POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

The efficiency of Italian lower secondary schools: combination of DEA with a graphical machine learning approach

TESI DI LAUREA MAGISTRALE IN
MANAGEMENT ENGINEERING

Author: **Luigi Iorio**

Student ID:	10794393
Advisor:	Prof. Tommaso Agasisti
Co-advisor:	Prof. Mara Soncin
Academic Year:	2022-23

Ai miei genitori:

Mariarosaria e Raffaele

Alla mia famiglia:

Rita, Ivan, Asia e i Nonni

Abstract

This study illustrates the ability of Machine Learning approaches to overcome classical regression techniques in identifying nonlinear relationships and interaction effects of factors influencing the efficiency of educational systems. The efficiency scores of 4264 Italian public lower secondary schools are computed by using Data Envelopment Analysis (DEA) with a double bootstrap procedure, while Random Forest is adopted to identify the variables that are associated with higher scores and visualise their effects in an easily interpretable way. The results confirm the differences in efficiency assessed in previous studies between schools located in different areas of the country, with schools in the South and the islands performing worse than the others. School size, class size and the percentage of immigrant students are the most important factors influencing efficiency. Evaluating interaction effects, it is also possible to consider in the assessment of efficiency the years of experience of the school head, the days of absence of teachers (not due to illness or maternity) and the percentage of funds allocated to pay salaries. Policy implications are presented in the last part of the study.

Keywords: school efficiency, data envelopment analysis, machine learning, random forest

Abstract in lingua italiana

Questo studio illustra la capacità degli approcci di Machine Learning di superare le classiche tecniche di regressione nell'identificare relazioni non lineari ed effetti di interazione dei fattori che influenzano l'efficienza dei sistemi educativi. I punteggi di efficienza di 4264 scuole secondarie inferiori pubbliche italiane sono calcolati utilizzando la Data Envelopment Analysis (DEA) con una procedura con doppio bootstrap, mentre un algoritmo di Random Forest è adottato per identificare le variabili che sono associate a punteggi più elevati e visualizzarne gli effetti in modo facilmente interpretabile. I risultati confermano le differenze di efficienza riscontrate in studi precedenti tra le scuole situate in diverse aree del Paese, con le scuole del Sud e delle isole che ottengono risultati peggiori rispetto alle altre. Le dimensioni della scuola, le dimensioni della classe e la percentuale di studenti immigrati sono i fattori più importanti che influenzano l'efficienza. Valutando gli effetti di interazione, è possibile includere nella valutazione dell'efficienza anche gli anni di esperienza del dirigente scolastico, i giorni di assenza dei docenti (non per malattia o maternità) e la percentuale di fondi destinati al pagamento degli stipendi. Le implicazioni politiche sono presentate nell'ultima parte dello studio.

Parole chiave: efficienza scolastica, data envelopment analysis, machine learning, random forest

List of Figures

List of Tables

Contents

1. Introduction

One of the main objectives of the schools is to promote the acquisition of knowledge, skills and competences, so that all students can develop their personal civic and professional identity. The economic system is inevitably linked to such knowledge, since, before being based on the exploitation of tangible resources and capital, is based on ideas and values that lead individuals to seek a higher degree of well-being, of the individual as well as of the community.

Education plays an important role in the process of economic growth, with all countries, not only the rich ones, investing a non-negligible share of their Gross Domestic Product in this sector. According to estimates of Organization for Economic Cooperation and Development (OECD), the average education spending for OECD countries in 2019 was 4.9% of GDP.

Countries spend so much on education, even though assessing the return on that investment is a very complex task, often dependent on factors that are not immediately or obviously measurable. The vast literature of cross-country growth regressions has provided support for the claim of a significant positive association between quantitative measures of schooling and economic growth (OECD 2010).

For a long time, the link between education and economic development of a country has been measured by the number of years of schooling attended, whereas now this relationship is established by considering the knowledge and skills acquired during the school years, the measurement of which has been made available by the spread of international learning surveys, such as the OCDE PISA tests.

Using data from these tests, Hanushek and Kimko (2000) demonstrated a statistically significant positive effect of cognitive skills on economic growth. A country-level standard deviation higher test performance, equivalent to 47 test-score points in the PISA 2000 mathematics assessment, would have yielded around 1% higher annual growth rate, which is a very large value if considering that the average annual growth of OECD countries was at that time roughly 1.5%.

Having established the importance for countries to invest in education, it remains to be seen how efficiently these resources are used by schools. Scholars have long been looking for ways to measure the efficiency of educational systems and identify the key aspects that influence those measures. A common objective in the literature has been to investigate school performance by looking for technical relationships between schools' outputs and inputs.

1.1 Motivation and research questions

In this context, this thesis has been carried out, precisely by examining the Italian situation, where, compared to the international scene, there are fewer studies assessing the efficiency of the education system.

The Italian education system has been under discussion since the early 2000s, and several reforms have been undertaken since then. In recent years, not least because of the pandemic, schools have once again become the focus of public debate, revealing all the organisational difficulties in a phase of extraordinary emergency, but also highlighting the problems that have accumulated over time.

When comparing the results of Italian students with peers in other countries, international assessments such as OECD PISA tests consistently report lower scores. While it is true that its education system is often spoken of for its quality and professionalism, data portray Italy as a State far from the standards of other OECD countries, also facing the problem of brain drain across borders.

In 2019, Italy invested 3.8% of GDP in education, quantifiable in about 70 billion euros and equal to only 8% (last country in Europe) of total public spending. This represents a long-term trend, considering that the drop in education spending followed the 2008 recession, with the consequent contraction of public budgets dedicated to this item.

However, looking at expenditures per student at primary and secondary school, Italy reported values in line with those of OECD countries, thus questioning the efficiency of the entire education system. Similar amount of resources per student, associated to poorer performance, are a sign that these resources have not often been used in the proper way, and many were even lost.

Keeping track of the relative position occupied by the Italian education sector in the international scenario is essential to understand how to overcome the difficulties and shortcomings that currently exist.

Policy makers generally seek information to justify budget for results and they want to ensure that schools use efficiently the resources to improve. If the gains from change are not too large, it may not make sense to politicians and decision-makers to take any moves unpopular with the existing educational establishment.

In this regard, this work aims to provide insights for possible policy implications by answering the following two main research questions:

1. What is the level of relative efficiency of lower secondary schools in Italy?
   This evaluation is made using Data Envelopment Analysis (DEA) on 2019 data from INVALSI and the Ministry of Education, with the results showing some differences in efficiency between different areas of the country.
2. What are the factors that are most able to explain these differences in efficiency? Specifically, this other analysis is conducted adopting advanced Machine Learning techniques, such as Random Forest, as an alternative to the classic two-stage approaches.

In recent years, to answer the second research question, several studies have adopted double bootstrap procedures, as not only do they make it possible to obtain school efficiency values that are statistically robust, but they also evaluate a consistent data generation process. However, modelling an educational system is very complex due to the multiplicity of its components and the interactions between them (Rebai et al., 2020), which cannot be fully explained even by those procedures.

The adoption of Machine Learning allows to reveal pertinent non-linear interactions and find hidden insights for a more complete picture of the school environment. This thesis contributes to extend the union between the fields of education and ML, since, despite a recognised potential and a widespread use in other various sectors of the economy, ML applications to education count very few studies. In addition, it is the first to apply ML to the evaluation of Italian schools.

1.2 Structure

The remainder of the work is organized as follows: the second chapter proposes a review of the literature related to school efficiency, while the third provides a description of the Italian school system and its performance compared to that of the other countries taking part in the 2018 OECD PISA tests.

The fourth chapter describes the methods adopted to measure the relative efficiency of each school and to investigate the discriminating factors of such efficiency, with the variables used at both stages of the analysis described in the fifth chapter.

The sixth presents the main results, and the seventh and final chapter provides some policy indications that can be drawn from the empirical analysis, as well as conclusions of a more general nature.

2. Efficiency studies on education

2.1 Introduction

The first step of the analysis was a review of the literature about the efficiency of the educational sector. Going through the wide range of works on educational productivity and efficiency, allowed to have a clear picture of the state-of-the-art, and to understand the main points on which to focus the analysis and the most used methodologies.

To assess school efficiency, it was first necessary to understand how to measure this efficiency. As schools can be considered as organizations whose main objective is to increase the knowledge of students, typically, student achievement was chosen as a possible assessable output to measure efficiency in an objective way.

Since Hanushek (1986), there has been the intent to evaluate these outputs by looking for their relationships with possible input resources provided by the schools. Sustaining that education could influence economic growth, he emphasised that school policies should focus on the quality of education and the evaluation of student performance, rather than on increasing school resources, such as money and teaching staff.

This performance also depends on other aspects such as students' learning capabilities or the social environment in which they live, and it is not always easy to separate the specific role of the school in improving student learning, from such other factors. That is why it is relevant, and at the same time challenging, to consider which are the variables to include into the analysis and determine what the true impact of the school itself is.

Different methodologies have been adopted to evaluate school efficiency. According to Bessent et al. (1982), the most employed method at that time was the least squares linear regression with a single output. A variant of that model instead, included the repetitive use of the linear regression to evaluate the effects of the same inputs on different outputs, taken one at a time. However, over the years, most of the studies converged towards the use of Data Envelopment Analysis (DEA), as the most suitable method to adopt, with more and more authors adding important contributions.

DEA is a non-parametric method used to estimate the efficiency of a set of homogeneous decision-making units (DMUs) (e.g., companies, hospitals, schools, banks). Its objective is to evaluate the relative efficiency of those decision-making units in exploiting some inputs to generate outputs, identifying the best practices in each sector. This is done by pointing out efficiency benchmarks among DMUs to identify those that are less efficient and could improve their performance.

It is crucial to consider and define the level of the analysis in the evaluation of school efficiency. Due to the difficulties of calculation and data availability, early studies on this topic focused on large decision-making units, such as school districts. Over the years, scholars have been able to choose the most appropriate DMU for their analyses, but in general, these have been carried combining channels to make them multilevel. In any case, it is important to recognise the importance of defining the decision unit to ensure accurate and meaningful results in the evaluation of school performance.

Since this thesis relies on the use of this practice, DEA will be explored in more detail in a dedicated section. The whole review of the literature was restricted to papers that used it as method for the evaluation of the efficiency and was mainly conducted following five different perspectives:

- application of DEA to the educational sector;
- evaluation of Italian schools' efficiency through use of DEA;
- comparison between public and private schools' performance;
- origins and development of the methodology;
- combination of machine learning techniques with DEA.

Then, within the section regarding the use of DEA in the educational sector, a classification of the different variables employed by scholars to assess academic efficiency was provided.

Particular attention was paid to Italian studies to understand which aspects had already been explored and which areas required further investigation and future development. Defining the limits of current knowledge and reflecting on the paths that this work should follow to expand knowledge, were fundamental steps to advance research.

Studies concerning competition were also analysed, even though this work does not consider any distinction in terms of public and private schools' performance, due to the lack of data concerning private schools.

Most of the authors who added relevant contributions to the assessment of school efficiency with DEA are mentioned in this chapter, while papers strictly regarding the origins and the development of the DEA itself, included its bootstrapped version, are reported in the chapter explaining the methodology.

The selection was conducted through Scopus, an online bibliographic database owned by the publishing company Elsevier, which collects information on scientific articles, book reviews, conference proceedings and other academic material published on various disciplines.

2.2 School efficiency through DEA

Bessent et al. (1982) was one of the first works to underline the advantages of DEA over other techniques. While comparing 167 elementary schools in the urban school district of Houston in California, out of which 89 were found to effectively utilize their resources, they pointed out some deficiencies in prior methods, which are listed below.

These methods did not consider multiple output simultaneously; they relied on regression coefficients that did not necessarily indicated the most efficient way to produce an impact on outputs; it was not well clear the difference of meaning between technical efficiency and other forms of efficiency when referring to school; the relationship between inputs and outputs could not be linear or independent, with also possibility of multicollinearity problems; previous experiments where it was possible to manipulate inputs were missing.

The sensitivity and the robustness of the efficiency measurement obtained through DEA were then tested by Sengupta (1987) by means of empirical applications, showing how stochastic variations of inputs and outputs could be considered. DEA does not assume a fixed production function and can handle variations in scale.

In the same year, Smith and Mayston (1987) used DEA to assess the performance of U.K. local education authorities and they found that the exclusion of an important output could significantly distort the results of the analysis. As it will be explained later, the efficiency of a decision-making unit is determined by the ratio of its outputs to its inputs. If an important output was not considered, the resulting efficiency score could not accurately reflect the DMU's true efficiency, leading to an underestimation of its performance. The study also investigated the impact of omitting certain inputs and selecting only the most efficient authorities on the analysis. It found that these factors could also affect the results of the analysis, but to a lesser extent than omitting an important output.

However, the methodology was employed by several authors like Ganley and Cubbin (1992), Thanassoulis and Dunstan (1994), and Bates (1997), on U.K. local education authorities, or Ruggiero (1996), Engert (1996), Duncombe et al. (1997), Chalos and Cherian (1995), and Chalos (1997), in studies of New York and Illinois school districts.

A turning point in the use of the model was constituted by Ray (1991). It was a study that examined resource utilisation efficiency in public schools in Connecticut, USA. The study used data from 1989-1990 to analyse several factors that could affect resource utilisation efficiency, including student to teacher ratios, spending per student, spending on support services, and school size. One of its objectives was to identify differences in efficiency in the use of resources between public schools and to identify the causes of these differences. The results of the study showed that public schools in Connecticut differed significantly in terms of resource utilisation efficiency.

For instance, the study found that larger schools tended to be more efficient in their use of resources than smaller schools. In addition, schools with a higher expenditure per student tended to be less efficient in the use of resources, which could indicate a possible waste of resources. The ratio of students to teachers was also identified as an important factor that might affect resource efficiency, as schools with a lower ratio tended to be more resource efficient.

The particularity of this publication, however, lied in the fact that it combined DEA with regression techniques. The result was a two-stage analysis in which it was possible to distinguish factors affecting the measurements in inputs provided by schools and some other socio-economic aspects. In the first stage, DEA relied on school inputs only, while, in the second stage, the efficiency scores previously returned were regressed against socio-economic factors.

This procedure allowed not only to assess objectively the efficiency of decision-making units, but also, with the second stage, to identify factors influencing efficiency. This could help educational authorities identify schools in need of improvement and identify best practices to share with other schools. From then on, the two-stage analysis has surely been the most adopted procedure when applying DEA, even if it followed several attempts to improve it.

For instance, McCarty and Yaisawarng (1993) observed that the two-stage approach could be problematic when there was strong correlation between the first-stage inputs and the second-stage independent variables. To measure the technical efficiency of 161 New Jersey school districts during the 1989-1990, they collected data on the inputs (such as total expenditures and number of teachers) and outputs (such as test scores and graduation rates) of each district and estimated technical efficiency scores using DEA.

Additionally, they examined the relationship between district characteristics (e.g., district size, per-pupil expenditure) and this efficiency. They found that larger districts tended to be more efficient, as did districts with higher per-pupil expenditures. However, the relationship between other district characteristics (such as student demographic characteristics) and efficiency was less clear.

In the second stage, they applied a Tobit regression. It is a regression model used to analyse continuous dependent variables that can take either censored or truncated values, where a variable is censored when its values are only known in a certain range, while it is truncated when values outside a certain range are unobservable. The adoption of this kind of regression was possible considering that efficiency scores returned by DEA in the first stage, or their reciprocal, assumes values between 0 and 1.

Arnold et al. (1996) experimented new combinations of DEA and regression models. DEA first allowed to identify the group of public secondary schools in Texas that were efficient, out of a total of 638 schools. Variables used to assess efficiency include the number of students, number of teachers, school budget, expenditure per student, student dropout rate, exam participation rate, class size and number of years of teacher experience.

The researchers then employed a single regression analysis that incorporated dummy variables to differentiate between schools that were efficient and those that were not. This enabled them to evaluate the entire collection of schools in a comprehensive manner. The analysis showed that some schools produced desired outputs using fewer inputs than others, while statistical regression revealed that school budget, expenditure per student and student dropout rate were among the factors that most influenced school performance. Efficient schools had positive input coefficients, while inefficient schools showed negative coefficients that were also statistically significant.

The two-stage model was also adopted by Kirjavainen and Loikkanent (1998), evaluating efficiency differences among Finnish senior secondary school, and Bradley et al. (2001), over all secondary schools in England between 1993 and 1998. These authors were among the first ones to include in the analysis the aspect of competition among public and private schools.

Kirjavainen and Loikkanent (1998) focused on analysing efficiency differences between Finnish high schools using DEA and Tobit regression, reinforcing the usefulness of this kind of regression model in education research. In this study, DEA considered costs and student performance, while Tobit analysis was used to analyse the relationship between school efficiency and contextual factors such as school size, percentage of students with special needs and geographical location of the school.

The second stage Tobit regression was widely adopted by other authors such as Duncombe et al. (1997), Afonso et Aubin (2006), Agasisti (2013).

The study of Bradley et al. (2001) instead, focused on the English school system reform of the 1980s, which introduced competition between schools to improve the quality of education offered. He used different measures of efficiency and considered the socio-economic level of students and the size of schools among the factors that could influence results.

Muniz (2002) described a new method to separate management inefficiency from external variables that influence the organisation's performance, such as technology and the competitive environment. In order to improve its results and interpretation, the method proposed a modification to the three-stage method developed by Fried and Lovell (1996), but exclusively basing on DEA model.

It involved several DEA sequential stages: in the first stage, only the variables controllable by the organisation, i.e., those dependent on management, were used, while the non-controllable inputs were ignored, and initial efficiency estimates for each DMU were produced. The inefficiency detected here was due to both the real technical inefficiency and the influence of the non-controllable inputs.

The aim of the second stage was to distinguish between and quantify the real technical inefficiency that the producer exhibited and the influence of the non-controllable inputs. The third stage consisted of a DEA in which adjusted values of outputs and controllable inputs were used for each DMU.

By comparing the results obtained in the first two stages of the analysis, it was possible to separate management inefficiency from variables external to management. In this way, it was possible to obtain a more accurate measure of the organisation's management efficiency and identify areas where performance could be improved.

This paper concluded with some application examples of Muniz's proposed method in the education sector, which demonstrated the usefulness of separating management inefficiency from external variables for an accurate assessment of organisations' efficiency. 62 public high schools located in the Spanish region of Asturias during the academic year 1996-1997 were evaluated and the results obtained showed that the student socio-economic and family status (a non-controllable input) had a direct influence on the school results.

Differently from previous works, Wanke et al. (2016) applied, for the first time, the DEA model to ensure that cost and learning efficiency levels were simultaneously optimized. As also pointed out by Agasisti (2013), according to the definition of efficiency provided by Farrell (1957), DEA was used to measure technical efficiency by assessing the ability to maximize outputs based on a given set of inputs.

This measurement did not consider the optimization of input usage based on their price, which is known as allocative efficiency. It was possible that schools deemed efficient using DEA may still be cost-inefficient, as a different combination of inputs at their respective prices could lead to a better solution. However, in the public sector, input prices may not always be available, making the focus on technical efficiency a necessity.

They collected data from a sample of 450 Australian schools and analysed the relationship between school costs and learning efficiency. Specifically, they sought to identify factors that influenced school costs and learning efficiency, such as the number of students, dropout rate and teacher qualification. The results of the study indicated that schools that had more students, qualified teachers and a low dropout rate had a lower unit cost of learning and higher learning efficiency than schools having fewer students, less qualified teachers and a higher dropout rate.

Worthington (2001) highlighted that the challenge of defining input costs in many public-sector contexts often makes it difficult to obtain this type of information. Consequently, many education efficiency studies tended to focus solely on measuring technical efficiency, which was particularly suitable for the DEA approach. This emphasis on technical efficiency measurement is the most common in the literature, as the lack of input cost information is a prevalent issue.

He examined different techniques for measuring efficiency in the education sector, including approaches based on DEA, analysing their applications, advantages and limitations. He observed that also other practices like Stochastic Frontier Approach (SFA) could provide an assessment of the relative efficiency of different organisations or educational institutions, however, each technique had its own strengths and weaknesses.

For instance, DEA could be used to measure the relative efficiency of organisations that produced multiple outputs, but it did not consider quality differences between the outputs. On the other hand, SFA could be used to measure the relative efficiency of organisations producing a single output but required the assumption of a specific production model. Overall, Worthington concluded that both frontier efficiency measurement techniques could be useful in analysing efficiency in education, but that it was important to choose the appropriate approach according to the specific needs of the analysis.

The importance of the contribution of Worthington (2001) consisted in providing not only a useful overview of frontier efficiency measurement techniques and their application in education, but also a comparison between all the possible variables that were considered in the analysis of previous scholars, offering valuable guidance for researchers and analysts working in this field.

Indeed, while most of reviewed studies agreed on using students' achievements as output in the DEA model, they adopted different choices for what concerns input variables and, in the two-stage-analyses, non-discretionary variables. Once that the two-stage procedures took over among scholars, the choice of different variables could constitute one of the main determining factors in achieving different results.

The learning achievements of students could be measured in different ways, such as through standardised tests or teacher evaluations. It is important to note that although student achievement was a common output used in DEA, it was not the only measure of school performance. Other measures, such as drop-out rates, parental participation and community involvement, could also provide important insights into the quality of education provided by a school. The context in which a school operates and the resources at its disposal when assessing its overall performance should be considered.

The use of different inputs in the DEA instead, makes it possible to assess the efficiency of schools based on their ability to produce educational results using available resources. Inputs are therefore a fundamental element, as they allow the efficiency of schools to be assessed objectively, comparing the resources used with the results obtained.

Finally, non-discretionary variables are factors that are outside the control of the schools, such as student demographics, school size or regional economic factors. By analysing non-discretionary variables, researchers could better understand the factors influencing school efficiency and identify areas for improvement.

The analysis conducted in this thesis, as most of the papers, relied on student achievement as output. No classification of other possible outputs was provided, and the author suggests the reading of the mentioned references in case more details are needed. On the other hand, an evaluation of the different choices of variables made by previous studies, both as inputs and non-discretionary variables, was reported in the two following sections.

As Worthington (2001), also Bradley et al. (2001) made a comparison between the different set of variables considered over the years. However, De Witte and Torres (2017), provided an extensive review of the literature on efficiency in education. They summarized the earlier applied inputs, outputs and contextual variables, as well as the used data sources of papers in the field of efficiency in education, and, seen its recent date of publication, it was taken as the most relevant source at this stage of the analysis.

2.2.1 Input variables

The input variables to consider in the usage of DEA depend on the objective of the analysis. Typically, inputs may include variables that are under control of schools, such as the number of teachers, number of students, school budget, available technological resources, number of programmes offered, size of school buildings, teacher qualifications, number of teaching hours offered and so on. Their choice must be carefully made to ensure that they are representative of the variables that influence the efficiency of the school.

Following the classification provided by De Witte and Torres (2017), the reviewed papers were grouped into four distinct groups, according to the different categories of utilized inputs. It was possible to distinguish between: factors pertaining to individual students, variables relating to families, aspects of educational institutions, and community-related factors.

Among student-related variables, prior achievement, even if in different forms like marks or test scores, was the broader used. Since student achievement was the most adopted output variable to measure student efficiency, the use of prior achievement allowed to have a direct comparison. From Bessent et al. (1982) to Cordero-Ferrera et al. (2015) the adoption of that input was present in several works over the years, as reported in Table 1.

Other variables regarding the behaviour of the students or their proficiency with the language were employed. Sarrico and Rosa (2009) used enrolment in advanced classes as a proxy to measure students' motivation, Grosskopf et al. (2014) used peer groups to control the potential impact on individual performance, while De Witte et al. (2010) used a variable to measure the language proficiency of the students.

Table 1. Overview of inputs: student-related variables

Inputs	Examples
Prior achievement	Bessent et al. (1982), Thanassoulis (1996), Portela and Thanassoulis (2001), De Witte et al. (2010), Podinovski et al. (2014), Cordero-Ferrera et al. (2015)
Motivation	Sarrico and Rosa (2009)
Peer group	Grosskopf et al. (2014)
Language proficiency	De Witte et al. (2010), Mancebón et al. (2012)

For what concerns family-related variables, socio-economic status and parental education were the most widely used. Socio-economic status was recognised to be a factor that could affect the efficiency of schools. The dispute among scholars was on either to consider it as an input in the DEA analyses, or as a non-discretionary variable. In Table 2 several studies that elaborated on its use as an input were reported.

For instance, Agasisti (2011a) and Mancebón et al. (2012), applied DEA to assess respectively the efficiency of Italian and Spanish universities using as input the socio-economic status index of the region where the university was located, Bradley et al. (2001) computed the proportion of pupils qualified for free education, identifying families requiring assistance, while Mizala et al. (2002), not only appraised the socioeconomic position of families but also assessed their access to fundamental services.

Parental education instead, was used as an input since Charnes et al. (1981). In their analysis of high schools in New Jersey, they found that schools with more educated parents tended to be more efficient in the use of their resources. A positive relationship between parental education and school efficiency was also found in the studies by Kirjavainen and Loikkanent (1998), Sarrico and Rosa (2009), Mancebón et al. (2012).

Table 2. Overview of inputs: family-related variables

Inputs	Examples
Socio-economic status	Charnes et al. (1981), Ruggiero (1996), Thanassoulis (1996), Bradley et al. (2001), Mizala et al. (2002), Sarrico and Rosa (2009), Agasisti (2011a, 2013), Mancebón and Muñiz (2008), Mancebón et al. (2012), Podinovski et al. (2014)
Parental education	Charnes et al. (1981), Kirjavainen and Loikkanent (1998), Sarrico and Rosa (2009), Mancebón et al. (2012)

As far as education institution variables are concerned, they surely represented the largest group of variables adopted in the literature. As said for the socio-economic index, also for education institution variables scholars were divided between those that adopted them as inputs and those considering them as non-discretionary variables. In this section, the first group of authors is reported, remanding to the next section for the other.

Sengupta (1987) was one of the first studies to use school size as an input in DEA, referring to number of students and number of teachers. These same variables were then adopted by Athanassopoulos and Shale (1997), Mizala et al. (2002), Burney et al. (2013) and Podinovski et al. (2014).

Some of the education institution variables were related to teachers, whether comparing the number of teachers with that of the student or considering their experience and salary. The teacher-to-student ratio, consisting in the ratio between the number of teachers and the number of students, or its reciprocal, was present in the analyses of Ray (1991), McCarty and Yaisawarng (1993), Mizala et al. (2002), Afonso and Aubyn (2006), Cordero-Ferrera et al. (2010) and Agasisti (2011a, b) (2013) (2014).

The number of years of service as a teacher was adopted by Ruggiero (1996), Kirjavainen and Loikkanent (1998), Agasisti et al. (2010), Agasisti et al. (2011), whereas Ruggiero (1996) (2000), Duncombe et al. (1997), Grosskopf et al. (2014) investigated the effects of considering the teacher salary.

Reducing class sizes and lowering teacher-to-student ratios could potentially improve students' academic performance. This could be attributed to teachers being able to employ more effective educational practices or because they were better able to give individual attention to each student.

Smith and Mayston (1987) showed that the inclusion of expenditure as an input in the DEA is important because many public organisations have an obligation to provide a certain level of output, regardless of the resources available. They were used in different studies ranging from expenditure per student, as for Cordero-Ferrera et al. (2010), to absolute expenditure, e.g., Grosskopf et al. (2014).

Finally, educational resources that schools made available to students were considered. Athanassopoulos and Shale (1997) employed the number of laboratories, Mancebon et al. (2012) and Agasisti (2013) used the proportion of computers connected to the web, as a proxy for the qualitative level of the school facilities.

Table 3. Overview of inputs: education institution variables

Inputs	Examples
Size (number of students, student per class, proportion of boys and girls)	Sengupta (1987), Athanassopoulos and Shale (1997), Mizala et al. (2002), Agasisti and Dal Bianco (2006, 2009), Mancebón et al. (2012), Burney et al. (2013), Podinovski et al. (2014), Royo & Faiardo (2020)
Student/teacher ratio (or viceversa)	Charnes et al. (1981), Bessent et al. (1982), Ray (1991), McCarty and Yaisawarng (1993), Mizala et al. (2002), Afonso and Aubyn (2006), Cordero-Ferrera et al. (2010), Agasisti (2011, 2013, 2014), Royo & Fajardo (2020)
Teacher experience	Bessent et al. (1982), McCarty and Yaisawarng (1993), Ruggiero (1996), Kirjavainen and Loikkanent (1998), Mizala et al. (2002), Sarrico and Rosa (2009), Agasisti et al. (2011), Royo & Fajardo (2020)
Teacher salary	Ruggiero (1996, 2000), Duncombe et al. (1997), Grosskopf et al. (2014)
Expenditures (teaching, research, administrators, supporting staff)	Bessent et al. (1982), Sengupta (1987), Smith and Mayston (1987), McCarty and Yaisawarng (1993), Ruggiero (1996), Duncombe et al. (1997), Muñiz (2002), Agasisti and Johnes (2009, 2010, 2015), Alexander et al. (2010), Cordero-Ferrera et al. (2010), Grosskopf et al. (2014), Rojo & Faiardo (2020)
Educational resources (Books, building, computers, bus, grants)	Ruggiero (1996), Athanassopoulos and Shale (1997), Agasisti and Dal Bianco (2006, 2009), Gimenez et al. (2007), Agasisti and Johnes (2009), Agasisti et al. (2011, 2012), Lee (2011), Mancebón et al. (2012), Agasisti (2013), Rojo & Faiardo (2020)

However, as noted by Worthington (2001), there was a lack of compelling empirical evidence, and it was highly questionable whether educational inputs could have a substantial positive effect on outcomes. Hanushek (2003) contended that there was no significant correlation between school resources and student performance.

Conversely, Mayston (1996) provided a different explanation for this finding, concluding that the levels of educational attainment and expenditure observed were endogenously determined, while many of the commonly used factors, such as class size, were widely claimed in the literature to represent inefficient policy. Therefore, whether it was appropriate to include these variables as inputs in efficiency studies or not, really depended on the specific context of the analysis.

Last group of inputs, according to De Witte and Torres (2017), was the one including the variables related to the community such as the features of the neighbourhood, its geographical position, the intensity of rivalry, and indicators representing different demographic factors like death rate, criminal activity, and influx of immigrants.

2.2.2 Non-discretionary variables

The impact of non-discretionary or environmental factors on academic outcomes has proven to be significant. Mainstream literature on educational efficiency primarily focused on examining the effects of structural, institutional, and socioeconomic variables on efficiency scores using various methodological approaches (Worthington 2001).

Like input variables, non-discretionary ones were classified into different categories at the student and family level, education institution level, and community level (the latter was not explored). Several scholars attempted to consider these variables as both input and contextual factors.

With respect to variables pertaining to students, it was examined how factors associated with race, ethnicity, minority status, or nationality impacted academic performance. Bradley et al. (2010) reported that a greater proportion of students from non-white ethnic backgrounds was positively associated with higher efficiency scores. On the contrary, most of the studies reported lower student results for immigrants.

Considering student disabilities, as in Johnes et al. (2012) and Grosskopf et al. (2014), brought to lower achievements and higher costs, due to their additional educational needs, whereas higher prior achievements resulted in higher student performance (Bonesrønning and Rattsø (1994) and Cherchye et al. (2010)). The remaining variables related to students were associated with other traits, such as their gender, or their proficiency in language.

Among family variables, the socio-economic status, e.g., McCarty and Yaisawarng (1993), Giménez et al. (2007), and parental educational, e.g., Afonso and Aubyn (2006), played a crucial role in determining student performance. These were widely accepted as a strong predictor of academic achievement around the world. Children tended to perform better if their parents had a higher socio-economic status or educational level.

Other studies considered the parental interaction with children, or the resources available at home, such as books and computers.

Table 4. Overview of non-discretionary variables

Non-discretionary variables	Examples
1. Student variables	
Gender	Thanassoulis (1999), Bradley et al. (2010), De Witte and Rogge (2010), Johnes et al. (2012)
Prior achievement	Bonesrønning and Rattsø (1994), Thanassoulis (1999), Cherchye et al. (2010), De Witte and Rogge (2010), Agasisti (2014)
Race/ethnicity/minority/ nationality	Ray (1991), Chalos (1997), Ruggiero (1999), Thanassoulis (1999), Bradley et al. (2010), Johnes et al. (2012)
Language proficiency	Duncombe et al. (1997), Ruggiero (1999), Cherchye et al (2010), Grosskopf et al. (2014)
Disabilities	Duncombe et al. (1997), Bradley et al (2010), Johnes et al. (2012), Grosskopf et al. (2014)

Table 4. (continued)

Non-discretionary variables	Examples
2. Family variables	
Socio-economic status	Ray (1991), McCarty and Yaisawarng (1993), Chalos (1997), Duncombe et al. (1997), Muñiz (2002), Giménez et al. (2007), Cordero-Ferrera et al. (2010), Alexander et al. (2010), Cherchye et al. (2010), Grosskopf et al. (2014), Agasisti et al. (2019)
Parental education	Duncombe et al. (1997), Muñiz (2002), Afonso and Aubyn (2006), Cherchye et al. (2010)
Relationship with children/involvement at home	Muñiz (2002), Giménez et al. (2007), Cordero-Ferrera et al. (2010), Agasisti (2011a, 2013)
Resources available at home/internet use	Giménez et al. (2007), Agasisti (2014).
3. Education institution variables	
Size (number of students/ class size/students/teacher ratio)	Duncombe et al. (1997), Alexander et al. (2010), Bradley et al. (2010), Agasisti (2011a, b, 2013), Johnes et al. (2012), Grosskopf et al. (2014), Kounetas et al. (2023)
Structure (enrollment/proportion of *boys and girls)*	Bradley et al. (2001), Alexander et al. (2010), Agasisti (2011a, 2013), Johnes et al. (2012)

Table 4. (continued)

Non-discretionary variables	Examples
Teacher characteristics (age/gender/education/ experience/number/salary)	Chalos (1997), Alexander et al. (2010), Bradley et al. (2010), De Witte and Rogge (2010), Johnes et al. (2012), Agasisti (2014), Kounetas et al. (2023)
Ownership (public, private, charter). Type of institution	Duncombe et al. (1997), Portela and Thanassoulis (2001), Bradley et al. (2001, 2010), Mizala et al. (2002), Alexander et al. (2010), Agasisti (2013)

Finally, school-related variables included several aspects. The impact of variables, such as class size, number of students, and teacher-student ratios, produced mixed results in various studies. On one hand, some studies suggested that smaller class sizes and lower teacher-student ratios could lead to improved academic performance because of better teaching practices and increased focus on individual students. On the other hand, other authors argued that these factors did not have a significant impact on students' academic results, particularly in secondary education.

Mixed results were also obtained when considering the ownership of the institutes. These were due to country-specific heterogeneity of the economic situation of the families, the level of competition among schools, the admission policy, and so on (Mancebón and Muñiz, 2008).

2.3 Review of Italian studies

Unlike the international literature, in Italy there were less studies about school efficiency. Barbetta and Turati (2003) seems to be the first attempt to move the analysis on Italian students, even though it was limited to junior high schools located in Piemonte. They computed efficiency scores for schools by applying DEA and, through a second stage, they investigated the determinants of efficiency scores' differentials. They found out that foreign and disabled students negatively affected efficiency, and that size was also relevant.

Factors affecting student achievement were the object of the studies of Bratti et al. (2007). Although these studies did not directly aim at evaluating school efficiency, they explained how students' results were related to the socio-economic status, the macro-area of the country (Northern Italy schools performed much better than others), and the type of school (academic and technical schools performed better than vocational schools). Furthermore, they noticed that private schools performed worse than public ones.

This last result was consistent with what also Barbetta and Turati (2003) found about for-profit schools, while it was inconsistent for the conclusions they draw about non-profit schools, with private schools seeming to be more efficient than public ones. These discrepancies in terms of student achievement, when comparing public and private education, encouraged the debate over competition.

As stated by Hoxby (2003), competition among public and private schools increased productivity. There was the need not only to refer to the differences on the relative performances between private and public, but also to properly include competition into the analysis, investigating if and how much it could affect the results. Therefore, Agasisti (2009) added three proper indicators of competition to the list of covariates, such as a dummy variable eventually indicating the presence of another school, the percentage of students enrolled in private schools and the number of schools per student, all these three with regard to the same area.

In a subsequent paper, Agasisti (2013) extended the previous studies applying a two-stage analysis on a sample of Italian schools, no longer limited to a region as for Barbetta and Turati (2003). While aggregating data at school level, he went through different combinations of DEA inputs in the first stage, then considering a broader range of covariates into a Tobit regression in the second stage. The results showed that at least one indicator of competition was statistically associated with higher performances of schools.

According to the analysis of Bratti et al. (2007), Longobardi et al. (2009) remarked the fact that students of the South had lower performances. This was well explained through economic-social-cultural variables, associated with both individual and family background information. They stated that a high-centralized system as the Italian one was not finally able to guarantee homogenous students' performance.

A comparison between the different areas of the country was also present in the paper by Di Giacomo and Pennisi (2015). Data of students from over 2000 among primary and lower secondary schools were obtained combining average INVALSI test scores with other administrative data provided by the Ministry of Education, and statistical data about the school locations provided by the Italian national institute of statistics (ISTAT). The high variability of results between schools, reporting once again that efficiency gains were higher in the South Italy, suggested that there was room for improving efficiency at the school level.

Instead, the objective of the analysis of Agasisti et al. (2014) was to evaluate the efficiency of public spending. The paper proposed a procedure to calculate adjusted efficiency measures to assess the pure management efficiency of each school and so to avoid confusing the institution performance with the aspects relating to its background. They noticed that there was an inverse relationship between the efficiency scores obtained before and after the correction due to external factors. In this last case, the efficient schools were the ones with students from a better socio-economic background and a lower number of immigrant students.

Most recently, the findings of Masci et al. (2018) revealed that the factors exerting the greatest influence on reading performance were primarily associated with the demographic features of the students, whereas the level of achievement in mathematics was somewhat linked to the administrative strategies employed by the school principal or head teacher. Despite this, it was important to note that schools and the educational process accounted for only a small portion of the variability in achievement scores, and the characteristics inherent to the students themselves played a significantly more substantial role.

2.4 Studies comparing public and private education

As noted for the Italian studies, the comparison between performances of public and private schools, has gained importance in the years, although the findings are mixed. Apart from those already mentioned, here is provided a summary of the studies on this topic, even if it was not explored in this thesis, due to the lack of information about private schools.

Ahn et al. (1988) focused on comparing the efficiencies of public and private doctoral-granting universities in the United States. The findings revealed significant differences in behaviour between the two groups. Overall, public universities were found to be more efficient than private universities in terms of managerial and program inefficiencies. However, when managerial inefficiencies were removed, and medical schools were not present, private universities were found to have more efficient programs.

Bertola e Checchi (2002) found that private schools in Italy served a distinct purpose compared to their American counterparts. While the latter enhanced the academic achievement of underprivileged students, the former seemed to cater the needs of students with limited academic ability from relatively rich family backgrounds.

In the analysis of Mizala et al. (2002) in assessing the technical efficiency of schools in Chile, the school type was highly impacting. They even distinguished between feepaying and subsidized private schools, finding that private feepaying schools outperformed all the others.

The aim of Mancebon and Muniz (2008) was to evaluate the effectiveness of various public and private high schools in Spain, analysing the sources of inefficiencies in each school. They distinguished these causes between factors due to managerial performance and factors due to structural differences of management models. The findings indicated that while private schools tended to achieve better academic outcomes than public schools overall, this was not necessarily due to superior management practices, but rather to the fact that their students tended to have a more advantageous educational background.

In contrast to prior research in Australia that concentrated on a particular group of schools, Nghiem et al. (2016) investigated both public and private schools. Their studies revealed that Catholic and independent primary schools exhibited lower efficiency levels than public schools. Conversely, at the secondary school level, public schools demonstrated lower efficiency levels than their non-public counterparts.

Overall, the findings suggest that private schools could have higher levels of academic achievement, but also serve a more advantaged student population and have more limited access. Public schools, on the other hand, served a more diverse student population and were often more cost-effective. However, the results varied depending on the country, the specific context, and the measures used to compare schools.

2.5 Combination of DEA and ML

The last step in the literature review on studies of school efficiency by means of Data Envelopment Analysis concerned the possibility of combining this methodology with more advanced Machine Learning techniques that could make significant improvements in the evaluation of educational system performance. Specifically, research conducted via Scopus returned the presence of such techniques in three previous works.

Sreekumar and Mahapatra (2011) developed an integrated approach combining DEA and Neural Network (NN) for assessment and prediction of performance of Indian Business schools for effective decision making.

NN is a computational model inspired by the workings of the human brain, in particular the way biological neurons communicate with each other to process information. It consists of a set of interconnected nodes, called artificial neurons or perceptrons, which process numerical input and produce numerical output. Communication between neurons is based on the transmission of signals via weights, which indicate the importance of the connection between two neurons.

The procedure they adopted, followed the steps of the well-known two-stage analysis, but school efficiency obtained through DEA was used as output variable in a generalized NN regression during training phase.

Schiltz et al. (2017) and Rebai et al. (2020) instead, focused their works on ML approaches such as regression trees and Random Forest (RF). A regression tree is an algorithm that repeatedly divides the input data set into smaller subsets based on a series of binary answer questions, until each leaf of the tree represents a relatively homogeneous subset of data in which a numerical value is predicted.

RF is a set of (in this case regression) trees that work together to make a more accurate prediction. Each tree is created randomly by selecting a random subset of the input data and a random subset of the features to divide the data. In this way, each tree can learn from a different part of the data and random variations in the tree creation process can reduce the risk of overfitting the training data.

Schiltz et al. (2017) adopted a two-stage methodology to evaluate the efficiency of Hungarian primary schools. They first constructed a measure of school value-added applying a DEA model with prior achievement in mathematics, reading and socio-economic background as inputs, and students' performance as outputs. Then, they noted that classical regression approaches were to some extent able to capture the important variables but failed to identify interesting nonlinear interactions between them. Therefore, they used regression trees and RF to capture these nonlinearities and visualize them to improve interpretability.

Same procedure was adopted by Rebai et al. (2020) in assessing Tunisian secondary schools. The results obtained from the regression trees suggested that several factors contribute to higher performance in schools, including school size, competition, class size, parental pressure, and proportion of girls, whereas school location did not appear to have any impact on school efficiency.

On the other hand, according to the outcomes of the RF algorithm, the proportion of girls at school and school size had the most significant impact on the predictive accuracy of the model, and thus could have a greater influence on school efficiency.

As Schiltz et al. (2017), also this study revealed the high non-linearity of the relationships between the key factors and school performance, highlighting the importance of modelling their interactions in influencing efficiency scores.

The use of ML methods is becoming increasingly widespread in various fields of the economy. For example, the use of ML algorithms can improve the accuracy of market forecasts and risk analyses in the field of finance, or it can be used to analyse large amounts of data on consumers, such as their purchasing habits or online behaviour, in order to improve marketing strategies and sales policies.

Despite these successful applications, the use of ML techniques in education has so far been relatively limited, even if it could represent a significant opportunity for improvement. More research and application of these techniques in this field is needed. There appears to be no specific information on authors who have used DEA and advanced Machine Learning such as Neural Networks or Random Forests together to evaluate Italian schools.

The use of ML techniques in combination with DEA could represent an active and expanding area of research, helping to improve the accuracy of efficiency estimates, identifying key factors influencing performance, the strengths and weaknesses of evaluated schools and providing a basis for improvement decisions.

3. Education system in Italy

3.1 Introduction

In this chapter an overview about the structure of the Italian education system is provided, mainly referring to official sources from the ministry, INVALSI and the OECD.

Education in Italy is regulated by the Ministry of Education and Merit (Ministero dell'Istruzione e del Merito – MIM) and the Ministry of University and Research (Ministero dell'Università e della Ricerca - MUR), and it is mainly provided by the State. However, private entities may establish independent educational institutions that, in some cases, can be equal to state schools and even issue qualifications with the same value. When this happens, they are named as parity schools, but in this thesis these schools are simply referred to as private schools.

There are also private non-parity schools, where regular attendance by pupils constitutes fulfilment of the education obligation, but they cannot issue qualifications with legal value nor intermediate or final certificates with legal certification value. Pupils must therefore take an aptitude test at the end of each school course or if they want to transfer to a state or parity school.

MIM has decentralised offices, which in turn are divided into the territorial entities at provincial level. These ensure the application of general provisions and compliance with minimum performance requirements in each region. Regions have joint responsibility with the State in some areas such as the organisation of the early childhood education and care (ECEC) till 3 years of age, the school calendar, the territorial distribution of schools and the right to study at higher levels. On the other hand, regions have exclusive legislative competence in the organisation of the regional vocational education and training system.

Compulsory education lasts 10 years, from 6 to 16 years of age. Parents of pupils, or anyone acting on their behalf, are responsible for the fulfilment of their children's educational obligation, while the supervision of the fulfilment of that obligation is the responsibility of the municipalities of residence and the school heads of the schools where the pupils are enrolled. There are no decentralised offices of the ministry at municipal level.

The choice of school by families is free. However, in state schools, limitations may be set in relation to the accommodation capacities of the facilities, or the number of staff assigned to individual schools by the school administration. In case of applications for enrolment exceeding the available places, the criteria for accepting of applications are established by the individual educational institutions. For compulsory levels of education, the right to study must in any case be guaranteed for all, through forms of coordination at territorial level between schools and local authorities.

3.2 Historical background

In this section, a review of the most relevant steps in the history of the Italian state school system was conducted. The beginning of the school system can be identified in 1859, the year in which the Casati Law (Legge Casati) was approved. The main aim of the law was that children should be able to read, write and count, and the same law sanctioned the compulsory and free elementary education for the lower course, provided by the state through the municipalities, which were also responsible for hiring teachers. Secondary education was under the responsibility of the provinces, and the universities were managed by the State.

This law was intended to take away the Church's centuries-old dominance in the field of education, since traditionally Italy education was provided, both at elementary and higher levels, by ecclesiastical institutes.

Primary school was divided into two periods of two years, but the second two-year period was only instituted in municipalities with more than four thousand inhabitants, or which had a secondary school in their territory. The municipalities, therefore, had to finance their own schools, and municipalities with fewer resources or those in poorer areas (characteristics that often coincided) had difficulty recruiting sufficiently qualified teachers for the primary schools.

This incentivised private education by wealthier families who often relied on a domestic tutor or private institutions. Far from becoming truly public, the Italian school did not even manage to become compulsory. The Casati Law itself did not provide for sanctions for parents who did not send their children to school and so many families preferred to keep their children at home for work in the fields.

A subsequent important law regarding the Italian education system was the Gentile Law (Legge Gentile) in 1923, which was issued with Mussolini and his National Fascist Party in power. It was a set of decrees published without parliamentary debate that remained largely in force, unchanged even after the advent of the Republic, until the Italian Parliament gave birth to the unified middle school in 1962. The primary school with the Gentile Law had a more authoritarian and hierarchical structure. The Catholic religion was placed as the basis of the education and morality of the child, in order they could be an example to future generations.

This law raised compulsory schooling to the age of 14. After the first five years of primary schools equal for all, the pupil had to choose between middle schools and work training. Only the middle school allowed access to the lyceums and in turn only the classical lyceum allowed enrolment in all university faculties. In 1962 the possibility for work training was abolished.

The various types of educational institutions, state, private and parochial, were regulated and the Magistral Institute was created for the training of future primary teachers. A great innovation came with the establishment of special schools for handicapped pupils, although something similar had already been done in the 19th century, and the pupils' learning of civic sense and fairness towards their neighbours.

Since 1962, the structure of the Italian education system has just undergone some modifications: the gap between males and females was narrowed merging the two distinct programmes for technical education and introducing optional mixed-gender gym classes; new lyceums, technical institutes and professional institutes were created, with a broader range of options for the students.

In 1999, as consequence of the Bologna Process, an international reform process of higher education systems in the European Union, the university system moved from the old 5-year system to a new system, which distinguished between the bachelor’s degree that lasted 3 years and the master’s degree lasting 2 years, and a credit system was adopted to quantify the work needed by each exam. There was just a 5-year degree for fields like Law and a 6-year one for Medicine.

The reform promoted by Moratti in 2003 opted for a three-year pre-school, followed by an eight-year first cycle with primary school and secondary school grade, ending with a state exam. The second cycle was divided into two distinct sections: that of the lyceums and that of vocational education, preparatory to work. Both these educational channels ended with a state exam, but only the lyceums, as for the school designed by Gentile, allowed direct access to university.

With the minister Fioroni, after several modifications, compulsory schooling was established to last 10 years and, in any case, until the age of 16. Consequently, the age for access to employment was raised to 16.

3.3 Public and private schools

The recognition of equality between private and public schools guarantees the equality of students' rights and duties, the same methods of conducting state examinations and the entitlement to issue qualifications with the same legal value as state schools.

The substantial difference lies in the fact that the public school is financially supported by the Italian State. In contrast, private schools administer themselves, and receive only a minimal contribution from the state coffers, which in some cases is totally absent. The lack of a significant economic contribution from the State obliges private schools to charge students a more or less substantial fee, depending on the type of structure and educational path.

Regardless of their financial management, they are equal in all respects to state public schools, and must comply with precise requirements, high quality standards and legal provisions guaranteeing their seriousness and reliability. The State must verify that private schools act in accordance with the law by carrying out rather strict periodic checks.

Moreover, thanks to their usually smaller size, limited number of students and the possibility of carefully selecting teaching staff, private schools could guarantee a level of quality that proves to be higher than that offered by public education. This represents an important point for reflection, as previously explained in the analysis of the literature.

According to data provided by the ministry Statistics Office, the percentage of students attending private schools is slightly less than 10%, with the highest number of students concentrated on the preschool.

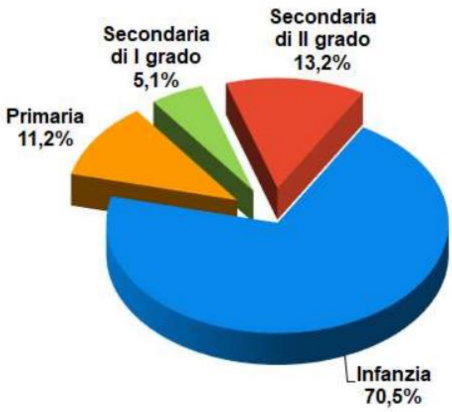

Figure 1. Distribution of private school students (Source: MIUR, a.y. 2021-2022)

3.4 Course structure

The structure of the educational system consists of 3 main levels: primary, secondary and tertiary education.

Services for children till 3 years of age are offered by educational services like nurseries, playgrounds or home-based services. These constitute the first part (0-3) of the ECEC. The preschool, known as scuola materna o dell'infanzia, completes the integrated system of ECEC (3-6) and, together with the primary school, represents the primary education.

Primary school represents the start of compulsory education, at 6 years of age, and lasts 5 years. It has the task of acquiring basic learning and laying the basis for the exercise of active citizenship, guiding each pupil to elaborate the meaning of their own experience.

The external assessment of pupils' learning is entrusted to INVALSI, the National Institute for the Evaluation of the Education and Training System, through specific tests. The results of these tests represent the main data source for the analysis conducted in this thesis and will be better explained in the chapter five about data.

For the primary level, the surveys must normally be carried out in the second and fifth classes. The results are returned to schools in both aggregate and disaggregated form, question by question, in order to provide managers and teachers useful tools for self-evaluation and improvement of teaching. Students move from primary to lower secondary education without any exam.

Secondary education starts with lower secondary school that lasts 3 years. It is generally attended by children between the ages of 11 and 14. Unlike in primary school, there are several specialist teachers in a class, who teach one subject or several subjects, and generally follow the class for the entire three-year course. The specific aim of the lower secondary school is the acquisition of fundamental knowledge and skills to develop basic cultural competences.

At this stage of the school career, the focus is on giving pupils points of view on reality and ways of knowing, interpreting and representing the world. For each discipline, targets are set for the development of competences considered indispensable for the achievement of the competence development goals. The third classes of the lower secondary level participate to the tests provided by INVALSI. To be admitted to the upper secondary school, students must sit in a state exam and successfully pass it.

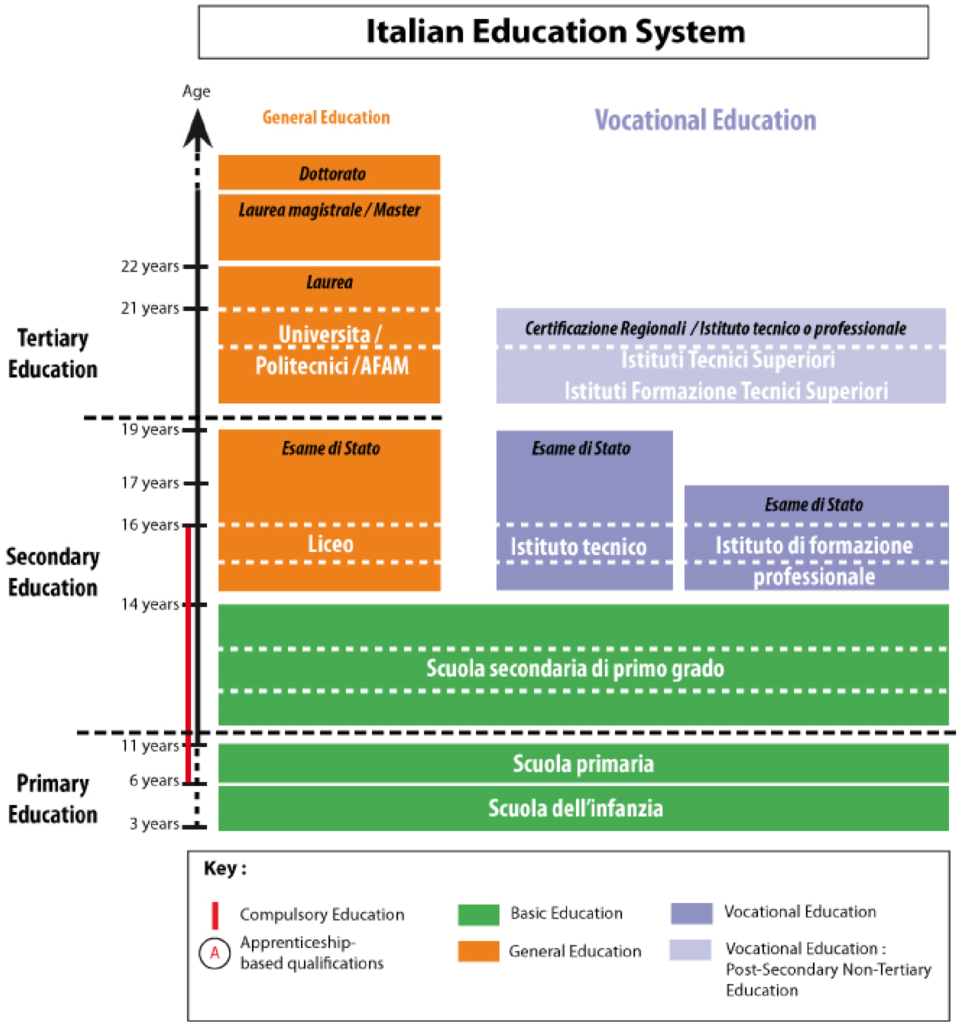

Figure 2. Structure of the Italian education system (Source: EOLSHE, 2020)

To ensure a process of educational continuity, there is the possibility of creating, under the direction of a single school head, comprehensive institutes consisting of preschool, primary and lower secondary schools.

Students have compulsory access to the upper secondary school since, being education compulsory until the age of 16, it also covers the first two years of the upper secondary school. At this level, basic education is over, and students can choose between the general (lyceums) and vocational (technical institutes) pathways offered as part of state secondary education, and the three- and four-year vocational education and training courses under regional competence (professional institutes).

To ensure equivalent training for all education, knowledge and competences that all students must have acquired by the end of compulsory education have been defined, beyond the specific programmes for the different upper secondary education pathways. They are oriented towards the acquisition of key competences that prepare young people for adult life and for lifelong learning such as communicating, collaborating and participating, acting autonomously and responsibly, planning, solve problems, identify connections and relationships, and acquiring and interpreting information.

At the end of the upper secondary school courses, both general and technical and vocational, student take a state examination. The state examination is aimed at ascertaining the knowledge and skills acquired in the last year of the course of study in relation to the general and specific objectives of each general cultural foundations and the candidate's critical capabilities. Passing this exam, students receive a certificate that gives them access to tertiary education.

Tertiary education is offered by universities or equivalent, institutes of the higher education for the fine arts, music and dance (Alta formazione artistica, musicale e coreutica - AFAM) and higher technological institutes (Istituti tecnologici superiori - ITS). The specific conditions for admission are established by the MUR. Higher education promotes scientific progress and provides the scientific culture necessary for the professional practice.

3.5 Performance analysis

In line with the main objective of this thesis, such as the evaluation of the Italian school efficiency, in this section the performances of the education system in Italy were compared with those of other countries.

The analysis was conducted considering the results of international standardized tests created by the OECD, an international organisation that encompasses the most economically advanced countries, within the specific Programme for International Student Assessment (PISA).

The OECD PISA tests are the largest international educational survey that every three years analyses the reading, mathematics and science skills of 15-year-old students in OECD and non-OECD member states, with the aim to understand what the quality of education in a country is.

Each survey focuses particularly on one of the subjects, while the other two domains are surveyed in less detail. To be able to do the tests, it is not necessary to remember what one has studied by appealing to memorisation, but it is indispensable to have understood it well, to face a new problem in the same domain. They focus on the concept of competence, which requires full understanding and not a superficial acquisition.

The tests also include optional domains, and questionnaires addressed not only to students but also to parents, teachers and schools. While the cognitive domains measure the children's ability to read, use mathematics and science, the questionnaires seek to understand how they feel about school and parents, monitoring the economic and social situation of the family. The aim is to define how the socio-economic environment in which the children live influences their school performance.

The first edition of the tests took place in 2000 and the most recent available results are the ones of 2018 tests, since 2021 tests were postponed to 2022, due to the uncertainty caused by the pandemic situation. OECD member states are not obliged to participate in the PISA test, which is why the number of participating states changes. In addition to the member states of the OECD, external countries can also participate. In 2018, out of 79 participating states, 36 were OECD members.

Figure 3. OECD member and partner countries in 2018 PISA tests (Source: OECD, 2018)

Italy participated to all the editions of the tests. In 2018, when the main domain was reading, 11785 Italian students from 550 schools took part in the tests. They scored below the OECD average in reading and science and in line with the OECD average in mathematics.

Looking at the trends of the results over the years in Figure 4, Italy's average performance declined after 2012 in reading and in science, while it remained stable in mathematics. However, the average reading score in 2018 was lower than in PISA 2000 and PISA 2009 (the two previous surveys with reading as the main domain), but close to the level observed in most of the remaining surveys; it was therefore not possible to determine a clear direction of change.

Performances in science in 2018 were significantly below the level observed in 2009-15, returning to the level of 2006. In mathematics, they improved between 2006 and 2009, and then remained stable after 2009.

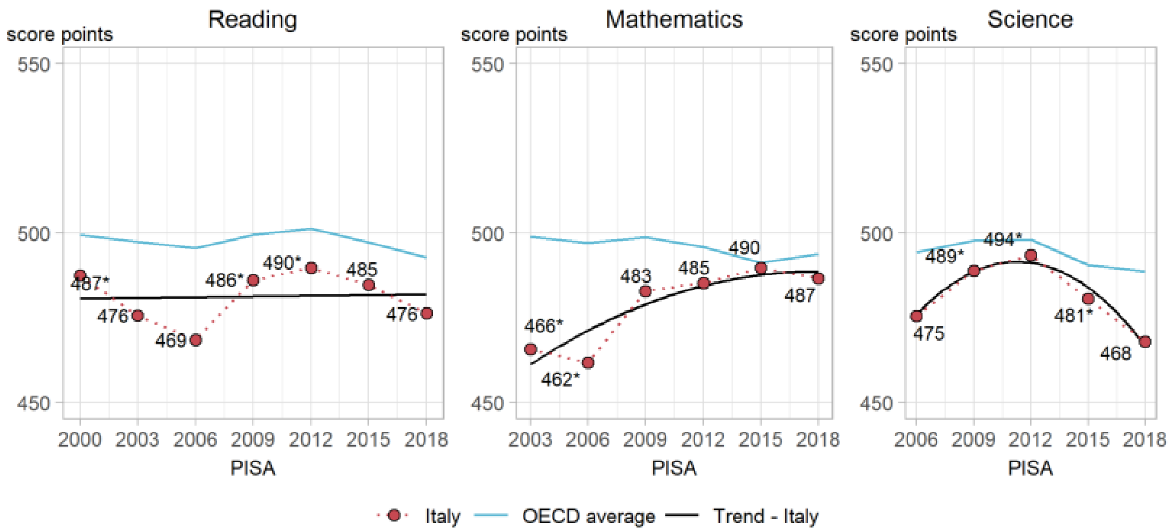

Figure 4. Trends in PISA scores of Italian students (Source: OECD, 2018)

Reading these numbers, Italian schools clearly offers less preparation for 15-year-olds than the average of other countries. In a first approach, they should focus on reading and science skills, while maintaining the good trend of the recent years in mathematics.

Moreover, as stated by INVALSI while providing an interpretation of these results, the quality of preparation varied greatly whether one lived in the North, the Centre or the South. Students from the North achieved the best results, while their peers in the South were more in difficulties.

These geographical differences should be reduced and all students in Italian schools, wherever they live and whatever their personal history, should be allowed to make the most of their potential.

The choice of high school also had a very big impact on children's reading, mathematics and science skills and the gender was another discriminant factor. The author recommends the reading of that analysis provided by INVALSI for further details. However, all these aspects represented interesting refection points and terms of comparison for the work conducted in this thesis.

4. Methods

4.1 Introduction

As said, schools can be considered as organizations whose main objective is to increase the knowledge of students. For this reason, their level of efficiency is often evaluated through student achievement. Understanding the extent to which a school output could be improved by enhancing efficiency alone, without requiring additional resources, is crucial, and requires a proper definition of the concept of efficiency.

In the past, measuring labour productivity was a common way to evaluate efficiency, but this approach failed to consider other inputs besides labour. This method was widely used by economic statisticians, but it proved to be inaccurate. Later, attempts were made to create efficiency indices that considered multiple inputs and compared them to output, but this approach encountered problems commonly associated with creating indices.

The objective of Farrell (1957) was to create a reliable measure of productive efficiency that considered all relevant inputs, while avoiding numerical issues related to indices. Farrell's approach aimed to determine how effectively firms were using their resources to produce goods and services. The idea was to compare the actual output produced by a firm with the maximum possible output that could be produced given its inputs, technology, and the market environment.

To do this, Farrell adopted the concept of a production frontier, which represents the boundary between the feasible and unfeasible production combinations of inputs and outputs. The production frontier is determined by the most efficient firms in the market, which produce the maximum amount of output given a set of inputs.

Using this production frontier, Farrell proposed two measures of efficiency: technical efficiency and allocative or price efficiency. Technical efficiency measures how well a firm uses its inputs to produce output relative to the production frontier. Allocative efficiency measures how well a firm allocates its inputs among different outputs to maximize profits, given the prices of inputs and outputs.

Farrell's definition of efficiency has been widely adopted in the field of productivity and efficiency analysis, and it has been applied to various sectors, including education, to identify the most efficient firms and potential sources of inefficiencies.

To explain it in a simpler way, the case of a firm employing two inputs to produce a single output is reported, under the initial assumption of constant returns to scale. This assumption permits all the relevant information to be presented in simple isoquant diagram.

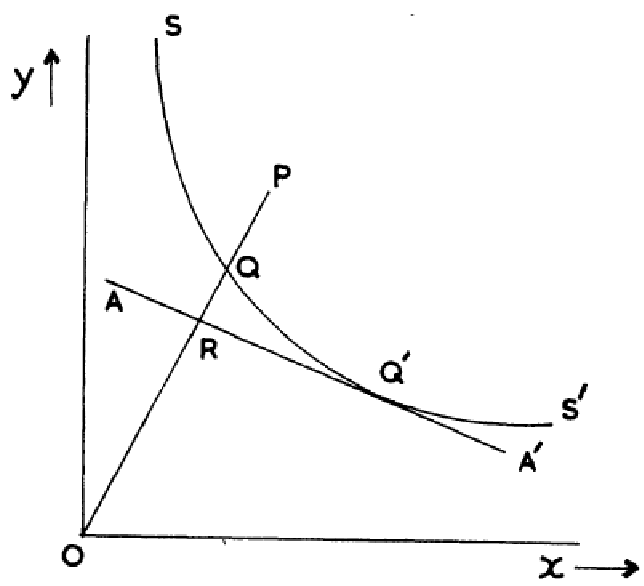

Figure 5. Graphical construction of Farrell's definition of efficiency (Source: Farrell, 1957)

The point P in Figure 5 indicates the inputs required by the two factors to produce one unit of output. Each firm can be represented by a point on the isoquant diagram, and the efficiency production function will be represented by an isoquant itself.

The efficient production function is initially taken as known: the isoquant SS' can be regarded as showing the various combinations of the two factors that a perfectly efficient company might use to produce one unit of output.

The point Q on SS' represents a firm that uses the same ratio of two factors of production as the point P but produces the same output using only a fraction OQ/OP of each factor. Therefore, the firm at point Q is more technically efficient than the firm at point P.

The ratio OQ/OP was defined by Farrell as the technical efficiency of point P, where a unitary value or 100% indicates perfect efficiency, and a value approaching zero indicates inefficiency.

If the slope of the line AA' represents the ratio of prices of the two factors of production, Farrell also defined the ratio OR/OQ as a measure of the price efficiency of point Q. Indeed, the cost of producing the output at point Q' is only a fraction OR/OQ of the cost of producing it at point Q. Therefore, the optimal production method is represented by Q' rather than Q.

In summary, he explained the concepts of technical efficiency and price efficiency using the example of two points on a production possibility frontier graph. The ratio of input to output (OQ/OP) was used to measure technical efficiency, while the ratio of prices of factors (OR/OQ) was used to measure price efficiency.

At this point, the challenge consists in determining an efficient production function based on the input and output values of several firms. Assuming that the isoquant is convex towards the origin and has no positive slope anywhere, the SS' curve in Figure 6 represents the most suitable estimation of the production function, since it meets both these assumptions and it is the least exacting standard of efficiency that is consistent with the observed points.

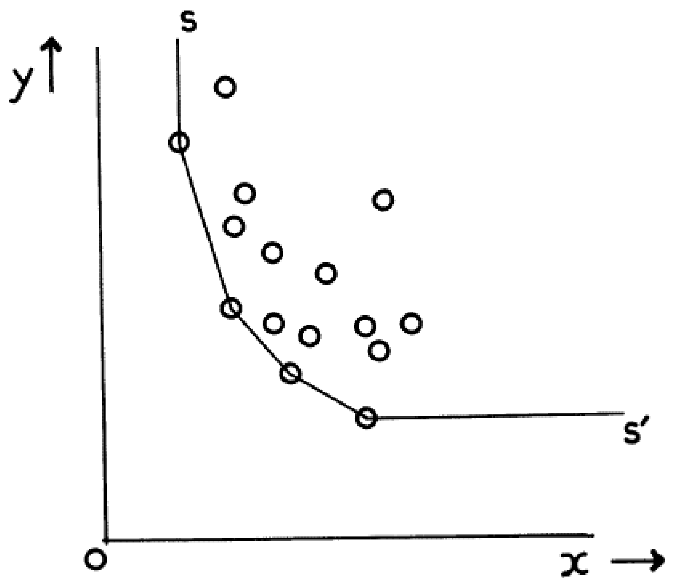

Figure 6. Production function in the case of two inputs and one output (Source: Farrell, 1957)

In economic theory, the assumption of convexity is a standard assumption, meaning that the production function is assumed to be curved, rather than linear. This is because in most cases, as more inputs are added to the production process, the increase in output becomes less and less.

On the other hand, the assumption of nowhere positive slope is necessary to ensure that increasing the application of both factors does not result in a decrease in output. This assumption is important because if the slope is positive, then increasing the application of both factors could lead to a point where further increases in input would cause output to decrease.

To measure the technical efficiency of a firm, a common approach is to compare it to a hypothetical firm that uses the same proportion of inputs. This hypothetical firm is constructed as a weighted average of two observed firms, where each input and output are a weighted average of the inputs and outputs of the observed firms. The weights are chosen to achieve the desired factor proportions, and the resulting efficiency measure can be used to identify areas for improvement in the firm's production process.

The isoquant diagram is abandoned when generalizing to cases with many inputs and outputs, and the curve SS' can be geometrically defined. In these cases, Farrell's efficiency concept can still be expressed as the ratio of weighted outputs to weighted inputs and there are two primary approaches to determine those weights: parametric and non-parametric models.

Parametric models rely on the assumption that a specific function that explains how inputs are transformed into outputs is known. By setting fixed weights, it is possible to create virtual inputs and outputs, which can be used to calculate efficiency. The Cobb-Douglas production function is an example of a more sophisticated function that academics used, rather than simple linear equations.

However, specifying the functional form beforehand can creates various issues with estimation and specification. This work instead, relied on a non-parametric method, such as Data Envelopment Analysis. In DEA, each observation is compared to a group of similar observations and the weights used in the calculation of the efficiency ratio are determined based on the best-performing observations in the group. The origins, the improvements and even any possible drawbacks of the methodology have been explored in the following sections.

4.2 Data Envelopment Analysis

4.2.1 Theoretical aspects

The first publication on DEA methodology dates to 1978, in the work entitled "Measuring the efficiency of decision making units" by Charnes, Cooper and Rhodes. Its aim was to extend Farrell's definition of efficiency, transforming the analysis initially conducted by Farrell on enterprises using only one input and output, into a linear programming problem to be adapted to a multitude of outputs, with the aim of improving the methodology inherent in the study of productivity.

Prior to 1978, measurements of productive efficiency were very accurate, precise and punctual, but too restrictive, ignoring the possibility of combining multiple inputs and outputs to achieve a measure of total efficiency.

The name Data Envelopment Analysis refers to its method of enveloping observations to create a frontier that evaluates the performance of entities. These entities can be diverse, such as businesses, government and non-profit agencies, schools, hospitals, and countries, and the term "Decision Making Unit" (DMU) is used to cover any such entity that produces similar outputs using similar inputs.

DEA is a frontier, linear programming technique, aimed at providing a non-parametric approach to compute efficiency scores, which indicate the degree of efficiency for each evaluated entity. It determines the DMUs located on the efficiency frontier, which serve as benchmarks for future improvements in the evaluated DMUs' performance, identifying the sources and amounts of inefficiency for each of their inputs and outputs.

Charnes, Cooper and Rhodes proposed the most basic DEA model, which allows for evaluation both through an input orientation as well as an output orientation. Their proposal was based on the evaluation of the efficiency of an economic entity as the ratio between the quantity of output and the corresponding quantity of input employed, both appropriately weighted. Here an exhaustive explanation of that model was given, with reference to textbook by Cooper et al. (2006).

The analysis sample must present characteristics of homogeneity, independence and autonomy and it is accepted that, within such a reference sample, there may exist several ideal DMUs whose efficiency is greater than or equal to that of all other DMUs in the sample.

The assumptions of the CCR model, which assumes the existence of n DMUs, with m inputs and s outputs, are as follows:

- there must be strictly positive numerical data available for all inputs and outputs considered;
- in general, it is preferable to consider a small number of inputs and a high amount of outputs;
- the inputs, outputs and choice of DMUs must reflect an interest in the evaluation of the relative efficiency of DMUs;
- efficiencies of scale are constant (CRS);
- all inputs and all outputs are reduced to a single virtual input (expressed as a weighted sum of the inputs) and to a single virtual output (expressed as a weighted sum of the outputs);
- the weights must be non-negative and such that the ratio is not greater than 1 for all DMUs.

For each DMU, weights are chosen for both the inputs (v_i) (i= 1, ..., m), and for the outputs (u_r) (r = 1, ..., s):

$$\mathit{virtual\ input} = v_1 x_{10} + \cdots + v_m x_{m0}$$

$$\mathit{virtual\ output} = u_1 y_{10} + \cdots + u_s y_{s0}$$

Using linear programming, it is possible to determine the input and output weights that maximise the ratio

$$\frac{\mathit{virtual\ output}}{\mathit{virtual\ input}}$$

The weights then become the unknown variables of the maximisation problem, while the ratio between the weighted quantity of output and input constitutes the objective function to be maximised. Optimal weights, not predetermined in advance by the decision maker but derived from the data, are variables that the model calculates to maximise the efficiency ratio and that generally vary from one DMU to another.

The resulting efficiency of each DMU is thus given by the ratio of the weighted sum of the outputs and the weighted sum of the inputs. Each economic unit is free to choose the weights in such a way as to maximise its own level of efficiency: the smaller the weights assigned to the inputs, the greater their availability, and vice versa.

The only constraint, inherent to the decision on the weights, is that they do not take on negative values and they are such that the efficiency ratio for all other DMUs is positive but not greater than 1. Once the data has been selected, the efficiency for each DMU is evaluated.

Choosing DMU_0 as the analysis entity whose efficiency one wish to know with respect to all other DMU_j, the following fractional programming (FP) problem is solved, through which it is obtained the optimal values of the input weights (v_i) (i= 1, …, m) and of the output weights (u_r) (r =1, ..., s):

(FP_0)

$$\max_{v,u} \theta = \frac{u_1 y_{10} + u_2 y_{20} + \cdots + u_s y_{s0}}{v_1 x_{10} + v_2 x_{20} + \cdots + v_m x_{m0}} \tag{1}$$

s.t.

$$\frac{u_1 y_{1j} + \cdots + u_s y_{sj}}{v_1 x_{1j} + \cdots + v_m x_{mj}} \leq 1 \quad (j = 1, \dots, n) \tag{2}$$

$$v_1, v_2, \dots, v_m \geq 0 \tag{3}$$

$$u_1, u_2, \dots, u_s \geq 0 \tag{4}$$

The non-negativity of the weights (3) (4) is not sufficient to guarantee a positive value in the (2). But, since the values of the inputs and outputs are assumed to be non-zero, then this condition associated with the non-negativity of the weights guarantees a value equal to or greater than zero to the ratio of virtual output over virtual input.

The fractional programming problem (FP_0) can be transformed into an equivalent linear programming problem, in the input-oriented form, through some mathematical steps.

First, it is noted that the denominator of the constraint (2), under the conditions of non-negativity of the variables of the denominator x and v, results also non-negative for each j. In this respect, it is possible to multiply the denominator in both sides of the inequation without changing the sign of the latter, thus obtaining the equation (1.4).

$$u_1 y_{1j} + \cdots + u_s y_{sj} \leq v_1 x_{1j} + \cdots + v_m x_{mj} \quad (j = 1, \dots, n)$$

The fractional problem (FP_0) turns out to be equivalent to the linear problem (LP_0) since, to bring the fractional form to the linear form, it is sufficient to set the denominator of the objective function (1) equal to 1 and set it as a constraint, obtaining (6). The solution will be obtained by maximising the numerator as in (5), under the above constraints:

(LP_0)

$$\max_{v,u} \theta = u_1 y_{10} + \cdots + u_s y_{s0} \tag{5}$$

s.t.

$$v_1 x_{10} + \cdots + v_m x_{m0} = 1 \tag{6}$$

$$u_1 y_{1j} + \cdots + u_s y_{sj} \leq v_1 x_{1j} + \cdots + v_m x_{mj} \quad (j = 1, \dots, n)$$

$$v_{1,} v_2, \dots, v_m \geq 0$$

$$u_{1,} u_2, \dots, u_s \geq 0$$

By solving it, DEA would provide the best set of weights to this combination of several inputs and outputs.

The optimal solution of the linear programming problem (v = v*; u = u*) is also the optimal solution for the corresponding fractional programming problem as the transformation performed is invertible according to the assumptions made. The two problems therefore share the same optimal solution θ* and it can be stated that, for the input-oriented version of the CCR model, a DMU_0 is efficient if θ*=1 and there exists at least one optimal vector-solution (v*, u*) with (v*> 0) and (u* >0). Otherwise, in all other cases, DMU_0 is considered inefficient.

The problem can be revisited from a dual point of view, by representing the inputs and outputs as a matrix (X, Y), and the weights v_m and u_s as vectors *v* and *u*, respectively. The new problem involves expressing θ as a real variable and introducing a non-negative vector of variables $\lambda = (\lambda_1, \ldots, \lambda_n)^T$:

(DLP_0)

$$\min_{\theta,\lambda} \theta$$

s.t.

$$\theta x_0 - \lambda X \geq 0$$

$$\lambda Y \geq y_0$$

$$\lambda \geq 0$$

As pointed out by Charnes, Cooper and Rhodes (1978) the importance of rewriting the problem in dual form is given by several aspects.

The dual form allows for less computational effort since the complexity of a linear problem increases as the number of constraints increases. While LP_0 requires a number of constraints equal to the number of DMUs of the sample examined (n), the problem in dual form only needs (m + s) constraints, equal to the number of inputs and outputs, which are generally less than the number of entities examined.

Moreover, the dual problem favours the interpretation of the results since in the DLP_0 problem the solutions are represented by the inputs and outputs, while in the LP_0 problem the weights are considered, which, although very useful, require further evaluation.

If the analysis conducted so far using the CCR model was aimed at input minimisation, with the aim of producing at least the same level of output, the complementary point of view is based on the output-oriented approach, the aim of which is to verify the presence of efficiency conditions in the maximisation of output.

The output-oriented model for DEA focuses on maximising outputs given existing input levels. In other words, it is assumed that DMUs try to obtain the maximum possible output given the available input level. Consequently, the model measures the efficiency of DMUs according to their ability to produce output from a given level of input.

Both the formulation and the solution of the output-oriented model are directly derivable from the input-oriented one, through simple linear transformations, such as:

$$\lambda = \frac{\mu}{\tau}; \quad \theta = \frac{1}{\tau}$$

By means of these two transformations, the problem can be rewritten as:

$$\max_{\tau,u} \tau$$

s.t.

$$x_0 - \mu X \geq 0$$

$$\tau y_0 - \mu Y \leq 0$$

$$\mu \geq 0$$

Adopting an input-orientation assigns a score ranging from 0 to 1, where a score of 1 is given to efficient DMUs and all other DMUs are considered inefficient. The score indicates that the DMU could reduce its inputs by (1- score value) percent and still achieve the same output level. Alternatively, an output-orientation assigns a score of 1 to efficient DMUs and a score greater than 1 to inefficient ones. A DMU can improve its production by (score value -1) percent while maintaining the same input level.

Later, Banker, Charnes, Cooper (1984) introduced a model, which represented a further extension of the CCR model. The difference with respect to the latter concerned mainly the returns of scale, which were no longer considered constant (CRS) with a production frontier represented by a half-line passing through the origin.

The BCC model includes the scale effect in its formulation, leading to a frontier represented by a convex function, expression of variable returns of scale (VRS). Considering variable returns to scale, it estimates the pure technical efficiency, stripped of the effect of the components of efficiency of scale.

From a mathematic point of view, the distinctive element of the BCC model that is not present in the CCR model is the inclusion of a new constraint representing the convexity condition, thus admitting the variability of returns to scale:

$$\sum_{j=1}^{n} \lambda_j = 1\,,\ \lambda_j \geq 0,\ \forall j$$

With the VRS approach, the efficiencies of scale can be increasing, constant, or decreasing, depending on the ratio of output to input.

It is possible to graphically represent a situation with 2 inputs and 1 output (i=2, r=1), in case of an input-oriented model, or, vice versa, 1 input and 2 output (i=1, r= 2), in case of an output-oriented model, both with VRS, noting the differences in terms of the efficiency frontier. The efficiency can be expressed as the ratio $\frac{\overline{OA'}}{\overline{OA}}$ in the input-oriented model, and as $\frac{\overline{OD'}}{\overline{OD}}$ in the output-oriented one, where O is the origin.

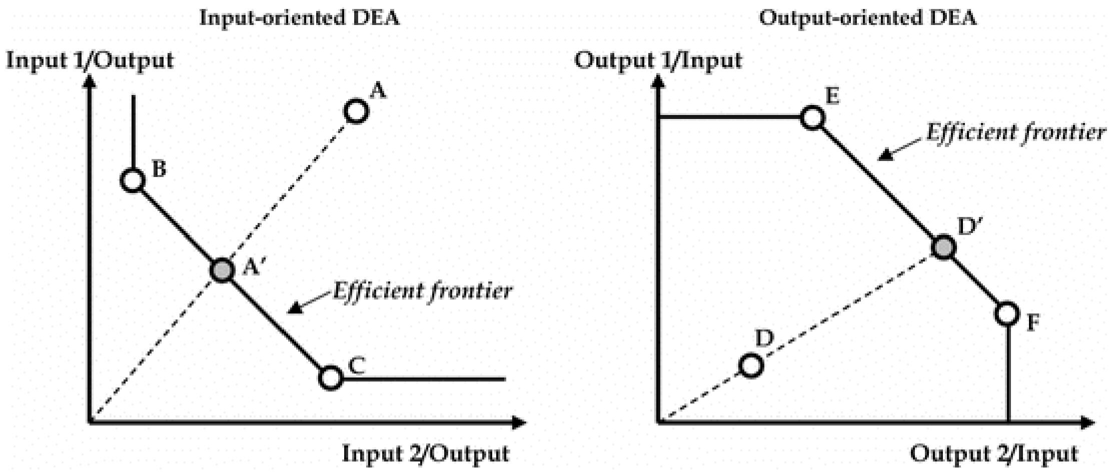

Figure 7. Efficiency frontiers for input- and output-oriented DEA (Source: Yang, 2017)

As well as the BCC model, other extensions of the CCR model were proposed, among which the multiplicative model (Charnes, Cooper, Seiford e Stutz, 1983), which provided a frontier alternatively log-linear and Cobb-Douglas, and the additive model (Charnes, Cooper, Golany, Seiford e Stutz, 1985), which was characterised by variable returns to scale and not necessary defined a priori the analysis orientation (input or output oriented), as both aspects were considered simultaneously.

However, these alternative models were out of the scope of the analysis, and the author suggests reading Cooper et al. (2006) textbook for more details.

4.2.2 Model adopted

First, it was necessary to define the model on which to base the analysis: the choice varied according to the scaling performance of the analysis sample, either constant (CRS) or variable (VRS).

Next, based on specific needs, it had to be established whether to conduct an analysis aimed at input minimisation, output maximisation or both. However, as said in the previous section, additive models were not taken into account, so it was possible to exclude the cases with both input and output orientation.

	Input-oriented	*Output-oriented*
Constant returns to scale	CCR_{INPUT}	CCR_{OUTPUT}
Variable returns to scale	BCC_{INPUT}	BCC_{OUTPUT}

Table 5. Variants of DEA models

In the review of the literature, most of the scholars chose to use a VRS approach, as it is widely recognized that continually increasing the amount of input can eventually lead to a saturation point. The notion that each additional unit of input increases the amount of output linearly and has the same effect as the first unit can be regarded as outdated.

Then, the schools' aim is to achieve the best possible results from students, with the number of resources provided. For these reasons, the entire analysis based on an output-oriented model with variable returns to scale.

Another important aspect of DEA concerns the calculation and interpretation of slack variables. Input and output slack can be respectively defined as the differences in terms of inputs and outputs of the DMU under analysis with respect to the efficient ones.

It is possible to define 2 vectors of non-negative slacks: s_r, which represents the amount of output slack that a DMU could increase while remaining on the efficient frontier with the same input, and s_i, which represents the amount of input slack that a DMU could decrease while getting the same output and remaining on the efficient frontier.

Considering s outputs and m inputs, with respect to the analysed DMU_0, the desired efficiency scores Φ come out of the following model:

$$\max \Phi + \varepsilon \sum_{r=1}^{s} s_r + \varepsilon \sum_{i=1}^{m} s_i$$

s.t.

$$\Phi_0 y_{r0} - \sum_{j=1}^{n} \lambda_j y_{rj} + s_r = 0, \quad r = 1, \dots, s$$

$$x_{i0} - \sum_{j=1}^{n} \lambda_j x_{ij} + s_i = 0, \quad i = 1, \dots, m$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j, s_r, s_i \geq 0 \quad \forall\, j = 1, \dots, m\,; \quad r = 1, \dots, s\,; \quad i = 1, \dots, m$$

where ε represents a tiny value used to accommodate errors without significantly impacting efficiency.

No DMU can be rated as efficient unless the following conditions are both satisfied: $\Phi = 1$ and the slack variables are all zero.

Slacks are often used as an indicator of inefficiency in DEA models. They indicate by how much the input and output values associated with them should be adjusted to eliminate all inefficiencies. By identifying and reducing slack, a DMU can improve its efficiency and move closer to the efficient frontier.

4.2.3 Advantages and drawbacks of DEA

This section refers to the advantages and disadvantages of using DEA methodology. The literature review has already highlighted some characteristics that led DEA to be preferred to more traditional regression techniques (Bessent et al., 1982). However, other benefits of the methodologies over alternative stochastic approaches, which have also been addressed by Worthington (1991), can be listed.

DEA does not require specific production functions to be explicitly defined, since it is a non-parametric technique. This makes it useful when detailed information on production functions is not available or when decision-making units have several production functions. When compared to models that need to estimate the production function, the analysis performed is relatively cheap in terms of cost and time.

As well as not being necessary to establish a functional form to represent production processes, it is not necessary to know the prices of inputs and outputs in advance. For this reason, DEA has been widely used in the evaluation of the technical efficiency, especially in those sectors where prices are not always available, such as education. It can use any type of unit of measure to analyse the efficiency of the DMUs, not being limited to monetary units.

The methodology can consider a multiplicity of inputs and outputs simultaneously, enabling efficiency to be assessed more comprehensively than with other techniques, and succeeds in producing a single measure of overall efficiency, without the need of predefining weighting factors.

The efficient frontier is formed from the most efficient DMUs actually measured, and not using estimated data: it uses individual observations, not their average or other functions, and produces an aggregate measure of efficiency for each DMU.

Furthermore, it can be considered a true benchmarking practice: DEA uses the most efficient decision-making units as benchmarks to evaluate other units. This allows to evaluate the efficiency of homogeneous DMUs, highlighting which inputs and outputs of inefficient DMUs should be changed to achieve efficiency. This practice is useful to the management as it suggests which DMUs are less efficient and in which direction they should work to improve their performance.

On the other hand, besides the positive aspects, it should be noted that in using the DEA methodology there are also some drawbacks, which if not considered could lead to an alteration of the validity of the entire analysis.

The efficiency measure elaborated is based on the only considered variables. It may therefore lead to a distorted representation, as some inputs and outputs may not have been included among the variables of the model. The correct identification of inputs and outputs is crucial to obtain reliable results.

Moreover, it solely takes into consideration discretionary variables, i.e., those which can be controlled and allocated by the DMUs. This limited approach fails to address significant factors contained in non-discretionary variables.

To address these limitations, many scholars have suggested a two-stage analysis (since Ray 1991). In this approach, efficiency is estimated in the first stage, while scores are regressed against covariates representing the environmental variables in the second stage. As seen above, part of the literature review focused on the scholars' choices about the variables to adopt both as input or output in the first stage, and non-discretionary variables in the second stage.

Then, the methodology does not foresee the possibility of all DMUs being inefficient and the efficiency of each DMU depends on the efficiency of the other DMUs. As a result, the efficiency assessed is relative, with the impossibility of giving an absolute assessment of the efficiency of each individual DMU. If one DMU has even slightly higher output than another, it is considered efficient and therefore able to change the production frontier.

A minimum number of DMUs is required for valid results. It requires a sample of decision-making units to be analysed much larger than the sum of the inputs and outputs so that a meaningful distinction can be made between efficient and inefficient units. To simplify the concept, having too few DMUs and too many inputs and outputs can easily lead to the evaluation of all DMUs as efficient, which would make DEA analysis of little use for discrimination and benchmarking. The units considered must be homogeneous, namely they must use the same type of resources to produce the same type of outputs.

However, the most common criticism levelled to DEA is the fact that the method is deterministic. Any deviation from the frontier is seen as an inefficiency in the production process, without considering random factors or external disturbances that may have influenced the findings (Førsund, 1992). It does not look at the variability in the data or other factors that could influence the results.

This means that the DEA could not consider measurement errors, market variability, variation in the price of goods and services, or other external factors that could influence the efficiency of the decision-making units. So, all the deviance from the frontier is attributed to inefficiency, without considering the possibility of random noise.

To overcome this main shortcoming, bootstrapped procedures were used to improve the validity and reliability of the DEA analysis, reducing the uncertainty associated with the results and increasing the robustness against outlier data.

4.3 Double Bootstrap DEA

Several studies have utilized two-stage procedures to account for exogenous factors that may impact firms' productivity. However, according to Simar & Wilson (2007) many of these studies have failed to provide a clear and coherent data-generating process (DGP), and the conventional inference approaches used have been found to be invalid due to unknown serial correlation among the estimated efficiencies.

As a result, the conventional inference techniques used in almost all the studies employing the two-stage approach were invalid. Xue and Harker (1999) and Hirschberg and Lloyd (2002), acknowledging the serial correlation issue, attempted to remedy it using a naive bootstrap method based on resampling from an empirical distribution, but this was demonstrated inconsistent in the context of nonparametric efficiency estimation by Simar and Wilson (1999a, b).

Moreover, neither of these studies provided a data generating process for which their second-stage regressions would be appropriate, and hence, it was unclear what they were estimating. To address that problem, it was necessary for the second stage analysis to incorporate a DGP that was logically congruent with the estimation technique employed in the initial phase.

Bootstrap, introduced by Efron (1979), and first applied to the production frontier framework of parametric, nonparametric and semiparametric model by Simar (1992), offers a promising approach for analysing the sensitivity of measured efficiency scores to variations in sampling.

This technique involves repeatedly simulating the data-generating process, typically through resampling, and applying the original estimator to each simulated sample. By mimicking the sampling distribution of the original estimator, the resulting estimates provide insight into how the efficiency scores would be affected by changes in the sampling process.

When using bootstrap methods in complex situations, such as nonparametric frontier estimation, a major challenge is to establish a clear model of the data-generating process. Without a well-defined DGP, it is impossible to determine whether the bootstrap is accurately mimicking the sampling distributions of the estimators, or some other distribution entirely.

Simar and Wilson (2007) presented a sensible DGP for such models and proposed two algorithms to implement respectively a single (algorithm #1) and a double (algorithm #2) bootstrap procedure, which permit valid inference. In particular, the double bootstrap procedure of the second algorithm was shown to improve statistical efficiency in the second-stage regression and it was the one adopted for the analysis of this thesis.

4.3.1 Simar and Wilson's (2007) algorithm #2

To make it simple, Simar and Wilson's (2007) algorithm #1 excludes those DMUs whose efficiency scores are equal to 1 and only uses the remaining scores as dependent variable in a truncated regression on environmental factors. Robust coefficients are then assessed through the bootstrap procedure.

Differently from algorithm #1, algorithm #2 employs a parametric bootstrap also in the first-stage problem, to produce bias-corrected estimates of the efficiency scores, and it is structured in the following steps:

STEP 1

Following the model illustrated in the section 4.2.2, DEA is performed on original data to compute a vector of efficiency scores $\hat{\delta}_i$, according to the production set $\mathcal{P}$, defined as:

$$\hat{\delta}_i = \hat{\delta}(x_i, y_i | \hat{\mathcal{P}}) \quad \forall\, i = 1, \dots, n$$

$$\mathcal{P} = \{(x, y) | x\ can\ produce\ y\}$$

STEP 2

Using the observations for which $\hat{\delta}_i > 1$, a truncated regression of $\hat{\delta}_i$ on z_i, a vector with the set of non-discretionary variables, is conducted, obtaining an estimate $\hat{\beta}$ of β and an estimate $\hat{\sigma_\varepsilon}$ of σ_ε, and it is expressed as follows:

$$\hat{\delta}_i = z_i\beta + \varepsilon_i \geq 1$$

STEP 3

This step is conducted $L1$ times to obtain a series of bootstrap estimates $\beta_i = \{\ \hat{\delta}^*_{ib}\}_{b=1}^{L1}$, and consists of the next four steps:

[3.1] For each $i = 1, \dots, n$, draw ε_i from the normal distribution $N(0, \widehat{\sigma}_\epsilon^2)$ with left-truncation at $(1 - z_i\hat{\beta})$;

[3.2] With these values, for each $i = 1, \dots, n$, compute $\delta_i^* = \ z_i\widehat{\beta} + \varepsilon_i$;

[3.3] Consider n new sets of inputs and outputs $(x*, y*)$, so that $x_i^* \ = x_i$ and $y_i^* = y_i\widehat{\delta_\iota}/\delta_i^*$;

[3.4] Compute $\delta_i^* = \delta(x_i, y_i|\widehat{\mathcal{P}}^{\ *}) \quad \forall\ i = 1, \dots\ , n$, where $\widehat{\mathcal{P}}^{\ *}$ is obtained by replacing x and y with x_i^* and y_i^* in $\mathcal{P}$.

STEP 4

For each $i = 1, \dots, n$, calculate a bias-corrected estimator $\hat{\hat{\delta}}_i$, using the bootstrap estimates present in β_i and the original estimate $\widehat{\delta_\iota}$, as shown below:

$$\hat{\hat{\delta}}_i = \widehat{\delta_\iota} - \widehat{BIAS}(\widehat{\delta_\iota})$$

$$\widehat{BIAS}(\widehat{\delta_\iota}) = BIAS(\widehat{\delta_\iota}) + \ v_i$$

$$BIAS(\widehat{\delta_i}) = E(\widehat{\delta_i}) - \delta_i$$

where v_i represents the residual having a null expected value and is reasonably smaller than $BIAS(\widehat{\delta_\iota})$ for samples of reasonable size.

STEP 5

Use the method of maximum likelihood to estimate the truncated regression of $\hat{\hat{\delta}}_i$ on z_i, to obtain the estimates $(\widehat{\widehat{\beta}}, \widehat{\widehat{\sigma}})$.

STEP 6

This step is necessary to obtain a set of bootstrap estimates $\mathcal{C} = \{(\hat{\beta}^*, \hat{\sigma}_\varepsilon^*)_b\}_{b=1}^{L2}$ and is conducted L2 times over the following three steps:

[6.1] For each $i = 1, \dots, n$, draw ε_i from the normal distribution $N(0, \hat{\sigma}_\varepsilon^2)$ with left-truncation at $(1 - z_i\hat{\hat{\beta}})$;

[6.2] Again, for each $i = 1, \dots, n$, compute $\delta_i^{**} = z_i\widehat{\widehat{\beta}} + \varepsilon_i$;

[6.3] Employ the method of maximum likelihood to estimate the truncated regression of δ_i^{**} on z_i, to obtain the estimates $\hat{\hat{\beta}}^*$ and $\hat{\hat{\sigma}}^*$.

STEP 7

Use the bootstrap values in $\mathcal{C}$, and the original estimates $\hat{\hat{\beta}}$ and $\hat{\hat{\sigma}}$, to construct estimated confidence intervals for each element of β and for σ_ε, with the desired level of confidence.

To determine the confidence interval for a given element of β, at a specified confidence level α, it is needed to find the values a_α and b_α such that the probability that ($\hat{\hat{\beta}}_j - \beta_j$) falls within the range $[-b_\alpha, -a_\alpha]$ is equal to $1-\alpha$:

$$Pr\,[-b_\alpha \leq (\hat{\hat{\beta}}_j - \beta_j) \leq -a_\alpha] = 1 - \alpha$$

However, since the distribution of ($\hat{\hat{\beta}}_j - \beta_j$) is unknown, Simar and Wilson proposed using the jth element of each bootstrap value $\hat{\hat{\beta}}^*$ to estimate a_α^* and b_α^* such that the probability that ($\hat{\hat{\beta}}_j^* - \beta_j$) falls within the range $[-b_\alpha^*, -a_\alpha^*]$ is approximately $1-\alpha$:

$$Pr\,[-b_\alpha^* \leq (\hat{\hat{\beta}}_j^* - \beta_j) \leq -a_\alpha^*] \cong 1 - \alpha$$

This approximation becomes more accurate as $L2$ approaches infinity. Once the values of a_α^* and b_α^* are obtained, they can be used to define the confidence interval for β_j at the α level as

$$[\,\hat{\hat{\beta}}_j + a_\alpha^*\,, \hat{\hat{\beta}}_j + b_\alpha^*]$$

The implementation of this algorithm was conducted by means of the *dea.env.robust* function included in the R package *rDEA*. A choice had to be made about the number of replications, L1 and L2, employed in the algorithm. L1, which determines the quantity of bootstrap replications required to compute the bias-corrected estimates $\hat{\hat{\delta}}_i$, was set at 100 replications. Simar and

Wilson (2007), showed that this is sufficient since $\hat{\hat{\delta}}_i$ necessitates only the computation of a mean followed by a difference.

On the other hand, L2 controls the number of bootstrap replications necessary for generating estimates of confidence intervals. Since estimating confidence intervals is akin to determining the tails of distributions, it necessitates more information. To this end, Hall (1986) suggested employing 1000 replications for estimating confidence intervals, whereas Simar and Wilson (2007) simulations employed 2000 replications, which was also the number adopted here.

Larger numbers of replications result in more precise estimates, and for confidence interval estimation, this occurs at a slow rate of diminishing returns. Nevertheless, the trade-off lies in the waiting time incurred when increasing the number of replications.

4.4 Random Forest

The double bootstrap procedure was used to improve the level of statistical significance of regression coefficients with respect to classical regression approaches. However, the fact that these coefficients are assumed to remain constant across all the values of the covariates, represents a further limitation when assessing the effects of environmental factors on school efficiency.

According to Schiltz et al. (2017), classical regressions may fail to identify nonlinear relationships between previously estimated efficiency scores $\hat{\hat{\delta}}$ and covariates, as well as possible interaction effects, if these are not considered a priori and included in the model phase.

Although the Italian school system is quite centralised, it was seen how schools have become more flexible in the management of their own resources. Considering the full context is, therefore, necessary to avoid making decisions based on over- or underestimations of effects related to each variable.

While classical regression techniques can identify important variables, advanced Machine Learning techniques, such as regression trees and Random Forests, can also capture nonlinear relationships and interaction effects between them, representing an alternative to the traditional two-stage approach (Rebai et al, 2020). By allowing data to catch these relationships without any prior assumptions, these techniques can provide a better understanding of the factors that influence school performance.

The first algorithm for random decision forests was developed by Ho (1995). Later, an extension of this algorithm was proposed by Breiman (2001), to which the R package *randomForest* used for this analysis refers.

RF is an ensemble method that combines several regression trees built over bootstrapped samples from the population. Unlike linear regression models, regression trees recursively partition the data into M smaller and non-overlapping subgroups, based on the values of the independent variables. The mean of the output values $\overline{\hat{\hat{\delta}}_m}$ of the observations within the subgroup, is then used to predict the value of each data point, as follows:

$$f(X) = \sum_{m=1}^{M} \overline{\hat{\hat{\delta}}_m}\, I_{(X \in R_m)}$$

The different subgroups are chosen, until a stopping criterion is reached (such as minimum number of samples for a node split, minimum number of samples for a terminal node, maximum number of terminal nodes, and so on) to minimize the residual sum of squares (RSS), which measures the variability of the dependent variable within each subgroup:

$$\sum_{m=1}^{M} \sum_{i \in R_m} (\hat{\hat{\delta}}_i - \overline{\hat{\hat{\delta}}_m})^2$$

Among the advantages of using regression trees, there is the fact that their structure provides a graphical representation of the decision-making process, as well as the identification of the most important features, as they are located at the top of the tree. Additionally, they can handle categorical predictors, without the need to generate dummy variables.

However, trees have their limitations, being sensitive to outliers and suffering from high variance. To overcome these problems, many trees can be aggregated into a ‘forest’, combining the predictions of each tree and obtaining an output with lower variance. What is more, RF selects a random sample of predictors at each split, so that trees do not rely on the same information and their outputs are less correlated, also avoiding overfitting and increasing generalization performance.

The Random Forest algorithm consists of the following steps:

STEP 1

Draw as many bootstrap samples from the original data as the desired number of trees.

STEP 2

For each bootstrapped sample, develop a regression tree randomly sampling m predictors at each split and choosing the best splits minimizing the RSS.

STEP 3

Predict each observation by averaging the predictions of all the trees in which that observation is contained in the sample.

As for Schiltz et al. (2017), the Random Forest adopted in this analysis based on 2000 regression trees and a number $m = \sqrt{p}$ of predictors selected as split candidates at each split, where p is total number of predictors. The method allowed to assess the importance of the predictors, considering how much the accuracy decreased when a single variable was excluded.

In fact, the algorithm first uses the out-of-bag (OOB) sample of data left out from each tree, to compute the mean prediction error of each observation over all trees that do not have the observation in their sample. Then, the values of the variable in the OOB sample are permuted, while keeping all other variables unchanged, and the error is computed again. The importance of variables is measured as the difference between OOB errors before and after the permutation, with variables having large values being more important than others, since not considering them decrease accuracy.

5. Data

5.1 Introduction

Since the mid-nineties, Italian schools have been granted administrative, organizational, and educational autonomy, and have been able to manage their own budgets. In fact, even though they have limited control over input amounts, schools have now some flexibility in managing resources, providing incentives to staff, and organizing service and education initiatives that can affect student performance (Di Giacomo and Pennisi 2015). This allows to consider them as true DMUs.

In DEA, DMUs are evaluated on their ability to make the best use of the resources at their disposal to produce outputs. Similarly, schools can be evaluated based on their ability to balance available resources with student needs and learning objectives. In this way, it is possible to identify schools that use resources efficiently and effectively and to provide guidance on how to improve the performance of less efficient schools.

For this thesis, data was initially collected at student-level. A sample of 502483 students, in the third year of lower secondary school in 2018-2019, was taken with reference to the sample built by INVALSI for the national student assessment. However, in accordance with Bradley et al. (2001), Mizala et al. (2002), Alexander et al. (2010), Agasisti (2011a, 2013), Podinovski et al. (2014) and others, the analysis was conducted at school-level, since it tended to provide good insights to be given to school management. Information was so aggregated, and from an initial sample of 4895 schools, some were excluded, mainly due to the presence of missing data, leading to a final number of 4264 schools.

The nonetheless satisfactory number of schools made it possible to proceed with the analysis without any imputation of data, thus avoiding any possibility of data distortion. This is a crucial aspect in any research project, as data imputation involves making assumptions about missing values, which may negatively affect the results, making them biased or unreliable.

Data was available for 18 different regions, as no information was present for the regions of Valle d’Aosta and Trentino Alto Adige, as well as for the two autonomous provinces of Trento and Bolzano. The schools sampled were distributed by region as follows.

Region	*Number of schools*	*Region*	*Number of schools*
Lombardia	645 (15.13%)	Calabria	198 (4.64%)
Campania	483 (11.33%)	Marche	131 (3.07%)
Sicilia	412 (9.66%)	Sardegna	119 (2.79%)
Lazio	368 (8.64%)	Abruzzo	108 (2.53%)
Veneto	325 (7.62%)	Liguria	102 (2.39%)
Puglia	311 (7.29%)	Friuli Venezia Giulia	76 (1.78%)
Emilia Romagna	296 (6.94%)	Basilicata	66 (1.55%)
Piemonte	295 (6.92%)	Umbria	56 (1.31%)
Toscana	252 (5.91%)	Molise	21 (0.49%)
Total number of schools		4264 (100%)	

Table 6. Number of sampled schools per region

Then, to also have a representation at macro-area level, it was possible to group regions into 5 different geographical areas:

- Northwest: Liguria, Lombardia, Piemonte
- Northeast: Emilia Romagna, Friuli Venezia Giulia, Veneto
- Centre: Lazio, Marche, Toscana, Umbria
- South: Abruzzo, Campania, Molise, Puglia
- South and isles: Basilicata, Calabria, Sardegna, Sicilia

The number of schools per geographical area was reported below:

Geographical area	*Number of schools*
Northwest	1042 (24.44%)
Northeast	697 (16.35%)
Centre	807 (18.93%)
South	923 (21.65%)
South and isles	795 (18.64%)
Total number of schools	4264 (100%)

Table 7. Number of sampled schools per geographical area

Note that, as already mentioned, there was no data from private schools, and the analysis focused exclusively on public schools.

5.2 Data Sources

The source dataset was obtained combining INVALSI test scores with additional information provided by the Statistics Office of the Ministry of Education. In this section both entities were described in more detail.

5.2.1 INVALSI

INVALSI is an Italian public body, established in 1999 and overseen by the Ministry of Education, responsible for assessing the quality of the national education system through the development of standardised tests.

Its objectives include: carrying out periodic and systematic checks on the learning outcomes of Italian students and processing their results, to give a general picture of the results at national, macro-area (Northwest, Northeast, Centre, South, South and Islands), regional and even provincial level for each grade and school address tested; to update and improve the assessment activities of schools and the school system as a whole; to analyse the reasons for school dispersion and the causes of failure. It also takes care of Italian participation in international evaluation surveys, like the OECD PISA tests mentioned in chapter 3 about the Italian education system.

After two years of experimentation, the first national INVALSI tests (in mathematics and Italian) were held in the 2005-2006 school year.

Today, such tests are administered at different stages of students' school careers that are:

- the second and fifth years of primary school (when pupils are 7 and 10 years old, respectively);
- the third year of lower secondary school (at 11 and 13 years old);
- the second and the fifth year of upper secondary school (at 15 and 18 years old).

They are used to measure the level of student learning in various disciplines, such as Italian, mathematics and English, but, while the first two cover all the abovementioned classes, the English test, introduced in 2018, is only compulsory for the final classes and is divided in two parts, reading and listening.

INVALSI tests are not memory tests. Instead, they require students to have the ability to reason to make the best use of what they have learnt, above all because the ability to reason is fundamental throughout life. Secondary school students take the tests online thanks to the computers provided by the schools, and, for the final classes, taking the tests is a requirement for admission to the State examination.

Test results are one element of a school's self-evaluation activities. They do not evaluate teachers' work, nor do the mechanisms of rewards and sanctions for institutions and teachers depend on test results. However, they provide information to teachers and educational institutions on the quality of their work. The information obtained from the tests is intended to be used as a tool to help teachers and schools improve their teaching methods and the educational experience of their students. By analysing the results, it is possible to identify areas where students are struggling and adjust the teaching strategies accordingly.

For several years, schools in Italy had the option to participate in INVALSI education assessments voluntarily. However, beginning with the 2009-2010 school year, all schools, both public and private, are required to participate and test all students in class. INVALSI provides information and the testing is administered directly by the schools on a predetermined date. Nonetheless, a representative sample of schools is selected from various regions each year to provide more expedient and impartial results and external observers are sent to these selected schools.

5.2.2 Sources from the Ministry of Education

The Statistics Office is an internal structure of the Ministry of Education that is responsible for collecting, processing and analysing data on the Italian education system. These processes allow to have a precise knowledge of the phenomena underlying the education system and are indispensable for administrative action and the performance of the institutional tasks of planning, monitoring and governance.

This entity may collaborate with other government agencies and organisations in the education sector to collect data and share information, and is involved in the development of national standards for the collection and presentation of education-related statistical data.

Among its activities, there are the collection of information on schools and students, the production of studies and analyses about the education system, the production of official statistics and the publication of reports and monitoring indicators.

Information on Italian schools concerns the number of students enrolled, the number of teachers, the equipment available and the activities carried out, whereas data on students refers to their geographical origin, age, gender, level of education and school performance. It also manages, at every stage of the process, the national student registry, for the realisation of the right to education.

Collecting data from primary and secondary sources, statistical information is provided through the publication of reports and the creation of databases. Most of the surveys carried out are included in the PSN (National Statistical Programme), which sets out the statistical surveys of public interest entrusted to SISTAN, the National Statistical System, which is the network of public and private entities providing official statistical information in Italy.

The Statistics Office is responsible for producing information and statistical data used to formulate public policies and make decisions. The processing and analysis of the data collected makes it possible to monitor the effectiveness of the Italian education system, identify any critical issues and define strategies for improving the quality of education. The statistics produced represent an important source of information for policy makers, scholars, practitioners and even the public.

Indeed, the ministry wants to permanently guarantee the accessibility and reusability of public data of the national education and training system, and publish them in open format. The online publication of data enhances the information heritage and promotes a deep and authentic knowledge of Italian schools.

Making data available to citizens, without the need for any authentication or identification, means guaranteeing administrative transparency and participation in the improvement of the school system as well as promoting the emergence of new services for students, teachers, families, research institutes or anyone interested in learning about the school world.

5.3 Variables

Combining data from the two sources, 230 variables were initially available. A selection process was made according to the purposes of the analysis and the insights emerged from the literature review.

5.3.1 Input variables

The proposed DEA model based on four different input variables. The first two were related to the prior achievement of the students: *Ita_scores_5* and *Mat_scores_5*, respectively reported the INVALSI test scores of Italian and mathematics that the analysed students achieved in the fifth year of the primary school.

Their values were expressed in a proficiency scale provided by the Rasch model, according to which the national average score for all levels of education was set at 200. That model assumes that the pupils’ responses are governed by two components: the first component is their ability or skill in choosing the right answer, the greater their preparation, the greater the probability of giving the correct answer; the second component is the difficulty of the questions. It is a probabilistic model, in which the response is always the result of the interaction between these two elements, ability and difficulty.

In addition to grammar, the Italian test assesses the ability to understand an authentic text, literary or otherwise, measuring several aspects of linguistic competence regarding students' ability to reflect on the text, evaluate it, understand its logical organisation and internal connections.

The mathematics test instead, verifies the most important knowledge, problem-solving and argumentation skills in four areas: probability and statistics, arithmetic or algebra, geometry, relations and functions. Of the three tests, it is the one that depends most on the possession of disciplinary knowledge, but the questions often start from real-world problems, and ask pupils to be able to think about why they make choices, thus being able to use deductive logic correctly.

Data from the English test results was not available, and it was therefore not possible to include it in the analysis.

The third input variable was the *ESCS* (Economic, Social and Cultural Status). It is an index about the socio-economic cultural background of the student’s family.

The literature review showed unequivocally the importance of this index, both as an input and explanatory variable of achievement levels, for its implications in terms of possible school policies. It was chosen to take it as an input, as it allowed to catch a higher variability in the computation of the efficiency scores, as it will be explained when presenting the results in the next chapter.

The ESCS is based on 3 indicators:

1) employment status of parents;
2) level of education of parents;
3) possession of certain specific good materials, intended as variables of proximity of an economic and cultural context favourable to learning.

Its actual calculation is carried out through a principal components analysis (PCA) of the 3 indicators introduced: in line with the OECD-PISA proposal, the factorial scores associated with the first principal component (normally able to explain at least 50% of the total variance) is assumed as ESCS values.

The ESCS by construction is an indicator with zero mean and unitary standard deviation. Thus, a student with a strictly positive individual ESCS is a student with a more favourable socio-economic and cultural background than the Italian average.

The aggregation of these first three variables at school level was done by averaging them for each school.

The last input considered was strictly related to school’s characteristics and was the *Teacher/student* ratio. Differently from the others, this variable did not represent information present in the original dataset but was constructed by considering the ratio between the number of teachers and the number of students of the school. In particular, the number of teachers was obtained summing the number of tenured and non-tenured teachers.

5.3.2 Output variables

In line with the choice of inputs, two outputs concerning students’ achievement were used. The variables *Ita_scores_8* and *Mat_scores_8* referred to the INVALSI test scores of Italian and mathematics that the analysed students achieved in the third year of the lower secondary school. The values they assumed can be explained in the same way as for the respective inputs.

Input and output variables were listed below.

Variable name	*Type*	*Value Description*
Ita_scores_5	Input	Avg. test scores per school
Mat_scores_5	Input	Avg. test scores per school
ESCS	Input	Avg. ESCS per school
Teacher/student	Input	Ratio
Ita_scores_8	Output	Avg. test scores per school
Mat_scores_8	Output	Avg. test scores per school

Table 8. DEA input and output variables

5.3.3 Non-discretionary variables

Simar and Wilson (2007) showed that all the previous attempts to accurately measure the effects of non-discretionary variables on efficiency measures, using non-parametric methods, were inconsistent due to the inability to construct confidence intervals. Nonetheless, the double-bootstrapped method overcame this problem enabling the use of explanatory variables and assessing the magnitude and significance of their impact.

First, according to one of the objectives of the analysis as that of obtaining a comparison of school efficiency by geographical areas, a set of dummies identifying schools located in the North, Centre, South, and South and islands was created.

Then, all the other non-discretionary variables adopted in the regression part could be classified in four different groups referring to school demographic variables, funding, teachers and school head. A further classification was made for funding, since it was possible to distinguish between the way they were used and the sources from which they came. Following this classification allowed to implement the different model alternatives.

Starting with school demographic variables, *school_size* and *class_size* referred to the number of students attending respectively that school or that class, while *% of girls* and *% of immigrants* accounted for the percentage of female and immigrant students in each school.

The calculation of *% of girls* and *% of immigrants*, given the original structure of the data, was possible only considering the number of observations, and thus of students, subject to the analysis, and not the values pertaining to the whole school. So, the assumption was made that schools maintained the same percentages regarding the presence of female and immigrant students.

Moreover, in the 2007-2008 school year, a fundamental distinction was introduced between pupils with non-Italian citizenship born in Italy and pupils of recent immigration, believing that such distinction indicated two very different types of students with different educational needs. However, this difference was not considered in the analysis, and *% of immigrants* included all students with foreign parents.

For what concerns funding, *%general funds*, *%salaries*, *%educational offer,* represented the variables related to how schools used the funds received, while *%private*, *%families*, *%municipality*, *%province*, *%region*, *%State*, *%UE* referred to the respective sources. All the adopted variables were expressed as percentage of the total funding.

More in detail, *%salaries* included salaries of permanent and substitute staff as well as resources for staff compensation, whereas the values of the two variables *%educational offer* and *%UE* were all equal, meaning that the funds coming from the European Union were totally allocated to the expansion of the educational offer.

About teachers, the years of service in the school and the number of days of absences were considered. Specifically, the variables *%teach1*, *%teach2-3*, *%teach4-5* and *%teach5more* indicated the percentages in terms of number of teachers respectively with 1, 2 or 3, 4 or 5, more than 5, years of service, instead, the variable *other_absences* expressed the average annual per capita absences of teachers not due to illness or maternity.

Last two variables concerned the figure of the school head. Respectively, *dir_esp* and *dir_serv* were related to the years of experience and years of service in the school of the school head.

Table 9. Non-discretionary variables considered

	Variable name	*Value Description*
School location	centre	Dummy {0,1}
	south	Dummy {0,1}
	south_and_isles	Dummy {0,1}
Demographic variables	school_size	Number of students
	class_size	Number of students
	% of girls	Percentage
	% of immigrants	Percentage
Funding use	%general funds	Percentage
	%salaries	Percentage
	%educational offer	Percentage
Funding sources	%private	Percentage
	%families	Percentage
	%municipalities	Percentage
	%province	Percentage
	%region	Percentage
	%State	Percentage
	%UE	Percentage
Teachers	%teach1	Percentage
	%teach2-3	Percentage
	%teach4-5	Percentage
	%teach5more	Percentage
	other_absences	Avg. per capita days/year
School head	dir_esp	Number of years
	dir_serv	Number of years

5.3 Descriptive statistics

The statistical description of variables was necessary to provide a clear and concise overview of the data being used. In this section, the main statistics of all considered variables were explored, to allow a better understanding of the context and, afterward, of the results illustrated in the next chapter.

The following table shows the mean, standard deviation, minimum and maximum of each variable.

Table 10. Descriptive statistics of variables

	Variable name	*Mean*	*Standard deviation*	*Min*	*Max*
Inputs	Ita_scores_5	208.388	14.498	130.061	294.195
	Mat_scores_5	213.134	19.576	113.042	339.807
	ESCS	0.112	0.327	-1.740	1.303
	Teacher/student	0.462	0.209	0.078	5.077
Outputs	Ita_scores_8	201.299	11.852	119.317	243.643
	Mat_scores_8	200.867	14.785	135.795	248.239
Non-discretionary variables	north	0.408	0.491	0.000	1.000
	centre	0.189	0.392	0.000	1.000
	south	0.216	0.412	0.000	1.000
	south_and_isles	0.186	0.389	0.000	1.000
	school_size	324.021	155.691	13.000	1130.000
	class_size	20.379	3.141	7.000	30.000
	% of girls	0.485	0.061	0.200	0.756
	% of immigrants	0.103	0.096	0.000	0.877
	%general funds	0.067	0.061	0.000	0.672
	%salaries	0.904	0.084	0.270	1.000

Non-discretionary variables	%educational offer	0.029	0.054	0.000	0.607
	%private	0,005	0.012	0.000	0.234
	%families	0.007	0.009	0.000	0.130
	%municipality	0.025	0.045	0.000	0.378
	%province	0.0002	0.002	0.000	0.036
	%region	0.010	0.031	0.000	0.639
	%State	0,923	0.080	0.281	1.000
	%UE	0.029	0.054	0.000	0.607
	%teach1	0.094	0.084	0.000	1.000
	%teach2-3	0.150	0.099	0.000	1.000
	%teach4-5	0.106	0.091	0.000	1.000
	%teach5more	0.650	0.157	0.000	1.000
	other_absences	20.016	5.731	3.000	52.000
	dir_esp	3.633	0.629	1.000	4.000
	dir_serv	2.658	1.152	1.000	4.000

Table 10. (continued)

6. Results

This chapter presents the results of the entire analysis and is organized in three sections: the first refers to the use of traditional DEA for the calculation of school efficiency scores; the second illustrates the results of the double bootstrap procedure, implemented through Simar and Wilson's (2007) second algorithm; the third reports the findings of the Random Forest.

Conclusions and policy implications for school management, based on such results, will follow in the next chapter.

6.1 Efficiency scores through traditional DEA

Based on available data, to assess if different combinations of inputs and outputs could affect efficiency scores returned by DEA, 6 different alternative models were evaluated:

- Model 1: Inputs are *Ita_scores_5*, *Mat_scores_5*, *Teacher/student*, Outputs are *Ita_scores_8, Mat_scores_8*;
- Model 2: Inputs are *Ita_scores_5*, *Mat_scores_5*, *Teacher/student, ESCS*, Outputs are *Ita_scores_8*, *Mat_scores_8*;
- Model 3: Inputs are *Ita_scores_5*, *Mat_scores_5*, *Teacher/student*, Output is *Ita_scores_8*;
- Model 4: Inputs are *Ita_scores_5*, *Mat_scores_5*, *Teacher/student*, Output is *Mat_scores_8*;
- Model 5: Inputs are *Ita_scores_5*, *Mat_scores_5*, *Teacher/student*, *ESCS*, Output is *Ita_scores_8*;
- Model 6: Inputs are *Ita_scores_5*, *Mat_scores_5*, *Teacher/student*, *ESCS*, Output is *Mat_scores_8*.

As stated in section 4.2.2, a traditional DEA with output orientation and variable returns to scale was applied to these different sets of input and output variables. Although the choice of an output orientated model returned efficiency scores assuming values in the range [1, +inf), it was preferred to report the scores considering their reciprocal and have them ranging in [0,1].

The efficiency scores obtained were analysed looking at the summary statistics and histograms below.

Model	*Min*	*1st Quartile*	*Median*	*Mean*	*3rd Quartile*	*Max*
Model 1	0.5785	0.8532	0.8921	0.8847	0.9222	1.000
Model 2	0.6561	0.8685	0.9046	0.8984	0.9330	1.000
Model 3	0.5036	0.8476	0.8821	0.8763	0.9112	1.000
Model 4	0.5470	0.8200	0.8713	0.8603	0.9070	1.000
Model 5	0.5607	0.8529	0.8862	0.8816	0.9145	1.000
Model 6	0.6596	0.8374	0.8828	0.8747	0.9157	1.000

Table 11. Summary statistics of efficiency scores for six alternative DEA models

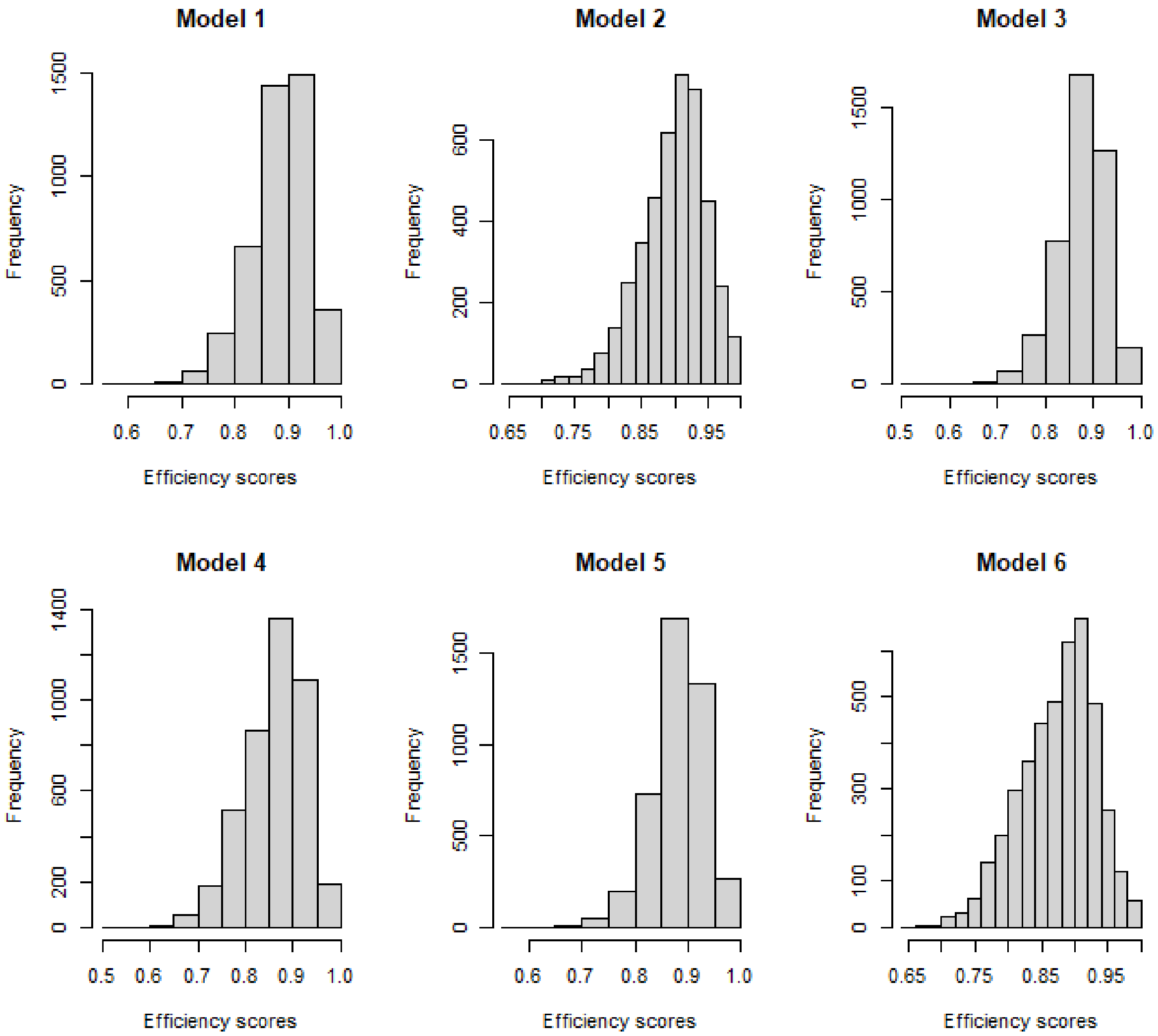

Figure 8. Histograms of efficiency scores

According to the displayed values, an overall good level of efficiency was shown, with median scores all exceeding the value of 0.87. However, also looking at the distributions of the scores, a certain dispersion of data could be noted, with some schools reaching minimum efficiency levels just above 0.5, as in Model 3. All distributions were in fact asymmetrical, showing long tails on the left side, and there was room for further investigation at a later stage of the analysis, to look for any discriminating factors for such differences in efficiency.

Afterwards, to further exploring possible consequences of changing model specifications, the correlation between the efficiency scores of the different models was checked.

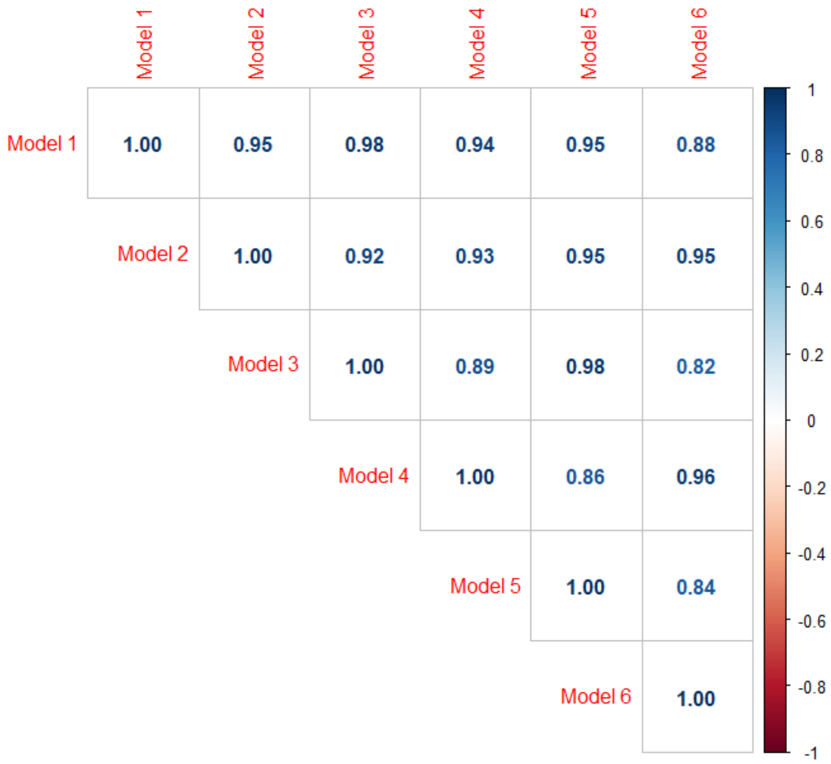

Figure 9. Correlation plot of efficiency scores of the different models

The scores were found to be all positively correlated with each other, with a minimum correlation value of 0.82 between Models 3 and 6. This showed that, whether to consider the *ESCS*, as much as the choice of a single output instead of two, did not affect the efficiency scores in a meaningful way and the choice of a model to run the double bootstrap procedure had to be based on further considerations.

In the first instance, models considering the *ESCS*, such as Models 2 and 6, allowed for a better explanation of the variability present in the scores, as can be seen from the narrower intervals in the respective histograms. Then, it had to be considered whether to adopt a model with only one or both outputs, since *Ita_scores_8* and *Mat_scores_8* were high correlated (0.89), and this could affect the results.

From a theoretical point of view, school could have a specificity in the production efficiency of one output instead of the other, even if they are highly correlated, due to the quality of the teachers, their experience, the teaching methods adopted. In this sense, it would be preferable to explore this possibility by means of an analysis that keeps them separate. However, the correlation plot showed that the results were very similar, and even making a different choice, they would not change so much.

Therefore, Model 2 with *Ita_scores_5*, *Mat_scores_5*, *Teacher/student* and *ESCS* as inputs, and *Ita_scores_8* and *Mat_scores_8* as outputs, was the one selected to run the double bootstrap procedure.

6.2 Results of Simar and Wilson's (2007) algorithm #2

Evaluating the results of traditional DEA models, allowed to select the best combination of inputs and outputs among those proposed. The use of Simar and Wilson's (2007) second algorithm enabled first to improve the accuracy of the efficiency scores of the chosen model and then to use the robust scores as dependent variable in a regression analysis to assess any factors that might affect those scores.

6.2.1 Robust efficiency scores

As mentioned earlier, the double bootstrap procedure overcame what was described by many as the main limitation of DEA, namely the fact that it is deterministic, and it allowed to derive more robust efficiency scores. Considering the possibility of random noise in the DEA model itself, a more accurate and reliable assessment of the efficiency of each school was obtained, as well as a better understanding of the variability in the efficiency scores.

The table below shows the original estimation $\hat{\delta}$ of the efficiency scores, which, as can be seen, is the same as that computed using traditional DEA, the bias-corrected estimation $\hat{\hat{\delta}}$ and the upper and lower bounds of the bias-corrected estimation with a 95% confidence level, where all these values refer to average values.

		Confidence Interval ($\hat{\hat{\delta}}$)	
$\hat{\delta}$	$\hat{\hat{\delta}}$	*Lower Bound*	*Upper Bound*
0.8984	0.8847	0.8777	0.8914

Table 22. Estimated efficiency scores and confidence intervals

According to the explanation of the steps of the algorithm in section 4.3.1, the bias-corrected estimation $\hat{\hat{\delta}}$ was calculated by subtracting the bias, which is a positive value, from the original estimation $\hat{\delta}$. For this reason, the bias-corrected values were lower than the original ones. Specifically, this difference was equal to 1.3%.

To fully grasp the results concerning these estimations, it was not enough to look at these numbers, and a histogram was helpful to inspect the dispersion of said scores.

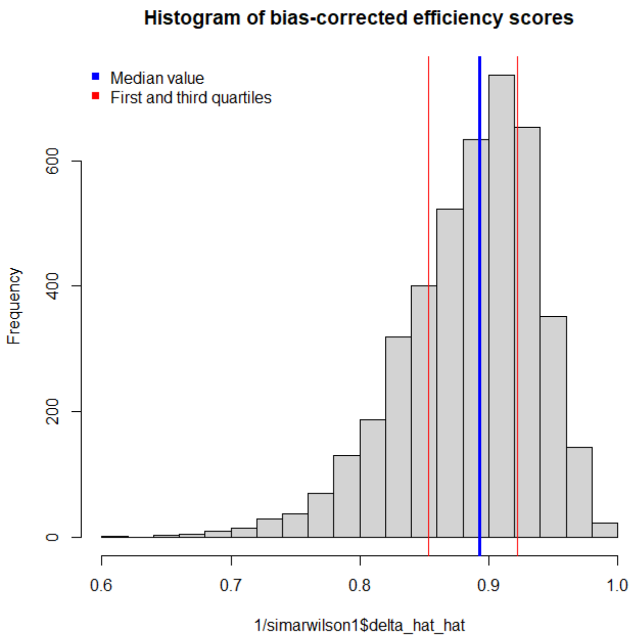

Figure 10. Bias-corrected efficiency scores' distribution

As for the traditional DEA model, the overall level of efficiency was high, with 75% of the schools having a score above 0.85, and almost half of the schools exceeding the value of 0.89.

However, the bias-corrected efficiency scores still presented a left-skewed distribution, with few schools having low efficiency values down to a minimum of 0.61. Full statistics were reported below.

Bias-corrected efficiency scores ($\hat{\hat{\delta}}$)					
Min	*1st Quartile*	*Median*	*Mean*	*3rd Quartile*	*Max*
0.6112	0.8537	0.8932	0.8847	0.9224	0.9893

Table 13. Summary statistics of bias-corrected efficiency scores

Note that the maximum value was not equal to 1 anymore, due to the bootstrap procedure, where no school performed as totally efficient in each of the replications of the algorithm.

Then, to have an idea of how schools allocated their input resources, generating respective outputs, and to capture differences in efficiency, some of the best and worst performing schools were reported below.

Table 14. Different ways to be efficient: some of the best and worst schools

		Inputs				*Outputs*	
School	$\hat{\hat{\delta}}$	*Ita_ scores_5*	*Mat_ scores_5*	$\frac{Teacher}{student}$	*ESCS*	*Ita_ scores_8*	*Mat_ scores_8*
(Top schools)							
40808055	0.9893	220.813	208.740	0.305	0.674	225.945	224.457
40401149	0.9887	217.305	223.967	0.424	-0.026	222.790	229.545
20206262	0.9879	212.077	219.806	0.131	0.502	216.406	212.256
40804041	0.9866	197.420	203.162	0.395	-0.135	205.773	225.230
40407234	0.9865	223.021	214.683	0.359	0.983	222.004	236.921
40402112	0.9864	214.402	208.030	0.464	-0.041	216.466	228.823
42008088	0.9855	229.439	258.020	0.667	0.376	235.129	236.149
41204487	0.9855	226.351	222.421	0.274	1.066	233.045	221.340
40703030	0.9849	208.155	213.273	0.207	0.654	218.017	223.693
41503238	0.9848	225.252	242.396	0.328	0.321	232.851	222.908
41103042	0.9848	236.006	231.013	0.299	0.648	231.863	236.987
40403041	0.9843	216.556	217.342	0.438	0.394	226.125	231.915

(Worst schools)							
41906182	0.6870	203.707	215.464	0.646	-0.623	158.320	153.773
41906094	0.6862	204.808	203.012	0.703	-1.156	157.092	153.538
41504332	0.6858	189.163	201.760	0.582	-0.933	159.492	152.644
41906181	0.6832	176.294	178.174	0.498	-0.876	156.886	148.731
41906205	0.6830	203.294	230.419	0.785	-0.515	161.087	149.419
41504692	0.6764	180.088	196.642	0.492	-1.174	152.707	152.193
41804061	0.6733	181.884	205.343	0.617	-0.580	152.932	152.926
41504382	0.6697	198.148	168.520	0.588	-1.052	154.665	145.153
41803011	0.6617	168.238	197.045	0.722	-0.806	149.087	150.966
41906199	0.6559	155.767	158.064	0.670	-0.818	147.263	138.882
41906178	0.6434	213.604	228.447	1.472	-1.572	119.317	142.938
41906180	0.6112	233.755	251.394	0.729	-1.179	140.944	135.795

Table 15. (continued)

From these values, it could be seen that there can be several ways to perform efficiently, so, not only achieving relatively high output values but also employing a relatively lower number of inputs.

For instance, comparing the schools identified by the codes 40804041 and 42008088, the latter turned out to be efficient, despite employing high amount of input resources, i.e., highest *Teacher/student* and *Mat_scores_5* than all the other schools shown, because of the relevant student achievement as outputs. The former, on the other hand, even with relatively lower output values, was considered efficient due to lower input values, with a negative *ESCS*, and a far lower *Teacher/student*.

With regard to the worst schools instead, the school identified by the code 41504332, despite receiving students with very similar test results in input, again compared to the efficient school 40804041, and moreover having a much lower *ESCS* (although higher *Teacher/student*), was classified as inefficient due to poor output results. All the worst schools reported here had negative *ESCS*, and higher *Teacher/student* values than the best performing schools.

As explained previously in this work, efficient schools are located on the efficiency frontier created by the DEA model. A graphical representation of such a frontier in a two-dimensional context, would only be possible for three different situations, i.e., cases with only 1 input and 1 output, or cases with 2 inputs for given outputs and vice versa.

However, it was possible to have an approximation of the frontier even for the considered model with 4 inputs and 2 outputs, using on the x-axis the sum of all inputs and on the y-axis the sum of all outputs (Figure 11). This facility could only be considered knowing that, using to plot multi input and multi output, then fully multi efficient schools may no longer be placed on the two-dimensional frontier (see the R function *dea.plot* in the *Benchmarking* package).

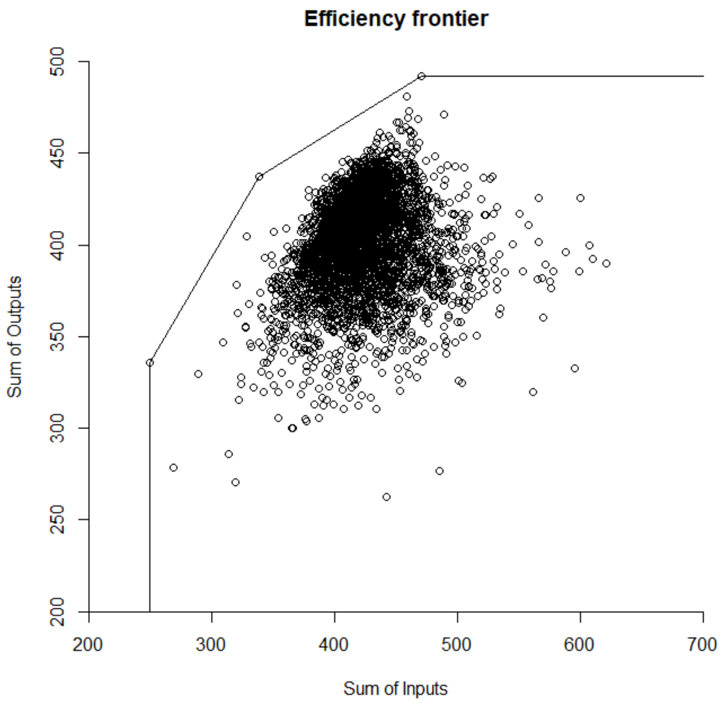

Figure 11. Approximation of the efficiency frontier created by DEA

The reason why this graph was used anyway, although it only referred to an approximation of the efficiency frontier, was that it was useful for an initial differentiation of schools according to geographical area. In fact, reproposing the same, but with the colouring of the different areas, it was possible to note that the schools analysed were gradually further away from the efficiency frontier as one moved from the North to the Centre to the South and the Islands.

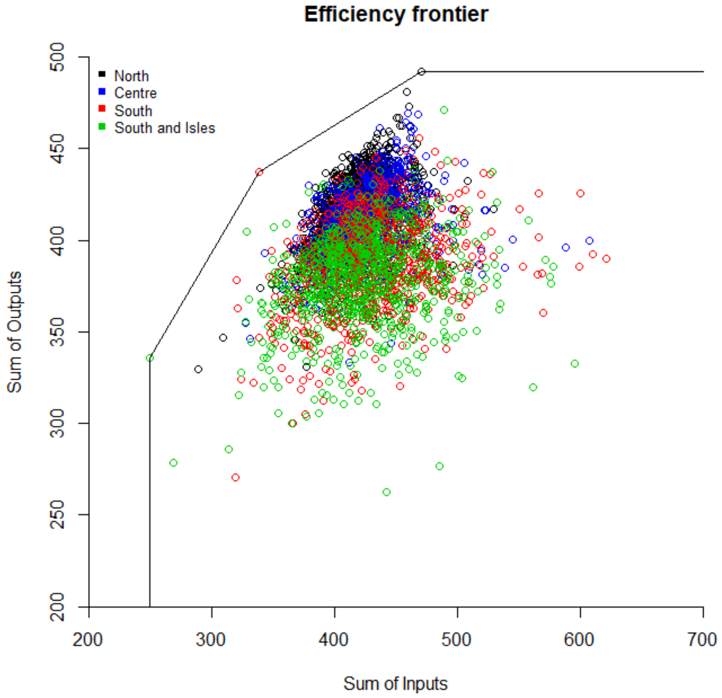

Figure 12. Approximation of the efficiency frontier with schools coloured by geographical area

Unlike from the initial classification, which was derived from INVALSI and distinguished between north-eastern and north-western regions, at this stage of the analysis it was preferred to consider all these regions as belonging to a single geographical area, namely the North. After all, both areas Northeast and Northwest referred to schools with high level of efficiency and combining them simplified the presentation and the interpretation of results.

After this initial insight, a more comprehensive analysis of possible differences in school efficiency at territorial level was conducted, first by geographical area, and then extending the level of the analysis to the single regions.

Box plots of the bias-corrected scores in Figure 13 confirmed what emerged in the first instance. Score distributions, and so the medians, tended to shift towards lower values as one moved from northern to southern regions and islands, as well as presenting a wider dispersion of values, resulting from larger boxes and longer whiskers.

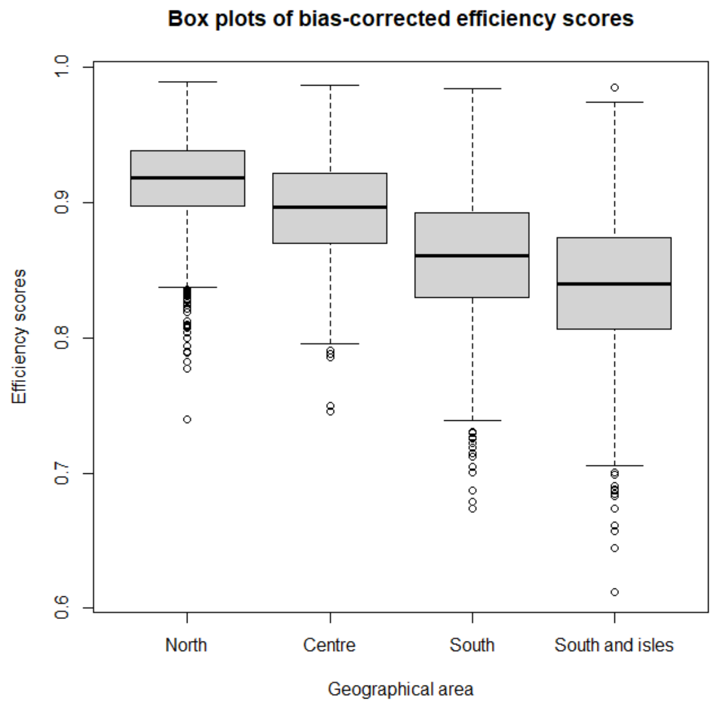

Figure 13. Box plots of bias-corrected efficiency scores by geographical area

A better understanding of these distributions was possible looking at the statistics below.

	Bias-corrected efficiency scores ($\hat{\hat{\delta}}$)					
Area	*Min*	*1st Quartile*	*Median*	*Mean*	*3rd Quartile*	*Max*
North	0.7396	0.8977	0.9184	0.9161	0.9382	0.9890
Centre	0.7454	0.8695	0.8968	0.8939	0.9213	0.9868
South	0.6740	0.8295	0.8609	0.8588	0.8921	0.9843
South and isles	0.6123	0.8062	0.8399	0.8366	0.8741	0.9854

Table 15. Summary statistics of bias-corrected efficiency scores by geographical area

Furthermore, data were also reported on a regional level, so that the school performance in each region could be investigated.

		Bias-corrected efficiency scores ($\hat{\hat{\delta}}$)					
Area	*Region*	*Min*	*1st Quartile*	*Median*	*Mean*	*3rd Quartile*	*Max*
North	Friuli Venezia Giulia	0.8458	0.9063	0.9294	0.9262	0.9453	0.9849
	Veneto	0.8254	0.9081	0.9255	0.9236	0.9429	0.9819
	Lombardia	0.7895	0.9021	0.9219	0.9202	0.9418	0.9881
	Emilia Romagna	0.8259	0.8947	0.9158	0.9141	0.9355	0.9890
	Piemonte	0.7396	0.8847	0.9079	0.9044	0.9265	0.9879
	Liguria	0.7888	0.8772	0.9016	0.8993	0.9252	0.9773
Centre	Marche	0.8479	0.8935	0.9107	0.9107	0.9280	0.9844
	Umbria	0.8325	0.8846	0.9052	0.9036	0.9282	0.9613
	Toscana	0.7908	0.8850	0.9014	0.9007	0.9240	0.9776
	Lazio	0.7454	0.8558	0.8820	0.8817	0.9088	0.9868
South	Abruzzo	0.7750	0.8617	0.8801	0.8842	0.9003	0.9787
	Puglia	0.7297	0.8446	0.8780	0.8743	0.9055	0.9776
	Molise	0.8077	0.8582	0.8751	0.8693	0.8864	0.9199
	Campania	0.6740	0.8167	0.8458	0.8427	0.8738	0.9843
South and Isles	Sardegna	0.7106	0.8361	0.8657	0.8607	0.8900	0.9854
	Basilicata	0.7534	0.8364	0.8637	0.8587	0.8823	0.9433
	Sicilia	0.6123	0.8071	0.8417	0.8361	0.8752	0.9740
	Calabria	0.6612	0.7880	0.8190	0.8160	0.8465	0.9608

Table 16. Summary statistics of bias-corrected efficiency scores by region

Within each geographical area, regions were listed by decreasing median efficiency level, while it is recalled that data for the regions of Valle d'Aosta and Trentino Alto Adige were not available.

The statistics showed that Veneto, Friuli Venezia Giulia and Lombardia had 75% of their schools with efficiency values above 0.90, while the school with the highest peak in efficiency was in Emilia Romagna (school 40808055). There was a difference of almost 11% between median scores of schools in Friuli Venezia Giulia and the ones of schools in Calabria, this gap becoming even much greater if considering the minimum values, close to 19%.

In general, the scores of the schools in the Centre maintained a distribution not too distant from those of the schools in the North, central and northern regions all had median values above 0.90, the only exception being the region of Lazio, with values closer to those of schools in the South. There was instead, a clear difference moving to the curves concerning the schools in the South and the South and the islands. These areas presented curves shifted towards lower values of the scores, flatter and larger, indicating a higher variability in the data. The worst performing school was in Sicily (school 41906180).

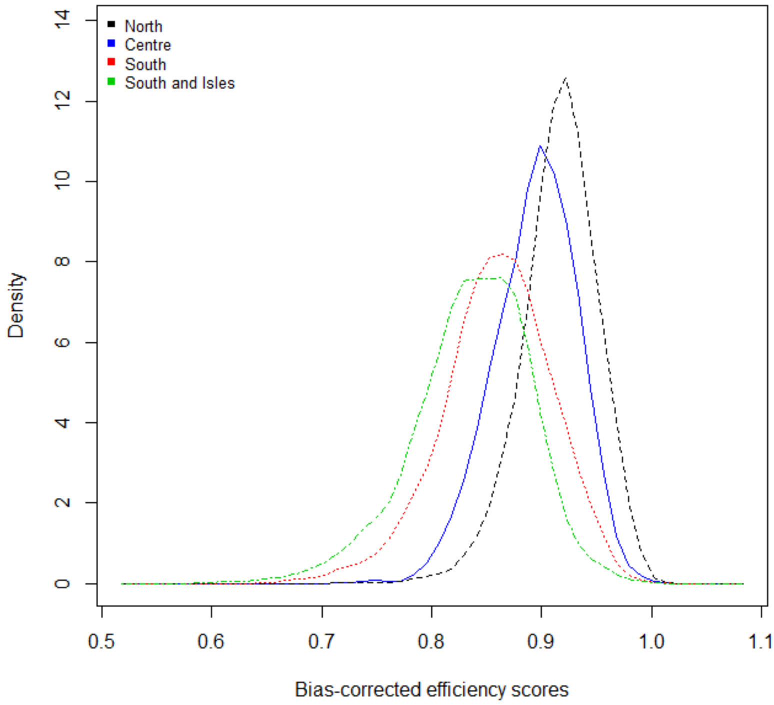

Figure 14. Density plots of bias-corrected efficiency scores by geographical area

At this point of the analysis, it was useful to make an important reflection. The calculation of efficiency scores with DEA, allowed a careful assessment of the contribution that schools had on student performance. The fact that the schools that turned out to be efficient from the analysis, were those located in northern regions of the country and receiving more prepared students, allowed to understand that their contribution in such preparation, was greater than the schools not considered as efficient.

In other words, the results said that schools with higher student performance were also those most capable to manage the available resources in such a way as to impact on that performance. This made for a less ordinary analysis, which might otherwise have been seen as the usual North-South divide that has been in vogue in the national arena for years.

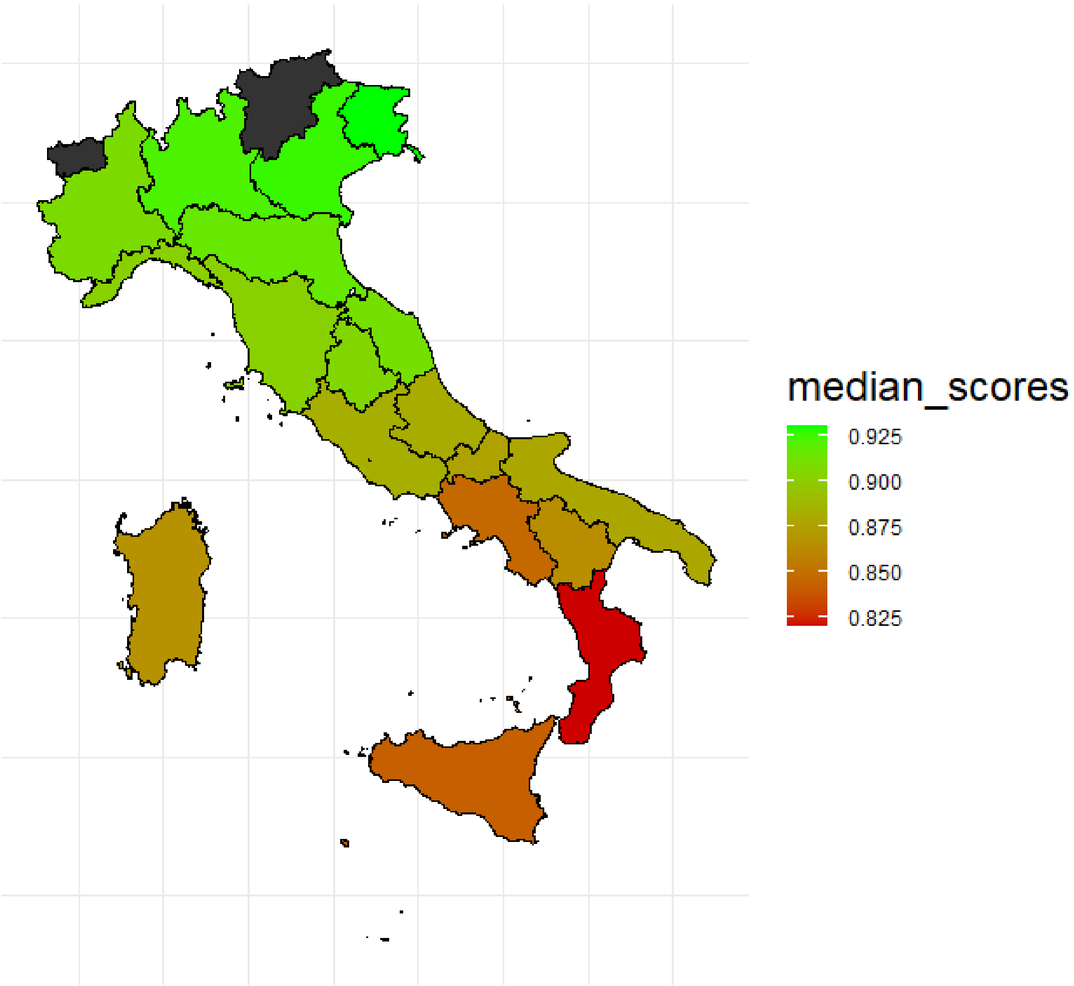

Figure 15. Map of Italy colored by regional median values of bias-corrected efficiency scores

6.2.2 Factors influencing efficiency

Once a complete picture of the efficiency score estimations was provided, the focus of the analysis shifted to inspecting the presence of any factors that may affect the values of those estimations. Specifically, the bias-corrected scores were used, in the second part of the algorithm, as the dependent variable within a truncated regression on the non-discretionary variables listed in the previous chapter.

Based on the classification made for these variables, seven different models were evaluated, obtained by a forward stepwise procedure, which allowed to gradually consider more and more variables, until all the variables under consideration were add to the model.

Two clarifications in this regard. First, the dummy variables concerning the location of schools were always included, for obvious reasons of interpretation, considering that one of the main objectives of the analysis was precisely to assess possible differences in efficiency between different geographical areas. Furthermore, as for the variables expressing complementary percentages, i.e., the ones regarding funding and teachers, it was possible to disregard one variable in the set. Thus, the variable *north*, as well as *%educational offer*, *%UE* and *%teach1*, were kept out of the various models.

Second, variables concerning the use and sources of funding were included alternatively, since, as can be seen from the correlation plot below, several of them had significant correlation values. Just think of the fact that European Union funds were entirely used for the expansion of the educational offer, while a large part of State funds were used to pay teachers' salaries. Holding them together within a single model, could have led to multicollinearity problems. On the other hand, the same R function adopted (*dea.env.robust*) already took this possibility into account, returning an error in that case.

Note that, from here on, all the presented results referred to the original values of the efficiency scores, in the range [1, +inf), and not to their reciprocal values as before.

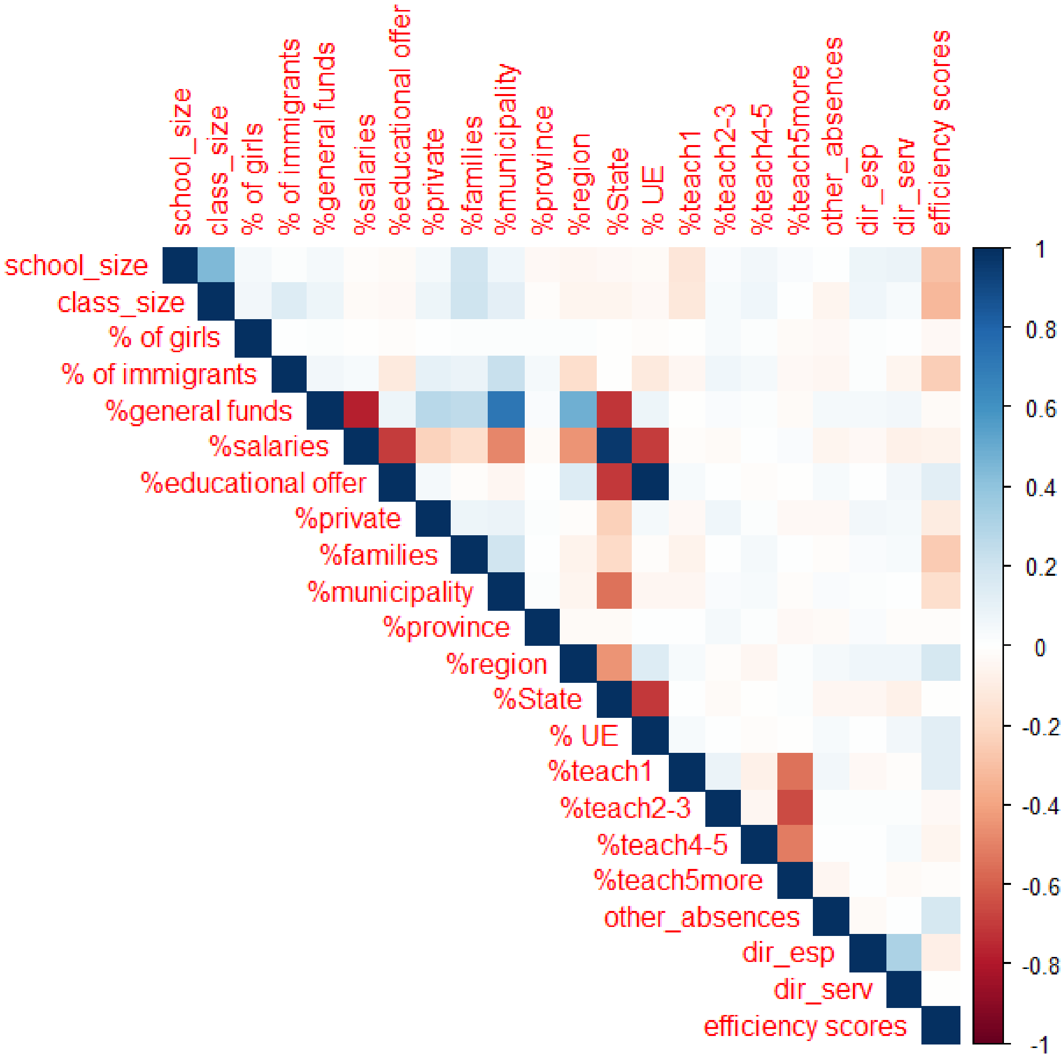

Figure 16. Correlation plot of non-discretionary variables

A careful analysis of this plot also revealed important information about the impact of some of the non-discretionary variables. It will be possible to see, in fact, in a while, that the variables with significant regression coefficients were exactly those with the highest correlation with the efficiency scores.

The table below shows the sets of non-discretionary variables adopted for each of the seven models. These models were referred to as SW Models, to distinguish them from the previous ones evaluated for the traditional DEA.

Variables	SW Model 1	SW Model 2	SW Model 3	SW Model 4	SW Model 5	SW Model 6	SW Model 7
School location							
centre	x	x	x	x	x	x	x
south	x	x	x	x	x	x	x
south_and_isles	x	x	x	x	x	x	x
Demographic variables							
school_size	x	x	x	x	x	x	x
class_size	x	x	x	x	x	x	x
% of girls	x	x	x	x	x	x	x
% of immigrants	x	x	x	x	x	x	x
Funding use							
%salaries	-	x	-	x	-	x	-
%educational offer	-	x	-	x	-	x	-
Funding sources							
%families	-	-	x	-	x	-	x
%municipality	-	-	x	-	x	-	x
%province	-	-	x	-	x	-	x
%region	-	-	x	-	x	-	x
%State	-	-	x	-	x	-	x
%UE	-	-	x	-	x	-	x
Teachers							
%teach2-3	-	-	-	x	x	x	x
%teach4-5	-	-	-	x	x	x	x
%teach5more	-	-	-	x	x	x	x
other_absences	-	-	-	x	x	x	x
School manager							
dir_esp	-	-	-	-	-	x	x
dir_serv	-	-	-	-	-	x	x

Table 17. Sets of non-discretionary variables adopted for seven alternative models

The results of the regression analysis for all the models were shown below. The tables reported, for each variable, the bias-corrected value $\hat{\hat{\beta}}$, the values $\hat{\hat{\beta}}^*$ corrected also for correlation, and the 95% confidence intervals of the regression coefficients. In addition, estimates of the standard deviations of the residuals were given. All these parameters were fully explained in the section 4.3.1.

Those variables not having the value 0 included in the interval, were found to be significant and were highlighted in grey. In this section, to allow a clear understanding of the values, the number of decimal places was increased to 6.

Table 18. Results of Simar and Wilson's second algorithm: SW Model 1

			Confidence Interval	
Variable	$\hat{\hat{\beta}}$	$\hat{\hat{\beta}}^*$	*Lower Bound*	*Upper Bound*
intercept	1.143841	1.143835	1.123600	1.163673
school_size	-0.000103	-0.000103	-0.000117	-0.000089
class_size	-0.001609	-0.001605	-0.002276	-0.000897
% of girls	-0.018218	-0.018178	-0.050136	0.013083
% of immigrants	0.104627	0.104535	0.079570	0.130301
centre	0.036361	0.036229	0.030898	0.041947
south	0.094958	0.094855	0.088847	0.101311
south_and_isles	0.120110	0.120064	0.113610	0.126524
	$\hat{\hat{\sigma}}_\varepsilon = 0.058631$	$\hat{\hat{\sigma}}_\varepsilon^* = 0.058560$		

Table 19. Results of Simar and Wilson's second algorithm: SW Model 2

			Confidence Interval	
Variable	$\hat{\hat{\beta}}$	$\hat{\hat{\beta}}^*$	*Lower Bound*	*Upper Bound*
intercept	1.153043	1.153205	1.113337	1.191238
school_size	-0.000103	-0.000103	-0.000116	-0.000088
class_size	-0.001607	-0.001601	-0.002320	-0.000912

% of girls	-0.018596	-0.018534	-0.049393	0.011086
% of immigrants	0.104372	0.104599	0.079692	0.128374
%general funds	-0.001401	-0.001605	-0.048714	0.046082
%salaries	-0.009499	-0.009809	-0.042574	0.024619
centre	0.036206	0.036058	0.030701	0.041864
south	0.094444	0.094526	0.088316	0.100643
south_and_isles	0.119636	0.119567	0.113344	0.126261
	$\hat{\hat{\sigma}}_\varepsilon = 0.058482$	$\hat{\hat{\sigma}}_\varepsilon^* = 0.058392$		

Table 20. Results of Simar and Wilson's second algorithm: SW Model 3

			Confidence Interval	
Variable	$\hat{\hat{\beta}}$	$\hat{\hat{\beta}}^*$	*Lower Bound*	*Upper Bound*
intercept	1.161504	1.160912	1.121060	1.203161
school_size	-0.000098	-0.000098	-0.000112	-0.000083
class_size	-0.001463	-0.001457	-0.002149	-0.000777
% of girls	-0.018716	-0.018962	-0.049866	0.012588
% of immigrants	0.100470	0.100137	0.076103	0.125880
%private	-0.106517	-0.106141	-0.277024	0.060405
%families	-0.639321	-0.636097	-0.876294	-0.402976
%municipality	-0.006126	-0.005924	-0.063318	0.052373
%province	0.821051	0.822626	-0.298679	1.947979
%region	-0.029218	-0.027396	-0.109202	0.040411
%State	-0.016223	-0.015500	-0.053225	0.018204
centre	0.036275	0.036149	0.030783	0.041755
south	0.092294	0.092333	0.085854	0.098750
south_and_isles	0.116912	0.116792	0.109883	0.123488
	$\hat{\hat{\sigma}}_\varepsilon = 0.058513$	$\hat{\hat{\sigma}}_\varepsilon^* = 0.058412$		

Table 21. Results of Simar and Wilson's second algorithm: SW Model 4

			Confidence Interval	
Variable	$\hat{\hat{\beta}}$	$\hat{\hat{\beta}}^{*}$	*Lower Bound*	*Upper Bound*
intercept	1.140892	1.141102	1.092472	1.187978
school_size	-0.000106	-0.000106	-0.000121	-0.000092
class_size	-0.001593	-0.001601	-0.002268	-0.00089
% of girls	-0.014698	-0.014835	-0.045270	0.016051
% of immigrants	0.100623	0.100862	0.075941	0.125971
%general funds	-0.004362	-0.003973	-0.056044	0.044521
%salaries	-0.008371	-0.008667	-0.043902	0.028944
%teach2-3	0.001025	0.001253	-0.032192	0.033919
%teach4-5	-0.000308	-0.000143	-0.029795	0.029790
%teach5more	-0.001122	-0.000835	-0.025199	0.021616
other_absences	0.002119	0.002118	0.001435	0.002794
centre	0.033741	0.033686	0.028088	0.039371
south	0.093572	0.093511	0.087332	0.099914
south_and_isles	0.116590	0.116509	0.109759	0.123394
	$\hat{\hat{\sigma}}_{\varepsilon} = 0.058895$	$\hat{\hat{\sigma}}_{\varepsilon}^{*} = 0.058787$		

Table 22. Results of Simar and Wilson's second algorithm: SW Model 5

			Confidence Interval	
Variable	$\hat{\hat{\beta}}$	$\hat{\hat{\beta}}^{*}$	*Lower Bound*	*Upper Bound*
intercept	1.147786	1.147840	1.104106	1.193397
school_size	-0.000101	-0.000101	-0.000115	-0.000087
class_size	-0.001402	-0.001390	-0.002115	-0.000705
% of girls	-0.014556	-0.014710	-0.046159	0.016156
% of immigrants	0.094843	0.094796	0.071071	0.119600
%private	-0.095957	-0.096965	-0.261768	0.072961
%families	-0.649222	-0.647366	-0.886346	-0.409918

%municipality	-0.013964	-0.014755	-0.070148	0.041982
%province	0.846115	0.856069	-0.257292	1.912739
%region	-0.026992	-0.027293	-0.100626	0.045018
%State	-0.014390	-0.014926	-0.048408	0.021323
%teach2-3	0.001052	0.001731	-0.031574	0.033373
%teach4-5	0.000611	0.001024	-0.029716	0.029431
%teach5more	-0.000621	-0.000177	-0.023634	0.021567
other_absences	0.002135	0.002129	0.001461	0.002832
centre	0.033582	0.033522	0.028129	0.039140
south	0.090301	0.090291	0.083761	0.097102
south_and_isles	0.112965	0.112967	0.106022	0.120001
	$\hat{\hat{\sigma}}_\varepsilon = 0.058548$	$\hat{\hat{\sigma}}^*_\varepsilon = 0.058388$		

Table 23. Results of Simar and Wilson's second algorithm: SW Model 6

			Confidence Interval	
Variable	$\hat{\hat{\beta}}$	$\hat{\hat{\beta}}^*$	*Lower Bound*	*Upper Bound*
intercept	1.154268	1.153413	1.107324	1.201825
school_size	-0.000105	-0.000105	-0.000118	-0.000090
class_size	-0.001571	-0.001576	-0.002276	-0.000870
% of girls	-0.014234	-0.014224	-0.043313	0.014696
% of immigrants	0.099379	0.099307	0.074838	0.123548
%general funds	-0.004690	-0.003865	-0.053720	0.042455
%salaries	-0.009719	-0.009207	-0.043283	0.025440
%teach2-3	0.002221	0.002663	-0.029705	0.032112
%teach4-5	0.000529	0.000939	-0.029705	0.029948
%teach5more	-0.000158	0.000195	-0.022549	0.021927
other_absences	0.002061	0.002067	0.001411	0.002731
dir_esp	-0.003426	-0.003407	-0.006531	-0.000289
dir_ser	-0.000287	-0.000277	-0.002129	0.001544
centre	0.033708	0.033677	0.028162	0.039284
south	0.091793	0.091812	0.085510	0.098215

south_and_isles	0.116179	0.116160	0.109515	0.12266
	$\hat{\hat{\sigma}}_\varepsilon = 0.058369$	$\hat{\hat{\sigma}}_\varepsilon^* = 0.058256$		

Table 24. Results of Simar and Wilson's second algorithm: SW Model 7

			Confidence Interval	
Variable	$\hat{\hat{\beta}}$	$\hat{\hat{\beta}}^*$	*Lower Bound*	*Upper Bound*
intercept	1.161036	1.161465	1.115107	1.207714
school_size	-0.000100	-0.000101	-0.000115	-0.000085
class_size	-0.001404	-0.001397	-0.002067	-0.00073
% of girls	-0.013861	-0.014084	-0.044266	0.017034
% of immigrants	0.094334	0.094501	0.070390	0.119455
%private	-0.091872	-0.091558	-0.253343	0.083490
%families	-0.650549	-0.649137	-0.900494	-0.396637
%municipality	-0.015656	-0.016902	-0.071413	0.044304
%province	0.869899	0.865669	-0.225067	2.001747
%region	-0.021343	-0.021273	-0.097259	0.049072
%State	-0.015259	-0.015674	-0.052457	0.019937
%teach2-3	0.001842	0.001681	-0.029287	0.033139
%teach4-5	0.001029	0.001426	-0.023894	0.030599
%teach5more	-0.000256	-0.000105	-0.023894	0.022202
other_absences	0.002125	0.002129	0.001454	0.002811
dir_esp	-0.003586	-0.003608	-0.006800	-0.000384
dir_ser	-0.000025	-0.000011	-0.001831	0.001756
centre	0.033745	0.033606	0.028264	0.039659
south	0.089277	0.089181	0.083068	0.096127
south_and_isles	0.113152	0.113064	0.106623	0.120260
	$\hat{\hat{\sigma}}_\varepsilon = 0.058596$	$\hat{\hat{\sigma}}_\varepsilon^* = 0.058455$		

As the efficiency scores were considered here ranging in [1, +inf), with schools being the more efficient the closer the values were to 1, negative regression coefficients increased the level of efficiency, while positive coefficients lowered it.

Variables found to be significant in the first models were confirmed to be significant with the addition of further factors in subsequent models. What is more, their coefficients retained the same sign. This was a key aspect in assessing the consistency of the results.

The following table shows these variables, indicating the effect they had on efficiency scores and the relative impact of each coefficient, with all other variables remaining at their values.

Significant Variables	*Effect*	*Absolute value coefficients*
school_size	Positive	0.000101
class_size	Positive	0.001397
% of immigrants	Negative	0.094501
%families	Positive	0.649137
other_absences	Negative	0.002129
dir_esp	Positive	0.003608
centre	Negative	0.033606
south	Negative	0.089181
south_and_isles	Negative	0.113064

Table 25. Significant variables according to Simar and Wilson's second algorithm

The significance of variables concerning the location of schools, allowed to start the analysis of the covariates by confirming what was said earlier about differences in efficiency between schools located in different areas in Italy. The dummies *centre*, *south*, and *south_and_isles*, all had a reducing effect on efficiency, with this effect becoming greater the further one moved towards the South and the islands, suggesting that schools located in these areas might be less efficient than those in other parts of the country.

These results were consistent with previous studies, such as Longobardi et al (2009), Agasisti (2013), Di Giacomo and Pennisi (2015). If a school was in Sicily, for instance, as much as in Calabria, efficiency scores were expected to fall by around 11%.

Moving on analysing the effect of all the other factors, variations of 1 standard deviation were taken as a reference to give an idea of corresponding variations of efficiency scores.

Differently from Agasisti (2013), for which size had no effects on performance, variables *school_size* and *class_size* affected school efficiency level. An increase in size, both in terms of the number of students in the school, and the number of students per class, had a positive effect on efficiency. Considering an increase in *school_size* by the amount of the average number of students in a class, namely 20 students, would rise by 0.2% the scores values (similar results for Alexander et al 2010), whereas with respect to the value of approximately 156 students, equivalent to 1 standard deviation, there would be a gain in efficiency of 1.6%. Same behaviour for *class_size*, for which with 3 more students per class, an increase of 0.4% would be achieved.

However, in studies by Masci et al (2018) and Kounetas et al (2023), *class_size* was found to have a negative effect, and smaller classes had a positive effect on student achievement. The question of resolving the ambiguity of the behaviour of this variable could touch several aspects. Larger classes could mean less attention from teachers towards students, on the other hand it would imply cost benefits. Anyways, the positive effect on efficiency was reinforced by the fact that *school_size* and *class_size* are necessarily positively correlated with each other, and *school_size* was considered, by most scholars, as a factor of efficiency promotion.

About the percentage of immigrants, a higher presence of foreign students negatively affected performance. A 10% increase in this parameter would lower the efficiency scores by about 0.9%. While most scholars agreed with this result, this was not the case of Bradley et al (2010), who found the opposite when considering the percentage of students born outside the UK.

This discrepancy could be explained by the fact that the contribution of foreign students could vary from country to country, depending on the context and circumstances of the immigration situation. For example, in some countries, foreign students may bring diversity and fresh perspectives, which may positively contribute to the learning environment and enhance academic outcomes. On the other hand, in other countries, an influx of foreign students may strain resources and create cultural and linguistic barriers, leading to negative impacts on academic performance.

Less common in the literature, but still significant for this analysis, were the variables regarding funding, teacher absence days and school head characteristics. *%families* was the only variable among the several regarding financing to have a significant coefficient. Specifically, being able to rely on 1% higher funds from families, schools would raise their efficiency level by 0.6%.

Based on this, variables related to sources of funds, even if only due to this single variable, prevailed over the ones concerning the use of this source. Nevertheless, having realised that schools could benefit from higher amount of money from families, further investigation should be done to understand how best to use these funds to increase performance.

The presence of teachers with more years of experience, contrary to what might have been expected, was not significant in predicting the values of the efficiency scores, while the number of days of absence of teachers was. 6 additional average per capita days, per year, of teachers' absences, would have a negative impact on efficiency slightly more than 1%.

Initially, the analysis was conducted by summing up all the days of absences due to different causes, including maternity and illness. Then, the study found that a more specific analysis was possible by considering the variable *other_absences*, which excluded absences due to maternity and illness.

Finally, the analysis suggested that the experience of the school head could have a positive impact on the school performance. For every half year of additional experience, there would be an increase in the value of scores close to 0.2%. This means that the longer a school head has been in her position, the more likely she was to have a positive impact on the efficiency level of her school.

This last result could be due to several factors, such as greater familiarity with teaching methods and curricula, more effective management of staff and resources, and a better understanding of how to motivate and engage students. Additionally, experienced school heads may be more skilled at identifying and addressing problems that arise in the school.

6.3 Nonlinear relationships and interaction effects through Random Forest

When evaluating the effects of environmental factors on school efficiency with Simar and Wilson's second algorithm, this did not consider that the relationship between those factors and the efficiency scores may also be complex and nonlinear.

The truncated regression on which the algorithm is based, determined in the first instance which variables had a significant effect on efficiency. However, with regression coefficients assumed constant across all the values of the covariates, its ability to assess any nonlinear relationships was limited.

Moreover, interaction effects due to the simultaneous changes in more covariates, was not included a priori in the model, since any assumption about interaction terms was made, and standard regression approaches, like the truncated regression here, do not consider these terms if not specified.

On this basis, Random Forest method was adopted to improve the accuracy of the results illustrated in the previous section and provide a more comprehensive understanding of the factors influencing school performance. In this section, no regression coefficients associated with any predictor were reported, but rather a purely graphical explanation was provided.

The focus was first on assessing the importance of the variables, checking whether the significant ones were the same as before, and then, on looking for possible nonlinear relationships as well as interaction effects between them.

Since regression trees, and thus also RF, work with categorical independent predictors, it was possible to include the variable *area_geo* in the analysis, thus being able to disregard dummy variables. Furthermore, making RF the several trees less correlated with each other, sampling not only over the observations but also over the features, all variables concerning funds and teachers were kept in the model.

The model, despite adopting the same parameter values as Schiltz et al. (2017), such as 2000 regression trees and a number of predictors at each split equal to the square root of the total, reported much better results, given the higher percentage of explained variance, i.e., 44.38% compared to 24.91%, and the lower MSE, i.e., 0.00284 compared to 0.01007.

As said, importance of predictors was first evaluated, and the corresponding plot was shown below.

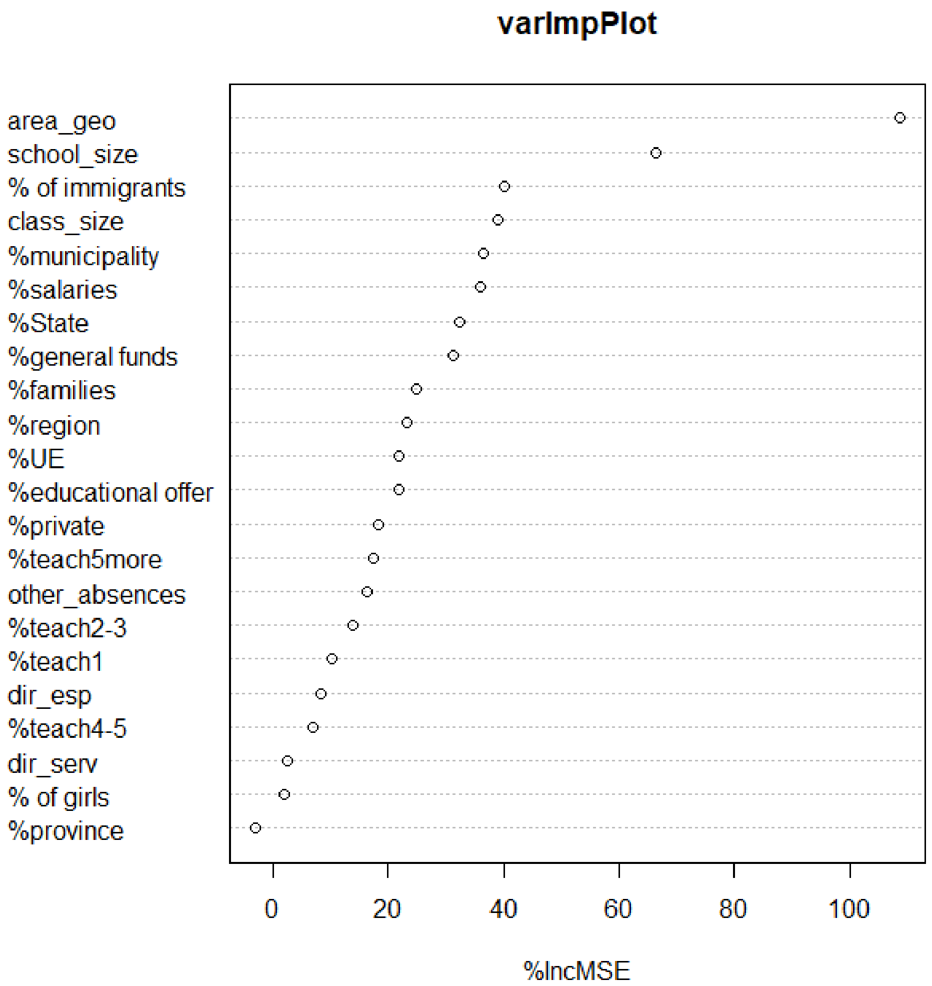

Figure 17. Plot of variable importance as measured by Random Forest

The outcome referred to variable importance measured as the percentage increase in MSE, when each variable was excluded from the model. The higher these values, the more important the variables, as excluding them increased error.

What emerged was that the geographical area where schools were located, still represented the main factor in the estimation of efficiency scores, further justifying the interest of this thesis in assessing differences in school efficiency at territorial level. Removing this variable, in fact, would lead to an increase in error of more than 100%.

The variables *school_size* and *class_size*, (Schiltz et al, 2017 and Rebai et al, 2020), in addition to *% of immigrants*, all related to school structure, also confirmed to be relevant in the analysis.

The following table reported the relative importance of each covariate, expressed with respect to the most important one. Significant variables for SW Models were highlighted in grey.

Variables	*Relative importance*
area_geo	100.00%
school_size	60.92%
% of immigrants	36.94%
class_size	35.74%
%municipality	33.49%
%salaries	32.96%
%State	29.75%
%general funds	28.70%
%families	22.77%
%region	21.46%
%UE	20.17%
%educational offer	20.15%
%private	16.87%
%teach5more	15.94%
other_absences	15.07%
%teach2-3	12.61%
%teach1	9.40%
dir_esp	7.62%
%teach4-5	6.36%
dir_ser	2.14%
% of girls	1.77%
%province	-2.08%

Table 25. Relative importance of variables

Looking at these values, some considerations could be made. To begin with, the variables *other_absences* and *dir_esp*, which were considered significant by the double bootstrap procedure, had lower importance values than the others.

Then, almost all the predictors concerning funding, with the exception of *%provinces*, were grouped in levels close in importance and greater than the ones of variables referring to both teachers and school head. This deviated from what was stated by Simar and Wilson's algorithm, which identified the only covariate *%families* to be significant, suggesting further investigation in the effects due to these variables.

The fact that *%provinces* had a negative value of importance, was also reflected in the sign of the respective regression coefficient, being the only variable regarding funds with a negative effect on efficiency. Not considering this variable in the model even improved its performance.

Furthermore, proving that the random forest could limit multicollinearity problems due to the correlation between regressors, the pairs of variables *%salaries* and *%State*, as much as *%UE* and *%educational offer*, which were seen to be correlated to each other, were ranked close in the *varImpPlot*.

Finally, it should be noted that *% of girls*, not significant even before, represented instead the variable with the higher predictive power in the analysis conducted by Rebai et al (2020) on the Tunisian secondary schools.

Once the relative importance of variables was established, it was necessary to investigate how these variables impacted the efficiency scores. To achieve this, it was useful to create partial plots that could reveal the direction and shape of such effects, which could then be analysed and interpreted accordingly. As pointed by Schiltz et al. (2017), this graphical approach can be compared to plotting the coefficient of a linear regression without assuming that it is constant across values of a specified variable.

The partial plots below confirmed the correlation between *school_size* and *class_size*, as the two variables showed very similar patterns, both having the effect of increasing efficiency for higher values.

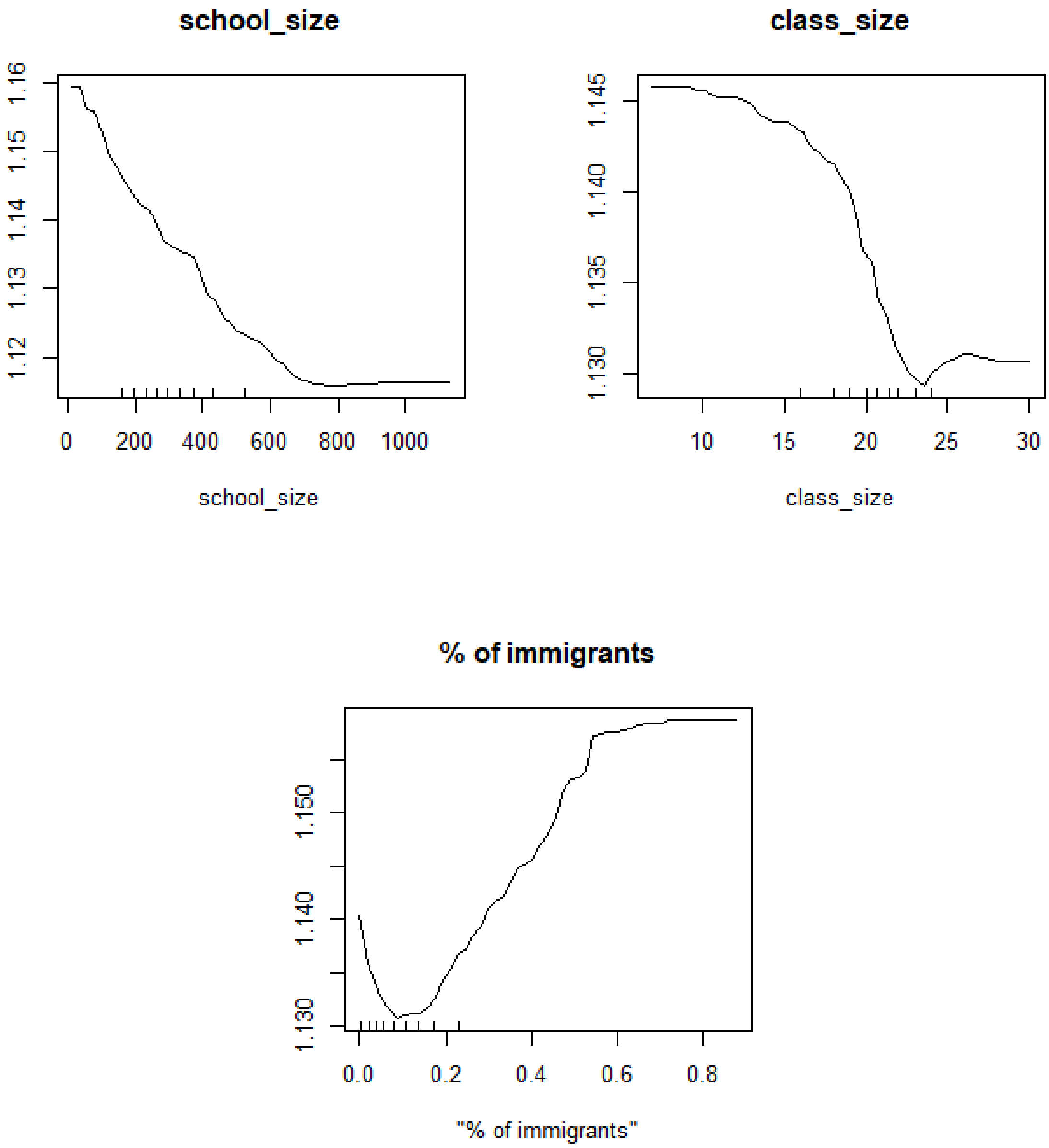

Figure 18. Partial plots of the variables *school_size*, *class_size* and *% of immigrants*

However, what could not be noticed before, was that a saturation point in efficiency was reached. School efficiency settled for schools receiving 700 and more students or having classes with at least around 24 students. Receiving many students, at a certain point, could stop creating positive externalities in the school environment, besides excessively intensify organisational aspects for the school head. Therefore, even if the variable *dir_esp* was no longer among the most important covariates, its possible interaction with the *school_size* should be evaluated.

The efficiency level due to *class_size* even slightly decreased for some values. Classes with too many students means that teachers could found it harder to dedicate the right level of attention to each of them.

For what concerns the *% of immigrants* instead, it could be seen that very low percentages might have a positive effect on efficiency, while starting from values close to 20%, a negative impact on school efficiency was confirmed, which lowered it almost linearly, till to become constant for percentages of immigrants close to 60%.

Moving to the variables about funding, it should be considered that all their distributions were asymmetrical, with a large proportion of the values going towards the one or the other tail. For this reason, to prevent a misleading understanding about the effect they had on school efficiency, their partial plots were showed disregarding those values.

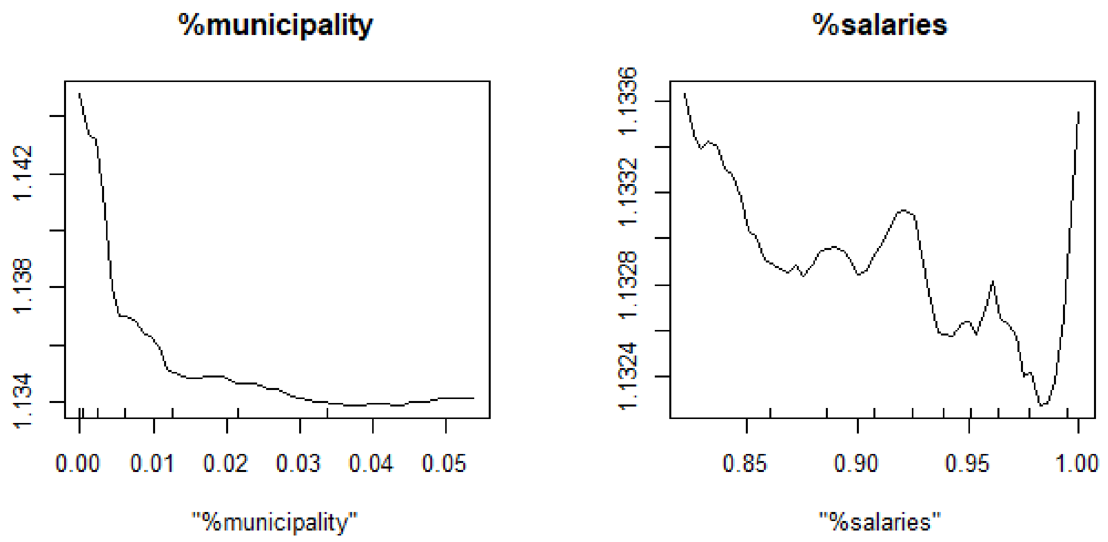

Figure 19. Partial plots of the variables *%municipality* and *%salaries*

The variables *%municipality* and *%salaries* both had positive effects on efficiency, according to the signs of the respective regression coefficients, despite having nonlinear patterns, reason why truncated regression failed to identify their importance. Efficiency reached a saturation level for percentages of funds received by municipalities slightly above 3%, whereas increasing the percentage of funds allocated to the payment of salaries, maintained, even though in an alternating way, the positive effect up to cover almost all the funds.

%families, which was instead significant before, also confirmed to increase efficiency for higher percentages, while different patterns had *%general funds*, *%region* and *%UE*, for which, as opposed to the sign of their regression coefficients, an increase in percentages had a negative impact on efficiency. However, the analysis of these variables could be affected by the very low values they assume.

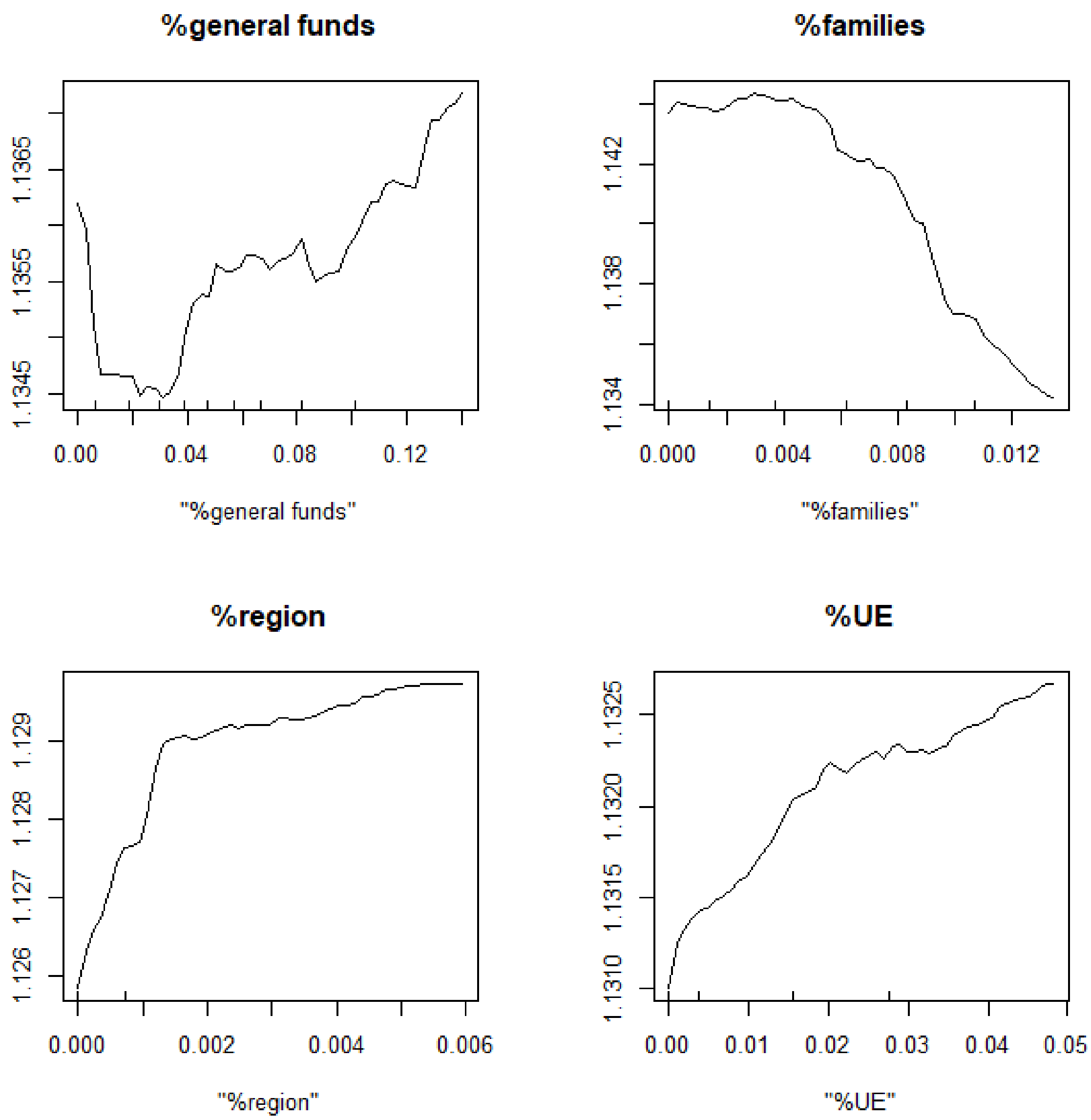

Figure 20. Partial plots of the variables *%general funds*, *%families*, *%region* and *%UE*

Finally, the two partial plots below referred to variables *other_absences* and *dir_esp* that were significant in the double-bootstrap procedure, but were at the bottom of the ranking of important variables provided by RF.

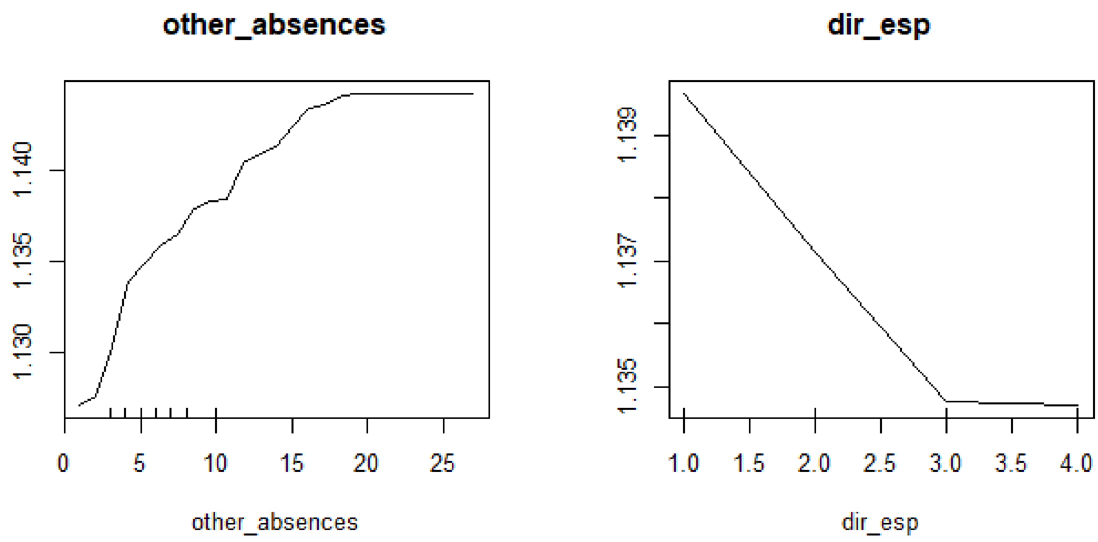

Figure 21. Partial plots of the variables *other_absences and dir_esp*

The two curves, especially the one of *dir_esp*, approximated linear trends, before settling to almost constant efficiency values, which explained why classical regression recognized some useful information in the two variables.

Later in the analysis, it was possible to also evaluate the effects on efficiency due to possible interactions between predictors, by using heatmaps. In each heatmap, the values of two covariates are plotted one against the others and the simultaneous effect on the efficiency scores is showed for each pair of values. It should be noted that it was not necessary to specify ex ante the effects of these interactions, since the structure of regression trees allowed to include them in the model anyway, and their estimation was possible relying on data.

The first of the interactions considered was the one between *school_size* and *dir_esp*, as previously mentioned. The heatmap below showed that although *school_size* was, after *area_geo*, the variable with the highest percentage of explained variability, its effect on school efficiency was mitigated to some extent by the number of years of experience of the school head.

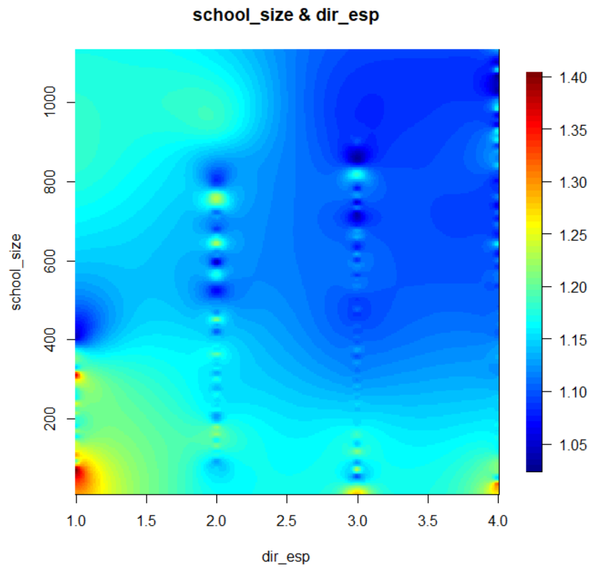

Figure 22. Heatmap of interaction effect of the variables *school_size and dir_esp*

School heads with up to 2 years of experience, only managed quite effectively schools with 400 to 600-800 students. For values below this range, the school context was probably too small, with inability of the school head to exploit economies of scale on available resources. On the other hand, for values above the range, the higher responsibilities were hardly managed by school heads at the beginning of their career.

The interaction of the two variables, led to the highest levels of efficiency when school heads with at least 3 years of experience were in charge of schools with 600 to 800 and more students. Note how, being *dir_esp* a discrete variable assuming only the values {1,2,3,4}, vertical patterns in correspondence of these values could be identified, highlighting the differences between various areas of efficiency.

Subsequently, the interaction of *class_size* with *other_absences* first, and *%salaries* later, was considered. The heatmap on the left panel showed that for classes with up to around 23 students, a higher number of absences of teachers, not due to illness or maternity, lowered the level of school efficiency. This negative effect seemed to describe an almost linear behaviour up to 20 days of absence, where this trend was interrupted due to the small number of schools assuming higher values.

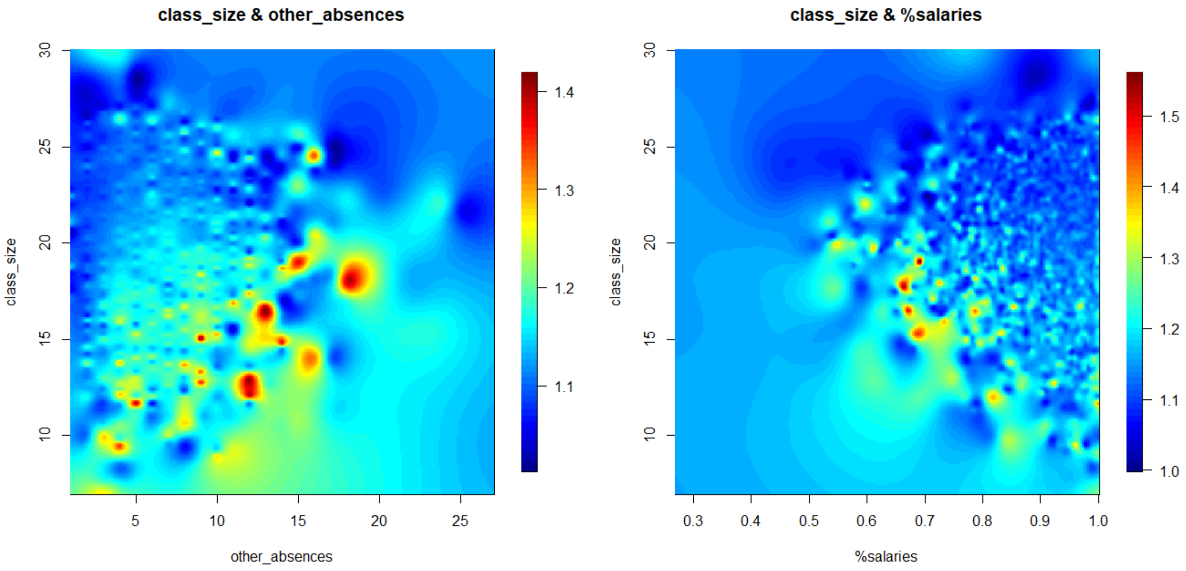

Figure 23. Heatmaps of interaction effects of the variables *class_size* and *other_absences* (on the left), and *class_size* and *%salaries* (on the right)

The heatmap on the right panel instead, referred to interaction effect of the number of students per class and the percentage of funds allocated to pay salaries. It could be seen that, for percentages greater than 50%, the allocation of more funds to salaries must be matched by more students per class for schools to be more efficient.

Increasing the percentages on the abscissa could in fact mean the hiring of teachers with more years of experience, and therefore capable of handling a higher number of students. Assigning such teachers to smaller classes would thus indicate an efficiency in the school's allocation of this resource.

The last interactions to be analysed, were those between *class_size* and *% of immigrants*, and *school_size* and *% of immigrants*. It was showed the negative effect on efficiency of immigrant percentages per school above 40-50%, whatever the number of students per class. This effect was mitigated at school level.

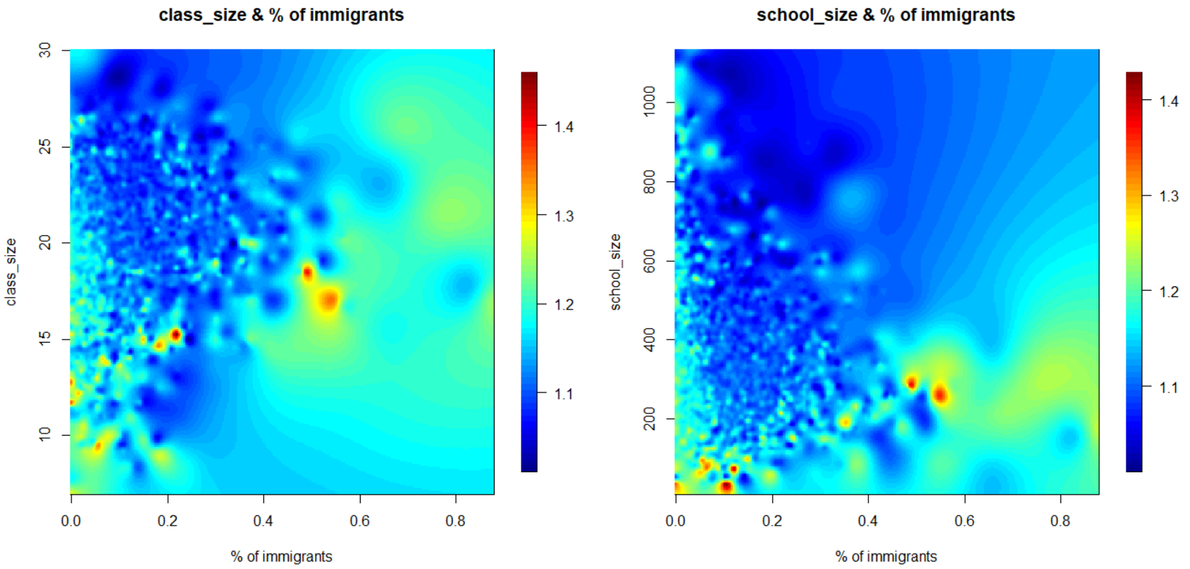

Figure 24. Heatmaps of interaction effects of the variables *class_size* and *% of immigrants* (on the left), and *school_size* and *% of immigrants* (on the right)

7. Conclusions and implications

The results illustrated above are here discussed with the intention of providing valuable insights for policymakers as well as outlining some guidelines that may lead to improvements in subsequent studies.

The first research question underlying the analysis, which asked to define the level of efficiency of Italian lower secondary schools, refers to a more purely descriptive part. The international panorama presents many studies referring to the subject of school efficiency, while there are only fewer studies in Italy. Of these, the present certainly stands out as the one relying on the highest availability of data, with a number of schools analysed equal to 4264. Before, Masci et al. (2018) conducted the analysis on 462 lower secondary schools, while Di Giacomo and Pennisi (2015) reached 1000 schools, including also primary ones. Over the years, the ever-increasing importance reserved for data and information systems has allowed for a higher availability of sources in the field of education, about both students' performance and schools' characteristics. Furthermore, Masci et al. (2018) adopted a granularity at the student level, whereas, in agreement with Agasisti (2013), but also with studies outside the Italian context such as Bradley et al. (2001), Mizala et al. (2002), and many others, it was chosen here to focus on the school level, with the idea that it was of greater relevance for the school management.

Moving to the application of DEA model, despite the careful analysis of the literature about the choice of input and output variables, it was seen that by changing the model specifications, the efficiency scores for different models showed high correlations with each other. Based on this, given the presence of two different outputs, although from a theoretical point of view, each school might tend towards the production efficiency of one output rather than another, the analysis was not carried out separately for each individual output. It was then possible to see how considering ESCS among the inputs, allows for a better explanation of the variability of the scores, leading to a better differentiation of schools in terms of efficiency.

In general, the level of relative efficiency of Italian lower secondary schools was quite high, with median values of the efficiency scores exceeding 89%, and it is interesting how there are different ways in which a school can be efficient, not only by presenting high values of outputs, but also by using less input resources. However, the efficiency scores showed a strong dispersion of data, with schools reaching minimum efficiency levels just above 60%. What is more, there were differences in efficiency when looking at schools located in different geographical areas of the country, namely North, Centre, South and South islands.

The most efficient schools were those in the North, with Friuli Venezia Giulia, Veneto and Lombardia having 75 percent of their schools with efficiency values above 90%. The central regions, apart from Lazio, still had quite high efficiency values, while lower values occurred in the regions of the South and South and islands. There were differences of respectively around 6% and 8% when comparing the median scores of the northern schools and those of the regions in the South, and in the South and islands, with the least efficient schools being in Calabria. These territorial differences are reflected in most previous studies, such as Longobardi (2009), Agasisti (2013), Di Giacomo and Pennisi (2015). The Italian system has always appeared as a highly centralised system unable to guarantee uniformity in student performance. At the same time, it is important to remark that the assessment of such differences see the schools with higher student performance being also those most capable to manage the available resources, and therefore the most efficient, not reducing the analysis to the usual North-South divide.

Trying to identify any environmental factors not under the school's control, that may affect those differences in efficiency, was the objective of the second research question. In the first instance, Simar and Wilson’s (2007) second algorithm allowed to evaluate the significance at 95% confidence level of these factors and their effect. It must be emphasised that, even though in other works such as Alexander et al. (2010) or Kounetas et al. (2023), were respectively considered 256 and 643 schools, this procedure can be applied to a sample of 4000 and more schools. The main computational effort, as pointed out by Simar and Wilson themselves, lies in generating the estimates of the confidence intervals, for which more information is needed. Anyways, with the number of replications adopted, it was possible to implement seven models for different set of variables, within a couple of working sessions, guaranteeing thus its replicability. Moreover, since the algorithm includes classical truncated regression techniques, all the usual assumptions made for such models should be evaluated. However, as stated by the same authors in a later work, Simar and Wilson (2011), their algorithm does not take into account the heteroscedasticity of the residuals, and, in this work, in no way their normality is verified. Problems of multicollinearity can be instead excluded, as the high correlation between some variables was considered, not including them together in the models.

The inclusion of dummy variables for the location of schools in these models confirmed the differences in efficiency showed in the first part of the analysis, with the negative effects in predicting efficiency scores especially for the schools located in the South and the islands. Beyond that, other variables were found to be significant, referring to various aspects such as the size and structure of the school, the teaching staff, the school head and the sources of funding received. Each of them represents a lever that can be moved by the policymakers to improve the efficiency of the school, and they are analysed one by one. The focus is on the sign rather than the magnitude of the coefficients, so to make comparison with other studies more straightforward. Furthermore, the use of Machine Learning, and specifically Random Forest, made it possible to evaluate nonlinear trends of these factors with respect to efficiency scores using a graphical approach. This makes it possible not to assume the effect of each of them to be constant, but rather to assess any changes for certain values, even leading to situations of stagnating efficiency. Also, the graphical approach accounts for the possibility to evaluate potential interaction effects due to simultaneous changes in two of the factors, without the need of specifying such interactions a priori.

With these in mind, the first two relevant factors refer to the number of students both in the school and in a class. As per Schiltz et al. (2017), increasing both these values leads to an increase in the level of efficiency. For what concern the school size, receiving many students gives the possibility of creating economies of scale and dividing fixed costs over a higher number of students, besides the fact that positive externalities are created in the school environment. It is possible improve efficiency of Italian secondary schools up to 700-800 students per school, after which the positive effects stop. Likewise, it is possible to assess how efficiency benefits from larger class sizes up to 23-24 students. For higher values, it may be more complicated for teachers to manage the class, with greater difficulty in devoting the right level of attention to each student. This last finding contrasts, for example, with the one of Masci et al. (2018), although it based on a much smaller sample of Italian lower secondary schools, which concluded that smaller classes promote better student learning.

Another lever that can be used to adjust the level of school efficiency is the percentage of immigrant students. With respect only to whether they have foreign parents, and without distinguishing between those born in Italy and those of recent immigration, increasing the percentage of immigrants in the school has a negative effect on efficiency, in agreement with Di Giacomo and Pennisi (2015). Bringing such percentages to around values of less than 20%, allows not to worsen the school efficiency, but rather to benefit from diversity and fresh perspectives brought by foreign students. This increases the focus on the importance of carefully assessing the percentage of non-nationals in the student body. However, it should be noted that, with respect to studies conducted in different countries, as for Bradley et al. (2010) in the UK, the effect may be different and strongly dependent on cultural aspects,

The teaching staff can be included among the factors impacting efficiency considering the number of days of absence. Reducing the average number of days of absence per year of teachers, not due to causes such as illness or maternity, shows an improvement in school efficiency. Students can develop healthy attendance habits and become more academically proficient, while the intensity of the educational process is drastically reduced when a teacher is absent. It should be noted that the total number of absences was also informative in predicting efficiency scores, however, excluding absences due to illness and maternity, which may be unavoidable, facilitates the work of policymakers in figuring out how to reduce these values. While one could immediately think of monetary compensation, such as salary increases or bonuses for high attendance, or even a sick leave buy-back programme, non-monetary measures could also be taken, such as implementing a code of conduct or limiting leave on specific dates such as close to holidays. However, with the information at hand it is difficult to determine which policies are best to implement, and a separate analysis should be made of the factors affecting teacher absenteeism.

The efficiency level can be also improved looking at the years of experience of the school head. Considered by Masci et al. (2018) in an extensive analysis involving different characteristics of the school head, the years of experience were not found to be significant. In this work, on the contrary, having a more experienced school head leads to an increase in the level of efficiency of the school. There could be several causes behind this, including a more effective management of staff and resources, and the capacity to analyse and plan for future need, or to improve the relationship between the school and the community. The increase of the efficiency is clear having more experienced school heads up to 3 years of experience, while there is almost no difference passing from 3 to 4 years of experience.

What has been said about all the above factors can be expanded reporting the relevant effects given by the interactions between some of them. For example, the limit of 700-800 students beyond which it has been seen that increasing the school size no longer affects efficiency, can be overcome in case of a school head with at least 3 years of experience. On the other side, to achieve high levels of efficiency, school heads at the beginning of their careers need contexts with at least around 400 students, and in any case no more than 700. Subsequently, it was seen how the choice of class size may depend on two factors such as teacher absenteeism and the percentage of funds allocated to salaries. In particular, the number of students per class is not sufficient to balance the negative effect that absences have on efficiency, and the higher it is, the greater the negative effect, considering that more students are affected. The class size also depends on how much money is allocated to salaries. In fact, it has been seen that for increasing percentages of such funds, a higher number of students per class is required for the school to be efficient. This can be explained by considering that more money spent on salaries may mean teachers with more years of experience, and this is reflected if one looks at the correlations of this variable with those related to the years of experience of teachers. More experienced teachers are thus able to manage larger classes, having this resource allocated in the best possible way. The last two interactions considered see the percentage of immigrants compared with both class and school size. If the percentage of immigrants is found to be around 50% for the whole school, its negative effect is not reduced by increasing the number of students per class, while it is less pronounced when assessed on the number of students in the whole school.

Looking at the results of the Random Forest it is possible to add further considerations to the analysis, namely that some of the funding variables, including the percentage of salaries mentioned before, explain a large part of the variability in the efficiency scores. However, the effects of these variables only refer to very low positive changes, as in the case of increased funds from the municipalities and the families, or even negative changes, in the case of increased funds from the regions and the EU. Then, the analysis could be affected by the highly skewed distributions that most of these variables present, with very low percentage values and many null values. It is therefore preferred to suggest a more detailed investigation of these items, considering that, to design adequate policies, it would also be necessary to better understand how funds from the different sources are used.

A possible improvement of this work could concern the collection of data covering a period of more years, since it only relies on data referring to the 2018-2019 school year. Conducting the analysis over several years, also given the increasingly easy accessibility of data, could support, or adjust, what has been seen for only one year. This is particularly important if one considers that in a highly centralised system such as the Italian one, where there are high levels of bureaucracy, the effect of measures taken by policymakers may require a 'time-lag' before manifesting itself (Kounetas et al., 2023). Furthermore, it would be significant to evaluate the effects of the pandemic, which has in any case had an impact on the school system and teaching.

Finally, this study confirms that the relationships between environmental factors and school efficiency are most often nonlinear, and while classical regression approaches fail to identify these relationships, the graphical Machine Learning approach adopted clearly display them. Despite the complexity of the model, through the partial plots and the heatmaps it is possible to have a clear understanding of the results. Subsequent studies could try not to limit the ML to this graphical aspect but push it to also assess the magnitude of the main factors influencing school efficiency, so to provide even more valid implications.

References

Afonso, A., & Aubyn, M. S. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic modelling*, *23*(3), 476-491.

Agasisti, T., (2009), The effect of competition on schools' performance. Preliminary evidence from Italy through OECD-PISA data, mimeo, Politecnico di Milano.

Agasisti, T. (2011a). How competition affects schools' performances: Does specification matter?. *Economics Letters*, *110*(3), 259-261.

Agasisti, T. (2011b). Performances and spending efficiency in higher education: a European comparison through non-parametric approaches. *Education Economics*, *19*(2), 199-224.

Agasisti, T. (2013). The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA2006 data. *Education Economics*, *21*(5), 520-544.

Agasisti, T. (2014). The efficiency of public spending on education: An empirical comparison of EU countries. *European Journal of Education*, *49*(4), 543-557.

Agasisti, T., Bonomi, F., & Sibiano, P. (2014). Measuring the "managerial" efficiency of public schools: a case study in Italy. *International Journal of Educational Management*.

Agasisti, T., Catalano, G., Landoni, P., & Verganti, R. (2012). Evaluating the performance of academic departments: An analysis of research-related output efficiency. *Research Evaluation*, *21*(1), 2-14.

Agasisti, T., & Dal Bianco, A. (2006). Data envelopment analysis to the Italian university system: theoretical issues and policy implications. *International Journal of Business Performance Management*, *8*(4), 344-367.

Agasisti, T., & Dal Bianco, A. (2009). Reforming the university sector: effects on teaching efficiency—evidence from Italy. *Higher education*, *57*, 477-498.

Agasisti, T., Dal Bianco, A., Landoni, P., Sala, A., & Salerno, M. (2011). Evaluating the efficiency of research in academic departments: An empirical analysis in an Italian region. *Higher Education Quarterly*, *65*(3), 267-289.

Agasisti, T., & Johnes, G. (2009). Beyond frontiers: comparing the efficiency of higher education decision-making units across more than one country. *Education economics*, *17*(1), 59-79.

Agasisti, T., & Johnes, G. (2010). Heterogeneity and the evaluation of efficiency: the case of Italian universities. *Applied economics*, *42*(11), 1365-1375.

Agasisti, T., & Johnes, G. (2015). Efficiency, costs, rankings and heterogeneity: the case of US higher education. *Studies in Higher Education*, *40*(1), 60-82.

Agasisti, T., Munda, G., & Hippe, R. (2019). Measuring the efficiency of European education systems by combining Data Envelopment Analysis and Multiple-Criteria Evaluation. *Journal of Productivity Analysis*, *51*, 105-124.

Ahn, T., Charnes, A., & Cooper, W. W. (1988). Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. *Socio-economic Planning sciences*, *22*(6), 259-269.

Alexander, W. R. J., Haug, A. A., & Jaforullah, M. (2010). A two-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools. *Journal of productivity analysis*, *34*, 99-110.

Arnold, V. L., BARDHAN, R., Cooper, W. W., & Kumbhakar, S. C. (1996). New Uses of DEA and Statistical Regressions for Efficiency Evaluation and Estimation–with an Application to Public Secondary Schools in Texas. *Annals of Operations Research, 66*.

Athanassopoulos, A. D., & Shale, E. (1997). Assessing the comparative efficiency of higher education institutions in the UK by the means of data envelopment analysis. *Education economics*, *5*(2), 117-134.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, *30*(9), 1078-1092.

Barbetta, G. P., & Turati, G. (2003). Efficiency of junior high schools and the role of proprietary structure. *Annals of Public and Cooperative Economics*, *74*(4), 529-552.

Bates, J. M. (1997). Measuring predetermined socioeconomic'inputs' when assessing the efficiency of educational outputs. *Applied Economics*, *29*(1), 85-93.

Bertola, G., & Checchi, D., (2002). Sorting and private education in Italy. CEPR Discussion Papers 3198, C.E.P.R.

Bessent, A., Bessent, W., Kennington, J., & Reagan, B. (1982). An application of mathematical programming to assess productivity in the Houston independent school district. Management science, 28(12), 1355-1367.

Bonesrønning, H., & Rattsø, J. (1994). Efficiency variation among the Norwegian high schools: consequences of equalization policy. *Economics of education review*, *13*(4), 289-304.

Bradley, S., Johnes, G., & Millington, J. (2001). The effect of competition on the efficiency of secondary schools in England. *European Journal of Operational Research*, *135*(3), 545-568.

Bratti, M., Checchi, D., & Filippin, A. (2007). *Da dove vengono le competenze degli studenti? I divari territoriali nell'indagine OCSE PISA 2003*. Il mulino.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Burney, N. A., Johnes, J., Al-Enezi, M., & Al-Musallam, M. (2013). The efficiency of public schools: the case of Kuwait. *Education Economics*, *21*(4), 360-379.

Chalos, P. (1997). An examination of budgetary inefficiency in education using data envelopment analysis. *Financial Accountability & Management*, *13*(1), 55-69.

Chalos, P., & Cherian, J. (1995). An application of data envelopment analysis to public sector performance measurement and accountability. *Journal of accounting and public policy*, *2*(14), 143-160.

Charnes, A., Cooper, W. W., Golany, B., Seiford, L., & Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of econometrics*, *30*(1-2), 91-107.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429-444.

Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. *Management science*, *27*(6), 668-697.

Charnes, A., Cooper, W. W., Seiford, L., & Stutz, J. (1983). Invariant multiplicative efficiency and piecewise Cobb-Douglas envelopments. *Operations Research Letters*, *2*(3), 101-103.

Cherchye, L., De Witte, K., Ooghe, E., & Nicaise, I. (2010). Efficiency and equity in private and public education: A nonparametric comparison. *European Journal of Operational Research*, *202*(2), 563-573.

Cooper, W. W., Seiford, L. M., & Tone, K. (2006). *Introduction to data envelopment analysis and its uses: with DEA-solver software and references*. Springer Science & Business Media.

Cordero-Ferrera, J. M., Pedraja-Chaparro, F., & Santín-González, D. (2010). Enhancing the inclusion of non-discretionary inputs in DEA. *Journal of the Operational Research Society, 61*(4), 574-584.

Cordero-Ferrera, J. M., Santín, D., & Sicilia, G. (2015). Testing the accuracy of DEA estimates under endogeneity through a Monte Carlo simulation. *European Journal of Operational Research*, *244*(2), 511-518.

De Witte, K., & López-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, *68*(4), 339-363.

De Witte, K., & Rogge, N. (2010). To publish or not to publish? On the aggregation and drivers of research performance. *Scientometrics, 85*(3), 657-680.

De Witte, K., Thanassoulis, E., Simpson, G., Battisti, G., & Charlesworth-May, A. (2010). Assessing pupil and school performance by non-parametric and parametric techniques. *Journal of the Operational Research Society*, *61*(8), 1224-1237.

Di Giacomo, G., & Pennisi, A. (2015). Assessing primary and lower secondary school efficiency within Northern, Central and Southern Italy. *Italian Economic Journal*, *1*(2), 287-311.

Duncombe, W., Miner, J., & Ruggiero, J. (1997). Empirical evaluation of bureaucratic models of inefficiency. *Public Choice*, *93*(1-2), 1-18.

Efron, B. (1979). Bootstrap methods: another look at the jackknife annals of statistics 7: 1–26.

Engert, F. (1996). The reporting of school district efficiency: The adequacy of ratio measures. *Journal of Public Budgeting, Accounting & Financial Management*, 8(2), 247-271.

EOSLHE. Albanesi, C. (2020). Higher Education Framework in Italy. *Learning with commitment.* Newsletter #3.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the royal statistical society: series A (General)*, *120*(3), 253-281.

Førsund, F. R. (1992). A comparison of parametric and non-parametric efficiency measures: The case of Norwegian ferries. *Journal of Productivity Analysis*, *3*(1-2), 25-43.

Fried, H. O., & Lovell, C. A. K. (1996). Searching for the Zeds. In *II Georgia productivity workshop*.

Ganley, J. A., & Cubbin, J. S. (1992). *Public sector efficiency measurement: Applications of data envelopment analysis*. Elsevier Science Inc..

Giménez, V., Prior, D., & Thieme, C. (2007). Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. *Journal of the Operational Research Society*, *58*(8), 996-1007.

Grosskopf, S., Hayes, K. J., & Taylor, L. L. (2014). Efficiency in education: Research and implications. *Applied Economic Perspectives and Policy*, *36*(2), 175-210.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics*, 1453-1462.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. Journal of economic literature, 24(3), 1141-1177.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The economic journal*, *113*(485), F64-F98.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American economic review*, *90*(5), 1184-1208.

Hirschberg, J., & Lloyd, P. (2001). Does the technology of foreign-invested enterprises spill over to other enterprises in China? An application of post-DEA bootstrap regression analysis. In *Models of the Chinese Economy* (pp. 88-110). Edward Elgar Publishing.

Ho, T. K. (1995, August). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.

Hoxby, C. M. (2003). *The Economics of School Choice. A National Bureau of Economic Research Conference Report*. Marketing Department, University of Chicago Press, 1427 East 60th Street, Chicago, IL 60637-2954.

INVALSI. (2019). Sintesi dei risultati Italiani di OCSE PISA 2018. *Area indagini internazionali INVALSI*.

Jill, J., Steve, B., & Allan, L. (2012). Efficiency in the Further Education Sector in England. *Open Journal of Statistics, 2012*.

Kirjavainen, T., & Loikkanent, H. A. (1998). Efficiency differences of Finnish senior secondary schools: an application of DEA and Tobit analysis. *Economics of Education Review*, *17*(4), 377-394.

Kounetas, K., Androulakis, G., Kaisari, M., & Manousakis, G. (2023). Educational reforms and secondary school's efficiency performance in Greece: a bootstrap DEA and multilevel approach. *Operational Research*, *23*(1), 1-29.

Lee, B. L. (2011). Efficiency of research performance of Australian universities: A reappraisal using a bootstrap truncated regression approach. *Economic Analysis & Policy, 41*(3).

Longobardi, S., Quintano, C., Castellano, R., 2009, The influence of social and economic factors on the competencies of Italian students. A multilevel analysis of data PISA2006, Rivista di Economia e Statistica del Territorio, 2, …

Mancebón, M. J., Calero, J., Choi, Á., & Ximénez-de-Embún, D. P. (2012). The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006. *Journal of the operational research Society*, *63*(11), 1516-1533.

Mancebón, M. J., & Muñiz, M. A. (2008). Private versus public high schools in Spain: disentangling managerial and programme efficiencies. *Journal of the operational Research Society*, *59*(7), 892-901.

Masci, C., De Witte, K., & Agasisti, T. (2018). The influence of school size, principal characteristics and school management practices on educational performance: An efficiency analysis of Italian students attending middle schools. *Socio-Economic Planning Sciences*, *61*, 52-69.

Mayston, D. J. (1996). Educational attainment and resource use: Mystery or econometric misspecification?. *Education Economics*, *4*(2), 127-142.

McCarty, T. A., & Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. *The measurement of productive efficiency: Techniques and applications*, 271-287.

Mizala, A., Romaguera, P., & Farren, D. (2002). The technical efficiency of schools in Chile. *Applied Economics*, *34*(12), 1533-1552.

MIUR. Ufficio di Statistica. Palmini, F & Di Ascenzo, D. (2022). Focus "Principali dati della scuola – Avvio Anno Scolastico 2022/2023".

Muñiz, M. A. (2002). Separating managerial inefficiency and external conditions in data envelopment analysis. *European Journal of Operational Research*, *143*(3), 625-643.

Nghiem, S., Nguyen, H. T., & Connelly, L. B. (2016). The Efficiency of Australian Schools: A nationwide analysis using gains in test scores of students as outputs. *Economic Papers: A journal of applied economics and policy*, *35*(3), 256-268.

OECD. (2010). The High Cost of Low Educational Performance. In Scenario. OECD.

OECD. PISA 2018 results.

Podinovski, V. V., Ismail, I., Bouzdine-Chameeva, T., & Zhang, W. (2014). Combining the assumptions of variable and constant returns to scale in the efficiency evaluation of secondary schools. *European Journal of Operational Research*, *239*(2), 504-513.

Portela, M. C. A. S., & Thanassoulis, E. (2001). Decomposing school and school-type efficiency. *European journal of operational Research*, *132*(2), 357-373.

Ray, S. C. (1991). Resource-use efficiency in public schools: A study of Connecticut data. *Management science*, *37*(12), 1620-1628.

Rebai, S., Yahia, F. B., & Essid, H. (2020). A graphically based machine learning approach to predict secondary schools performance in Tunisia. *Socio-Economic Planning Sciences*, *70*, 100724.

Royo, C. C., & Fajardo, M. T. M. (2020). Using DEA to Evaluate the Efficiency of Public High Schools for STEM Offering. *American Journal of Educational Research*, *8*(3), 155-161.

Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European journal of operational research*, *90*(3), 553-565.

Ruggiero, J. (1999). Nonparametric analysis of educational costs. *European Journal of Operational Research*, *119*(3), 605-612.

Sarrico, C. S., & Rosa, M. J. (2009). Measuring and comparing the performance of Portuguese secondary schools: A confrontation between metric and practice benchmarking. *International Journal of Productivity and Performance Management*, *58*(8), 767-786.

Schiltz, F., Masci, C., Agasisti, T., & Horn, D. (2017). *Using machine learning to model interaction effects in education: A graphical approach* (No. BWP-2017/4). Budapest Working Papers on the Labour Market.

Sengupta, J. K. (1987). Data envelopment analysis for efficiency measurement in the stochastic case. *Computers & operations research*, *14*(2), 117-129.

Simar, L. (1992). *Estimating efficiencies from frontier models with panel data: a comparison of parametric, non-parametric and semi-parametric methods with bootstrapping* (pp. 167-199). Springer Netherlands.

Simar, L., & Wilson, P. W. (1999a). Some problems with the Ferrier/Hirschberg bootstrap idea. *Journal of productivity Analysis*, *11*, 67-80.

Simar, L., & Wilson, P. W. (1999b). Of course we can bootstrap DEA scores! But does it mean anything? Logic trumps wishful thinking. *Journal of productivity Analysis*, 93-97.

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

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

Smith, P., & Mayston, D. (1987). Measuring efficiency in the public sector. *Omega*, *15*(3), 181-189.

Sreekumar, S., & Mahapatra, S. S. (2011). Performance modeling of Indian business schools: a DEA-neural network approach. *Benchmarking: An International Journal*.

Thanassoulis, E. (1996). Altering the bias in differential school effectiveness using data envelopment analysis. *Journal of the Operational Research Society*, *47*(7), 882-894.

Thanassoulis, E. (1999). Setting achievement targets for school children. *Education Economics*, *7*(2), 101-119.

Thanassoulis, E., & Dunstan, P. (1994). Guiding schools to improved performance using data envelopment analysis: An illustration with data from a local education authority. *Journal of the Operational Research Society*, *45*(11), 1247-1262.

Wanke, P., Blackburn, V., & Barros, C. P. (2016). Cost and learning efficiency drivers in Australian schools: a two-stage network DEA approach. *Applied Economics*, *48*(38), 3577-3604.

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education economics*, *9*(3), 245-268.

Xue, M., & Harker, P. T. (1999). Overcoming the inherent dependency of DEA efficiency scores: a bootstrap approach. *Unpublished Working Paper, Wharton Financial Institutions Center, University of Pennsylvania*.

Yang, F. C. (2017). Integrating corporate social responsibility and profitability into best practice selection: the case of large Taiwanese firms. *Quality & Quantity*, *51*(4), 1493-1512. (Scientific diagram only).

Acknowledgements

I would like to thank Prof. Tommaso Agasisti for giving me the opportunity to work on this thesis in line with my studies and interests, and for his invaluable insights during its preparation.

I would also like to thank Prof. Mara Soncin for her support during this journey.