have reduced it by 34,29% overall. In 2014, the direct effect indicates that an increase of 1 day could have reduced the phenomenon by 23,77% in the province of origin, the indirect effect indicates that an increase of 1 day could have reduced it by 17,31% in the other provinces and the total effect indicates that an increase of 1 day could have reduced it by 41,08% overall;

- **Rate of medical equipment** – In 2012, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 4,20% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 3,19% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 7,39% overall. In 2013, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 3,99% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 2,81% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 6,80% overall. In 2014, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 3,86% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 2,81% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 6,67% overall;

- **Rate of nurses** – In 2012, the effects were not statistically significant. In 2013, the effects were not statistically significant, but the p-values decreased. In 2014, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 1,95% in the province of origin and the total effect indicates that an increase of 1 unit could have incremented it by 3,36% overall.

4.1.2 Day admissions

Overview

In the same manner used for the previous portion, the first part of the analysis involves procuring information from the data to discern how the phenomenon had been taking effect in the country. First of all, the following table summarises the main information on the data concerning regional patient immigration for day admissions, for each year during the period 2012-2014:

Variable	Minimum	Mean	Maximum
RHIDAP12	0,700	8,673	36,190
RHIDAP13	0,760	8,772	39,500
RHIDAP14	0,780	8,840	39,190

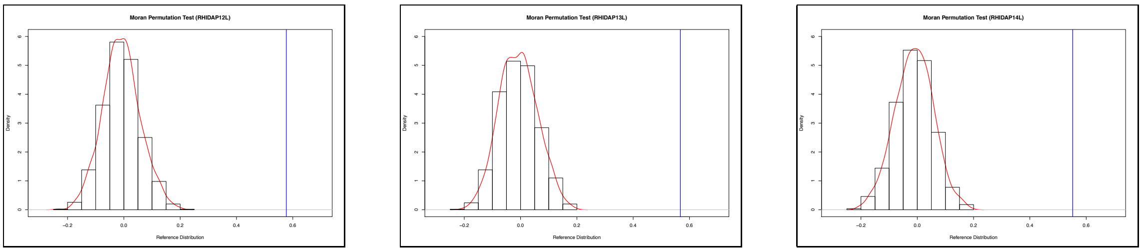
Table 4.16: Summary of regional patient immigration (day admissions) (2012-2014)

The table illustrates that the percentage of patients gathering health treatments for day admissions from a province in a particular region, coming from another region, had featured increases of its minimum and maximum values over time, with a consequently increasing average percentage. Therefore, it can be stated that the occurrence of regional patient immigration for day admissions had incremented during that period on average in the country, making the phenomenon of interest for further research. Employing the log-transformed dependent variables, the Moran's I tests for RHIDAPxxL calculated the following Moran's I values for each year, excluding 3 observations without information in the data:

Variable	Moran's I	p-value
RHIDAP12L	0,577102696	$2{,}2e^{-16}$
RHIDAP13L	0,567573674	$2{,}2e^{-16}$
RHIDAP14L	0,552091204	$2{,}2e^{-16}$

Table 4.17: Moran's I values for RHIDAPxxL (2012-2014)

The following images display various density plots on the reference distribution for the Moran's I values related to each year, which demonstrate how every observed value is statistically significant and quite distant from the expected value $E(I) = \frac{-1}{1-N} = \frac{-1}{1-107} = -0{,}009433962$:



(a) Moran permutation test for RHIDAP12L

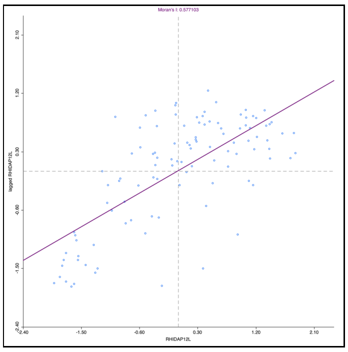
(b) Moran permutation test for RHIDAP13L

(c) Moran permutation test for RHIDAP14L

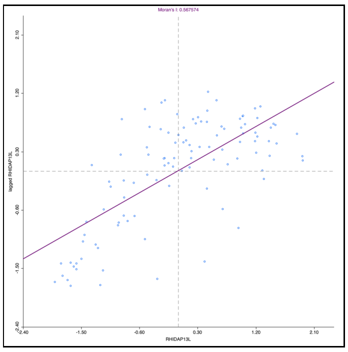
Figure 4.5: Moran permutation tests for RHIDAPxxL (2012-2014)

Taking the low p-values and the significant differences with the expected value into account, it is possible to reject the null hypothesis of absence of spatial autocorrelation and to declare that positive spatial autocorrelation in the data is observed for each year in the period 2012-2014. The underlying meaning is that the phenomenon of patient immigration for day admissions had not been occurring in a random fashion across the country, but rather had tended to be clustered among its various areas, with provinces having high patient immigration percentages being closer to one another and provinces with low patient immigration percentages displaying the same disposition. This result is significant, since it illustrates that the behaviour of patients towards the treatment offers in a province was not independent from that of other patients found in close provinces, violating the assumption of independence of observations in a linear regression model and suggesting the need to conduct some sort of spatial analysis.

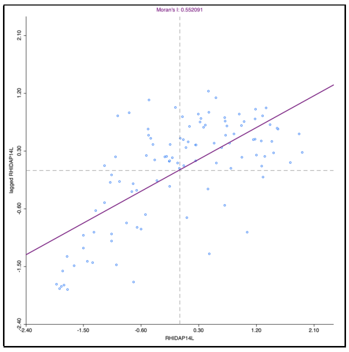
This situation can be more thoroughly discerned with the support of supplementary instruments that communicate further information. For instance, the following Moran scatter plots, obtained from the programme GeoDa, can assist with the identification of the presence and direction of spatial autocorrelation related to the dependent variables of patient immigration for day admissions, for each year in the period 2012-2014:



(a) Moran scatter plot for RHIDAP12L



(b) Moran scatter plot for RHIDAP13L

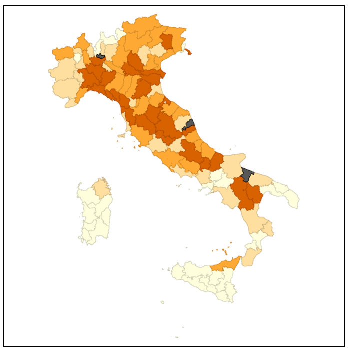


(c) Moran scatter plot for RHIDAP14L

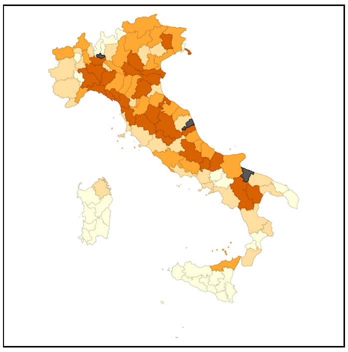
Figure 4.6: Moran scatter plots for RHIDAPxxL (2012-2014)

The Moran scatter plots portray the presence of a positive spatial autocorrelation of the phenomenon in each year between 2012 and 2014, driven by the observations in the lower-left and upper-right quadrants: some provinces with high patient immigration rates had tended to be close to others with high patient immigration rates as well (upper-right quadrant), while some provinces with low patient immigration rates had tended to be near others with low patient immigration rates too (lower-left quadrant). Considering the information from the data, it is possible to declare that the phenomenon had become slightly less clustered from 2012 to 2014, although while retaining a significant number of clusters of provinces with similar patient behaviour.

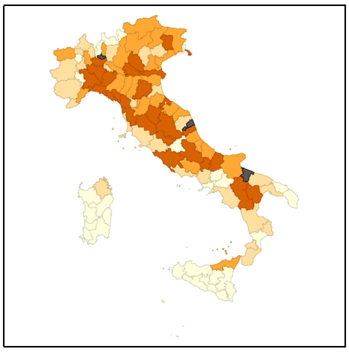
In addition, the following quartile maps depict how the percentage values of patient immigration for day admissions are distributed when grouped into four classes:



(a) Quartile map for RHIDAP12L



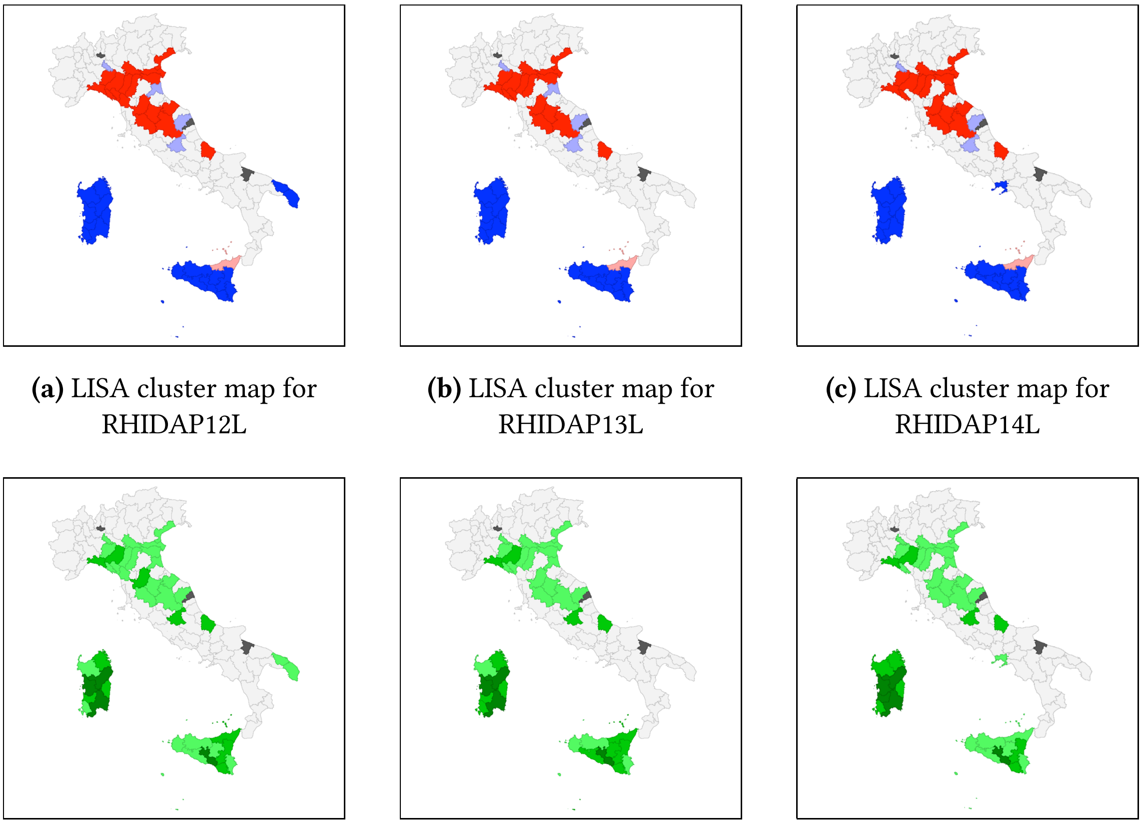
(b) Quartile map for RHIDAP13L



(c) Quartile map for RHIDAP14L

Figure 4.7: Quartile maps for RHIDAPxxL (2012-2014)

The phenomenon of regional patient immigration for day admissions seemed to take place for the most part in provinces of Northern and Central Italy, with some outliers in Southern and Insular Italy. The following LISA cluster maps and LISA significance maps are also employed to further discern the aspects of its occurrence in the country:



(a) LISA cluster map for RHIDAP12L

(b) LISA cluster map for RHIDAP13L

(c) LISA cluster map for RHIDAP14L

(d) LISA significance map for RHIDAP12L

(e) LISA significance map for RHIDAP13L

(f) LISA significance map for RHIDAP14L

Figure 4.8: LISA cluster and significance maps for RHIDAPxxL (2012-2014)

In the LISA cluster maps, a province that is marked with a colour represents the core of a cluster of neighbouring provinces, as defined by the specified weights matrix, which has percentages of patient immigration that are either similar or dissimilar to those of nearby provinces. A province is marked in red if it has a high percentage of patient immigration and is surrounded by neighbouring provinces with a high percentage, while it is marked in blue if it has a low percentage of patient immigration and is surrounded by neighbouring provinces with a low percentage. A light-red province consists of an outlier with a high percentage of patient immigration that is surrounded by neighbouring provinces with a low percentage, while a light-blue province consists of an outlier with a low percentage of patient immigration that is surrounded by neighbouring provinces with a high percentage. All the marked provinces reached statistical significance and their significance levels are mirrored in the LISA significance maps with various degrees below $\alpha = 0{,}05$. For this subtopic, values for three observations are missing as shown by the provinces marked in grey. In this situation, the cluster maps show a concentration of slightly more clusters with high patient immigration percentages around Northern and Central Italy and low patient immigration percentages in Insular Italy, with a lower number of outliers present around them, compared to the previous case.

Analysis framework

The second part of the analysis involves the definition of a specific analysis framework and the illustration of the diverse analysis procedures that depend upon it. In particular, the framework features a multiple linear regression equation and a set of variables that, to allow the data to be examined through various statistical models, are defined for the subtopic in question according to the following specifications (where “xx” corresponds to a specific year in the period 2012-2014):

$$Y_i = \alpha \iota_n + \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 X_{3_i} + \beta_4 X_{4_i} + \beta_5 X_{5_i} + \epsilon_i \quad \text{for } i = 1, \dots, n \tag{4.2}$$

Equation variable	Specific variable
Y	RHIDAPxxL
X_1	BedDARxxC
X_2	AvgDHCLxxC
X_3	MedEqRxxC
X_4	DocDenRxxC
X_5	NursesRxxC

Table 4.18: Specific variables in equation 4.2 for regional patient immigration (day admissions) (2012-2014)

Analysis procedure (2012)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedDAR12C	1,276195	3,369
AvgDHCL12C	1,211212	
MedEqR12C	2,416280	
DocDenR12C	2,622583	
NursesR12C	3,060269	

Table 4.19: VIFs and condition number of the predictors in equation 4.2 (2012)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 16{,}91$ and $p\text{-}value = 4{,}1e^{-12}$) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}339869706$ is significantly diverse from the expected value $E(I) = -0{,}023056909$ ($p\text{-}value = 1{,}783e^{-8}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	39,91	2,659e^{-10}
LMerr	24,513	7,38e^{-7}
RLMlag	15,692	0,00007454
RLMerr	0,29508	0,587
SARMA	40,205	1,86e^{-9}

Table 4.20: Results of the specification tests for equation 4.2 (2012)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	AIC	BIC	Log Likelihood	R^2	LR Test
LM	263,8407	282,5506	-124,9204	0,4287	–
SLX	253,242	285,3159	-114,6210	0,5042	–
SAR	231,5824	252,965	-107,7912	0,633686	–
SEM	239,3305	260,7131	-111,6652	0,6166564	–
SDM	236,3067	271,0534	-105,1533	0,6440394	SAR
SDEM	237,2077	271,9545	-105,6039	0,6418641	–
SARAR	229,3236	253,3791	-105,6618	0,6903272	–

Table 4.21: Measures of goodness of fit for equation 4.2 (2012)

The SAR model has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SEM), a result that aligns with the outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SAR model, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one; although it could be considered as well, the SARAR model is excluded when accounting for the results of the specification tests. The information from the two approaches indicates that the SAR model has the best goodness of fit and should be taken as the source for the results.

Analysis procedure (2013)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedDAR13C	1,180806	3,661
AvgDHCL13C	1,149577	
MedEqR13C	2,268178	
DocDenR13C	3,019336	
NursesR13C	3,414054	

Table 4.22: VIFs and condition number of the predictors in equation 4.2 (2013)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 15{,}88$ and $p\text{-}value = 1{,}615e^{-11}$) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}347850971$ is significantly diverse from the expected value $E(I) = -0{,}023466336$ ($p\text{-}value = 8{,}387e^{-9}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	37,179	1,078e^{-9}
LMerr	25,678	4,034e^{-7}
RLMlag	11,512	0,0006915
RLMerr	0,01114	0,9159
SARMA	37,19	8,4e^{-9}

Table 4.23: Results of the specification tests for equation 4.2 (2013)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	**AIC**	**BIC**	**Log Likelihood**	$\mathbf{R^2}$	**LR Test**
LM	261,2638	279,9736	-123,6319	0,4124	–
SLX	250,9989	283,0729	-113,4995	0,4884	–
SAR	229,6341	251,0167	-106,8170	0,6228858	–
SEM	233,8685	255,2511	-108,9342	0,6204632	–
SDM	232,0595	266,8063	-103,0297	0,6423357	SAR
SDEM	232,9152	267,662	-103,4576	0,642109	SEM
SARAR	230,9572	255,0127	-106,4786	0,642195	SAR

Table 4.24: Measures of goodness of fit for equation 4.2 (2013)

The SAR model has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SEM), a result that aligns with the outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SAR model, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. The information from the two approaches indicates that the SAR model has the best goodness of fit and should be taken as the source for the results.

Analysis procedure (2014)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedDAR14C	1,184219	3,677
AvgDHCL14C	1,120495	
MedEqR14C	2,633239	
DocDenR14C	2,661435	
NursesR14C	3,521040	

Table 4.25: VIFs and condition number of the predictors in equation 4.2 (2014)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 13{,}78$ and $p\text{-}value = 2{,}975e^{-10}$) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}364526636$ is significantly diverse from the expected value $E(I) = -0{,}022538781$ ($p\text{-}value = 2{,}102e^{-9}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	38,266	6,173e^{-10}
LMerr	28,199	1,095e^{-7}
RLMlag	10,07	0,001507
RLMerr	0,0033187	0,9541
SARMA	38,269	4,897e^{-9}

Table 4.26: Results of the specification tests for equation 4.2 (2014)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	**AIC**	**BIC**	**Log Likelihood**	**R^2**	**LR Test**
LM	265,5379	284,2477	-125,7690	0,376	–
SLX	250,3731	282,4471	-113,1866	0,4811	–
SAR	233,583	254,9657	-108,7915	0,6014319	–
SEM	235,5903	256,973	-109,7952	0,6090084	–
SDM	231,6269	266,3737	-102,8135	0,6352724	–
SDEM	233,6317	268,3784	-103,8158	0,6294917	–
SARAR	234,5415	258,5969	-108,2707	0,6311416	SAR / SEM

Table 4.27: Measures of goodness of fit for equation 4.2 (2014)

The SAR model has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SEM), a result that aligns with the outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDM as the most appropriate one and the likelihood ratio test recommends that it should not be reduced to any other model, as the decrease in log likelihood is statistically significant even when accounting for the additional complexity of the model compared to a nested one. The information from the two approaches indicates that the SDM model has the best goodness of fit and should be taken as the source for the results.

Results

The third part of the analysis involves the presentation and explanation of the outcomes resulting from the outlined procedures of data analysis. First of all, to provide them in a clear manner, the following three tables illustrate the results for each considered year in the period 2012-2014, with p-values in parentheses and asterisks indicating which of them are statistically significant:

Variable	Direct impact	Indirect impact	Total impact
BedDAR12C	0.020963348 (0.65613)	0.018727949 (0.6651101)	0.039691297 (0.65619039)
AvgDHCL12C	0.073211229 (0.37590)	0.065404449 (0.4338619)	0.138615678 (0.39549678)
MedEqR12C	0.051562393* (0.00003475)	0.046064107* (0.0074324)	0.097626500* (0.00023356)
DocDenR12C	0.004946617 (0.85440)	0.004419141 (0.8148363)	0.009365758 (0.83248571)
NursesR12C	0.007323766 (0.48505)	0.006542806 (0.5379784)	0.013866573 (0.50361194)

Table 4.28: Impacts in the SAR model for RHIDAP12L (2012)

Variable	Direct impact	Indirect impact	Total impact
BedDAR13C	-0.004618702 (0.95225)	-0.004337478 (0.9891959)	-0.008956180 (0.97088751)
AvgDHCL13C	0.077665874 (0.32494)	0.072936939 (0.3786874)	0.150602812 (0.34159478)
MedEqR13C	0.047098529* (0.000072168)	0.044230785* (0.0073978)	0.091329314* (0.00036361)
DocDenR13C	0.017611153 (0.40696)	0.016538842 (0.4420112)	0.034149995 (0.41776375)
NursesR13C	0.004079423 (0.71533)	0.003831035 (0.7564705)	0.007910458 (0.73343690)

Table 4.29: Impacts in the SAR model for RHIDAP13L (2013)

Variable	Direct impact	Indirect impact	Total impact
BedDAR14C	0.02496807 (0.73810361)	-0.30353934* (0.041891)	-0.2785713 (0.1037109)
AvgDHCL14C	0.04561864 (0.60137343)	0.30626045 (0.167909)	0.3518791 (0.1119946)
MedEqR14C	0.04532690* (0.00025341)	0.05616212 (0.078588)	0.1014890* (0.0062121)
DocDenR14C	0.02793807 (0.23210749)	-0.13787857* (0.018818)	-0.1099405 (0.0908658)
NursesR14C	-0.01340610 (0.27244170)	0.03676750 (0.297869)	0.0233614 (0.5717126)

Table 4.30: Impacts in the SDM for RHIDAP14L (2014)

Since the outcomes have been retrieved from spatial models, the procedures of data analysis generated various types of effect concerning the independent variables that are represented by three types of impact. With regards to this particular subtopic of patient immigration for day admissions, the impacts can be defined as follows:

- **Direct impact**: it measures the average effect that a factor in a province has on patient immigration for day admissions in the same province;
- **Indirect impact**: it measures the average effect that a factor in a province has on patient immigration for day admissions in the other provinces, in a direct manner or through its influence on the phenomenon in the same province;
- **Total impact**: it measures the average effect that a factor in a province has on patient immigration for day admissions in all provinces in a global fashion, by merging the direct and indirect impacts.

Establishing a distinction between these effects permits to see whether the various impacts differ in terms of statistical significance (e.g. the direct or indirect impact may be statistically significant, while the total may not) and to evaluate the strengths of the direct and indirect impacts, which may be hidden if solely looking at the total impact.

In addition to the results for the independent variables, the analysis outcomes for each year also involve the following spatial coefficients:

- **RHIDAP12L** (SAR model): $\rho = 0{,}51431$ (with $p\text{-}value = 4.826e^{-9}$);
- **RHIDAP13L** (SAR model): $\rho = 0{,}52856$ (with $p\text{-}value = 6{,}6667e^{-9}$);
- **RHIDAP14L** (SDM): $\rho = 0{,}46205$ (with $p\text{-}value = 5{,}2436e^{-6}$), $\theta_1 = -0{,}200427$ (with $p\text{-}value = 0{,}035297$) (spatial lag of BedDAR14C), $\theta_4 = -0{,}098710$ (with $p\text{-}value = 0{,}006463$) (spatial lag of DocDenR14C).

The results for the years 2012 and 2013 are gathered from the SAR model, which provides a spatial coefficient ρ, while those for the year 2014 are taken from the SDM, which produces various spatial coefficients ρ and θ for the independent variables, all of significant importance. In fact, ρ denotes the average influence that factors in a province have on patient immigration for day admissions in all the other provinces in a global manner, through endogenous interactions occurring in the phenomenon itself that affect neighbouring and non-neighbouring provinces through spatial spillovers (e.g. one factor in a province influences the phenomenon there, which influences it in a neighbouring province, which in turn affects it in a province that is close only to the latter); furthermore, these spatial spillovers can return back and influence the phenomenon in the province of origin. In addition, θ denotes the effect that a factor in a province directly produces on the phenomenon in another province neighbouring it as defined by the weights matrix, without passing through an influence on the phenomenon in the province of origin. As the results show, the coefficient ρ had remained significantly high during that period, although it decreased in 2014, indicating the continuous occurrence of indirect effects of factors that from a province had globally spilled over the other neighbouring and non-neighbouring provinces in the entire country, in addition to direct influences over the phenomenon in the province of origin. The coefficients θ also show an effect of the rate of beds for day admission and the rate of doctors and dentists in a province on the phenomenon in nearby provinces, as defined by the weights matrix.

Returning to the three main tables with the outcomes for the independent variables and considering just the statistically significant results, highlighted by an asterisk, the following statements on their relation to the phenomenon of patient immigration for day admissions can be made:

- **Rate of beds for day admissions** – In 2012, the effects were not statistically significant. In 2013, the effects were not statistically significant and the p-values increased. In 2014, following a decrease of the p-values, the indirect effect indicates that an increase of 1 unit could have reduced the phenomenon by 30,35% in the other provinces;

- **Rate of medical equipment** – In 2012, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 5,16% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 4,61% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 9,76% overall. In 2013, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 4,71% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 4,42% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 9,13% overall. In 2014, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 4,53% in the province of origin and the total effect indicates that an increase of 1 unit could have incremented it by 10,15% overall;

- **Rate of doctors and dentists** – In 2012, the effects were not statistically significant. In 2013, the effects were not statistically significant, but the p-values decreased. In 2014, following another decrease of the p-values, the indirect effect indicates that an increase of 1 unit could have reduced the phenomenon by 13,79% in the other provinces.

4.2 Regional patient emigration

4.2.1 Ordinary admissions

Overview

The first part of the analysis involves obtaining information from the data to understand how the phenomenon had been occurring in the country. First of all, the following table summarises the main information on the data regarding regional patient emigration for ordinary admissions, for each year during the period 2012-2014:

Variable	Minimum	Mean	Maximum
RHEOAP12	1,860	9,224	28,230
RHEOAP13	1,790	9,330	29,300
RHEOAP14	1,940	9,406	27,260

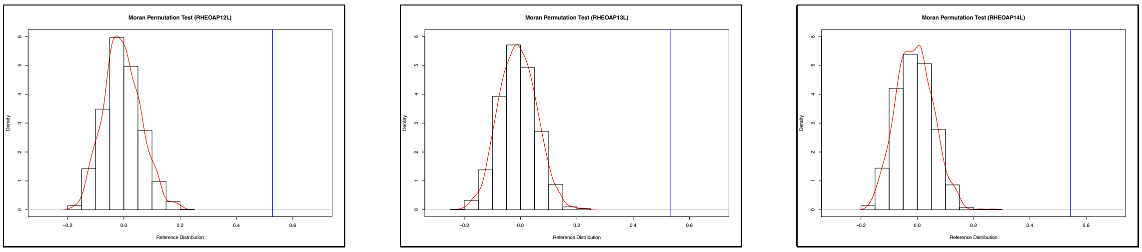
Table 4.31: Summary of regional patient emigration (ordinary admissions) (2012-2014)

The table portrays that the percentage of patients going from a province in a region to another region to attain health treatments for ordinary admissions had decreased in certain areas and increased in others over time, with reduced differences in 2014 but still an overall raising average percentage. Therefore, it can be asserted that the occurrence of regional patient emigration for ordinary admissions had incremented during that period on average in the country, making the phenomenon of interest for additional research. Employing the log-transformed dependent variables, the Moran's I tests for RHEOAPxxL calculated the following Moran's I values for each year:

Variable	Moran's I	p-value
RHEOAP12L	0,527955242	$2{,}269e^{-16}$
RHEOAP13L	0,533675936	$2{,}2e^{-16}$
RHEOAP14L	0,544158307	$2{,}2e^{-16}$

Table 4.32: Moran's I values for RHEOAPxxL (2012-2014)

The following images display various density plots on the reference distribution for the Moran's I values related to each year, which highlight how every observed value is statistically significant and quite distant from the expected value $E(I) = \frac{-1}{1-N} = \frac{-1}{1-110} = -0{,}009174312$:



(a) Moran permutation test for RHEOAP12L

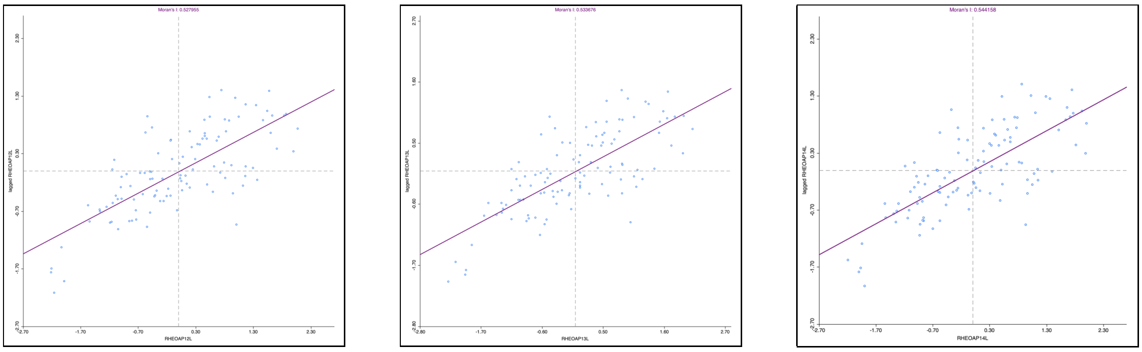
(b) Moran permutation test for RHEOAP13L

(c) Moran permutation test for RHEOAP14L

Figure 4.9: Moran permutation tests for RHEOAPxxL (2012-2014)

Taking the low p-values and the significant differences with the expected value into account, it is possible to reject the null hypothesis of absence of spatial autocorrelation and to declare that positive spatial autocorrelation in the data is observed for each year in the period 2012-2014. The underlying meaning is that the phenomenon of patient emigration for ordinary admissions had not been occurring in a random fashion across the country, but rather had tended to be clustered among its various areas, with provinces having high patient emigration percentages being closer to one another and provinces with low patient emigration percentages displaying the same disposition. This result is significant, since it illustrates that the behaviour of patients towards the treatment offers in a province was not independent from that of other patients found in close provinces, violating the assumption of independence of observations in a linear regression model and suggesting the need to conduct some sort of spatial analysis.

This situation can be more thoroughly discerned with the support of supplementary instruments that communicate further information. For instance, the following Moran scatter plots, obtained from the programme GeoDa, can assist with the identification of the presence and direction of spatial autocorrelation related to the dependent variables of patient emigration for ordinary admissions, for each year in the period 2012-2014:



(a) Moran scatter plot for RHEOAP12L

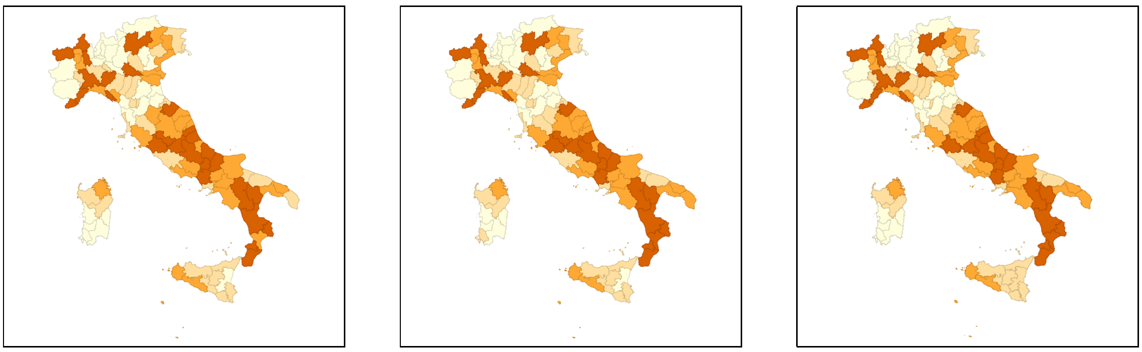
(b) Moran scatter plot for RHEOAP13L

(c) Moran scatter plot for RHEOAP14L

Figure 4.10: Moran scatter plots for RHEOAPxxL (2012-2014)

The Moran scatter plots portray the presence of a positive spatial autocorrelation of the phenomenon in each year between 2012 and 2014, driven by the observations in the lower-left and upper-right quadrants: some provinces with high patient emigration rates had tended to be close to others with high patient emigration rates as well (upper-right quadrant), while some provinces with low patient emigration rates had tended to be near others with low patient emigration rates too (lower-left quadrant). Considering the information from the data, it is possible to state that the phenomenon had become slightly more clustered from 2012 to 2014, highlighting a greater presence of clusters of provinces with similar patient behaviour.

In addition, the following quartile maps depict how the percentage values of patient emigration for ordinary admissions are distributed when grouped into four classes:



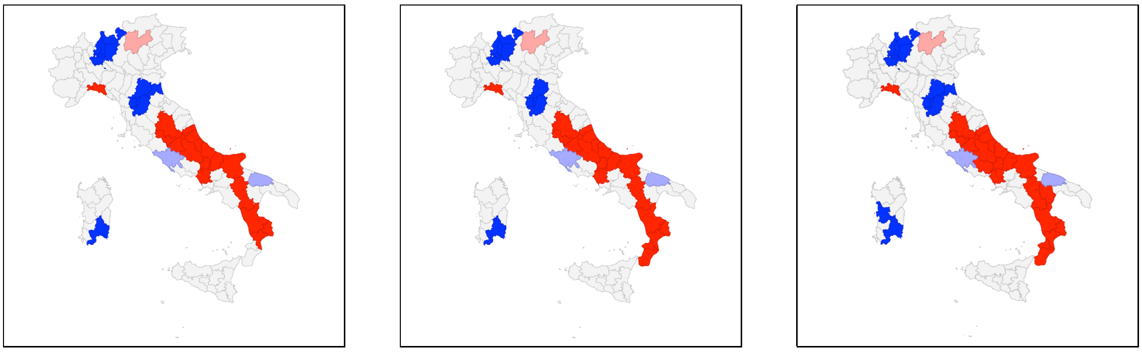
(a) Quartile map for RHEOAP12L

(b) Quartile map for RHEOAP13L

(c) Quartile map for RHEOAP14L

Figure 4.11: Quartile maps for RHEOAPxxL (2012-2014)

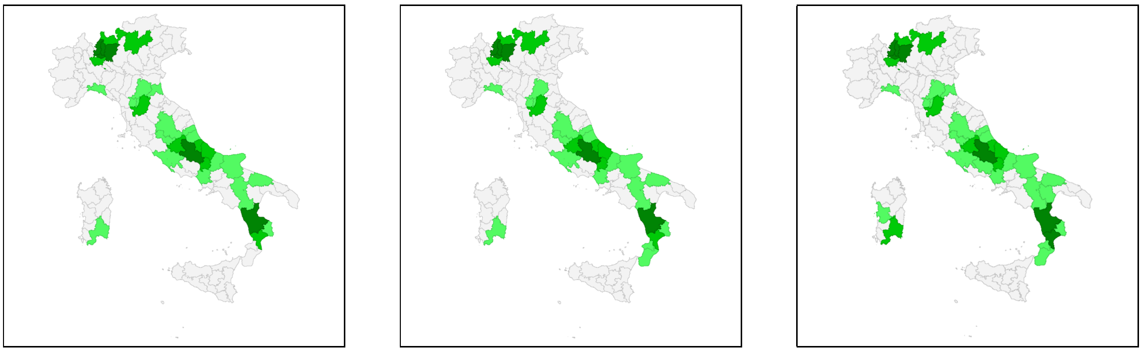
The phenomenon of regional patient emigration for ordinary admissions seemed to occur mainly in provinces of Central and Southern Italy, with a few outliers in Northern Italy. The following LISA cluster maps and LISA significance maps are also employed to further discern the aspects of its occurrence in the country:



(a) LISA cluster map for RHEOAP12L

(b) LISA cluster map for RHEOAP13L

(c) LISA cluster map for RHEOAP14L



(d) LISA significance map for RHEOAP12L

(e) LISA significance map for RHEOAP13L

(f) LISA significance map for RHEOAP14L

Figure 4.12: LISA cluster and significance maps for RHEOAPxxL (2012-2014)

In the LISA cluster maps, a province that is marked with a colour represents the core of a cluster of neighbouring provinces, as defined by the specified weights matrix, which has percentages of patient emigration that are either similar or dissimilar to those of nearby provinces. A province is marked in red if it has a high percentage of patient emigration and is surrounded by neighbouring provinces with a high percentage, while it is marked in blue if it has a low percentage of patient emigration and is surrounded by neighbouring provinces with a low percentage. A light-red province consists of an outlier with a high percentage of patient emigration that is surrounded by neighbouring provinces with a low percentage, while a light-blue province consists of an outlier with a low percentage of patient emigration that is surrounded by neighbouring provinces with a high percentage. All the marked provinces reached statistical significance and their significance levels are mirrored in the LISA significance maps with various degrees below $\alpha = 0{,}05$. For this subtopic, values are present for all the observations and thus no province is marked in grey. In this situation, the cluster maps illustrate a concentration of clusters with high patient emigration percentages around Central and Southern Italy and low patient emigration percentages in Northern Italy and the island of Sardegna, with an overall low number of outliers.

Analysis framework

The second part of the analysis involves the definition of a specific analysis framework and the illustration of the diverse analysis procedures that depend upon it. In particular, the framework features a multiple linear regression equation and a set of variables that, to allow the data to be examined through various statistical models, are defined for the subtopic in question according to the following specifications (where "xx" corresponds to a specific year in the period 2012-2014):

$$Y_i = \alpha \iota_n + \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 X_{3_i} + \beta_4 X_{4_i} + \beta_5 X_{5_i} + \epsilon_i \qquad \text{for } i = 1, \ldots, n \tag{4.3}$$

Equation variable	**Specific variable**
Y	RHEOAPxxL
X_1	BedOARxxC
X_2	AvgOHDxxC
X_3	MedEqRxxC
X_4	DocDenRxxC
X_5	NursesRxxC

Table 4.33: Specific variables in equation 4.3 for regional patient emigration (ordinary admissions) (2012-2014)

Analysis procedure (2012)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedOAR12C	4,225806	4,785
AvgOHD12C	1,178224	
MedEqR12C	2,418608	
DocDenR12C	3,632225	
NursesR12C	4,481831	

Table 4.34: VIFs and condition number of the predictors in equation 4.3 (2012)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic (F = 2,863 and $p\text{-}value$ = 0,01835) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}5254127$ is significantly diverse from the expected value $E(I) = -0{,}0182972$ ($p\text{-}value = 2{,}2e^{-16}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	66,498	$3{,}331\text{e}^{-16}$
LMerr	60,963	$5{,}773\text{e}^{-15}$
RLMlag	5,675	0,01721
RLMerr	0,1402	0,7081
SARMA	66,638	$3{,}331\text{e}^{-15}$

Table 4.35: Results of the specification tests for equation 4.3 (2012)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	**AIC**	**BIC**	**Log Likelihood**	$\mathbf{R^2}$	**LR Test**
LM	210,9863	229,8897	-98,49316	0,07873	–
SLX	211,2354	243,6412	-93,61771	0,1143	–
SAR	150,8698	172,4736	-67,43489	0,5798283	–
SEM	152,6238	174,2277	-68,31192	0,576837	–
SDM	158,6619	193,7681	-66,33093	0,5844861	SAR / SEM
SDEM	158,667	193,7732	-66,33350	0,5938929	SEM
SARAR	151,6061	175,9105	-66,80306	0,5625552	SAR / SEM

Table 4.36: Measures of goodness of fit for equation 4.3 (2012)

The SAR model has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SEM), a result that aligns with the outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SAR model or SEM, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. The information from the two approaches indicates that the SAR model has the best goodness of fit and should be taken as the source for the results.

Analysis procedure (2013)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedOAR13C	3,981335	4,649
AvgOHD13C	1,225448	
MedEqR13C	2,335382	
DocDenR13C	4,371061	
NursesR13C	4,311720	

Table 4.37: VIFs and condition number of the predictors in equation 4.3 (2013)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic (F = 2,561 and $p\text{-}value$ = 0,03155) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}561740274$ is significantly diverse from the expected value $E(I) = -0{,}019565575$ ($p\text{-}value = 2{,}2e^{-16}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	71,108	2,2e^{-16}
LMerr	69,685	2,2e^{-16}
RLMlag	1,842	0,1747
RLMerr	0,419	0,5174
SARMA	71,527	3,331e^{-16}

Table 4.38: Results of the specification tests for equation 4.3 (2013)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant; even though their robust forms are not, the LMlag test has a higher value and its robust version has a lower p-value, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	AIC	BIC	Log Likelihood	R^2	LR Test
LM	212,5315	231,4349	-99,26575	0,06682	–
SLX	216,7924	249,1981	-96,39618	0,06953	–
SAR	149,3005	170,9044	-66,65027	0,5876633	–
SEM	149,2556	170,8594	-66,62778	0,591999	–
SDM	157,047	192,1532	-65,52349	0,5970916	SEM / SAR
SDEM	156,6134	191,7196	-65,30669	0,6081584	SEM
SARAR	149,3894	173,6938	-65,69472	0,5691427	SEM / SAR

Table 4.39: Measures of goodness of fit for equation 4.3 (2013)

The SAR model and SEM have a similar goodness of fit for the data that is better than that of the linear model and the other that considers a single spatial effect (SLX), a result that aligns with the uncertain outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDEM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SEM, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. Given the similarities between the SAR model and SEM, the results of the specification tests and the literature advice on preferring the spatial effects in the dependent variable instead of those in the error term, the SAR model should be taken as the source for the results.

Analysis procedure (2014)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedOAR14C	3,040995	4,451
AvgOHD14C	1,203415	
MedEqR14C	2,725743	
DocDenR14C	3,275205	
NursesR14C	4,519485	

Table 4.40: VIFs and condition number of the predictors in equation 4.3 (2014)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic (F = 4,086 and $p\text{-}value$ = 0,001993) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value I = 0,53444915 is significantly diverse from the expected value $E(I) = -0{,}01914007$ ($p\text{-}value = 2{,}2e^{-16}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	70,901	$2{,}2\text{e}^{-16}$
LMerr	63,078	$1{,}998\text{e}^{-15}$
RLMlag	7,942	0,00483
RLMerr	0,11919	0,7299
SARMA	71,02	$3{,}331\text{e}^{-16}$

Table 4.41: Results of the specification tests for equation 4.3 (2014)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	AIC	BIC	Log Likelihood	R^2	LR Test
LM	204,2251	223,1285	-95,11256	0,124	–
SLX	203,5741	235,9799	-89,78707	0,1647	–
SAR	142,4569	164,0608	-63,22846	0,602906	–
SEM	146,1541	167,7579	-65,07704	0,5939369	–
SDM	152,0267	187,1329	-63,01334	0,6025141	SAR / SEM
SDEM	152,4202	187,5264	-63,21009	0,610998	SEM
SARAR	144,1349	168,4392	-63,06744	0,5927438	SAR

Table 4.42: Measures of goodness of fit for equation 4.3 (2014)

The SAR model has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SEM), a result that aligns with the outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SAR model or SEM, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. The information from the two approaches indicates that the SAR model has the best goodness of fit and should be taken as the source for the results.

Results

The third part of the analysis involves the presentation and explanation of the outcomes resulting from the outlined procedures of data analysis. First of all, to provide them in a clear manner, the following three tables illustrate the results for each considered year in the period 2012-2014, with p-values in parentheses and asterisks indicating which of them are statistically significant:

Variable	Direct impact	Indirect impact	Total impact
BedOAR12C	-0.032728361* (0.0382495)	-0.06297222 (0.100446)	-0.09570058 (0.068746)
AvgOHD12C	0.137562685* (0.0227332)	0.26468258 (0.077608)	0.40224526* (0.048593)
MedEqR12C	0.024770610* (0.0048488)	0.04766081* (0.041308)	0.07243142* (0.019325)
DocDenR12C	0.008470388 (0.6146643)	0.01629776 (0.651263)	0.02476815 (0.636146)
NursesR12C	-0.016095227 (0.0884557)	-0.03096862 (0.147475)	-0.04706384 (0.117774)

Table 4.43: Impacts in the SAR model for RHEOAP12L (2012)

Variable	Direct impact	Indirect impact	Total impact
BedOAR13C	-0.050787473* (0.0018148)	-0.09978046* (0.027326)	-0.15056794* (0.010819)
AvgOHD13C	0.138952157* (0.0470484)	0.27299469 (0.101268)	0.41194685 (0.074770)
MedEqR13C	0.021506810* (0.0137815)	0.04225372* (0.041107)	0.06376053* (0.024921)
DocDenR13C	0.016444597 (0.3915538)	0.03230815 (0.448132)	0.04875275 (0.423500)
NursesR13C	-0.008024112 (0.3467495)	-0.01576471 (0.390140)	-0.02378882 (0.370080)

Table 4.44: Impacts in the SAR model for RHEOAP13L (2013)

Variable	Direct impact	Indirect impact	Total impact
BedOAR14C	-0.055715792* (0.000063821)	-0.101319404* (0.021031)	-0.15703520* (0.0045221)
AvgOHD14C	0.113728322 (0.0747726)	0.206815435 (0.135576)	0.32054376 (0.1050338)
MedEqR14C	0.022140836* (0.0067493)	0.040263204* (0.042919)	0.06240404* (0.0205641)
DocDenR14C	0.003806299 (0.8145630)	0.006921771 (0.834458)	0.01072807 (0.8264625)
NursesR14C	-0.003936929 (0.6790318)	-0.007159321 (0.697946)	-0.01109625 (0.6895355)

Table 4.45: Impacts in the SAR model for RHEOAP14L (2014)

Since the outcomes have been retrieved from spatial models, the procedures of data analysis generated various types of effect concerning the independent variables that are represented by three types of impact. With regards to this particular subtopic of patient emigration for ordinary admissions, the impacts can be defined as follows:

- **Direct impact**: it measures the average effect that a factor in a province has on patient emigration for ordinary admissions in the same province;
- **Indirect impact**: it measures the average effect that a factor in a province has on patient emigration for ordinary admissions in the other provinces, in a direct manner or through its influence on the phenomenon in the same province;
- **Total impact**: it measures the average effect that a factor in a province has on patient emigration for ordinary admissions in all provinces in a global fashion, by merging the direct and indirect impacts.

Establishing a distinction between these effects permits to see whether the various impacts differ in terms of statistical significance (e.g. the direct or indirect impact may be statistically significant, while the total may not) and to evaluate the strengths of the direct and indirect impacts, which may be hidden if solely looking at the total impact.

In addition to the results for the independent variables, the analysis outcomes for each year also involve the following spatial coefficients:

- **RHEOAP12L** (SAR model): $\rho = 0{,}72313$ (with $p\text{-}value = 3{,}2196e^{-15}$);
- **RHEOAP13L** (SAR model): $\rho = 0{,}72824$ (with $p\text{-}value = 6{,}6613e^{-16}$);
- **RHEOAP14L** (SAR model): $\rho = 0{,}70909$ (with $p\text{-}value = 1{,}4433e^{-15}$).

The results for every year are gathered from the SAR model, which provides a spatial coefficient ρ of significant importance. In fact, ρ denotes the average influence that factors in a province have on patient emigration for ordinary admissions in all the other provinces in a global manner, through endogenous interactions occurring in the phenomenon itself that affect neighbouring and non-neighbouring provinces through spatial spillovers (e.g. one factor in a province influences the phenomenon there, which influences it in a neighbouring province, which in turn affects it in a province that is close only to the latter); furthermore, these spatial spillovers can return back and influence the phenomenon in the province of origin. As the results show, the coefficient had remained significantly high during that period, apart from slight fluctuations, indicating the continuous occurrence of indirect effects of factors that from a province had globally spilled over the other neighbouring and non-neighbouring provinces in the entire country, in addition to direct influences over the phenomenon in the province of origin.

Returning to the three main tables with the outcomes for the independent variables and considering just the statistically significant results, highlighted by an asterisk, the following statements on their relation to the phenomenon of patient emigration for ordinary admissions can be made:

- **Rate of beds for ordinary admissions** – In 2012, the direct effect indicates that an increase of 1 unit could have reduced the phenomenon by 3,27% in the province of origin. In 2013, the direct effect indicates that an increase of 1 unit could have reduced the phenomenon by 5,08% in the province of origin, the indirect effect indicates that an increase of 1 unit could have reduced it by 9,98% in the other provinces and the total effect indicates that an increase of 1 unit could have reduced it by 15,06% overall. In 2014, the direct effect indicates that an increase of 1 unit could have reduced the phenomenon by 5,57% in the province of origin, the indirect effect indicates that an increase of 1 unit could have reduced it by 10,13% in the other provinces and the total effect indicates that an increase of 1 unit could have reduced it by 15,70% overall;

- **Average duration of an ordinary admission** – In 2012, the direct effect indicates that an increase of 1 day could have incremented the phenomenon by 13,76% in the province of origin and the total effect indicates that an increase of 1 day could have incremented it by 40,22% overall. In 2013, the direct effect indicates that an increase of 1 day could have incremented the phenomenon by 13,90% in the province of origin. In 2014, the effects were not statistically significant;

- **Rate of medical equipment** – In 2012, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 2,48% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 4,77% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 7,24% overall. In 2013, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 2,15% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 4,23% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 6,37% overall. In 2014, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 2,21% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 4,03% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 6,24% overall.

4.2.2 Day admissions

Overview

In the same manner employed for the previous segment, the first part of the analysis involves retrieving information from the data to comprehend how the phenomenon had been taking place in the country. First of all, the following table summarises the main information on the data concerning regional patient emigration for day admissions, for each year during the period 2012-2014:

Variable	Minimum	Mean	Maximum
RHEDAP12	2,070	10,328	36,050
RHEDAP13	2,610	10,531	37,320
RHEDAP14	2,530	10,903	37,410

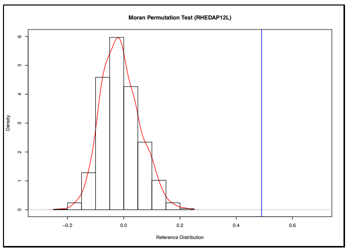
Table 4.46: Summary of regional patient emigration (day admissions) (2012-2014)

The table depicts that the percentage of patients going from a province in a region to another region to gather health treatments for day admissions had featured increases of its minimum and maximum numbers over time, with a reduction of differences but still a consequently raising average percentage. Therefore, it can be said that the occurrence of regional patient emigration for day admissions had incremented during that period on average in the country, induced from the increase of the minimum percentage for the most part, making the phenomenon of interest for further research. Employing the log-transformed dependent variables, the Moran's I tests for RHEDAPxxL calculated the following Moran's I values for each year:

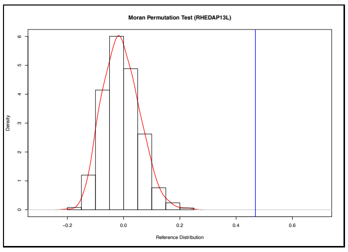
Variable	Moran's I	p-value
RHEDAP12L	0,489947696	2,349e^{-14}
RHEDAP13L	0,468473032	2,713e^{-13}
RHEDAP14L	0,460966893	6,016e^{-13}

Table 4.47: Moran's I values for RHEDAPxxL (2012-2014)

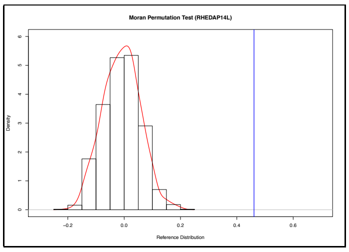
The following images display various density plots on the reference distribution for the Moran's I values related to each year, which delineate how every observed value is statistically significant and quite distant from the expected value $E(I) = \frac{-1}{1-N} = \frac{-1}{1-110} = -0{,}009174312$:



(a) Moran permutation test for RHEDAP12L



(b) Moran permutation test for RHEDAP13L

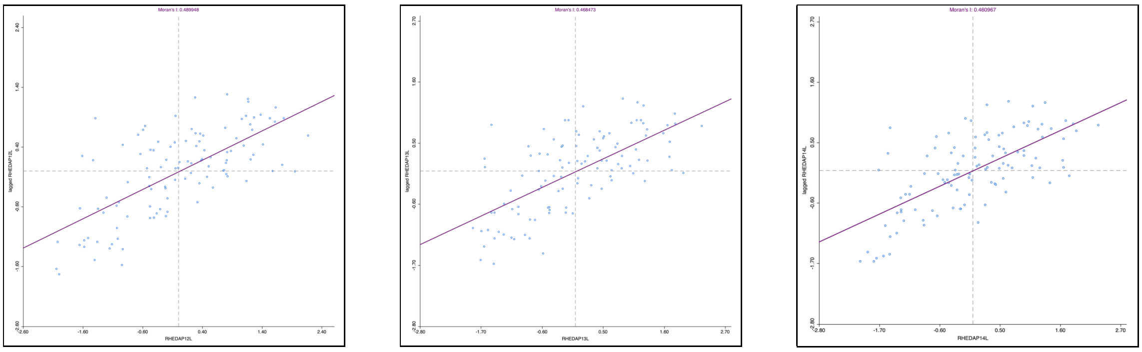


(c) Moran permutation test for RHEDAP14L

Figure 4.13: Moran permutation tests for RHEDAPxxL (2012-2014)

Taking the low p-values and the significant differences with the expected value into account, it is possible to reject the null hypothesis of absence of spatial autocorrelation and to declare that positive spatial autocorrelation in the data is observed for each year in the period 2012-2014. The underlying meaning is that the phenomenon of patient emigration for day admissions had not been occurring in a random fashion across the country, but rather had tended to be clustered among its various areas, with provinces having high patient emigration percentages being closer to one another and provinces with low patient emigration percentages displaying the same disposition. This result is significant, since it illustrates that the behaviour of patients towards the treatment offers in a province was not independent from that of other patients found in close provinces, violating the assumption of independence of observations in a linear regression model and suggesting the need to conduct some sort of spatial analysis.

This situation can be more thoroughly discerned with the support of supplementary instruments that communicate further information. For instance, the following Moran scatter plots, obtained from the programme GeoDa, can assist with the identification of the presence and direction of spatial autocorrelation related to the dependent variables of patient emigration for day admissions, for each year in the period 2012-2014:



(a) Moran scatter plot for RHEDAP12L

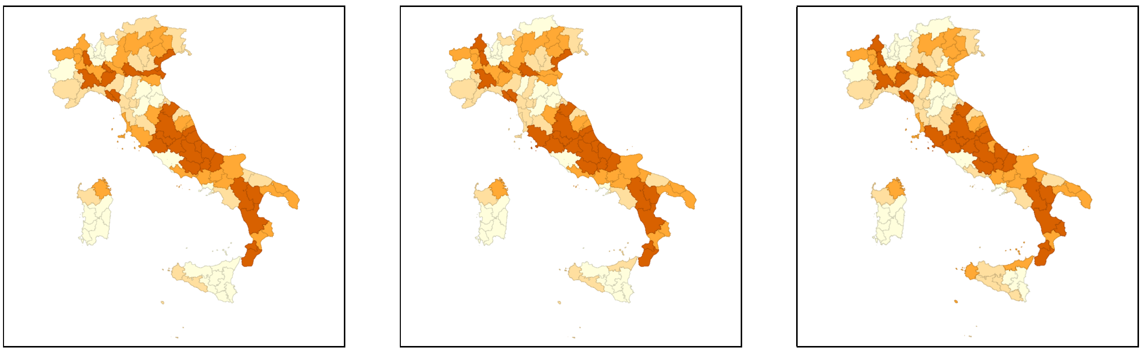
(b) Moran scatter plot for RHEDAP13L

(c) Moran scatter plot for RHEDAP14L

Figure 4.14: Moran scatter plots for RHEDAPxxL (2012-2014)

The Moran scatter plots portray the presence of a positive spatial autocorrelation of the phenomenon in each year between 2012 and 2014, driven by the observations in the lower-left and upper-right quadrants: some provinces with high patient emigration rates had tended to be close to others with high patient emigration rates as well (upper-right quadrant), while some provinces with low patient emigration rates had tended to be near others with low patient emigration rates too (lower-left quadrant). Considering the information from the data, it is possible to affirm that the phenomenon had become slightly less clustered from 2012 to 2014, although while retaining a significant number of clusters of provinces with similar patient behaviour.

In addition, the following quartile maps depict how the percentage values of patient emigration for day admissions are distributed when grouped into four classes:



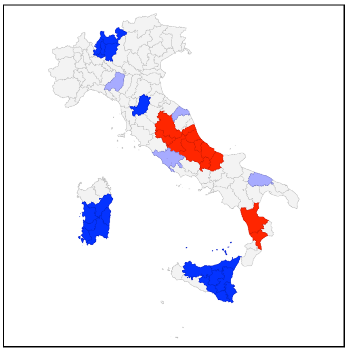
(a) Quartile map for RHEDAP12L

(b) Quartile map for RHEDAP13L

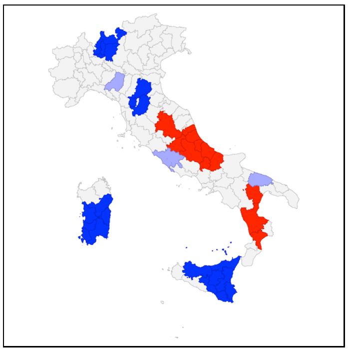
(c) Quartile map for RHEDAP14L

Figure 4.15: Quartile maps for RHEDAPxxL (2012-2014)

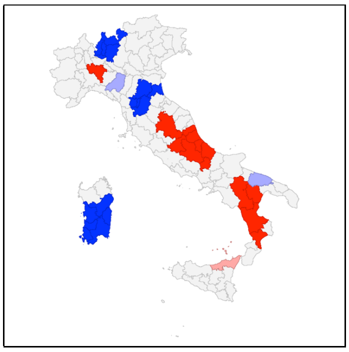
The phenomenon of regional patient emigration for day admissions seemed to take effect for the most part in provinces of Central and Southern Italy, with a few outliers in Northern Italy. The following LISA cluster maps and LISA significance maps are also employed to further discern the aspects of its occurrence in the country:



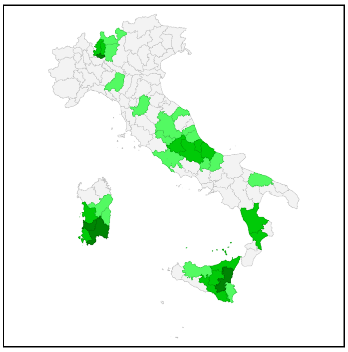
(a) LISA cluster map for RHEDAP12L



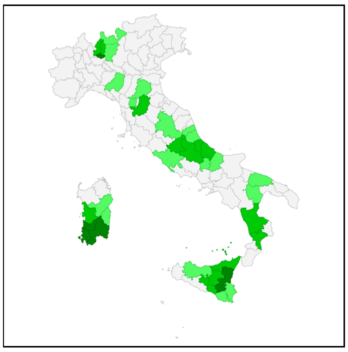
(b) LISA cluster map for RHEDAP13L



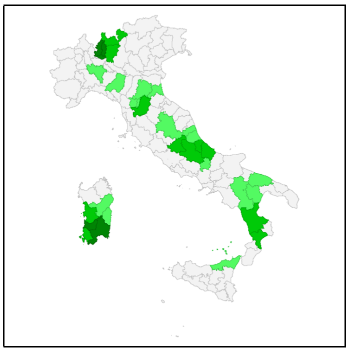
(c) LISA cluster map for RHEDAP14L



(d) LISA significance map for RHEDAP12L



(e) LISA significance map for RHEDAP13L



(f) LISA significance map for RHEDAP14L

Figure 4.16: LISA cluster and significance maps for RHEDAPxxL (2012-2014)

In the LISA cluster maps, a province that is marked with a colour represents the core of a cluster of neighbouring provinces, as defined by the specified weights matrix, which has percentages of patient emigration that are either similar or dissimilar to those of nearby provinces. A province is marked in red if it has a high percentage of patient emigration and is surrounded by neighbouring provinces with a high percentage, while it is marked in blue if it has a low percentage of patient emigration and is surrounded by neighbouring provinces with a low percentage. A light-red province consists of an outlier with a high percentage of patient emigration that is surrounded by neighbouring provinces with a low percentage, while a light-blue province consists of an outlier with a low percentage of patient emigration that is surrounded by neighbouring provinces with a high percentage. All the marked provinces reached statistical significance and their significance levels are mirrored in the LISA significance maps with various degrees below $\alpha = 0{,}05$. For this subtopic, values are present for all the observations and thus no province is marked in grey. In this situation, the cluster maps illustrate a concentration of clusters with high patient emigration percentages around Central and southern Italy and low patient emigration percentages in Northern and Insular Italy, with an overall low number of outliers.

Analysis framework

The second part of the analysis involves the definition of a specific analysis framework and the illustration of the diverse analysis procedures that depend upon it. In particular, the framework features a multiple linear regression equation and a set of variables that, to allow the data to be examined through various statistical models, are defined for the subtopic in question according to the following specifications (where "xx" corresponds to a specific year in the period 2012-2014):

$$Y_i = \alpha \iota_n + \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 X_{3_i} + \beta_4 X_{4_i} + \beta_5 X_{5_i} + \epsilon_i \qquad \text{for } i = 1, \dots, n \tag{4.4}$$

Equation variable	**Specific variable**
Y	RHEDAPxxL
X_1	BedDARxxC
X_2	AvgDHCLxxC
X_3	MedEqRxxC
X_4	DocDenRxxC
X_5	NursesRxxC

Table 4.48: Specific variables in equation 4.4 for regional patient emigration (day admissions) (2012-2014)

Analysis procedure (2012)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedDAR12C	1,295610	3,462
AvgDHCL12C	1,218713	
MedEqR12C	2,468500	
DocDenR12C	2,739882	
NursesR12C	3,182269	

Table 4.49: VIFs and condition number of the predictors in equation 4.4 (2012)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 3{,}103$ and $p\text{-}value = 0{,}01189$) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}42839934$ is significantly diverse from the expected value $E(I) = -0{,}02275153$ ($p\text{-}value = 1{,}421e^{-12}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	47,806	$4{,}704e^{-12}$
LMerr	40,529	$1{,}937e^{-10}$
RLMlag	8,0328	0,004594
RLMerr	0,75515	0,3849
SARMA	48,562	$2{,}851e^{-11}$

Table 4.50: Results of the specification tests for equation 4.4 (2012)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	AIC	BIC	Log Likelihood	R^2	LR Test
LM	224.4892	243.3926	-105.24461	0.08797	–
SLX	220.5342	252.94	-98.26711	0.1561	–
SAR	182.4512	204.0551	-83.22561	0.4882587	–
SEM	183.5673	205.1712	-83.78367	0.4938972	–
SDM	183.5568	218.663	-78.77840	0.5224921	SAR / SEM
SDEM	181.6064	216.7126	-77.80318	0.5369856	–
SARAR	184.395	208.6993	-83.19748	0.4801871	SAR / SEM

Table 4.51: Measures of goodness of fit for equation 4.4 (2012)

The SAR model has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SEM), a result that aligns with the outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDEM as the most appropriate one and the likelihood ratio test recommends that it should not be reduced to any other model, as the decrease in log likelihood is statistically significant. Given the outcomes of the specification tests and the literature advice on preferring the spatial effects in the dependent variable instead of those in the error term, as well as to avoid the risk of overfitting the data with a higher number of variables that are present in the SDEM, as indicated by the BIC, as a cautious choice the SAR model should be taken as the source for the results.

Analysis procedure (2013)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	**VIF**	**Condition number**
BedDAR13C	1,197818	3,746
AvgDHCL13C	1,152183	
MedEqR13C	2,307904	
DocDenR13C	3,126998	
NursesR13C	3,523863	

Table 4.52: VIFs and condition number of the predictors in equation 4.4 (2013)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 2{,}532$ and $p\text{-}value = 0{,}03325$) indicate that the model fits the data better than an intercept-only model without independent variables.

Before taking the model as valid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}42918109$ is significantly diverse from the expected value $E(I) = -0{,}02277746$ ($p\text{-}value = 1{,}226e^{-12}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	**Value**	**p-value**
LMlag	42,958	5,593e^{-11}
LMerr	40,677	1,796e^{-10}
RLMlag	2,3456	0,1256
RLMerr	0,064471	0,7996
SARMA	43,022	4,548e^{-10}

Table 4.53: Results of the specification tests for equation 4.4 (2013)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant; even though their robust forms are not, the LMlag test has a higher value and its robust version has a lower p-value, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	AIC	BIC	Log Likelihood	R^2	LR Test
LM	215.9556	234.859	-100.97780	0.06564	–
SLX	217.316	249.7218	-96.65801	0.0926	–
SAR	178.2705	199.8744	-81.13527	0.4500096	–
SEM	177.7038	199.3077	-80.85191	0.4618092	–
SDM	181.2157	216.3219	-77.60785	0.4830245	SEM / SAR
SDEM	179.2285	214.3347	-76.61424	0.5008487	SEM
SARAR	179.0347	203.339	-80.51735	0.436392	SEM / SAR

Table 4.54: Measures of goodness of fit for equation 4.4 (2013)

The SAR model and SEM have a similar goodness of fit for the data that is better than that of the linear model and the other that considers a single spatial effect (SLX), a result that aligns with the uncertain outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDEM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SEM, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. Given the outcomes of the specification tests and the literature advice on preferring the spatial effects in the dependent variable instead of those in the error term, the SAR model should be taken as the source for the results.

Analysis procedure (2014)

The procedure begins with the multiple linear regression model, which is analysed using the OLS method. The existence of collinearity between predictors is controlled with the VIFs and the highest condition number, which are shown in the following table:

Variable	VIF	Condition number
BedDAR14C	1,206972	3,770
AvgDHCL14C	1,125603	
MedEqR14C	2,685215	
DocDenR14C	2,769913	
NursesR14C	3,654801	

Table 4.55: VIFs and condition number of the predictors in equation 4.4 (2014)

The values suggest that severe collinearity is absent, since they are lower than the reference cutoff values of 10 for the VIFs and 30 for the condition number. The results of the F test statistic ($F = 1{,}957$ and $p\text{-}value = 0{,}09122$) indicate that the model does not fit the data better than an intercept-only model without independent variables.

Before taking the model as invalid, a global Moran's I test is executed to evaluate the presence of spatial autocorrelation in its residuals. The resulting value $I = 0{,}407702918$ is significantly diverse from the expected value $E(I) = -0{,}022243066$ ($p\text{-}value = 1{,}368e^{-11}$), leading to the conduction of further investigations with the specification tests for spatial dependence in the linear regression model, which give the following results:

Test	Value	p-value
LMlag	42,049	8,901e^{-11}
LMerr	36,707	1,373e^{-9}
RLMlag	6,4772	0,01093
RLMerr	1,1355	0,2866
SARMA	43,185	4,193e^{-10}

Table 4.56: Results of the specification tests for equation 4.4 (2014)

The specification tests for spatial effects in the dependent variable and in the error term are statistically significant, but only the robust version of the LMlag test reaches statistical significance, hence conducting a SAR model is the suggested next step. Taking this advice into account, all the other statistical models are also implemented to gather further information from the top-down approach with the purpose of merging it with the suggestion from the bottom-up procedure, so that it can be possible to choose the model that better fits the data among all, as described in the section on model selection. The following table summarises all the measures that can be used to compare the goodness of fit between the various statistical models:

Model	**AIC**	**BIC**	**Log Likelihood**	$\mathbf{R^2}$	**LR Test**
LM	210.2467	229.15	-98.12333	0.04205	–
SLX	210.7204	243.1262	-93.36021	0.07716	–
SAR	172.7662	194.37	-78.38309	0.4374015	–
SEM	174.0013	195.6051	-79.00063	0.438835	–
SDM	177.8241	212.9304	-75.91207	0.4571764	SAR / SEM
SDEM	174.4411	209.5474	-74.22056	0.484646	SEM
SARAR	174.4993	198.8036	-78.24964	0.4204128	SAR / SEM

Table 4.57: Measures of goodness of fit for equation 4.4 (2014)

The SAR model has a better goodness of fit for the data compared to the linear model and the others that consider a single spatial effect (SLX and SEM), a result that aligns with the outcome of the specification tests. Among the other more encompassing models, an overall view of the measures suggests the SDEM as the most appropriate one, but the likelihood ratio test recommends that it should be preferably reduced to a SEM, as the decrease in log likelihood is not statistically significant when accounting for the additional complexity of the model compared to a nested one. Given the outcomes of the specification tests and the literature advice on preferring the spatial effects in the dependent variable instead of those in the error term, the SAR model should be taken as the source for the results.

Results

The third part of the analysis involves the presentation and explanation of the outcomes resulting from the outlined procedures of data analysis. First of all, to provide them in a clear manner, the following three tables illustrate the results for each considered year in the period 2012-2014, with p-values in parentheses and asterisks indicating which of them are statistically significant:

Variable	Direct impact	Indirect impact	Total impact
BedDAR12C	0.02891863 (0.4513599)	0.04173358 (0.496879)	0.07065221 (0.473573)
AvgDHCL12C	-0.11426022 (0.0849824)	-0.16489328 (0.145631)	-0.27915350 (0.109773)
MedEqR12C	0.02860363* (0.0036043)	0.04127899* (0.038228)	0.06988261* (0.013555)
DocDenR12C	-0.03509363* (0.0309117)	-0.05064496 (0.066654)	-0.08573859* (0.042563)
NursesR12C	-0.01031192 (0.2386027)	-0.01488152 (0.310730)	-0.02519344 (0.272960)

Table 4.58: Impacts in the SAR model for RHEDAP12L (2012)

Variable	Direct impact	Indirect impact	Total impact
BedDAR13C	0.004335591 (0.915811)	0.005882694 (0.910934)	0.010218285 (0.911978)
AvgDHCL13C	-0.080926545 (0.188416)	-0.109804211 (0.254389)	-0.190730756 (0.216128)
MedEqR13C	0.021602521* (0.020831)	0.029311121 (0.082062)	0.050913642* (0.042919)
DocDenR13C	-0.034707455* (0.032350)	-0.047092393 (0.083575)	-0.081799848* (0.049578)
NursesR13C	-0.003608491 (0.673572)	0.004896138 (0.677746)	0.008504629 (0.672138)

Table 4.59: Impacts in the SAR model for RHEDAP13L (2013)

Variable	Direct impact	Indirect impact	Total impact
BedDAR14C	0.01564904 (0.721189)	0.02189972 (0.74787)	0.03754876 (0.734201)
AvgDHCL14C	-0.05127254 (0.370220)	-0.07175227 (0.41333)	-0.12302481 (0.389053)
MedEqR14C	0.01983944* (0.028362)	0.02776388 (0.08243)	0.04760332* (0.046982)
DocDenR14C	-0.02266266 (0.167885)	-0.03171477 (0.22712)	-0.05437743 (0.191925)
NursesR14C	-0.01039252 (0.245261)	-0.01454360 (0.31288)	-0.02493612 (0.276426)

Table 4.60: Impacts in the SAR model for RHEDAP14L (2014)

Since the outcomes have been retrieved from spatial models, the procedures of data analysis generated various types of effect concerning the independent variables that are represented by three types of impact. With regards to this particular subtopic of patient emigration for day admissions, the impacts can be defined as follows:

- **Direct impact**: it measures the average effect that a factor in a province has on patient emigration for day admissions in the same province;
- **Indirect impact**: it measures the average effect that a factor in a province has on patient emigration for day admissions in the other provinces, in a direct manner or through its influence on the phenomenon in the same province;
- **Total impact**: it measures the average effect that a factor in a province has on patient emigration for day admissions in all provinces in a global fashion, by merging the direct and indirect impacts.

Establishing a distinction between these effects permits to see whether the various impacts differ in terms of statistical significance (e.g. the direct or indirect impact may be statistically significant, while the total may not) and to evaluate the strengths of the direct and indirect impacts, which may be hidden if solely looking at the total impact.

In addition to the results for the independent variables, the analysis outcomes for each year also involve the following spatial coefficients:

- **RHEDAP12L** (SAR model): ρ = 0,64848 (with $p\text{-}value$ = 3,2206e^{-11});
- **RHEDAP13L** (SAR model): ρ = 0,63163 (with $p\text{-}value$ = 2,984e^{-10});
- **RHEDAP14L** (SAR model): ρ = 0,64011 (with $p\text{-}value$ = 3,3135e^{-10}).

The results for every year are gathered from the SAR model, which provides a spatial coefficient ρ of significant importance. In fact, ρ denotes the average influence that factors in a province have on patient emigration for day admissions in all the other provinces in a global manner, through endogenous interactions occurring in the phenomenon itself that affect neighbouring and non-neighbouring provinces through spatial spillovers (e.g. one factor in a province influences the phenomenon there, which influences it in a neighbouring province, which in turn affects it in a province that is close only to the latter); furthermore, these spatial spillovers can return back and influence the phenomenon in the province of origin. As the results show, the coefficient had remained significantly high during that period, apart from slight fluctuations, indicating the continuous occurrence of indirect effects of factors that from a province had globally spilled over the other neighbouring and non-neighbouring provinces in the entire country, in addition to direct influences over the phenomenon in the province of origin.

Returning to the three main tables with the outcomes for the independent variables and considering just the statistically significant results, highlighted by an asterisk, the following statements on their relation to the phenomenon of patient emigration for day admissions can be made:

- **Rate of medical equipment** – In 2012, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 2,86% in the province of origin, the indirect effect indicates that an increase of 1 unit could have incremented it by 4,13% in the other provinces and the total effect indicates that an increase of 1 unit could have incremented it by 6,99% overall. In 2013, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 2,16% in the province of origin and the total effect indicates that an increase of 1 unit could have incremented it by 5,09% overall. In 2014, the direct effect indicates that an increase of 1 unit could have incremented the phenomenon by 1,98% in the province of origin and the total effect indicates that an increase of 1 unit could have incremented it by 4,76% overall;

- **Rate of doctors and dentists** – In 2012, the direct effect indicates that an increase of 1 unit could have reduced the phenomenon by 3,51% in the province of origin and the total effect indicates that an increase of 1 unit could have reduced it by 8,57% overall. In 2013, the direct effect indicates that an increase of 1 unit could have reduced the phenomenon by 3,47% in the province of origin and the total effect indicates that an increase of 1 unit could have reduced it by 8,18% overall. In 2014, the effects were not statistically significant.

Chapter 5

Discussion

The theoretical model has portrayed how random events occurring on the side of either treatment-seeking patients (e.g. temporary residence in another region) or suppliers of health care treatments (e.g. temporary shortages in the health care supply) can fracture conditions of equilibrium among regional health care systems that retain equal features, leading to resulting outcomes affecting the situation of a region that depend upon which aspect of the phenomenon had previously taken place. In fact, as perceived at the end:

- Region 1, encountering only patient emigration, has a temporary increase in the potential of the health care supply, which however remains untouched and leads to the existence of underused resources, as well as obligations to reimburse the costs of its escaping patients to other regions of destination. Eventually, the outcomes would lead to disinvestments from the health care supply or tax increases, needed to cover the losses and to retain the same level of supply, causing further patient escapes and lower attraction rates that maintain it in a negative vicious cycle. The occurrence of favourable random events or the provision of additional funding by the central state would be required to balance the situation;

- Region 3, confronting only patient immigration, has a temporary decrease in the potential of the health care supply that may cause patient emigration if maximum capacity is reached, but will be favoured by cost reimbursements for treating incoming patients of other regions. Eventually, the outcomes would lead to investments in the health care supply or tax decreases, as profits can be used to retain the same level of supply, causing further patient immigration and higher retention rates that maintain it in a positive vicious cycle. The occurrence of unfavourable random events would cause the situation to return towards the initial equilibrium;

- Region 2, encountering patient emigration and immigration, stands in a situation that can remain stable or sway towards a negative or positive end, depending upon which random events occur at the beginning and through forthcoming stages.

To restore an equilibrium, the state should intervene through the outline of policies which target the aspects that appeared to be relevant for the issue, independently from retaining the same structure of the health care system or revisioning it, also due to the length of time that would be needed for structural changes. This scope can be achieved considering the results of the data analysis, which depict how patient immigration and emigration had been steadily occurring in the considered period, with the former being more relevant when approaching Northern Italy, the latter being more substantial when reaching Southern Italy and low rates for the islands that may stem from their isolation from the mainland of the country. In addition, the following statements can summarise the results on the various factors representing the resources in the health care supply:

- **Rate of beds for ordinary admissions** – An increase in a province could have reduced patient emigration for ordinary admissions from the province of origin and the other provinces. Hence, increasing the rate of beds for ordinary admissions could reduce patient escapes;

- **Rate of beds for day admissions** – An increase in a province could have reduced patient immigration for day admissions into the other provinces, suggesting the presence of competition among regions. Thus, increasing the rate of beds for day admissions could reduce patient escapes and favour competition among providers;

- **Average duration of an ordinary admission** – An increase in a province could have reduced patient immigration for ordinary admissions into the province of origin and the other provinces and could also have increased patient emigration for ordinary admissions mostly from the province of origin. Therefore, reducing the average duration of an ordinary admission could increase patient attraction into a province and decrease patient escapes from a province;

- **Rate of medical equipment** – An increase in a province could have increased patient immigration and emigration for ordinary and day admissions. Regarding patient immigration, the results could convey that the rate of medical equipment had mirrored adequate treatment abilities that had been upholding the attraction of patients into the province of origin and the other provinces. Concerning patient emigration, the results could suggest that, even though a province had possessed sufficient medical equipment, the resources might have not been used efficiently or for the intended purposes, leading patients escape from the province of origin and the other provinces. As a consequence, monitoring the usage of the resources in local health units could help discern if they compose a set of resources that is adequate for the needs of the local populations and are employed in appropriate and efficient manners;

- **Rate of doctors and dentists** – An increase in a province could have reduced patient immigration for day admissions into the other provinces, advocating for the presence of competition among regions, and could also have reduced patient emigration for day admissions primarily from the province of origin. Therefore, increasing the rate of doctors and dentists could reduce patient escapes and favour competition among providers;

- **Rate of nurses** – An increase in a province could have raised patient immigration for ordinary admissions mainly into the province of origin. Hence, increasing the rate of nurses could increase patient attraction into a province and decrease patient escapes from a province.

If the legislation and the organisation of the health care system are allowed to change, other more sensible arrangements may resolve the problem and provide more stability for the long term. With advice from the theoretical model and the concepts on collective action for the usage of common resources, an alternative overview of the system will be given and policies related to another potential solution will be illustrated.

As a general overview, it can be reminded that the whole Italian health care system can be considered as a common-pool resource that is organised by the central state and whose resource units, recognised as the provision of health care services, are given to treatment-seeking individuals by those working in the field; in this context, the involved actors can be described in the following manners:

- The state can be identified as the primary provider, which organises the National Health Service as a whole;
- The public and accredited private suppliers can be recognised as lower-level providers, since they are more directly in contact with the local populations, as producers, because their efforts sustain the system, and as appropriators, since they can achieve certain immaterial and monetary benefits from providing their patients with health treatments;
- The treatment-seeking patients can be seen both as appropriators, because they can obtain benefits from receiving health treatments, and as producers, because their decisions to be treated by certain providers indirectly lead them obtain the mentioned benefits, therefore giving importance to their passive decisions.

Being more specific, the entire system can be separated into several regional health care systems, defined as similar but divided structures which are organised and given their own degree of resources based upon the size and needs of the local populations in the regions. Under the current legislation, a federal structure in the supply of health care services is accompanied by a global national demand that is not adjusted into a similar nested framework; therefore, each federated regional supply structure may potentially have to confront a whole national demand of individuals that, given their self-interests and the size of the entire national group, cannot cooperate efficiently for the prosperity of every regional health care system. Furthermore, the effects of random events, as well as changes in the preferences of patients that can happen as a result of rationality and self-interest to maximise individual benefits, cannot be avoided.

As a consequence of these circumstances, the ending outcomes on free-riding and the tragedies of certain regional health care systems, as predicted in the theoretical model, will happen among the regions; nonetheless, the type of tragedy and the probability of its occurrence vary depending upon the outcomes of each regional context at a certain point in time. In particular, a standard tragedy of the common resource can be observed for a region that falls into the positive case, associated with region R_3 in the theoretical model, only in an unfavourable situation where, in a specific time period, the amount of patients that immigrate there to seek for health treatments is so high that all the available resources become strained, hence residents or other potential incoming patients need to emigrate somewhere else, to postpone their required treatments while waiting for some resources to be freed, to rely on offers from the private sector or to completely forgo their needs; even if the region is favoured by a positive cycle of continuous improvements of its health care system that is sustained by incoming patients, the risk of attracting too many individuals in relation to its capabilities at a certain point in time still remains. On the opposite end, an inverse tragedy of the common resource can be recognised for a region that falls into the negative case, associated with region R_1 in the theoretical model, in which its health care system loses the capacity to give sufficient treatments to the local population so continuously, or taxes are increased so steadily to maintain it to the initial levels, that the common resource becomes uncommon, among residents of the region as well as patients in other areas of the country, and drifts towards a decline. Concerning the most common situation where a region is susceptible to emigration and immigration of patients, associated with region R_2 in the theoretical model, the ending outcomes are undetermined, because they may be comparable to those of region R_1 in unfavourable conditions or align with those of region R_3 in favourable circumstances, with greater uncertainty on the risks of resource overuse. Accounting for these cases is fundamental when enquiring about the issue and reasoning on effective public policies to target it, since their differential details illustrate how the equilibrium is always in a fragile state and exposed to further disturbances.

The proposed solution relies on a policy that influences the inherent ideology of the current National Health Service to redefine the framework of the national demand into a nested structure that mirrors the federal structure of the supply: subsidising free patient choice of treatment to effective individual needs and capacities of regional health care systems. As a general rule, a patient who resides in a region will be allowed to seek for health treatments there, while doing so in other regions will be forbidden. This policy can ensure the achievement of the following outcomes:

- Resources in a region are always proportioned to the needs of its population;
- Investments can be easily made according to present and prospective needs of the population and conditions of the regional health care system;
- Monetary resources are maintained into the region, providing it with the ability to balance costs and benefits of treatments in its local health units;
- Monitoring mechanisms on the usage and investment of resources can be agreed upon at a provincial level with binding contracts between smaller groups of people, ensuring the formation of optimal agreements for the local contexts.

Beginning from a situation of equality, each regional health care system would not be influenced by the occurrence of random events, therefore vicious cycles that cause growing differences between regions over time cannot be created. However, given the health care system has already been running for several decades, there exists the need to level the quantity and quality of provision of health care services among regional health care systems. According to some information from the literature on the improvements of regional health care systems with significant rates of patient immigration, a region can efficiently fulfil this purpose through accreditations of private providers, that enable them to operate on behalf of the National Health Service to assist the regional public health care system with the provision of the essential services of the statutory benefits package to the local population, while charging either nothing or the same costs; given that the accreditation processes result from agreements that depend upon the needs of the local population and are aligned with the available monetary resources, the outlined policy outcomes ensure that every region can efficiently concur with private providers to reach its improvement aims. Nonetheless, even though processes of authorisation and accreditation can enhance the overall regional provision of health care services towards higher standards, the occurrence of unpredicted long-term outcomes resulting from a greater reliance on private providers should be taken into account. With regards to the provision of the essential levels of care, an excessive dependence for the achievement of regional requirements could confine most of the benefits of treating local patients into the realms of the private market; while a portion of monetary benefits could be extracted through regional taxes, this approach would not be able to touch the immaterial benefits (e.g. training of personnel, patient attraction) and may induce certain private providers to opt out from accreditation agreements if the taxation of accredited activities becomes too severe. In addition, further evaluation needs to be conducted when the discussion is extended to consider the provision of an entire array of health care services. In fact, in the presence of a dual provision of the same multitude of services in which a public system coexists with private alternatives, Epple and Romano (1996) underlined that a plurality of high-income and low-income households expresses similar preferences that advocate for reductions of public expenditures, contrasting those of middle-income households that favour their increases [13, 316]; in the framework of a dual provision of health care services, these same preferences are revealed since wealthy individuals would select the offers of private providers without contemplating public alternatives, poor people would not be willing to replace public health care services with private alternatives in any case and those in the middle class would choose to use public health care services when given the opportunity in alternative to private ones, therefore preferring them to be of higher quality. When a public system coexists alongside private alternatives, these contrasting group preferences could produce the formation of a majority coalition that endorses a reduction of public expenditures for health care services, whose occurrence would lead to a reduction in their quality and in turn cause individuals, even those from the middle class, to drop out from the public system because the effective quality would be unable to meet their demanded level of quality; a decline of support from all classes of individuals would cause the public health care system to lose funding and to confront further quality decreases due to the absence of sufficient resources to sustain sufficient developments, to the advantage of private providers. All the highlighted issues should be avoided to ensure that the reliance on accredited private providers can increase the regional supply and enhance the provision quality through greater degrees of competition in the market without weakening the support for the public health care system in the long term. For this purpose, considering that a region retains ruling authority on its resources and the features of accreditations, the following practical measures can be implemented:

- The accreditation of private providers, that could more efficiently offer better services compared to public suppliers in the region, occurs moderately to permit the regional public health care system to gradually meet the same higher qualitative standards over time, so that the private offering never substantially overtakes the public provision in the eyes of the local populations;
- The agreements for the accreditation of private providers should comprise a fair share of taxes, which is both sufficiently low for a private provider to be tolerable when accounting for the accreditation and the additional exposure to the general public, as well as sufficiently high for a region to extract an adequate portion of monetary profits that would be lost whenever a patient obtains a health treatment from a private provider rather than a public supplier;
- The agreements for the accreditation of private providers should require specific health treatments to be supplied in a public structure, equipped with appropriate means, so that immaterial benefits can support the enhancement of features of the public system; a region could grant a private provider certain concessions, such as specific tax reductions, to conclude suitable negotiations for both parties.

Returning to the proposed policy and assuming that all regional health care systems have reached an equal level of quality that is sufficient considering the needs of the local population that lives in each region, the structure of the new system would require that, in normal situations, an individual who wishes to obtain treatments for a supposedly better system would simply move to another region, causing permanent changes in the resource structure between two regional health care systems that will not influence any sort of equilibrium of costs and benefits; an individual would not be able to emigrate from a region to another to take advantage of higher quality health care resources that are present there, but for which he or she had not made any sort of direct contribution, therefore free-riding in a regional health care system that is shaped upon the needs of a local population that the individual does not precisely belong to; the individual would also contribute to the advancement of his or her regional health care system by gathering a health treatment locally, while the region would not have to cover the costs resulting from an optional decision on obtaining a health treatment elsewhere. In this context, the concerns that may arise due to a centralised structure of the demand, which surpasses the equilibrium between regional health care systems thanks to free patient choice of treatment, would never exist. However, the immediate need for health treatments due to the occurrence of a random event (e.g. incident in a region that is located far from the region of residence) can exempt from the general rule, but would be managed differently compared with the current system. First of all, in situations of resource constraints, local health units in a region would always consider its residents ahead of people coming from other areas. Secondly, two additional elements are established to handle the exceptions:

- On a national level, the state defines standard maximum costs that patients have to pay directly to local health authorities when being treated outside of their region of residence, which will depend upon the category of a health treatment but not on the regions of origin and destination; furthermore, regions and provinces are made aware about them to proceed with reimbursements to patients in the years following the payments for health treatments;

- On a provincial level, individuals that reside in a province of a region agree upon the definition of a common provincial fund, financed through progressive general taxation, to ensure that a province is always able to directly reimburse local health units of other regions with the treatment costs that these had to sustain in case of treating one of its residents and to secure the absence of further pressure on the financial resources of single patients, the province itself and the related region.

A patient always pays for the treatment costs up to the level defined by the national standards and, if the charges are higher, the needed additional amount is paid through the fund of the province of residence using a forgivable payment without any obligation towards the patient. To reflect the idea of progressive taxation on the total expenditure, the portion paid by the patient will be reimbursed by the province of residence through tax credits in the following years, according to collective arrangements agreed among provinces located in the same region. In general, the proposed policy ensures that, once the health care system returns to an improved equilibrium, every region will be able to retain its treatment-seeking residents and to provide them with the required health care services at the right time, without concerns on the availability of resources in the health care supply and the quality of health treatments. Furthermore, in special circumstances that involve individuals receiving health treatments outside their regions of residence under the realms of the mentioned rule exceptions, each region will not have neither to sustain unforeseen costs that are not accompanied by complementary benefits in case of escaping patients nor to endure temporary financial losses or potential resource overuses in case of incoming patients. Eventually, these outcomes will prevent the development of imbalances that could induce tragedies of common and uncommon regional health care systems in the long term, supporting the retention of a national equilibrium and safeguarding the principles which once established the Italian National Health Service and the concept of universal health as an individual right and a collective interest that was conceived in the Constitution of the Italian Republic.

Chapter 6

Conclusion

This research thesis has examined one issue of the Italian health care system, concerning individuals moving from a region in search for better health treatments, as it is deemed to be important for reasons of sustainability of the system and equality between individuals seeking for appropriate health services. The relevance of the matter has been depicted in the literature review, which has also given some considerations for the data analysis and details that highlighted the contribution of this thesis. Further research information has been described under the realms of the theoretical foundations for the development of a theoretical model and the conduction of the data analysis. The theoretical model has illustrated how the problem can arise from a situation of equality between regions, with further differences that widen over time, due to the occurrence of random events that are beyond any possible control. In the context of an already existent issue, the data analysis has contributed with illustrating how the phenomenon, in terms of immigration into a region and emigration from a region, has continued to be relevant over the considered time period, with clear differences when moving among macro areas of the country, as well as finding out how acting on certain factors related to resources in the supply could reduce its occurrence rate and improve the equilibrium between regions. However, the discussion section also included a straightforward solution, which involves subsidising free patient choice to actual individual needs and capacities of the regional health care systems to create a federal structure in the demand that mirrors the nested framework of the supply. With quality alignments among regional health care systems for the needs of the local populations and the creation of national cost standards and provincial funds, that are used without obligations and follow ideas of progressive taxation, each nested health care system will run efficiently without incurring the theoretical outcomes.

6.1 Limitations

This thesis has provided a simple theoretical model to illustrate the establishment and development of regional patient migration, as well as an examination of certain data to portray additional information for discussion. Nonetheless, the research retains specific inherent limitations, whose reveal is considered to be fundamental to comprehend not only the boundaries of the conclusions, but also potential opportunities for conducting further analyses of the matter. Therefore, with the uppermost intellectual honesty, the following limitations have been identified:

- Certain information concerning other indicators that could have been included in the statistical models was absent from the source, despite its completeness for the topic in question. First of all, some information was missing for a variety of years, thus the time period to examine had to be restricted according to the available data; in addition, more specific information was available at a regional level compared with the provincial one on which the analysis has been performed, a circumstance that may suggest the need for data on a more detailed territorial level;

- The type of analysis and the used data have been useful to look at the phenomena with regards to the provision and obtainment of short to medium health care. A more complete overview of the situation would also need to include examinations on long-term health care, which however requires a diverse analysis method;

- The essence of the analysis made use of data that was quantitative by nature. Still, it is important to recognise how qualitative information, such as direct interviews with selected patients or doctors working in various health facilities, can add more insights that, in combination with the analysis of quantitative data, may provide a more complete picture of the topic;

- The data analysis separately considered the years in the given time period without employing specific statistical techniques, such as those of time-series analysis;

- The methods and the data analysis of this research have been established on the geographical level of provinces of Italy, for which certain administrative data is available. This characteristic introduces some limits on the ability to infer results for the reality as seen at an individual level, since this level of examination does not permit to fully capture the interactions occurring between single individuals in the population, whose analysis through the same models may provide dissimilar outcomes. This problematic can be defined as an ecological fallacy, which occurs when deducting inferences on individuals from the examination of data related to the groups they belong to. Nonetheless, it can be declared that a valid reason for collecting data on a provincial level instead of a regional one was to mitigate the extent of this specific issue in this research.

6.2 Further research

A few potential opportunities for further research on the topic can be underlined when taking the mentioned limitations into account. For instance, the same research could be executed again in the future to observe whether the introduction of specific legislative policies has induced positive or negative changes; in this context, the analysis may be enhanced if additional data is provided by official sources, such as the one leveraged for this research, especially in terms of time availability and territorial detail. Moreover, the employment of other types of statistical techniques could provide a more comprehensive overview of the state of the matter and implement diverse functions, such as prediction of changes, to enhance an empirical analysis. In addition, as mentioned previously, the inclusion of qualitative research, either in combination with quantitative data or in an independent manner, could assist with discerning the subject using a distinct analytical approach. Finally, the usage of data on a more detailed geographical level, such as that of municipalities, can provide results that reduce the degree of the ecological fallacy that could exist in examinations based upon less accurate geographical scales.

Appendix A

Data set preparation

A.1 Repair of shapefile geometries

The control and repair of the geometries of the original shapefiles of Italian provinces were performed using the software SpatiaLite, an open source library that implements spatial features into the SQLite database engine. The information concerning the entire procedure is outlined in the following tables:

Layer	
Table	ProvCM01012016_WGS84
Geometry column	geometry
Geometry type	MULTIPOLYGON
Dimensions	XY
SRID	0

Table A.1: Details of the original shapefile layer

Statistics	
Total Rows	110
NULL Geometries	NONE
Valid Geometries (full valid)	106
Valid Geometries (minor issues)	NONE
Invalid Geometries	4
Suggested action	This layer contains invalid Geometries; a repair action is urgently required.

Table A.2: Statistics of the original shapefile geometries

ROWID	Error cause
65	Ring Self-intersection at or near point 1258539.7858999996 4539825.2137000002
66	Ring Self-intersection at or near point 1308926.0356999999 4476059.9891999997
107	Ring Self-intersection at or near point 528108.51559999958 4571486.8072999995
110	Ring Self-intersection at or near point 444979.67760000005 4320623.5316000003

Table A.3: List of invalid geometries

Statistics	
Invalid Geometries	4
Repaired Geometries (fully recovered)	4
Repaired Geometries (by discarding fragments)	NONE
Failures (Not Repaired Geometries)	NONE
Suggested action	This layer has been successfully repaired and is now completely valid; doesn't require any further corrective action.

Table A.4: Repair of the original shapefile geometries

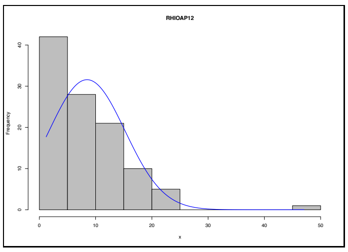
Statistics	
Total Rows	110
NULL Geometries	NONE
Valid Geometries (full valid)	110
Valid Geometries (minor issues)	NONE
Invalid Geometries	NONE
Suggested action	This layer is perfectly valid; doesn't require any corrective action.

Table A.5: Statistics of the repaired shapefile geometries

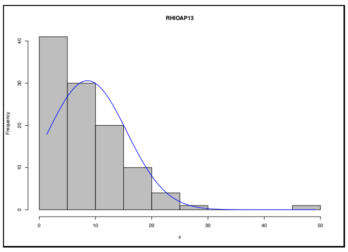
A.2 Data transformation

Recalling the passage on the transformation of the dependent and independent variables contained in the section about the preparation of the data set, the present section of the appendix presents additional histograms and probability plots that illustrate the effects of the logarithmic transformation and the mean centring procedure executed on the data for each variable and every year in the period 2012-2014. The histograms and probability plots showing the effects of the logarithmic transformation on the values and residuals of the dependent variables are considered at first, while the histograms depicting the effects of the mean-centring procedure on the values of the independent variables follow in a separate section. For each variable of a particular year, the histogram or probability plot referred to the original variable is shown on top of that of the transformed variable.

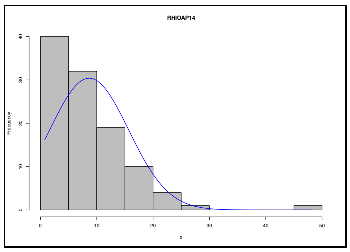
A.2.1 Dependent variables



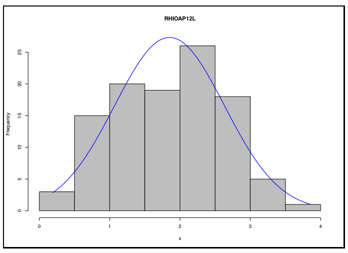
(a) Distribution of RHIOAP12



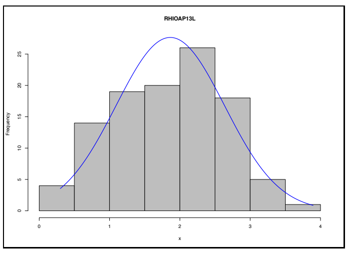
(b) Distribution of RHIOAP13



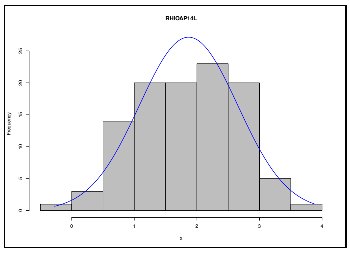
(c) Distribution of RHIOAP14



(d) Distribution of RHIOAP12L

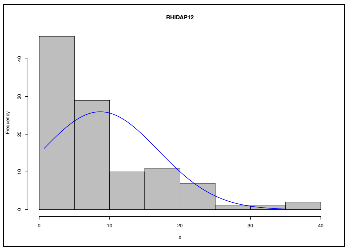


(e) Distribution of RHIOAP13L

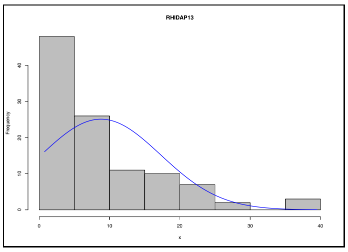


(f) Distribution of RHIOAP14L

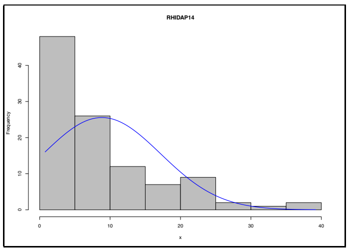
Figure A.1: Logarithmic transformation of RHIOAPxx (2012-2014)



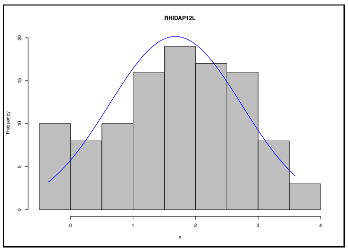
(a) Distribution of RHIDAP12



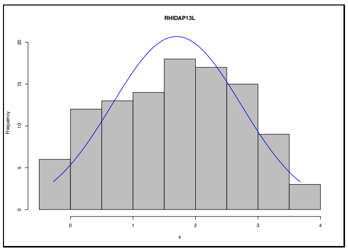
(b) Distribution of RHIDAP13



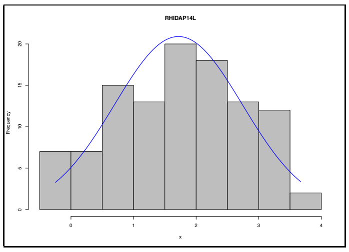
(c) Distribution of RHIDAP14



(d) Distribution of RHIDAP12L

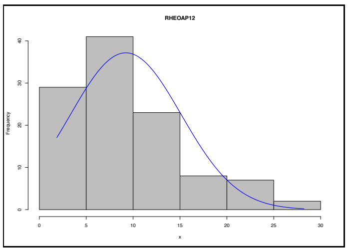


(e) Distribution of RHIDAP13L

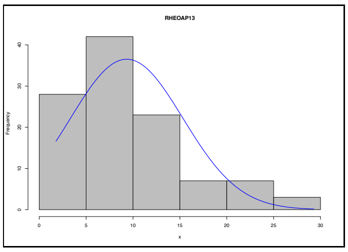


(f) Distribution of RHIDAP14L

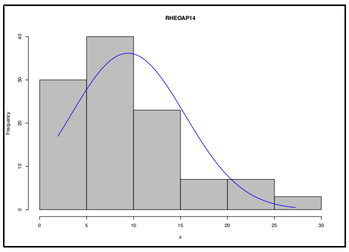
Figure A.2: Logarithmic transformation of RHIDAPxx (2012-2014)



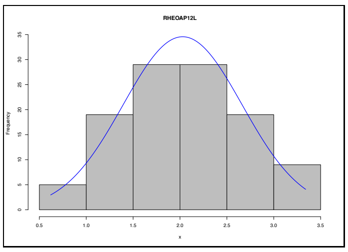
(a) Distribution of RHEOAP12



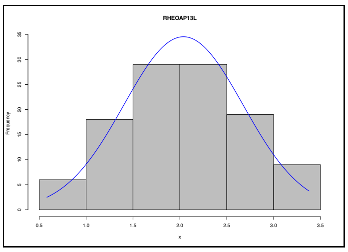
(b) Distribution of RHEOAP13



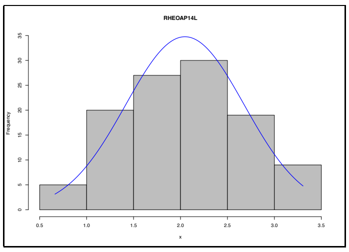
(c) Distribution of RHEOAP14



(d) Distribution of RHEOAP12L

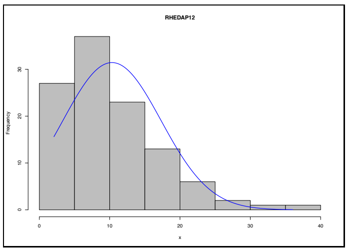


(e) Distribution of RHEOAP13L

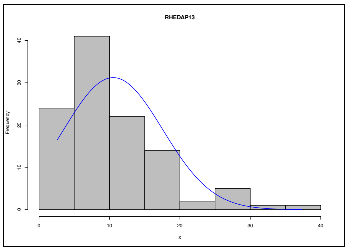


(f) Distribution of RHEOAP14L

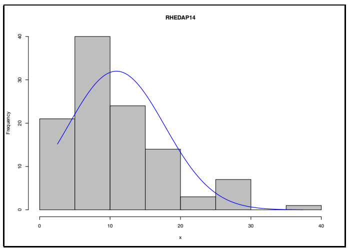
Figure A.3: Logarithmic transformation of RHEOAPxx (2012-2014)



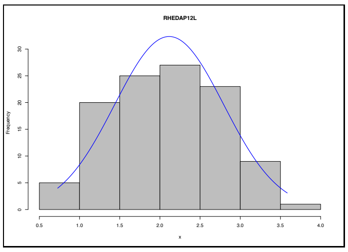
(a) Distribution of RHEDAP12



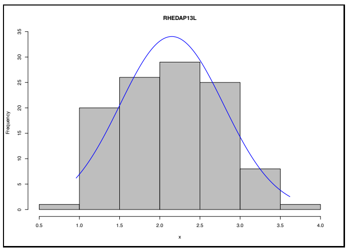
(b) Distribution of RHEDAP13



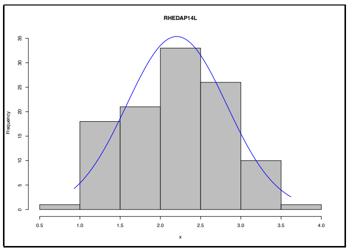
(c) Distribution of RHEDAP14



(d) Distribution of RHEDAP12L

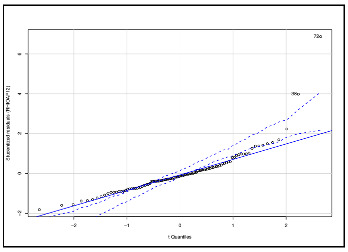


(e) Distribution of RHEDAP13L

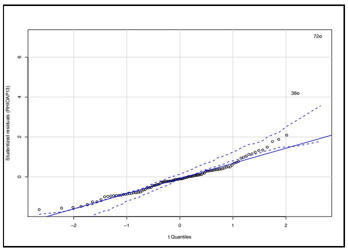


(f) Distribution of RHEDAP14L

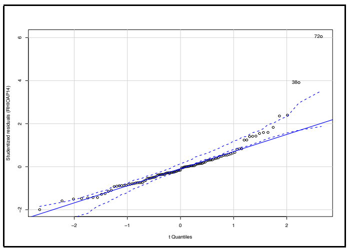
Figure A.4: Logarithmic transformation of RHEDAPxx (2012-2014)



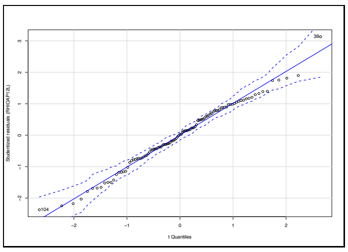
(a) Q-Q plot of residuals for RHIOAP12



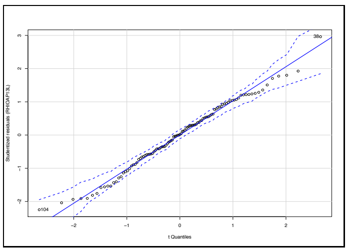
(b) Q-Q plot of residuals for RHIOAP13



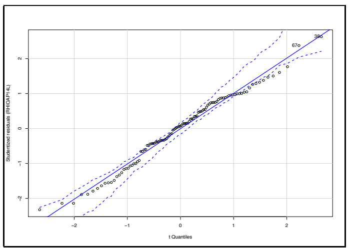
(c) Q-Q plot of residuals for RHIOAP14



(d) Q-Q plot of residuals for RHIOAP12L

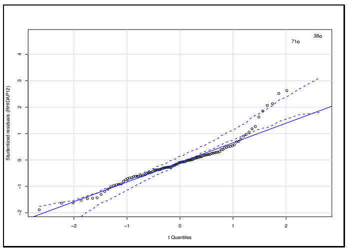


(e) Q-Q plot of residuals for RHIOAP13L

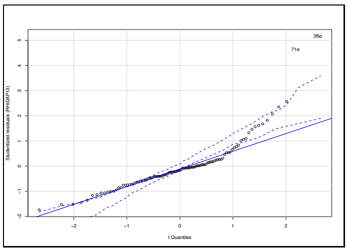


(f) Q-Q plot of residuals for RHIOAP14L

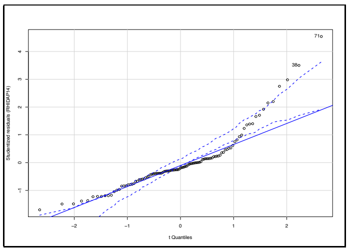
Figure A.5: Q-Q plot of residuals for RHIOAPxx (2012-2014)



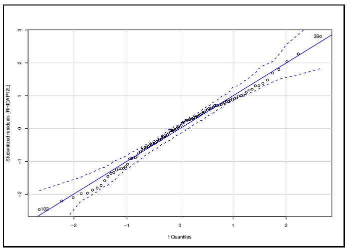
(a) Q-Q plot of residuals for RHIDAP12



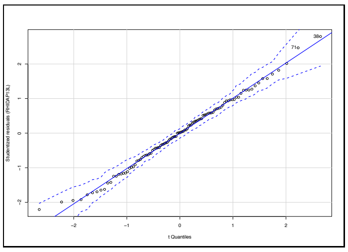
(b) Q-Q plot of residuals for RHIDAP13



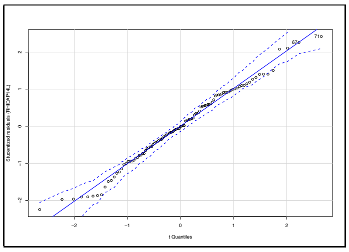
(c) Q-Q plot of residuals for RHIDAP14



(d) Q-Q plot of residuals for RHIDAP12L

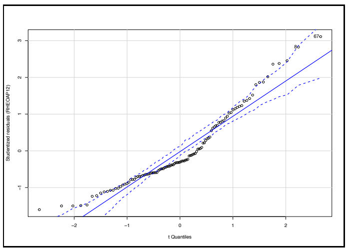


(e) Q-Q plot of residuals for RHIDAP13L

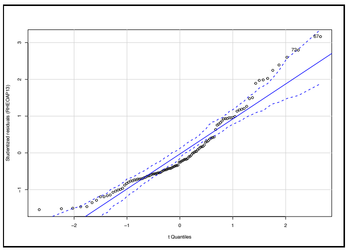


(f) Q-Q plot of residuals for RHIDAP14L

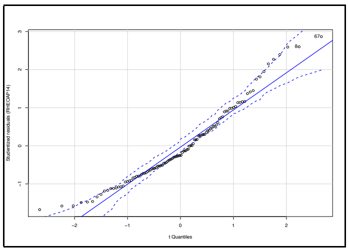
Figure A.6: Q-Q plot of residuals for RHIDAPxx (2012-2014)



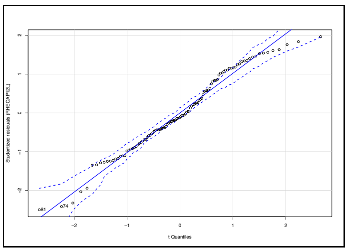
(a) Q-Q plot of residuals for RHEOAP12



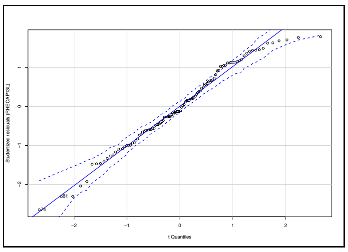
(b) Q-Q plot of residuals for RHEOAP13



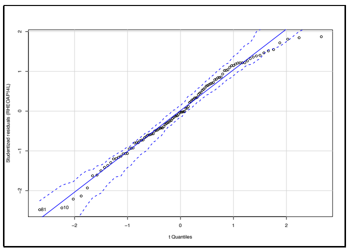
(c) Q-Q plot of residuals for RHEOAP14



(d) Q-Q plot of residuals for RHEOAP12L

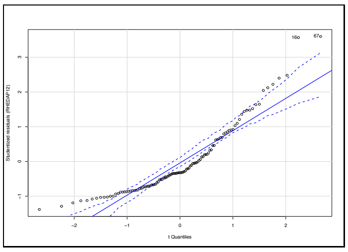


(e) Q-Q plot of residuals for RHEOAP13L

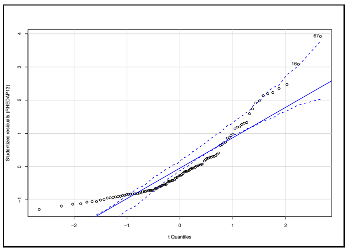


(f) Q-Q plot of residuals for RHEOAP14L

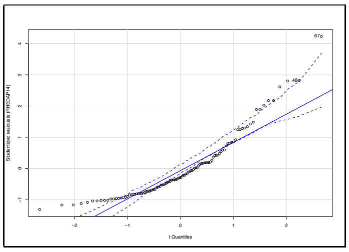
Figure A.7: Q-Q plot of residuals for RHEOAPxx (2012-2014)



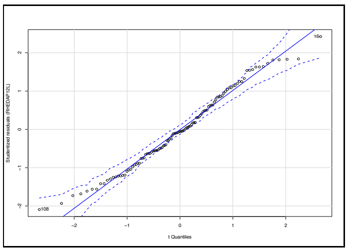
(a) Q-Q plot of residuals for RHEDAP12



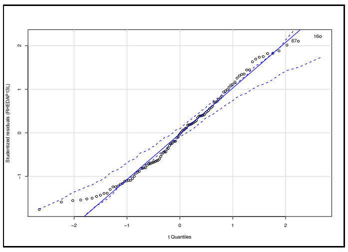
(b) Q-Q plot of residuals for RHEDAP13



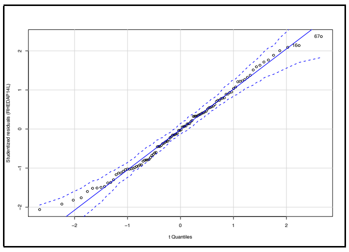
(c) Q-Q plot of residuals for RHEDAP14



(d) Q-Q plot of residuals for RHEDAP12L



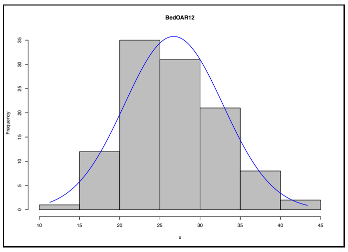
(e) Q-Q plot of residuals for RHEDAP13L



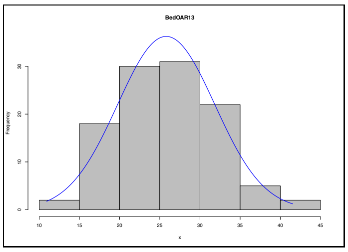
(f) Q-Q plot of residuals for RHEDAP14L

Figure A.8: Q-Q plot of residuals for RHEDAPxx (2012-2014)

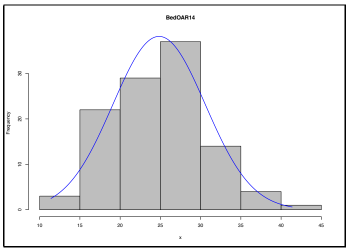
A.2.2 Independent variables



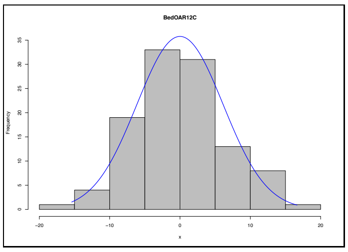
(a) Distribution of BedOAR12



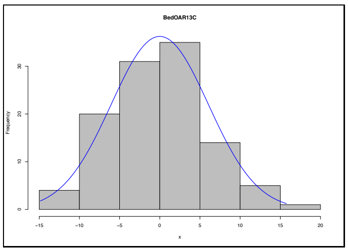
(b) Distribution of BedOAR13



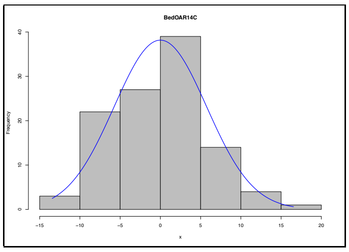
(c) Distribution of BedOAR14



(d) Distribution of BedOAR12C

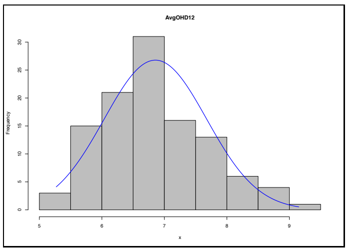


(e) Distribution of BedOAR13C

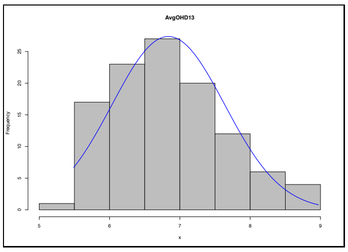


(f) Distribution of BedOAR14C

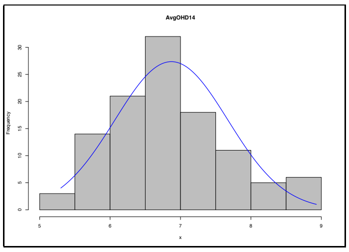
Figure A.9: Mean centring of BedOARxx (2012-2014)



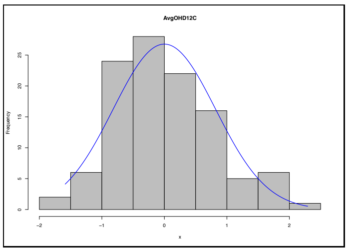
(a) Distribution of AvgOHD12



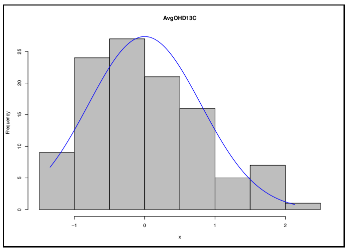
(b) Distribution of AvgOHD13



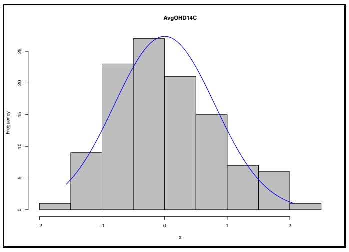
(c) Distribution of AvgOHD14



(d) Distribution of AvgOHD12C

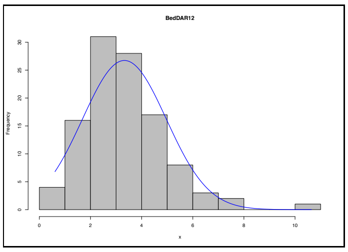


(e) Distribution of AvgOHD13C

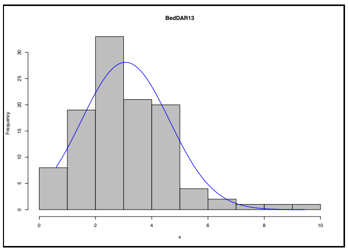


(f) Distribution of AvgOHD14C

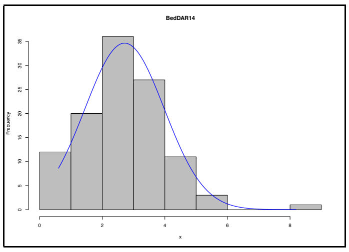
Figure A.10: Mean centring of AvgOHDxx (2012-2014)



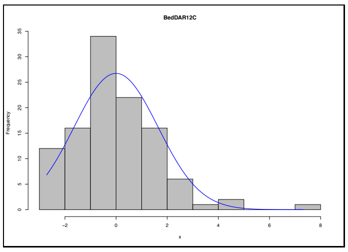
(a) Distribution of BedDAR12



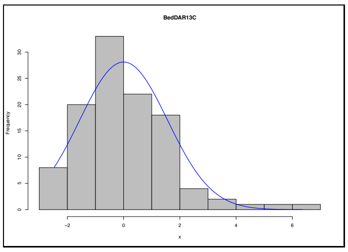
(b) Distribution of BedDAR13



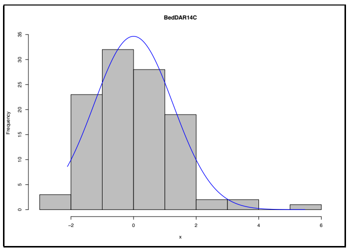
(c) Distribution of BedDAR14



(d) Distribution of BedDAR12C

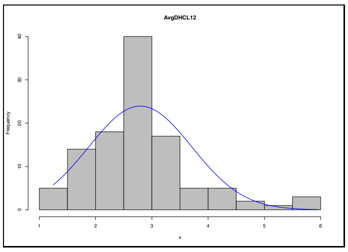


(e) Distribution of BedDAR13C

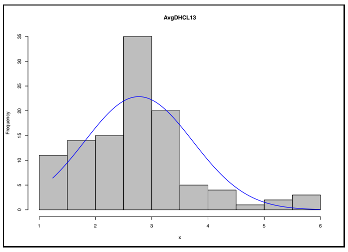


(f) Distribution of BedDAR14C

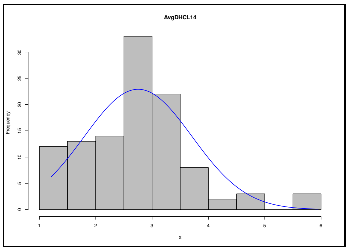
Figure A.11: Mean centring of BedDARxx (2012-2014)



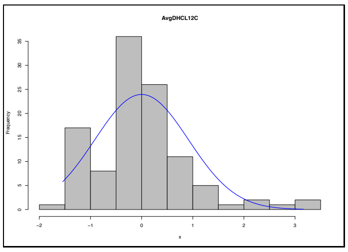
(a) Distribution of AvgDHCL12



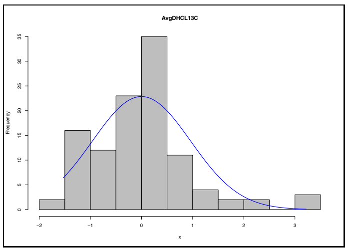
(b) Distribution of AvgDHCL13



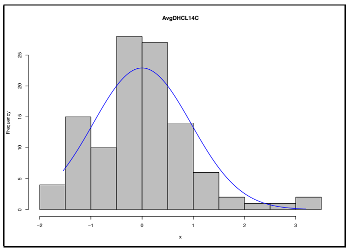
(c) Distribution of AvgDHCL14



(d) Distribution of AvgDHCL12C

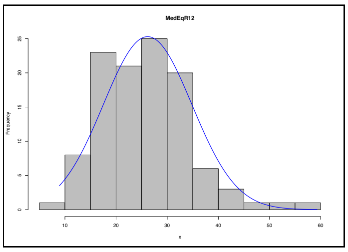


(e) Distribution of AvgDHCL13C

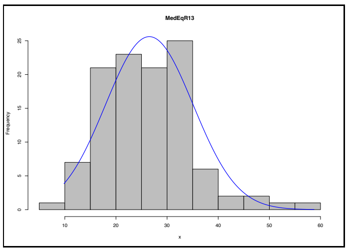


(f) Distribution of AvgDHCL14C

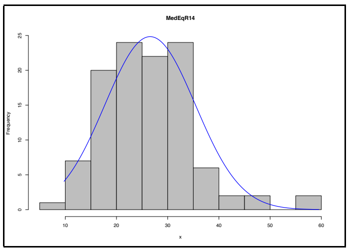
Figure A.12: Mean centring of AvgDHCLxx (2012-2014)



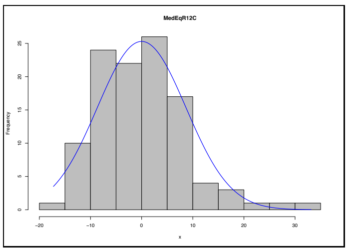
(a) Distribution of MedEqR12



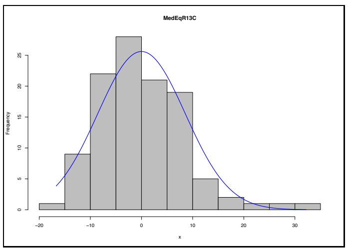
(b) Distribution of MedEqR13



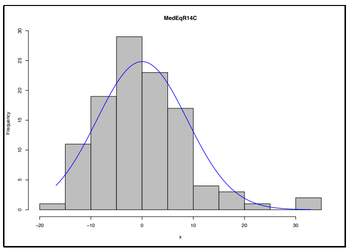
(c) Distribution of MedEqR14



(d) Distribution of MedEqR12C

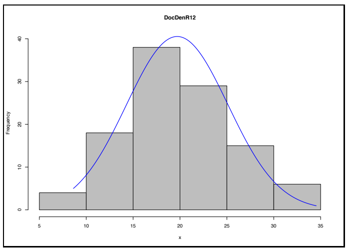


(e) Distribution of MedEqR13C

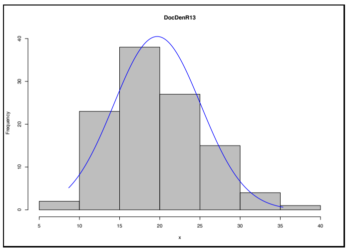


(f) Distribution of MedEqR14C

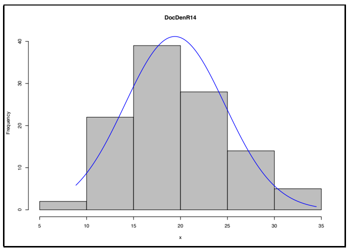
Figure A.13: Mean centring of MedEqRxx (2012-2014)



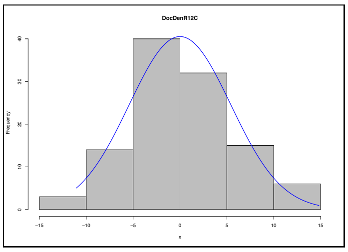
(a) Distribution of DocDenR12



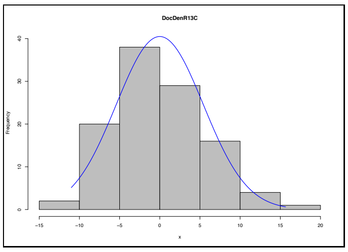
(b) Distribution of DocDenR13



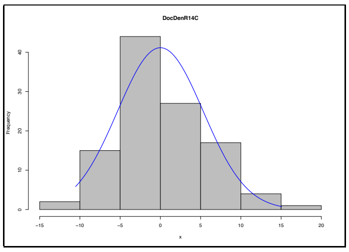
(c) Distribution of DocDenR14



(d) Distribution of DocDenR12C

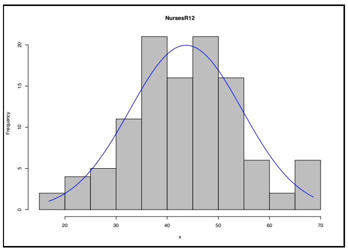


(e) Distribution of DocDenR13C

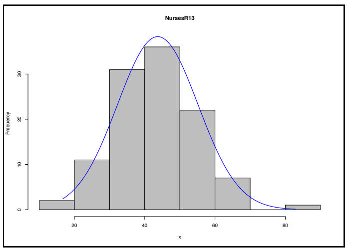


(f) Distribution of DocDenR14C

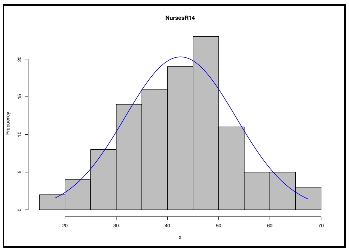
Figure A.14: Mean centring of DocDenRxx (2012-2014)



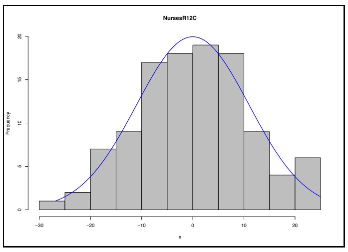
(a) Distribution of NursesR12



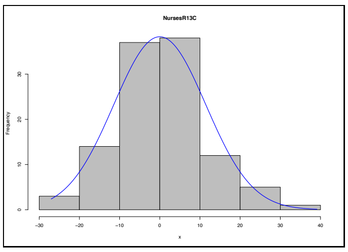
(b) Distribution of NursesR13



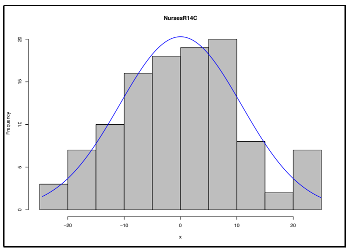
(c) Distribution of NursesR14



(d) Distribution of NursesR12C



(e) Distribution of NursesR13C



(f) Distribution of NursesR14C

Figure A.15: Mean centring of NursesRxx (2012-2014)

Appendix B

Programming code

B.1 Data transformation

The following extract portrays a portion of the programming code that was written with the R language to execute the preliminary logarithmic transformations on the dependent variables and mean-centring procedures on the independent variables, for the purpose of preparing the definitive data set before advancing with the data analysis, using the data concerning every subtopic in the year 2014 as an example:

```
### IT-PMC-RHM 2014 - Data transformation ###

# Install additional packages
install.packages("rcompanion")
install.packages("tseries")

# Load additional packages
library(rcompanion)
library(tseries)

# Change the settings of scientific notation
options(scipen = 6)

# Read the file containing the data
ITRHM2014.data.initial <- read.csv2("../Data/IT-RHM-2014-Data-Initial.csv", header = TRUE, encoding = "UTF-8")
attach(ITRHM2014.data.initial)
summary(ITRHM2014.data.initial)
```

```
# Histograms showing the distribution of every variable
plotNormalHistogram(RHIOAP14, prob = FALSE, col = "gray",
    main = "RHIOAP14", linecol = "blue", lwd = 2)
plotNormalHistogram(RHIDAP14, prob = FALSE, col = "gray",
    main = "RHIDAP14", linecol = "blue", lwd = 2)
plotNormalHistogram(RHEOAP14, prob = FALSE, col = "gray",
    main = "RHEOAP14", linecol = "blue", lwd = 2)
plotNormalHistogram(RHEDAP14, prob = FALSE, col = "gray",
    main = "RHEDAP14", linecol = "blue", lwd = 2)
plotNormalHistogram(BedOAR14, prob = FALSE, col = "gray",
    main = "BedOAR14", linecol = "blue", lwd = 2)
plotNormalHistogram(AvgOHD14, prob = FALSE, col = "gray",
    main = "AvgOHD14", linecol = "blue", lwd = 2)
plotNormalHistogram(BedDAR14, prob = FALSE, col = "gray",
    main = "BedDAR14", linecol = "blue", lwd = 2)
plotNormalHistogram(AvgDHCL14, prob = FALSE, col = "gray",
    main = "AvgDHCL14", linecol = "blue", lwd = 2)
plotNormalHistogram(MedEqR14, prob = FALSE, col = "gray",
    main = "MedEqR14", linecol = "blue", lwd = 2)
plotNormalHistogram(DocDenR14, prob = FALSE, col = "gray",
    main = "DocDenR14", linecol = "blue", lwd = 2)
plotNormalHistogram(NursesR14, prob = FALSE, col = "gray",
    main = "NursesR14", linecol = "blue", lwd = 2)

# Test the assumption of normality of the residuals for the
     dependent variables
jarque.bera.test(na.omit(RHIOAP14)) # Reject
jarque.bera.test(na.omit(RHIDAP14)) # Reject
jarque.bera.test(RHEOAP14) # Reject
jarque.bera.test(RHEDAP14) # Reject
```

```
# Remove the heteroscedasticity of residuals for every dependent variable with a logarithmic transformation
ITRHM2014.data.transformed = data.frame()[1:110, 0]
ITRHM2014.data.transformed $ PMC_Name = ITRHM2014.data.initial $ PMC_Name
ITRHM2014.data.transformed $ RHIOAP14L = log(ITRHM2014.data.initial $ RHIOAP14)
ITRHM2014.data.transformed $ RHIDAP14L = log(ITRHM2014.data.initial $ RHIDAP14)
ITRHM2014.data.transformed $ RHEOAP14L = log(ITRHM2014.data.initial $ RHEOAP14)
ITRHM2014.data.transformed $ RHEDAP14L = log(ITRHM2014.data.initial $ RHEDAP14)

# Mean centre the independent variables around 0
mean.centre <- function(x){scale (x, scale = FALSE)}
ITRHM2014.data.transformed $ BedOAR14C = mean.centre(ITRHM2014.data.initial $ BedOAR14)
ITRHM2014.data.transformed $ AvgOHD14C = mean.centre(ITRHM2014.data.initial $ AvgOHD14)
ITRHM2014.data.transformed $ BedDAR14C = mean.centre(ITRHM2014.data.initial $ BedDAR14)
ITRHM2014.data.transformed $ AvgDHCL14C = mean.centre(ITRHM2014.data.initial $ AvgDHCL14)
ITRHM2014.data.transformed $ MedEqR14C = mean.centre(ITRHM2014.data.initial $ MedEqR14)
ITRHM2014.data.transformed $ DocDenR14C = mean.centre(ITRHM2014.data.initial $ DocDenR14)
ITRHM2014.data.transformed $ NursesR14C = mean.centre(ITRHM2014.data.initial $ NursesR14)
```

```
# Histograms illustrating the distribution of the log-
   transformed dependent variables and the mean-centred
   independent variables
plotNormalHistogram(ITRHM2014.data.transformed $ RHIOAP14L,
    prob = FALSE, col = "gray", main = "RHIOAP14L", linecol
    = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ RHIDAP14L,
    prob = FALSE, col = "gray", main = "RHIDAP14L", linecol
    = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ RHEOAP14L,
    prob = FALSE, col = "gray", main = "RHEOAP14L", linecol
    = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ RHEDAP14L,
    prob = FALSE, col = "gray", main = "RHEDAP14L", linecol
    = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ BedOAR14C,
    prob = FALSE, col = "gray", main = "BedOAR14C", linecol
    = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ AvgOHD14C,
    prob = FALSE, col = "gray", main = "AvgOHD14C", linecol
    = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ BedDAR14C,
    prob = FALSE, col = "gray", main = "BedDAR14C", linecol
    = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ AvgDHCL14C
   , prob = FALSE, col = "gray", main = "AvgDHCL14C",
   linecol = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ MedEqR14C,
    prob = FALSE, col = "gray", main = "MedEqR14C", linecol
    = "blue", lwd = 2)
```

```r
plotNormalHistogram(ITRHM2014.data.transformed $ DocDenR14C, prob = FALSE, col = "gray", main = "DocDenR14C", linecol = "blue", lwd = 2)
plotNormalHistogram(ITRHM2014.data.transformed $ NursesR14C, prob = FALSE, col = "gray", main = "NursesR14C", linecol = "blue", lwd = 2)

# Test the assumption of normality of the residuals for the log-transformed dependent variables
jarque.bera.test(na.omit(ITRHM2014.data.transformed $ RHIOAP14L)) # Do not reject
jarque.bera.test(na.omit(ITRHM2014.data.transformed $ RHIDAP14L)) # Do not reject
jarque.bera.test(ITRHM2014.data.transformed $ RHEOAP14L) # Do not reject
jarque.bera.test(ITRHM2014.data.transformed $ RHEDAP14L) # Do not reject

# Write the data of the new dependent variables and independent variables in a separate file to be merged with the main shapefile
write.csv2(ITRHM2014.data.transformed, "../Data/IT-RHM-2014-Data-Transformed.csv", fileEncoding = "UTF-8")
```

Listing B.1: Data transformation (R)

B.2 Data analysis

The following excerpt shows a portion of the programming code that was written with the R language to perform the data analysis, using the data concerning regional patient immigration for ordinary admissions in the year 2014 as an example:

```r
### IT-PMC-RHIOA 2014 - Data analysis ###

# Install additional packages
install.packages("car")
install.packages("perturb")
install.packages("rgdal")
install.packages("spdep")

# Load additional packages
library(car)
library(perturb)
library(rgdal)
library(spdep)

# Change the settings of scientific notation
options(scipen = 6)

# Import the shapefile with the data and the weights matrix
    created with GeoDa
ITRHM2014.data = readOGR(dsn = "../Spatial", layer = "IT-
   RHM-2014")
attach(ITRHM2014.data@data)
summary(ITRHM2014.data)
PMC.neighbours.queen1 <- read.gal("../Spatial/IT-RHM-2014-WF
   -Queen1.gal", override.id = TRUE)
summary(PMC.neighbours.queen1)
PMC.neighbours.queen1.listw <- nb2listw(
   PMC.neighbours.queen1, glist = NULL, style = "W",
   zero.policy = FALSE)
ITRHM2014.coordinates <- coordinates(ITRHM2014.data)
plot(PMC.neighbours.queen1.listw, ITRHM2014.coordinates)
```

```
# Create a second listw excluding the observations without
   data for Y
ITRHM2014.listw.NAdrop <- c(82, 83, 84)
PMC.neighbours.queen1.listw.NAdrop <- subset(
   PMC.neighbours.queen1.listw, !(1:length(
   PMC.neighbours.queen1) %in% ITRHM2014.listw.NAdrop))
summary(PMC.neighbours.queen1.listw.NAdrop)

# Moran's I test for spatial autocorrelation (based on the
   normal assumption and permutations)
moran.test(RHIOAP14L, PMC.neighbours.queen1.listw,
   randomisation = TRUE, zero.policy = FALSE, alternative =
    "greater", rank = FALSE, na.action = na.omit)
RHIOAP14L.Moran.test.permutations.queen1 <- moran.mc(
   RHIOAP14L, PMC.neighbours.queen1.listw, 999, na.action =
    na.omit)

# Portray a density plot of the Moran's I on the reference
   distribution
RHIOAP14L.Moran.test.permutations.queen1.density <- density(
   RHIOAP14L.Moran.test.permutations.queen1 $ res[1:length(
   RHIOAP14L.Moran.test.permutations.queen1 $ res) - 1])
plot(RHIOAP14L.Moran.test.permutations.queen1.density, main
    = "Moran Permutation Test (RHIOAP14L)", xlab = "
   Reference Distribution", xlim = c(-0.3, 0.7), ylim = c
   (0, 6), lwd = 2, col = 2)
hist(RHIOAP14L.Moran.test.permutations.queen1 $ res[1:
   length(RHIOAP14L.Moran.test.permutations.queen1 $ res) -
    1], freq = F, add = T)
abline(v = RHIOAP14L.Moran.test.permutations.queen1 $
   statistic, lwd = 2, col = 4)
```

```r
# Define the multiple linear regression equation
RHIOA2014.regression = RHIOAP14L ~ (BedOAR14C + AvgOHD14C +
     MedEqR14C + DocDenR14C + NursesR14C)

### MLR (with OLS) (Y = αι_n + βX + ϵ)
RHIOA2014.regression.ols = lm(RHIOA2014.regression, data =
    ITRHM2014.data)
summary(RHIOA2014.regression.ols)
qqPlot(RHIOA2014.regression.ols, ylab = "Studentized
    residuals (RHIOAP14L)")

# Measures of collinearity
vif(RHIOA2014.regression.ols)
colldiag(RHIOA2014.regression.ols)

# Measures of goodness of fit
AIC(RHIOA2014.regression.ols)
BIC(RHIOA2014.regression.ols)

# Moran's I test for spatial autocorrelation in the
    residuals from the estimated linear regression model
lm.morantest(RHIOA2014.regression.ols,
    PMC.neighbours.queen1.listw) # Positive spatial
    autocorrelation

# Specifications tests to examine the spatial dependence
    from the linear regression model: LMlag, LMerr, RLMlag,
    RLMerr and SARMA
lm.LMtests(RHIOA2014.regression.ols,
    PMC.neighbours.queen1.listw, test = c("LMlag", "LMerr",
    "RLMlag", "RLMerr", "SARMA")) # RLMlag provides the main
     significant test result
```

```r
# Positive spatial autocorrelation is present in the residuals from the estimated linear regression model, therefore proceed with further statistical spatial models: SLX, SAR, SEM, SDM, SDEM and SARAR

### SLX (Y = αι_n + βX + θWX + ε)
RHIOA2014.regression.slx = lmSLX(RHIOA2014.regression, data = ITRHM2014.data, PMC.neighbours.queen1.listw)
summary(RHIOA2014.regression.slx)
impacts(RHIOA2014.regression.slx, listw = PMC.neighbours.queen1.listw)
summary(impacts(RHIOA2014.regression.slx, listw = PMC.neighbours.queen1.listw.NAdrop, R = 999), zstats = TRUE)

# Measures of goodness of fit
AIC(RHIOA2014.regression.slx)
BIC(RHIOA2014.regression.slx)

### SAR (Y = ρWY + αι_n + βX + ε)
RHIOA2014.regression.sar = lagsarlm(RHIOA2014.regression, data = ITRHM2014.data, PMC.neighbours.queen1.listw)
summary(RHIOA2014.regression.sar)
impacts(RHIOA2014.regression.sar, listw = PMC.neighbours.queen1.listw.NAdrop)
summary(impacts(RHIOA2014.regression.sar, listw = PMC.neighbours.queen1.listw.NAdrop, R = 999), zstats = TRUE)

# Spatial Breusch-Pagan test for heteroskedasticity
bptest.sarlm(RHIOA2014.regression.sar, studentize = TRUE)
```

```
# Measures of goodness of fit
AIC(RHIOA2014.regression.sar)
BIC(RHIOA2014.regression.sar)
RHIOA2014.regression.sar.pseudoR2 = 1 - ((RHIOA2014.regression.sar $ SSE) / (var(na.omit(ITRHM2014.data $ RHIOAP14L))*(length(na.omit(ITRHM2014.data $ RHIOAP14L)) - 1)))

### SEM ($Y = \alpha\iota_n + \beta X + \epsilon$, $\epsilon = \lambda W\epsilon + \mu$)
RHIOA2014.regression.sem = errorsarlm(RHIOA2014.regression, data = ITRHM2014.data, PMC.neighbours.queen1.listw)
summary(RHIOA2014.regression.sem)

# Spatial Hausman test for consistency of estimates
Hausman.test(RHIOA2014.regression.sem)

# Spatial Breusch-Pagan test for heteroskedasticity
bptest.sarlm(RHIOA2014.regression.sem, studentize = TRUE)

# Measures of goodness of fit
AIC(RHIOA2014.regression.sem)
BIC(RHIOA2014.regression.sem)
RHIOA2014.regression.sem.pseudoR2 = 1 - ((RHIOA2014.regression.sem $ SSE) / (var(na.omit(ITRHM2014.data $ RHIOAP14L))*(length(na.omit(ITRHM2014.data $ RHIOAP14L)) - 1)))

### SDM ($Y = \rho WY + \alpha\iota_n + \beta X + \theta WX + \epsilon$)
RHIOA2014.regression.sdm = lagsarlm(RHIOA2014.regression, data = ITRHM2014.data, PMC.neighbours.queen1.listw, type = "mixed")
summary(RHIOA2014.regression.sdm)
```

```
impacts(RHIOA2014.regression.sdm, listw = PMC.neighbours.queen1.listw.NAdrop)
summary(impacts(RHIOA2014.regression.sdm, listw = PMC.neighbours.queen1.listw.NAdrop, R = 999), zstats = TRUE)

# Likelihood ratio tests for restrictions to nested models
LR.sarlm(RHIOA2014.regression.sdm, RHIOA2014.regression.sar) # SDM to SAR
LR.sarlm(RHIOA2014.regression.sdm, RHIOA2014.regression.sem) # SDM to SEM
LR.sarlm(RHIOA2014.regression.sdm, RHIOA2014.regression.slx) # SDM to SLX
LR.sarlm(RHIOA2014.regression.sdm, RHIOA2014.regression.ols) # SDM to MLR

# Spatial Breusch-Pagan test for heteroskedasticity
bptest.sarlm(RHIOA2014.regression.sdm, studentize = TRUE)

# Measures of goodness of fit
AIC(RHIOA2014.regression.sdm)
BIC(RHIOA2014.regression.sdm)
RHIOA2014.regression.sdm.pseudoR2 = 1 - ((RHIOA2014.regression.sdm $ SSE) / (var(na.omit(ITRHM2014.data $ RHIOAP14L))*(length(na.omit(ITRHM2014.data $ RHIOAP14L)) - 1)))

### SDEM ($Y = \alpha\iota_n + \beta X + \theta WX + \epsilon$, $\epsilon = \lambda W\epsilon + \mu$)
RHIOA2014.regression.sdem = errorsarlm(RHIOA2014.regression, data = ITRHM2014.data, PMC.neighbours.queen1.listw, etype = "emixed")
```

```r
summary(RHIOA2014.regression.sdem)
impacts(RHIOA2014.regression.sdem, listw =
   PMC.neighbours.queen1.listw)
summary(impacts(RHIOA2014.regression.sdem, listw =
   PMC.neighbours.queen1.listw, R = 999), zstats = TRUE)

# Likelihood ratio tests for restrictions to nested models
LR.sarlm(RHIOA2014.regression.sdem,
   RHIOA2014.regression.sem) # SDEM to SEM
LR.sarlm(RHIOA2014.regression.sdem,
   RHIOA2014.regression.slx) # SDEM to SLX
LR.sarlm(RHIOA2014.regression.sdem,
   RHIOA2014.regression.ols) # SDEM to MLR

# Spatial Hausman test for consistency of estimates
Hausman.test(RHIOA2014.regression.sdem)

# Spatial Breusch-Pagan test for heteroskedasticity
bptest.sarlm(RHIOA2014.regression.sdem, studentize = TRUE)

# Measures of goodness of fit
AIC(RHIOA2014.regression.sdem)
BIC(RHIOA2014.regression.sdem)
RHIOA2014.regression.sdem.pseudoR2 = 1 - ((
   RHIOA2014.regression.sdem $ SSE) / (var(na.omit(
   ITRHM2014.data $ RHIOAP14L))*(length(na.omit(
   ITRHM2014.data $ RHIOAP14L)) - 1)))

### SARAR $(Y = \rho WY + \alpha\iota_n + \beta X + \epsilon,\ \epsilon = \lambda W\epsilon + \mu)$
RHIOA2014.regression.sarar = sacsarlm(RHIOA2014.regression,
    data = ITRHM2014.data, PMC.neighbours.queen1.listw,
   type = "sac")
```

```r
summary(RHIOA2014.regression.sarar)
impacts(RHIOA2014.regression.sarar, listw =
    PMC.neighbours.queen1.listw.NAdrop)
summary(impacts(RHIOA2014.regression.sarar, listw =
    PMC.neighbours.queen1.listw.NAdrop, R = 999), zstats =
    TRUE)

# Likelihood ratio tests for restrictions to nested models
LR.sarlm(RHIOA2014.regression.sarar,
    RHIOA2014.regression.sem) # SARAR to SEM
LR.sarlm(RHIOA2014.regression.sarar,
    RHIOA2014.regression.sar) # SARAR to SAR
LR.sarlm(RHIOA2014.regression.sarar,
    RHIOA2014.regression.ols) # SARAR to MLR

# Spatial Breusch-Pagan test for heteroskedasticity
bptest.sarlm(RHIOA2014.regression.sarar, studentize = TRUE)

# Measures of goodness of fit
AIC(RHIOA2014.regression.sarar)
BIC(RHIOA2014.regression.sarar)
RHIOA2014.regression.sarar.pseudoR2 = 1 - ((
    RHIOA2014.regression.sarar $ SSE) / (var(na.omit(
    ITRHM2014.data $ RHIOAP14L))*(length(na.omit(
    ITRHM2014.data $ RHIOAP14L)) - 1)))
```

Listing B.2: Data analysis (R)

Bibliography

[1] Gerard F. Anderson, Uwe E. Reinhardt, Peter S. Hussey, and Varduhi Petrosyan. It's The Prices, Stupid: Why The United States Is So Different From Other Countries. *Health Affairs*, 22(3):89–105, 2003. PMID: 12757275.

[2] Luc Anselin. Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, 20(1):1–17, 1988.

[3] Luc Anselin. Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2):93–115, 1995.

[4] Luc Anselin and Sergio J. Rey. *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL*. GeoDa Press, Chicago, 2014.

[5] Luc Anselin, Ibnu Syabri, and Youngihn Kho. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis*, 38(1):5–22, 2006.

[6] Silvia Balia, Rinaldo Brau, and Emanuela Marrocu. Interregional patient mobility in a decentralized healthcare system. *Regional Studies*, 52(3):388–402, 2018.

[7] David A. Belsley. *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. Wiley Series in Probability and Statistics. Wiley, 1991.

[8] Jonathan Bendor and Dilip Mookherjee. Institutional Structure and the Logic of Ongoing Collective Action. *The American Political Science Review*, 81(1):129–154, 1987.

[9] Elenka Brenna and Federico Spandonaro. Regional Incentives and Patient Cross-Border Mobility: Evidence from the Italian Experience. *International Journal of Health Policy and Management*, 4(6):363–372, 2015.

[10] Andrew D. Cliff and John K. Ord. *Spatial Autocorrelation*. Monographs in spatial and environmental systems analysis. Pion, London, 1973.

[11] Avedis Donabedian. Evaluating the Quality of Medical Care. *The Milbank Quarterly*, 83(4):691–729, 2005.

[12] J. Paul Elhorst. Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis*, 5(1):9–28, 2010.

[13] Dennis Epple and Richard E. Romano. Ends against the middle: Determining public service provision when there are private alternatives. *Journal of Public Economics*, 62(3):297 – 325, 1996.

[14] Giovanni Fattore, Giuseppina Petrarca, and Aleksandra Torbica. Traveling for care: Inter-regional mobility for aortic valve substitution in Italy. *Health Policy*, 117(1):90–97, 2014.

[15] Raymond J. G. M Florax, Hendrik Folmer, and Sergio J. Rey. Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Regional Science and Urban Economics*, 33(5):557 – 579, 2003.

[16] R. C. Geary. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5(3):115–146, 1954.

[17] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press, 2006.

[18] Garrett Hardin. The Tragedy of the Commons. *Science*, 162(3859):1243–1248, 1968.

[19] J. A. Hausman. Specification Tests in Econometrics. *Econometrica*, 46(6):1251–1271, 1978.

[20] Joop J. Hox and Hennie R. Boeije. Data Collection, Primary vs. Secondary. In Kimberly Kempf-Leonard, editor, *Encyclopedia of Social Measurement*, pages 593 – 599. Elsevier, New York, 2005.

[21] William C. Hsiao. Abnormal economics in the health sector. *Health Policy*, 32(1):125 – 139, 1995. Health Sector Reform in Developing Countries: Making Health Development Sustainable.

[22] Carlos M. Jarque and Anil K. Bera. A Test for Normality of Observations and Regression Residuals. *International Statistical Review / Revue Internationale de Statistique*, 55(2):163–172, 1987.

[23] Harry H. Kelejian and Ingmar R. Prucha. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121, July 1998.

[24] Michael H. Kutner, Chris Nachtsheim, John Neter, and William Li. *Applied Linear Statistical Models*. McGraw-Hill Irwin series Operations and decision sciences. McGraw-Hill Irwin, 2005.

[25] J. LeSage and R. K. Pace. *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. CRC Press, 2009.

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

[27] Thomas R. Malthus. An Essay on the Principle of Population, as it Affects the Future Improvement of Society with Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers. 1798.

[28] Charles F. Manski. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542, 07 1993.

[29] Gabriele Messina, Silvia Forni, Francesca Collini, Cecilia Quercioli, and Nicola Nante. Patient mobility for cardiac problems: a risk-adjusted analysis in Italy. *BMC Health Services Research*, 13(1):56, Feb 2013.

[30] Gabriele Messina, Nicola Vigiani, Lucia Lispi, and Nicola Nante. Patient migration among the Italian regions in 2003. *Italian Journal of Public Health*, 5:45–52, 03 2008.

[31] Marcello Montefiori. Spatial Competition for Quality in the Market for Hospital Care. *The European journal of health economics : HEPAC : health economics in prevention and care*, 6:131–5, 04 2005.

[32] Patrick A. P. Moran. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):243–251, 1948.

[33] Patrick A. P. Moran. Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2):17–23, 1950.

[34] John A. Nyman. American Health Policy: Cracks in the Foundation. *Journal of Health Politics, Policy and Law*, 32(5):759–783, 10 2007.

[35] OECD. *OECD Reviews of Health Care Quality: Italy 2014*. 2015.

[36] OECD / European Observatory on Health Systems and Policies. *Italy: Country Health Profile 2017*. 2017.

[37] Mancur Olson. *The Logic of Collective Action*. American studies collection. Harvard University Press, 1971.

[38] Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Political Economy of Institutions and Decisions. Cambridge University Press, 1990.

[39] R. Kelley Pace and James P. LeSage. A spatial Hausman test. *Economics Letters*, 101(3):282 – 284, 2008.

[40] E. Pierini, M. Pioppo, G. Troiano, P. Casucci, O. Checconi, F. Ruffini, G. Messina, and N. Nante. Patient mobility for bone marrow transplant: the experience of the Perugia Hospital, years 2000-2013. *Ann Ig*, 27(5):769–776, Sep-Oct 2015.

[41] Katri K. Sieberg and Olga Shevtsova. Cost efficiency and coverage problems in US Healthcare. *Hallinnon tutkimus*, 31(2):131–137, 2012.

[42] Charles M. Tiebout. A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64(5):416–424, 1956.

[43] W. R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(sup1):234–240, 1970.

[44] Federico Toth. How health care regionalisation in Italy is widening the North–South gap. *Health Economics, Policy and Law*, 9(3):231–249, 2014.

List of Figures

List of Tables