UNIL | Université de Lausanne

Faculté de droit, des sciences criminelles et d'administration publique

Institut de hautes études en administration publique

Essays on the measurement of school efficiency

THÈSE DE DOCTORAT

présentée à la Faculté de droit, des sciences criminelles et d'administration publique de l'Université de Lausanne pour l'obtention du grade de

Docteur en administration publique

par

Jean-Marc HUGUENIN

Directeur de thèse
Prof. Nils Soguel

LAUSANNE
2014

UNIL | Université de Lausanne

Faculté de droit, des sciences criminelles et d'administration publique

Institut de hautes études en administration publique

Essays on the measurement of school efficiency

THÈSE DE DOCTORAT

présentée à la Faculté de droit, des sciences criminelles et
d'administration publique de l'Université de Lausanne
pour l'obtention du grade de

Docteur en administration publique

par
Jean-Marc HUGUENIN

Directeur de thèse
Prof. Nils Soguel

LAUSANNE
2014

UNIL | Université de Lausanne
Faculté de droit,
des sciences criminelles
et d'administration publique

IMPRIMATUR

Le Décanat de la Faculté de droit, des sciences criminelles et d'administration publique, sur proposition d'un jury formé des professeurs Nils Soguel, Laure Athias, Jean-Blaise Wasserfallen et du docteur Jill Johnes, sans se prononcer sur les opinions du candidat, autorise l'impression de la thèse de Monsieur Jean-Marc Huguenin, intitulée :

Essays on the measurement of school efficiency

Lausanne, le 5 novembre 2014

Prof. Martial Pasquier

M. Pasquier

Vice-Doyen de la Faculté de droit,
des sciences criminelles
et d'administration publique

Year : 2014

Essays on the measurement of school efficiency

Huguenin Jean-Marc

UNIL | Université de Lausanne

Faculté de droit, des sciences criminelles et d'administration publique

Institut de hautes études en administration publique

Essays on the measurement of school efficiency

THÈSE DE DOCTORAT

présentée à la Faculté de droit, des sciences criminelles et d'administration publique de l'Université de Lausanne pour l'obtention du grade de

Docteur en administration publique

par

Jean-Marc HUGUENIN

Directeur de thèse
Prof. Nils Soguel

LAUSANNE

2014

Terms of reference

Author	Jean-Marc Huguenin - Senior lecturer, University of Lausanne, Swiss Graduate School of Public Administration – IDHEAP – - Chief executive officer, Economistes indépendants
Title	Essays on the measurement of school efficiency
Jury	Nils Soguel, IDHEAP, University of Lausanne Laure Athias, IDHEAP, University of Lausanne Jill Johnes, Management School, University of Lancaster Jean-Blaise Wasserfallen, Faculty of Business and Economics (HEC), University of Lausanne, and University Hospital of Lausanne (CHUV)
Starting date	May 2012
Filing date	November 2014
Pre-defense date	July 4th, 2013
Defense date	August 19th, 2014
Public defense date	November 5th, 2014
Acronym	PhD_JMH_DEA
Place	Lausanne

IDHEAP
Bâtiment IDHEAP
CH – 1015 Lausanne
T 021 557 40 00
F 021 557 40 09
jean-marc.huguenin@unil.ch
www.idheap.ch/fp

In loving memory of

Selma Kohli (Imhof) (1912 – 2012)
Roland Kohli (1914 – 1988)

Claude-Rolande Guerdat-Lojpur (Kohli) (1954 – 2011)
Slobodan Lojpur (1969 – 2011)

René Kohli (1942 – 2000)
Marianne Kohli (Keller) (1942 – 2006)

Francis Huguenin (1925 – 1998)

Olga Huguenin (Hildbrand) (1901 – 1983)
Georges Huguenin (1901 – 1986)

Emma Kohli (Wütrich) (1885 – 1976)
Fernand-Auguste Kohli (1889 – 1960)

Pauline Imhof (Heid) (1867 – 1960)
Jakob Imhof (1857 – 1944)

Antoinette Boillat (Sylvant) (1902 – 1936)
Maurice Boillat (1902 – 1964)

Contents

List of figures

Note that the figures contained in Annex 1 do not appear in this list of figures.

List of tables

Note that the tables contained in Annex 1 do not appear in this list of tables.

PhD thesis

Essays on the measurement of school efficiency

Jean-Marc Huguenin

Abstract

Introduction

School efficiency has become a major issue worldwide for different reasons, one of them being the consequent amount of public spending devoted to education in developed countries. In the United States of America, the Supreme Courts in several states have ruled that the system of financing public education was unconstitutional because these states have not provided public education in an efficient manner. In France, the Cour des comptes shows that the resources allocated to compulsory education are sufficient but are misused. In Switzerland, improving efficiency in compulsory education is one of four reforms recommended by a recent OECD analysis. Efficiency is also one of three criteria selected by the Swiss Conference of Cantonal Ministers of Education to assess the education system.

These examples lead to the necessity to measure school efficiency. Measuring school efficiency is a challenging task, especially considering that multiple inputs (such as capital, labour, energy, materials and services) are combined in order to produce multiple outputs (pupils, courses, competences, etc.). The potential lack of data is an additional problem. As a result, there is still an absence of an official analytical framework to measure school efficiency in Switzerland and abroad: which is therefore still to be designed. This framework has to accommodate the different characteristics of school environment (for example, schools have a different proportion of allophone or underprivileged pupils), and specifically the issue of positive discrimination.

Objectives

In Switzerland, the efficiency of primary schools has never been assessed. Moreover, the determinants of primary school efficiency have never been identified. As a result, the **first objective** of the current thesis consists in measuring primary school efficiency and in identifying the determinants of primary school efficiency. By achieving this first goal, this thesis provides to decision makers (and among them, education ministers) an analytical framework allowing them to measure school efficiency. It also identifies the determinants of school efficiency and highlights the importance of several environmental variables. The number of school sites, a variable not tested in previous studies, is included in the analysis. Information about the impact of multi-sites on school efficiency is particularly valuable in a context of school mergers observed in several Swiss states.

As with several other states, the State of Geneva has introduced a priority education policy, based on positive discrimination, in 2008. The **second objective** of the current thesis is to develop an analytical framework to measure school efficiency which is able to deal with positive discrimination. In other words, schools benefiting from additional resources in order to improve equity should not be penalized in the measurement of efficiency by the fact that they use more resources than others, all other things being equal (and particularly pupils' performance).

Within DEA, several alternative models allow for an environmental adjustment. These models lead to possible divergent results, leaving the decision makers in a delicate (not to say confusing) situation when the time comes to select one of the models. The **third objective** of this thesis is therefore to test how the diverging results can be narrowed using suitable techniques. The use of a technique that offers the option of selecting the most suitable model according to the preferences of the decision makers will be proposed.

In order to foster the mastering and the use of Data Envelopment Analysis – a performance measurement technique – in Switzerland, and especially in the public sector, a pedagogical guide about DEA is introduced. This is the **fourth and last objective** of the thesis.

Essay # 1
Determinants of school efficiency: the case of primary schools in the State of Geneva, Switzerland

The public primary school system in the State of Geneva, Switzerland, is characterized by centrally evaluated pupils' performance with the use of standardized tests. As a result, consistent data are collected among the system. The 2010-2011 dataset is used in a two-stage analysis of school efficiency. At the first stage, Data Envelopment Analysis (DEA) is employed to calculate an individual efficiency score for each school. It shows that, on average, each school could reduce its inputs by 7% and still provide the same quality of pupils' performance. At the second stage, efficiency is regressed on school characteristics and environmental variables, all of them being not controllable by headteachers. The model is tested for multicollinearity, heteroskedasticity and endogeneity. Four variables are identified as statistically significant. School efficiency is influenced negatively by *(1)* the provision of special education, *(2)* the proportion of disadvantaged pupils and *(3)* the fact of operating on several locations (or sites). School efficiency is influenced positively by school size (captured by the number of pupils). The proportion of allophone pupils, the fact to operate in urban areas and the provision of reception classes for immigrant pupils are not significant. Although the significant variables influencing school efficiency are not under the control of headteachers, it does not mean that nothing can be done to either boost their positive impact or curb their negative impact. Policy-related implications are discussed.

Essay # 2
DEA does not like positive discrimination: a comparison of alternative models based on empirical data

Due to the existence of free software and pedagogical guides, the use of Data Envelopment Analysis (DEA) has been further democratized in recent years. Nowadays, it is quite usual for practitioners and decision makers with no or little knowledge in operational research to run themselves their own efficiency analysis. Within DEA, several alternative models allow for an environmental adjustment. Five alternative models, each of them easily accessible to and achievable by practitioners and decision makers, are performed using the empirical case of the 90 primary schools of the State of Geneva, Switzerland. As the State of Geneva practices an upstream positive discrimination policy towards schools, this empirical case is particularly appropriate for an environmental adjustment. The majority of alternative DEA models deliver divergent results. It is a matter of concern for applied researchers and a matter of confusion for practitioners and decision makers. From a political standpoint, these diverging results could lead to potentially ineffective decisions.

Essay # 3
DEA and non-discretionary variables: selecting the right model (for you) using multi-criteria decision analysis

Within Data Envelopment Analysis (DEA), several alternative models allow for an environmental adjustment. The majority of them deliver divergent results. From a practical standpoint, but also from a political perspective, decision makers (i.e. top civil servants and ministers) face the difficult task of selecting the most suitable model. This study is performed to overcome this difficulty. By doing so, it fills a research gap. First, a two-step web-based survey is conducted. In the first step, the survey aims to collect general views from DEA scholars and practitioners to identify the selection criteria. In the second step, the survey aims to prioritize and weight the selection criteria identified in the first step with respect to the goal of selecting the most suitable model. But it also aims to collect the preferences of the respondents about which model is preferable to fulfil each selection criterion. Second, Analytic Hierarchy Process, a multi-criteria decision analysis method, is used to quantify the preferences expressed in the survey. Results show that the understandability, the applicability and the acceptability of the alternative models are valid selection criteria. When results are aggregated over the respondents, the categorical model developed by Banker and Morey (1986a) emerges as the most suitable model. However, individual results may vary and other models may be identified as the most suitable ones from an individual perspective.

Conclusion

Conducting an efficiency analysis allows decision makers to hold an open discussion about the way to improve entities' efficiency. In this way, the results of an efficiency analysis are, in themselves, not the most important part of the process. They represent rather a means which permits the reaching of an objective of continuous improvement within the organizations. As a result, conducting an efficiency analysis represents a step towards evidence-based management or policy.

An environmental variable of particular interest, tested in this thesis, consists of the fact that operations are held, for certain schools, on multiple sites. Results show that the fact of being located on more than one site has a negative influence on efficiency. A likely way to solve this negative influence would consist of improving the use of ICT in school management and teaching. Planning new schools should also consider the advantages of being located on a single site, which allows attainment of critical size in terms of pupils and teachers.

The fact that underprivileged pupils perform worse than privileged pupils has been public knowledge since the 1960s. As a result, underprivileged pupils have a negative influence on school efficiency. This is confirmed by this thesis for the first time in Switzerland. Several countries have developed priority education policies in order to compensate for the negative impact of disadvantaged socioeconomic status on school performance. In general, these policies have failed. As a result, other actions need to be taken.

In order to define these actions, one has to identify the social-class differences which explain why disadvantaged children underperform.

Childrearing and literary practices, health characteristics, housing stability and economic security influence pupil achievement. Rather than allocating more resources to schools, policymakers should therefore focus on related social policies. For instance, they could define pre-school, family, health, housing and benefits policies in order to improve the conditions for disadvantaged children.

Several alternative efficiency measurement approaches (or models within a particular approach) allow for an environmental adjustment. The majority of them deliver divergent results. Multi-criteria decision analysis methods can help the decision makers to select the most suitable approach (i.e. the approach which suits best their own preferences). The number of selection criteria should remain parsimonious and not be oriented towards the results of the approaches in order to avoid opportunistic behaviour. The selection criteria should also be backed by the literature or by an expert group. Once the most suitable approach is identified, the principle of permanence of methods should be applied in order to avoid a change of practices over time.

List of abbreviations

ADMIN	Number of full-time equivalent administrative and technical staff
AE	Allocative efficiency
ALLO	Percentage of allophone pupils (per school)
BUDGET	School budget in Swiss francs – excluding staff salaries and capital expenditures –
CLASS	Number of classes within a school
COLS	Corrected ordinary least squares
CRS	Constant returns to scale
CRSTE	Constant returns to scale technical efficiency
DEA	Data Envelopment Analysis
DMU	Decision Making Units
DRS	Decreasing returns to scale
EE	Economic efficiency
FDH	Free disposal hull
FTE	Full-time equivalent
IRS	Increasing returns to scale
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares
PUPIL	Number of pupils in a school
RECEPTION	Dummy variable referring to the fact that special reception classes for immigrants pupils are available at a particular school
SCORE2	Pupils' results at the French and mathematics standardized tests at the end of the second grade
SCORE4	Pupils' results at the French, German and mathematics standardized tests at the end of the fourth grade
SCORE6	Pupils' results at the French, German and mathematics standardized tests at the end of the sixth grade
SE	Scale efficiency
SFA	Stochastic frontier analysis
SITE	Dummy variable referring to the fact that a school is located on one or several spots (sites)
SOCIO	Percentage of pupils (per school) whose parents are blue-collar workers or unqualified workers

SPECIAL	Dummy variable referring to the fact that special education for special needs pupils is available at a particular school
SCCME	Swiss Conference of Cantonal Ministers of Education
TE	Technical efficiency
TEACHER	Number of full-time equivalent teaching staff
URBAN	Dummy variable referring to the fact that a school is located in an urban or a rural area
VRS	Variable returns to scale
VRSTE	Variable returns to scale technical efficiency

Acknowledgements

I would not have been able to produce the current thesis without the advice, help, support and encouragement of many people. They contributed to improve greatly its quality. Of course, any potential remaining mistakes are mine.

My first and greatest thank you goes to Nils Soguel, Professor at the Swiss Graduate School of Public Administration – IDHEAP – of the University of Lausanne. His patience, especially during the process *before* the PhD registration, was remarkable. When the time had really come, he gently and sensitively nudged me into the PhD process. I am deeply and forever indebted to him.

I also address a special thank you to the members of the PhD jury. Laure Athias, Professor at IDHEAP, made numerous valuable comments about the articles to improve their quality. Jill Johnes, Reader at the Management School of the University of Lancaster, helped me progress by raising critical issues. She also helped me stay focused on the objectives of the articles. Finally, Jean-Blaise Wasserfallen, Professor at the Faculty of Business and Economics (HEC) of the University of Lausanne, made me focus on the political and practical implications of my results.

I am grateful to Timothy Coelli, Adjunct Professor at the School of Economics of the University of Queensland, and Michel Deslierres, Professor at the School of Economics of the University of Moncton, for having let me reproduce screenshots of their respective software packages, DEAP and Win4DEAP.

My colleagues (and friends) at the Public Finance Chair of IDHEAP deserve a special mention. Béatrice Hausmann provided constant and unfailing support. I owe her a lot. Maxime Clemenceau wisely advised me on econometrical issues. Julie Silberstein gave me an initiation to STATA ®. And Evelyn Munier shared with me her interest in school organisation. Gilles Léchot, external consultant fellow at the Public Finance Chair of IDHEAP and CEO of Compas Management Services, gave me much valuable advice, especially about how to manage a PhD thesis. His experience was highly appreciated. I thank all of them.

Séverin Blanc, IT specialist at the IDHEAP, unlocks various situations along the PhD process. I express my gratitude to him for his help.

I also thank Faye Sizeland, Lecturer at the Economics and Finance Department of the University of Portsmouth, for having skillfully and diligently proofread most of the present thesis, and Robert Gausden, Principal lecturer at the Economics and Finance Department of the University of Portsmouth, for having tutored me about the Tobit regression.

Before and during the time of the thesis, I received the constant encouragement of my colleagues and friends of the Observatory on Primary Education of the State of Geneva. I express my gratitude to Georges Solaux, Professor at the University of Bourgogne, Jean-Paul Payet, Professor at the University of Geneva and José Ramirez, Professor at the School of Business Administration of the University of Applied Sciences and Arts Western Switzerland.

I also express my gratitude to my colleagues of the headteachers training program (www.fordif.ch): Frédéric Yvon, Professor at the University of Geneva, Olivier Perrenoud, Professor at the Teacher Education College of the State of Vaud and David Perrenoud, Senior lecturer at the Swiss Federal Institute for Vocational Education and Training.

I would not have been able to undertake PhD studies if I had not been properly trained in business administration and economics. Bruno Amsler, Professor at the High School of Biel/Bienne, was my professor when I was a high school student and my mentor when I was pursuing pedagogical studies. He is also a friend. If a prize for the best economics professor had existed in Switzerland, Bruno Amsler would undoubtedly have been its laureate year after year.

Paul-André Garessus, former Professor of mathematics and current Head of the teacher training program at the Teacher Education College of the states of Berne, Jura and Neuchâtel, wisely advised me on mathematical issues. I thank him for his help.

I also thank the many anonymous respondents of the Data Envelopment Analysis community who agreed to take part in my web-based survey. Without their commitment, I would not have been able to produce the last article of the current thesis.

Finally, I had the privilege to obtain the support of the Edmond de Rothschild Group. I sincerely thank it for its trust and for having believed in this project.

On a personal level, I would not be the person I am without the support and love of my parents, Mariette and Denis Huguenin. Each of them, in their own way, contributed to my development. My mother in particular is the one who helped me with my homework during my first school years. She put me on the right track. I owe to her much of my school performance. My father embodies the model of work and entrepreneurship I would like to reach. I owe to him a lot of my working and ethical values. I am deeply and forever indebted to them. Thank you.

Liliane Huguenin is my last living grand-parent. In many ways, I share with her common values and qualities. Thanks to her, I know where I come from, and where I have to go.

My sister Karine Saucy (Huguenin) and her children Joanie, Bastien and Clément provided me a great support, especially while I was staying in Portsmouth. I thank them all for having being present, while I was away.

And I thank Jörg Tillmann Neuscheler and Florian Eisele, my confederates from Texas A&M University, for having showed me the way to follow.

Finally, the time of the current thesis is inextricably linked to the city of Portsmouth. In spring 2012, I was invited by Alessio Ishizaka, Reader at the Operations and Systems Management Department of the University of Portsmouth, in order to write a chapter dedicated to Data Envelopment Analysis in his textbook about multi-criteria decision analysis methods. This represented the milestone which really launched my PhD. I have had the opportunity to stay again in Portsmouth in 2013 and 2014. Portsmouth is a surprising historical city. In Southsea, one finds a liquid silver sea and an infinite beach covered with shells and pebbles. In Commercial Road, there is always the next James Blunt playing guitar in the street. In Fratton stands the last ol'good stadium, home of Pompey. In the darkest corner of the Festing Pub, there is always a bunch of happy guys drinking pints of ale and watching football games. In the Rose garden at sunset, teenagers hide on the roof when the gates are closing. In Canoe Lake, kids fish small crabs and eat fluorescent blue ice creams. And in Albert Road, one finds the best food providers, among them the Bodrum restaurant and Shariff's take-away.

My last thank you goes to Jessana Devaux, who has been by my side for the last decade.

1 Summary report

1.1 Preliminary remarks

The current thesis consists of the sum of three scientific essays according to the regulations governing award of the degree of doctor of philosophy in public administration at the University of Lausanne (art. 15). In other words, this thesis is not a monograph. The structure of a thesis consisting of the sum of several scientific essays differs in some regards from the structure of a monographic thesis. The above cited regulations (art. 15) state that the essays are accompanied by a summary report (*rapport synthétique* in French). This summary report contains a presentation of the essays, the main issues of the thesis, the context of the thesis and the conclusion and perspectives[1]. Note that the regulations cited above do not mention that the summary report should contain a review of the literature. This makes sense as every essay is independent from each other and already contains its own literature review.

However, in this thesis, a selected literature review, which is complementary (and not redundant) to the literature reviews included in the essays, is provided in the summary report. It covers the recent use of DEA in the education sector and in Switzerland.

In a thesis consisting of the sum of several scientific essays, the essays have (obviously) to present some sort of link among them. Usually, this link is assured by the theme developed or the approaches used. But, as these essays are independent from each other, they do not necessarily have to present a logical sequence. It means, for instance, that the second essay is not necessarily the (logical) successor to the first one. As a result, the essays must not be considered as chapters (as in a monographic thesis) with a logical sequence. In this thesis, although a strong link appears between the essays (and, to some extent, even a logical sequence), readers must not be fooled. The essays of this thesis remain

1 Following the regulations cited above, the first part of this thesis is called 'summary report' and not 'introduction'.

independent essays and must be considered as such (and not as chapters with a logical sequence).

The first essay (Determinants of school efficiency: the case of primary schools in the State of Geneva, Switzerland) has been accepted for publication in the International Journal of Educational Management (Huguenin, forthcoming). The pedagogical guide about Data Envelopment Analysis contained in Annex 1 has been published in 2012 in English and in 2013 in French (Huguenin, 2012, 2013a) as a *cahier de l'IDHEAP*, the high quality line of publications of the Swiss Graduate School of Public Administration. A synthetized version of this guide has also been published as a chapter in Ishizaka and Nemery (2013) (Huguenin, 2013b). The second (DEA does not like positive discrimination: a comparison of alternative models based on empirical data) and third essays (DEA and non-discretionary inputs: how to select the most suitable model (for you) using multi-criteria decision analysis) are currently in the process of being published.

1.2 Context of the thesis

School efficiency

Measuring and comparing efficiency are not only major trends observed in the education sector; they are also at work across the whole public sector. The reasons why these trends have emerged are partly cultural and partly technical. The notion of efficiency is a key feature of the new public management philosophy (Hood, 1991). And the measurement of efficiency has been facilitated by the 'Big Data' phenomenon (Zikopoulos & Eaton, 2001). Public sector benchmarking is a relatively new issue (Dorsch & Yasin, 1998) but is of increasing importance (van Helden & Tillema, 2005). It has become a key and lasting feature of tomorrow's public management and policy.

School efficiency is a major concern worldwide. Based on the preliminary work of Sutherland, Price, Joumard and Nicq (2007), the Organisation for Economic Co-operation and Development (OECD, 2007, pp. 262-278) assesses the efficiency of compulsory education in the OECD countries[2]. It concludes that the mean efficiency is equal to 69.3%, meaning that, on average, each country could reduce the resources allocated to the education sector by 30.7% and still obtain the same educational performance. OECD (2007, p. 14) stresses the fact that "this efficiency indicator is exploratory at this stage; it covers only elementary and secondary schooling and it will require substantial further development over the years to come, (…)." School efficiency has become a major issue for different reasons, one of them being the amount of public spending devoted to education in developed countries. Comparisons across countries but also within countries, made possible by studies such as the

[2] The method used is Data Envelopment Analysis, as in this thesis.

Program for International Student Assessment, have unveiled the potential for improvement in most countries.

As illustrative examples, three cases (in three different countries) are briefly introduced below in order to stress the fact that school efficiency has become an essential issue around the world.

In the United States of America, Ruggiero (2004, p. 323) relates how the provision of public education has moved from equity "towards educational goals efficiently" by the ruling of the courts. Reschovsky (1994) provides a review of courts' rulings. The Supreme Courts in several states, including Ohio[3], Kentucky, Montana, New Jersey and Texas, have ruled that the system of financing public education (through the equalization of fiscal resources available to school districts) was unconstitutional because these states have not succeeded in providing public education in an efficient manner, as stated in their constitutions[4].

In France, a recent report from the *Cour des comptes* (Cour des comptes, 2013) is devoted to the management of teachers. It shows that France is ranked # 18 out of 34 members of the OECD concerning pupils' performance. However, France allocates as much or more public spending to compulsory education than countries which perform better. In other words, the efficiency of compulsory education in France is low. The *Cour des comptes* argues that the resources allocated to compulsory education are sufficient but are misused[5]. Basically, it advocates the improvement of efficiency in order to improve pupils' performance.

In Switzerland, improving efficiency in compulsory education is one of four reforms recommended by a recent OECD analysis to raise education outcomes (Fuentes, 2011). Efficiency happens to also be one of three criteria[6] selected by the Swiss Conference of Cantonal Ministers of Education (SCCME) to assess the education system (Wolter, 2010)[7]. Despite this, studies of the efficiency of Swiss schools are virtually non-existent (Olivares and Schenker-Wicki, 2012, 2010; Solaux, Huguenin, Payet & Ramirez, 2011; Meunier, 2008; Diagne, 2006; Schenker-Wicki & Hürlimann, 2006). As a result, decision makers still

---

3 The chronology of the Ohio School Funding litigation is available at http://www.bricker.com/services/resource-details.aspx?resourceid=412

4 For instance, the New Jersey State Constitution mentions explicitly that "The Legislature shall provide for the maintenance and support of a thorough and efficient system of free public schools for the instruction of all the children in the State between the ages of five and eighteen years" (article VIII, section IV, paragraph 1). The New Jersey State Constitution can be found at http://www.cityconnections.com/NJ_STATE_CONSTITUTION.pdf

5 In French: "Le ministère de l'éducation nationale ne souffre pas d'un manque de moyens budgétaire ou d'un nombre trop faible d'enseignants mais d'une utilisation défaillante des moyens existants" (Cour des comptes, 2013, p. 136).

6 Effectiveness and equity are the other two criteria.

7 Another example of the importance of efficiency is found in Finland. The Finnish National Board of Education considers that quality, efficiency, equity and internationalization are the key words in Finnish education policies (http://www.oph.fi/english/education).

rely on partial productivity ratios (mainly cost per pupil) to monitor the education system.

Measuring school efficiency

From a methodological standpoint, measuring school efficiency is a challenging task, especially considering that multiple inputs (such as capital, labor, energy, materials and services) are combined in order to produce multiple outputs (pupils, courses, competences, etc.). The potential lack of data is an additional problem. In the case of Switzerland, Sheldon (1995, pp. 67-68) points out that there has long been a lack of comparable data, especially output data, on a national scale and often on a cantonal scale. However, in a recent attempt to harmonize compulsory education, the SCCME has defined national objectives in terms of standardized competences to be acquired by pupils. These standardized competences will be assessed in the near future (probably starting in 2015), offering comparable data on compulsory education output.

However, there is still an absence of an official analytical framework to measure school efficiency in Switzerland; which is therefore still to be designed. Furthermore, this framework has to accommodate the different characteristics of school environment (for example, schools have a different proportion of allophone or underprivileged pupils). Several states have implemented positive discrimination policies, mainly in the form of upstream additional resources allocated to disadvantaged schools. Considering an identical level of output, the danger is that schools benefiting from upstream additional resources would obtain a lower efficiency score if the analytical framework was not adjusted for the environment.

The absence of official analytical frameworks is also noted in other countries, even in countries where comparable data are available[8]. For instance, the efficiency indicator developed by the OECD in 2007 (OECD, 2007) has not been repeated in the following editions of *Education at a Glance* published by the OECD. This could be interpreted as a political difficulty to agree on such an analytical framework (i.e. the choice of the method and the choice of the inputs and outputs). Note that Portugal has, however, developed a flexible analytical framework based on Data Envelopment Analysis. This framework is available to headteachers through a website. Headteachers can conduct their own efficiency analysis by selecting the inputs and the outputs to be included in the assessment (Portela, Camanho & Borges, 2011). Approaches to measure school efficiency are discussed in Section 1.6.

---

8 The absence of an analytical framework to measure school efficiency is not to be confounded with league tables ranking schools according to student achievement (effectiveness). Such league tables exist in several countries, such as England.

1.3 Issue of the thesis

In Switzerland, the efficiency of primary schools has never been assessed. Moreover, the determinants of primary school efficiency have never been identified. As a result, the **first aim** of the current thesis consists in measuring primary school efficiency and identifying the determinants of primary school efficiency. By achieving this first goal, the thesis provides to decision makers (and among them, education ministers) an analytical framework allowing to measure school efficiency. It also identifies the determinants of school efficiency and highlights the importance of several environmental variables. Some of these variables are well-known in the empirical literature, such as the socio-economic status of pupils. But an original variable, not tested in previous studies, is included in the analysis. This variable is the number of sites on which a particular school operates. Information about the impact of multi-sites on school efficiency is valuable in a context of school mergers observed in several Swiss states.

As in several other states, the State of Geneva introduced a priority education policy in 2008. The **second aim** of the current thesis is to develop an analytical framework to measure school efficiency which is able to deal with positive discrimination. In other words, schools benefiting from additional resources in order to improve equity should not be penalized in the measurement of efficiency by the fact that they use more resources than more privileged schools, all other things being equal (and particularly pupils' performance).

Within DEA, several alternative models allow for an environmental adjustment. These models lead to possible divergent results, leaving the decision makers in a delicate (not to say confusing) situation when time comes to select one of the models. The **third aim** of the current thesis is therefore to test how the diverging results can be narrowed using suitable techniques. The use of a technique that offers the option of selecting the most suitable model according to the preferences of the decision makers will be proposed.

The approach to measure school performance will be based on Data Envelopment Analysis (DEA), a non-parametric approach developed by Charnes, Cooper and Rhodes (1978). Although the use of DEA is widespread around the world (Emrouznejad, Parker & Tavarez, 2008), the number of published applications concerning Switzerland is limited to a restricted group of initiated authors: Olivares and Schenker-Wicki (2012, 2010), Solaux, Huguenin, Payet and Ramirez (2011), Schoenenberger, Mack and von Gunten (2009), Soguel and Huguenin (2008), Widmer and Zweifel (2008), Meunier (2008), Diagne (2006), Jeanrenaud and Vuilleumier (2006), Schenker-Wicki and Hürlimann (2006), Ferro-Luzzi, Flueckiger, Ramirez and Vassiliev (2006), Farsi and Filippini (2005), Steinmann and Zweifel (2003). In order to foster the mastering and the use of DEA in Switzerland, and especially in the public sector, a pedagogical guide about DEA will be introduced. This is the **fourth and last aim** of the current thesis.

1.4 Content of the thesis and methods

In addition to this summary report, the current thesis includes three essays and a pedagogical guide.

The first essay of the thesis (Determinants of school efficiency: the case of primary schools in the State of Geneva, Switzerland) is contained in Chapter 2. In this essay, a two-stage DEA model is applied in order *(1)* to measure school efficiency and *(2)* to identify the non-discretionary variables which impact school efficiency. The two-stage model combines an **efficiency analysis** and an **econometric analysis**. It consists in solving a DEA model with no environmental adjustment in the first stage. In the second stage, an econometric model is used as a corrective method in order to allow for environmental factors. The exogenous variables that impact the efficiency scores are identified. Regardless of the fact that primary school efficiency has never been measured in Switzerland, the originality and the strengths of this essay are as follows: first, the second stage regression model is tested for multicollinearity, heteroskedasticity and endogeneity; this is seldom done in two-stage DEA applications; second, the specification of the functional form of the second stage regression model is investigated; this is seldom done in two-stage DEA applications ; third, the second stage regression model tests the impact of the number of school sites on school efficiency; this variable has never been tested in two-stage DEA applications.

The second essay of the thesis (DEA does not like positive discrimination: a comparison of alternative models based on empirical data) is located in Chapter 3. In this essay, several models allowing for an environmental adjustment, within DEA, are applied to an identical case. It investigates if these models lead to possibly diverging results. School efficiency scores derived from these models are compared with the help of several indicators. The originality and the strengths of this essay are twofold: first, comparison analysis of DEA models using empirical data (and not simulated data) are rare in the literature; second, a new model is proposed in addition to the existing models; this model has been developed to answer the particular features of positive discrimination in educational policy.

The third essay of the thesis (DEA and non-discretionary inputs: how to select the right model (for you) using multi-criteria decision analysis) is contained in Chapter 4. As efficiency results diverge according to the alternative models tested, this essay performs a **multi-criteria analysis** in order to help decision makers to select the most suitable model from a practical standpoint. This essay argues that the choice of the 'right' model should reflect the preferences of the decision-makers. As a result, the preferences of the decision makers are expressed upon a chosen set of criteria. The Analytic Hierarchy Process method – AHP – (Saaty, 1980) is the multi-criteria analysis method of choice in this essay.

Finally, Annex 1 contains a pedagogical guide about DEA in the public sector.

1.5 Empirical field and data

Empirical field

In the State of Geneva, early childhood (corresponding to the international standard classification of education 0 – ISCED # 0 –) – duration of 2 years –, primary (ISCED # 1) – duration of 6 years – and lower secondary education (ISCED # 2) – duration of 3 years – is compulsory.

The empirical field of this thesis consists in the full population of 90 primary schools in the State of Geneva – school year 2010-2011 –[9]. These schools are funded by the State government (basically staff salary) and by local authorities – i.e. municipalities – (basically school infrastructure). Pupils' competences are assessed with the use of standardized tests at three different times in two or three subjects. At the end of the second grade, French (mother tongue) and mathematics are assessed; at the end of the fourth and sixth grade, French, German (first foreign language) and mathematics are assessed.

Primary schools are managed by headteachers (or principals) assisted by one or several teachers working part time as headteachers' assistants. The staff consists of teachers, secretaries and schoolkeepers (maintenance). In some schools, educators are also active.

In order to adjust to local environment, partial autonomy in management is decentralized to schools. For instance, headteachers define job profiles and recruit teachers; they are responsible for school quality (and hence pupils performance); they also chair the school board.

Every school has a school board composed by representatives of the school staff, parents and civil-servants of the municipality. School boards are headed by headteachers. They are instances of democracy where stakeholders are informed and consulted. They only have limited authority over school management, but can make propositions about day-to-day school life. School boards aim to develop better relationships between school, families and local communities.

Primary schools- main characteristics are the following:

- One school can be located on one or several spots (up to five) – which implies that school buildings can be spread over several locations (or sites)–;
- Special education is available only in a limited number of schools (21 schools out of 90); it means that pupils with special needs are grouped in the schools where special education is available;
- Special reception classes for immigrant pupils are available only in a limited number of schools (35 schools out of 90).

---

[9] The State of Geneva has been selected for two main reasons. First, it has implemented upstream positive discrimination measures since 2008. Second, access to a school database has been secured for the school year 2010-2011; this is crucial in a context where data *(1)* including input and output and *(2)* covering all schools of a State (and not just a sample) are particularly difficult to obtain (if not inexistent). Once the standardized competences of pupils have been assessed on a national scale, the analytical framework used in this thesis could easily be extended to match this scale.

The State of Geneva practices a policy of positive discrimination towards schools. Additional teaching resources are allocated to disadvantaged schools. Five school categories (A to E) are defined according to the percentage of pupils (per school) whose parents are blue-collar workers or unqualified workers – category # 9 of the International Standard Classification of Occupations – (Observatory on Primary Education, 2010).

Database

The consolidated database has been constructed from several unlinked sub-databases. These sub-databases are official sub-databases of the General Direction of Primary Schools. Note that they are not accessible by the public. The first sub-database provides information about pupils' performance. Data about pupil's performance can be considered as high quality data considering that test scores are totally standardized in the State of Geneva, from the design (by civil servants external to classes and schools) to the evaluation. As a result, test results provide perfectly comparable information over time and across schools. The second one contains data about school resources. The third one provides information about pupils, except concerning their performance. Finally, the fourth one contains financial data. These sub-databases are issued by the Service of Schooling, the Service of Human Resources, the Service of Schools and the Service of Finance respectively. All these Services are part of the General Direction of Primary Schools. The sub-databases contain data already aggregated at the school level. This data is constructed on the basis of the full populations of schools (90), pupils (34'324) and staff (116.8 full-time equivalent administrative staff and 2009.5 full-time equivalent teachers) in the State of Geneva.

The consolidated database has been structured in such a way to be exploitable. Data has been double checked. Certain variables have been standardized (e.g. pupils' scores on cantonal tests). The final database contains information about schools- resources (teaching staff, administrative and technical staff, and financial budget) and pupils' performance. It also contains information about schools- characteristics (number of sites, availability of special education and reception classes, urban or rural location) and school environment (socioeconomic status of pupils).

1.6 Efficiency measurement

Public policy analysis

Mény and Thoenig (1989, p. 9) define public policy analysis as the "study of the action of public authorities within society". Knoepfel, Larrue, Varone and Hill (2011) identify three major currents in public policy analysis:

- Policy analysis based on the theories of state;
- Explaining how public action functions;

- Evaluation of the effects of public action.

These three currents are not mutually exclusive. The approaches to measure efficiency are part of public policy analysis. By measuring efficiency, these methodological approaches are clearly situated in the last current (evaluation of the effects of public action). However, some approaches, such as DEA, also succeeds in explaining how public action works. In this sense, the approaches to measure efficiency are also part of the second current (explaining how public action functions).

Leading evaluators such as Cronbach (1963) or Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman and Provus (1971) define evaluation as providing information for decision-making. This thesis focuses on school evaluation. Nevo (1995, p. 11) defines educational evaluation as an "act of collecting systematic information regarding the nature and quality of educational objects". According to Nevo (2007), educational evaluation covers the following domains of practice: "student assessment, teacher evaluation, evaluation of instructional materials, program and project evaluation, and school evaluation" (p. 442). Decision-making, improvement, accountability, professionalization and certification are the five functions of educational evaluation defined by Nevo (2007, p. 443). Note that DEA, the method of choice of the current thesis, covers the first four functions of educational evaluation cited above. The essays of this thesis focus on school evaluation.

Knoepfel *et al.* (2011, p. 229-259) make a distinction between the output, the impacts and the outcomes of a public policy:

- The output "identifies the final products of political-administrative processes (that is, the tangible results of implementation)" (p. 229). For instance, the output of an educational policy could be the number of courses taught in schools.
- The impacts are defined as the "changes in the behaviour of target groups that are directly attributable to the entry into force of the political-administrative programmes" (p. 230). For instance, the impacts of an educational policy could be the technical and social competences acquired by pupils. The impacts are observable among target groups only.
- Finally, the outcomes are defined as "all of the effects (…) that are attributable to the policy and triggered in turn by implementation acts (outputs)" (p. 234). The outcomes are observable effects among the end beneficiaries. For instance, the outcomes of an educational policy could be the wealth of the society measured by gross domestic product or the crime rate in the society.

Note that DEA uses the term output in a generic way to cover indiscriminately the notions of output, impact and outcome defined by Knoepfel *et al.* (2011).

Notions of performance in the public sector

Bouckaert and Halligan (2008, p. 14) note that performance is not a unitary concept. This is consistent with de Bruijn (2002, p. 7), who considers performance as multiple. Other authors, such as Stewart and Walsh (1994, p. 45), consider performance in the public sector as an elusive concept. Summermatter and Siegel (2009) realize a meta-analysis of 320 studies dealing with the notion of performance in the public sector[10]. They note that “there is certainly no explicit or implicit consensus about performance of public institutions” (p. 6). They conclude that “obviously, the academic literature is far from applying a consistent interpretation of what performance in the public sector means” (p. 15).

Summermatter and Siegel (2009) extract the definitions of performance from the studies and deconstruct them into their components[11]. Efficiency is mentioned in 56% of the definitions (Summermatter & Siegel, 2009, p. 9). It is the third most cited component (after outcome, 68%, and output, 66%). Note that the notion of impact is not a cited component.

The notion of performance is often defined by its dimensions or components. For instance, van der Waldt (2004, p. 34), in his textbook devoted to the management of performance in the public sector, considers the 3-Es model to approach the notion of performance. This model has been developed by the British Audit Commission (Audit Commission, 2010). It divides performance into three aspects (Audit Commission, 2010, p. 8): economy, efficiency and effectiveness – the 3 Es –. Economy is achieved when the inputs are acquired at the minimal cost. Efficiency consists in “producing the maximum output for any given set of resource inputs or using the minimum inputs for the required quantity and quality of service provided” (Audit Commission, 2010, p. 9). Finally, effectiveness is achieved when an organization reaches its established goals, defined in terms of outputs or outcomes, regardless of the inputs used[12].

Defining performance is important because it determines “both what it is to be measured and how it can be measured” (O’Neill & West-Burnham, 2001, p. 7). In the current thesis, the notion of performance in the public sector can be narrowed to the single notion of efficiency[13], as considered in DEA, for the following reasons:

- The cost of the inputs (which allows measurement of the economy aspect), the established goals measured as outputs or outcomes (which are the components of the effectiveness aspect according to the British Audit

[10] For instance, Neely, Mills, Gregory and Platts (1995) define performance as “the process of quantifying the efficiency and effectiveness of action” (p. 80).

[11] The authors identify twelve components: input, throughput, output, outcome, efficiency, effectiveness, additional types of ratios, quality, requirements, stakeholder-related aspects, value and ethical aspects.

[12] The definition of the outcome by the Audit Commission (2010, p. 9) seems to mix the notions of outcomes and impacts as defined by Knoepfel *et al.* (2011, p. 234 and p. 236). It states that the outcome is “the actual impact and value of the service delivered”.

[13] When the condition of the environment has an influence on the production function of a public organization, the notion of efficiency has to take into account the impact of environmental variables. As shown by Andrews, Boyne and Enticott (2006), performance failure in the public sector is due to both misfortune (i.e. external circumstances) and mismanagement.

Commission) may be included in the measurement of efficiency in DEA; as a result, the aspect of efficiency covers both aspects of economy and effectiveness;

- More than half the definitions of public performance already explicitly consider the notion of efficiency (Summermatter & Siegel, 2009, p. 9);
- The other most cited components in the definitions of public performance are outcomes and outputs; as mentioned above, these two components may be included in the measurement of efficiency using DEA[14]; moreover, in a context of budget restriction (not to say austerity), to focus exclusively on outputs or outcomes (in other words, effectiveness), regardless of the quantity of inputs used to produce them, has become a less tenable position.

Summermatter and Siegel (2009) also note that "there is (…) no consensus about what constitutes appropriate performance measures, indicators or indices" (p. 6). By narrowing the notion of public performance to efficiency, and by using DEA, a unified measure of efficiency can be proposed[15]. Note that unlike efficiency, economy and effectiveness might be measured internally by an entity without referring to a production frontier (and to benchmark itself with other entities). When the production frontier (or efficiency frontier) is unknown, entities cannot measure their efficiency. However, they can measure their productivity which, as with efficiency, allows at the individual entity level links to monetary resources, input costs, physical inputs, physical outputs and output prices[16].

---

[14] Most Data Envelopment Analysis empirical studies consider only outputs and not outcomes or impacts. As a result, they focus on internal technical efficiency. Other studies consider outcomes but call them outputs. But some studies, such as Soguel and Huguenin (2008), distinguish outputs, outcomes and impacts in order to measure external technical efficiency.

[15] Note that performance may be adequately measured by means of other statistical or non-statistical approaches, such as citizen survey evaluations. For instance, a recent study lead by Favero and Meier (2013) compares parent and teacher evaluations to government records of schools' performance (mainly standardized test scores). It shows that parents and teachers are able to conduct meaningful evaluations of school quality. This is probably right in a transparent system where public information about school quality is available, i.e. where there is a low level of asymmetric information. In the case described by Favero and Meier (2013), the city of New York reduces this asymmetric information bias with the help of central accountability tools published on its website (D'Souza, 2013). But parents and teachers are probably unlikely to conduct meaningful evaluations of schools quality in an opaque system where public information about school quality is not divulgated by the central administration (in other words, with a high level of asymmetric information), as in Switzerland. Moreover, a citizen survey evaluation does not provide managerial information, such as the potential of schools' improvement (in terms of input reduction or output augmentation). As a result, it is not a valuable management tool.

[16] The productivity of an entity is defined as the ratio of its output to its inputs (Lovell, 1993, p. 3). This ratio is easy to calculate if the entity uses a single input to produce a single output. It is more difficult if the entity uses multiple inputs and multiple outputs. In this case, the inputs in the denominator have to be aggregated, as the outputs in the numerator.

Notions of efficiency

Pareto efficiency is used as a starting point in order to define the notion of efficiency used in DEA. Pareto efficiency is a state of allocation of resources in which it is impossible to make one party better off without making another party worse off (Varian, 2010, pp. 15-16). In DEA, Cooper, Seiford and Zhu (2004) extend the Pareto efficiency notion to define efficiency as follows: "Full (100%) efficiency is attained by any DMU[17] if and only if none of its inputs or outputs can be improved without worsening some of its other inputs or outputs" (p. 3). However, as the theoretically possible levels of efficiency are usually unknown, the above definition is enhanced to take into account only the empirical data available:

> A DMU is to be rated as fully (100%) efficient on the basis of available evidence if and only if the performance of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs (p. 3).

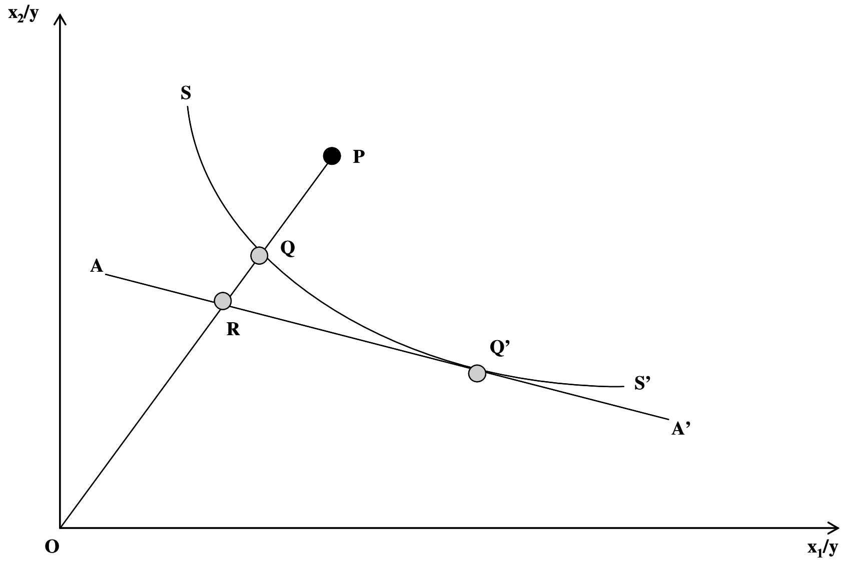
In order to measure efficiency, Charnes, Cooper and Rhodes (1978) develop an algorithm based on mathematical optimization. The initial DEA model proposed by these authors is built on the earlier work of Farrell (1957). Farrel (1957) defines three types of efficiency: **technical efficiency**, **price efficiency** and **overall efficiency**. To illustrate these three types of efficiency, consider that a set of entities use two inputs (x_1 and x_2) to produce one output (y), all of them being measured in quantities. Figure 1 represents the position of a particular entity, P, in an input-oriented representation. It also represents the isoquant line SS'. This isoquant line joins all possible combinations of x_1 and x_2 allowing an entity to produce a given quantity of output in an efficient way. An entity is technically efficient if it minimized the use of the inputs for a given set of output. For instance, the degree of technical efficiency (TE) of entity P is written as follows:

$$TE = \frac{OQ}{OP}$$

Note that the quantity (1 – TE) represents the technical inefficiency of an entity. It indicates by how much an entity must reduce its inputs while keeping its output constant in order to become technically efficient.

[17] DMU stands for Decision Making Units. A DMU is an entity which converts inputs into outputs. Some authors, such as Coelli, Prasada Rao, O'Donnell and Battese (2005), use the term firm instead of DMU. However, in the current thesis and following the practice of the International Public Sector Accounting Standards Board (2012), the term entity instead of DMU or firm is preferred.

Figure 1
The general notion of efficiency is deconstructed into three types by Farrell (1957)



Source: Figure reproduced from Farrell (1957, p. 254)

Both inputs and outputs are measured in physical terms (i.e. quantities) in the technical efficiency type. A further type of efficiency, called **price efficiency** by Farrell (1957) – but referred as **allocative efficiency** in the DEA literature – is defined by introducing the isocost line AA' into Figure 1. The isocost line represents the price ratio of x_1 and x_2. An entity is allocative efficient if it uses the proportion of inputs that minimized their cost given their respective price. An entity is technically efficient if it is located on the isoquant line; it is allocative efficient if it is located on the isocost line. For instance, point Q is technically efficient but allocative inefficient. Point Q' is technically and allocative efficient; it uses the two inputs in the best proportion given their respective price. As the costs of production of R are the same as those of Q' (both points are located on the isocost line), the degree of allocative efficiency (AE) of P is written as follows:

$$AE = \frac{OR}{OQ}$$

The third type of efficiency is the **overall efficiency** – sometimes referred as the economic efficiency in the DEA literature –[18]. The overall efficiency is defined as the product of technical efficiency and allocative efficiency. In the case of P, the overall efficiency (OE) is written as follows:

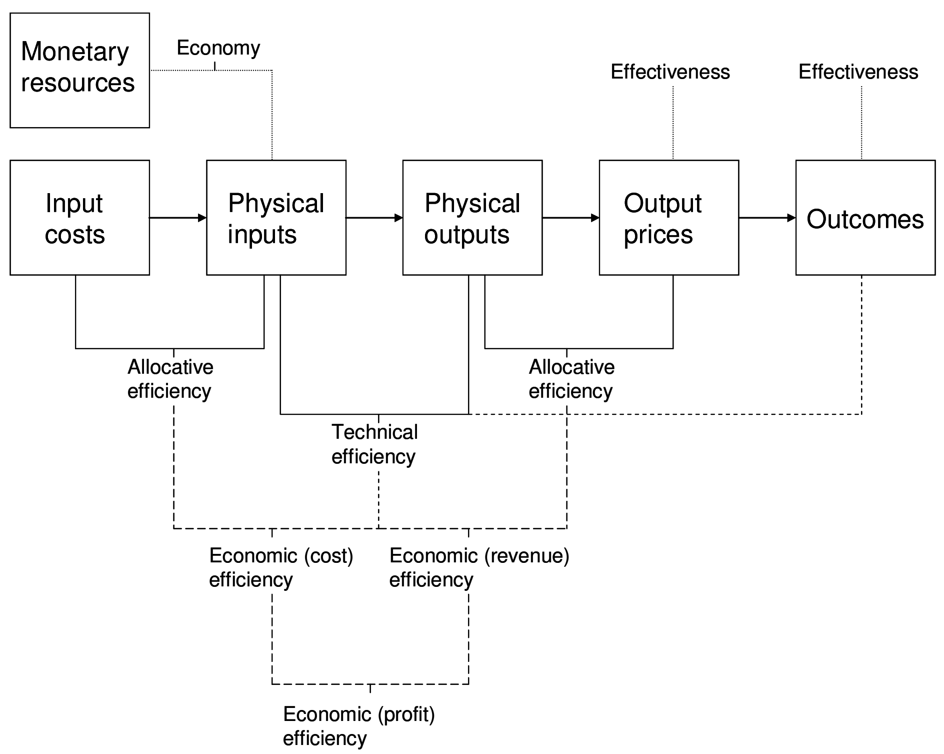
[18] Note that some authors, such as Coelli *et al.* (2005), use the terms of cost efficiency, revenue efficiency and profit efficiency instead of overall efficiency. In the illustrative case developed in this section, the overall efficiency corresponds to the cost efficiency because the prices (or costs) of the inputs are known.

$$OE = \frac{OR}{OP} \qquad [OE = TExAE = \frac{OQ}{OP} x \frac{OR}{OQ} = \frac{OR}{OP}]$$

Note that the quantity (1 – OE) represents the overall inefficiency (or cost inefficiency in this case) of an entity. If P was producing at the cost-minimizing point Q', it would reduce its costs of production by the distance RP.

Figure 2 establishes the link between the DEA notions of efficiency and the 3-Es model of performance. In the upper part of the figure, economy and effectiveness are represented. Economy measures the quantity of physical inputs which can be acquired with the available monetary resources. Effectiveness measures the degree of achievement of objectives in terms of both outputs and outcomes. The lower part of the figure pictures the DEA notions of efficiency. Technical efficiency links physical inputs and physical outputs (and possibly outcomes)[19]. Allocative efficiency links input costs and physical inputs or output prices and physical outputs. Cost efficiency is measured by taking into account inputs cost, physical inputs and physical outputs. Revenue efficiency is measured by taking into account output prices, physical outputs and physical inputs. Finally, profit efficiency is measured by taking into account input costs, physical inputs, output prices and physical outputs.

Figure 2
Linking the DEA notions of efficiency with the 3-Es model of performance



Source: Own representation.

---

[19] Technical efficiency could also take into account the impacts. As the impacts do not appear explicitly in the 3-Es model, they are not mentioned in Figure 2.

In this thesis, technical efficiency is the efficiency notion used throughout all the essays.

Note that by defining efficiency as the relation between resources and outcomes, Knoepfel *et al.* (2011, p. 234) consider a restrictive definition of efficiency which excludes de facto the outputs and the impacts. This restrictive definition will not be retained in the current thesis.

Other authors, such as Grin and Hanhart (2003), make a distinction between the notions of internal and external efficiency. Internal efficiency corresponds to the link between inputs and outputs. External efficiency takes additionally into account the outcomes and eventually the impacts. An example of external technical efficiency using DEA is found in Soguel and Huguenin (2008). In the current thesis, external technical efficiency will be measured, as the DEA 'outputs' will in fact be impacts (pupils' performance). The term of outputs will however be used in the essays, as empirical DEA applications do not distinguish output, impact and outcome. The term of output is therefore used as a generic term including impact and potentially outcome.

Techniques to measure efficiency

Johnes (2004, p. 624) considers two basic approaches to the measurement of efficiency: the statistical approach and the non-statistical approach.

The statistical approach uses econometric techniques to measure efficiency. Efficiency corresponds to the difference between the entity's observed output and the output which would be produced if the entity was operating on the production frontier. The statistical approach assumes a specific distribution for the error term (which represents the efficiencies).

The non-statistical approach uses linear programming techniques (in the case of DEA) or mathematical algorithms (in the case of Free Disposal Hull) to compute the production frontier and to measure efficiency. The non-statistical approach does not make any assumption regarding the distribution of efficiency.

Both statistical and non-statistical approaches could be either parametric or non-parametric. However, the statistical approach is often parametric and the non-statistical approach is often non-parametric[20]. In the statistical parametric approach, the production frontier is characterized by the formulation of a function. The function depends on various parameters (linear, Cobb-Douglas, quadratic, and so on). In the non-statistical non-parametric approach, mathematical algorithms (optimized or not) are used to define the production frontier. These algorithms depend on various parameters inherent to the technique used (DEA or Free Disposal Hull for instance). No function specification has to be formulated.

Finally, both statistical and non-statistical approaches can be either deterministic or stochastic. The deterministic approach assumes that the

[20] For a detailed taxonomy of approaches, the interested reader will refer to Daraio and Simar (2007, p. 27).

differences between the observed outputs and the outputs specified by the production frontier correspond exclusively to inefficiency. The stochastic approach assumes that "deviations from the production function are a consequence not just of inefficiency, but also of measurement errors, random shocks and statistical noise" (Johnes, 2004, p. 625). The aim of the stochastic approach is therefore to separate the residual into an inefficiency component and a random component.

The three main groups of approaches are the deterministic statistical parametric methods, the stochastic statistical parametric methods and the deterministic non-statistical non-parametric methods. Figure 3 displays these three main groups of approaches and the methods that each contains. For each method, the main advantages and disadvantages are listed. A limited selection of applications in the education sector is also indicated for each method.

Figure 3
Three main approaches to measure efficiency

Approach	Methods	Selected applications in the education sector	Main advantages	Main disadvantages
Deterministic statistical parametric	Ordinary Least Squares	Smith and Street (2006) Johnes and Taylor (1990)	The significance of the frontier's parameters can be tested	Risk of mis-specification of the functional form Risk of mis-specification of the error (inefficiency) distribution
	Corrected Ordinary Least Squares	Bifulco and Bretschneider (2001) Barrow (1991)		
	Modified Ordinary Least Squares	-		
	Maximum Likelihood Estimation	Chakraborty (1998)		
Stochastic statistical parametric	Stochastic Frontier Analysis	Blank, van Hulst, Koot and van der Aa (2012) Smith and Street (2006) Stevens (2001)	The significance of the frontier's parameters can be tested Inclusion of a stochastic term	Unsuitable for application where there are multiple inputs **and** multiple outputs
Deterministic non-statistical non-parametric	Data Envelopment Analysis	Sarrico, Rosa and Coelho (2010) Joumady and Ris (2005)	Suitable for application where there are multiple inputs **and** multiple outputs Does not require specification of the functional form	Moderatly vulnerable to small sample size
	Free Disposal Hull	De Witte, Thanassoulis, Simpson, Battisti and Charlesworth-May (2010) Lavado and Cabanda (2009)		

Data Envelopment Analysis

DEA is the method of choice of the first two essays of this thesis. A comprehensive pedagogical guide about the basic models of DEA is presented in Annex 1. This guide has been published by Huguenin (2012 – in English –, 2013a – in French –) and by Huguenin (2013b) in Ishizaka and Nemery (2013). It contains:

- An intuitive introduction to the basics of DEA;
- The mathematical approach behind the two principal DEA models;
- The description of the twin software package DEAP and Win4DEAP;
- Some extensions of DEA;
- The implementation of DEA using *Microsoft Excel* ® Solver.

Readers not familiar with DEA will be wise, at this point, to read Annex 1.

As it provides valuable managerial information to decision makers, DEA is more than just a performance measurement technique. It is often considered as a decision support method. DEA covers the functions of decision-making, improvement, accountability and professionalization as defined by Nevo (2007) in the context of educational evaluation. DEA includes the following managerial information:

- By calculating an efficiency score, it indicates if an entity is efficient or has capacity for improvement;
- By setting target values for input and output, it calculates how much input must be decreased or output increased in order to become efficient[21];
- By identifying the nature of returns to scale, it indicates if an entity has to decrease or increase its scale in order to minimize the average total cost;
- By identifying a set of efficient entities (called peers or benchmarks) for each inefficient entity, it specifies which other organizations' processes need to be analyzed in order to improve its own practices.

DEA is used to measure the performance of firms or entities (called Decision-Making Units – DMUs –) which convert multiple inputs into multiple outputs. Entity efficiency is defined as the ratio of the sum of its weighted outputs to the sum of its weighted inputs (Thanassoulis *et al.*, 2008, p. 264). DEA is suitable for the use of both private sector firms and public sector organizations (and even for entities such as regions, countries, etc.). It is formulated in Charnes *et al.* (1978, 1981) in order to evaluate a U.S. federal government program in the education system called 'Program Follow Through'. Since 1978, the use of DEA has spread to other public organizations (hospitals, aged-care facilities, social service units, unemployment offices, police forces, army units, prisons, waste management services, power plants, public transportation companies, forestry companies, libraries, museums, theatres, etc.) and to the private sector (banks, insurance companies, retail stores, etc.).

Each entity's efficiency score is calculated relative to an efficiency frontier (equivalent to the production frontier of statistical approaches). Entities located on the efficiency frontier have an efficiency score of 1 (or 100%). Entities operating beneath the frontier have an efficiency score inferior to 1 (or 100%) and hence have the capacity to improve future performance. Note that no entity can be located above the efficiency frontier because they cannot have an efficiency score greater than 100%. Entities located on the frontier serve as

[21] Note that the link between performance information and organizational performance is not as obvious as the literature suggests (Kroll, 2012). Hood (2012) argues that targets and rankings (two sets of performance information provided by DEA) can be effective in some conditions but ineffective in others. This is also the case for performance data used as intelligence, i.e. "numbers as background information for choice by users or for policy change or management intervention" (Hood, 2012, p. 86). DEA efficiency scores should probably be applied as intelligence (i.e. order of magnitude) rather than targets. In this sense, they are considered more as an objective basis to hold an open discussion about the way to improve entity efficiency rather than numbers to be strictly applied. As mentioned by Jennings (2012), "intelligence might use the same numbers but do so in a more nuanced way, considering the numbers as part of a broader discussion of organizational information, purpose, and performance" (p. 1). Hood (2012) also argues that intelligence is better suited in organizations with an egalitarian culture, which seems to be the case in Swiss schools.

benchmarks – or peers – to inefficient entities. These benchmarks (i.e. real entities with real data) are associated with best practices. DEA is therefore a powerful benchmarking technique.

Critical discussion about Data Envelopment Analysis

The basic models of DEA (Charnes, Cooper & Rhodes, 1978; Banker, Charnes & Cooper, 1984) are transparent and easy to apply due to the development of user-friendly software and pedagogical guides. Nowadays, DEA is used not only by fundamental and applied researchers but also by decision makers and practitioners. As with every performance measurement technique, DEA presents advantages but also disadvantages. The current section contains a critical discussion of the main features of DEA.

Unlike statistical approaches, DEA can accommodate multiple inputs **and** multiple outputs. It is a strength in the context of the public sector where multiple non-monetary outputs are generally provided. However, the **specification** (i.e. the choice and/or the quantity) **of inputs and outputs** to be included in the analysis impacts efficiency results. The exclusion of important inputs or outputs may bias results. This is especially the case when an omitted input or output is not correlated with the variables already included in the model. Moreover, the number of inputs and outputs which can be included in the model depends on the entities' sample size. As Cooper *et al.* (2006) point out,

> if the number of DMUs (n) is less than the combined number of inputs and outputs (m + s), a large portion of the DMUs will be identified as efficient and efficiency discrimination among DMUs is questionable due to an inadequate number of degrees of freedom. (…). Hence, it is desirable that n exceeds m + s by several times. A rough rule of thumb in the envelopment model is to choose n (= the number of DMUs) equal to or greater than max {m x s, 3 x (m + s)} (p 106).

Shelton (1986, p. 82) mentions that "there is no way of assessing the relative strengths of different model specifications (…). What criteria, then, do we use to choose among alternative model specifications, to include or reject inputs or outputs from the analysis?" To answer this shortage, some authors, such as Chalos and Cherian (1995), propose a process to select the right inputs. Others, such as Smith (2005), define important modelling criteria, as practicality, parsimony, accuracy, plausibility and freedom from perverse incentives. Shelton (1986, p. 86) shows that "variable selection can be either critical or inconsequential". It is therefore recommended to test several models with alternative variable selection. This recommendation is dependent on data availability. Smith and Mayston (1987) suggest also carrying out a sensitivity analysis by including or excluding variables in order to test the robustness of DEA results.

Multicollinearity linked to large numbers of inputs or outputs has long been underresearched in DEA (Johnes, 2004, p. 643), but does not seem problematic (Smith, 2005). This is confirmed by Hansen (2008), who shows that

> multicollinearity is not an influential factor for the Data Envelopment Analysis model. This is because the nonparametric DEA model produces technical efficiency indicators for each observation based on a linear-programming maximization routine that relies on the relationship among input and output quantities and not the covariance among them (p. 88).

Endogeneity has, on the contrary, implications for DEA, even if DEA does not model the relationships between inputs and outputs. Orme and Smith (1996) and Smith (2006) discuss the possibility that the level of inputs may be endogenous when feedback happens from the outputs to the inputs allocated to the activity. They show that endogeneity is likely to generate biased efficiency results with small sample sizes. However, this bias becomes less pronounced as sample sizes increase, and it impacts more inefficient entities using low levels of the endogenous input than entities using higher level of the endogenous input[22].

Unlike statistical parametric approaches, DEA does not need the **production function** and the **distribution of inefficiencies** to be specified. As a result, it avoids the potential problems of mis-specification. But it also means that "there are no familiar parametric tests with which to check the validity of the model" (Johnes, 2004, p. 643). DEA does not account for random noise.

DEA can accommodate both **discretionary** and **non-discretionary variables** in either one-stage or two-stage models. But in the two-stage model, non-discretionary variables have to be first identified in order to be included in the second stage (letting the first stage handle exclusively discretionary variables). As Smith (2005) points out, virtually no variable is discretionary (in other words, under the control of the management) in the short run. It is therefore necessary to explicitly state the assumptions under which variables are considered within the control of the entities.

DEA is oriented toward **managerial information**. It offers decision support to decision makers (managers, politicians, etc.) by fixing input and/or output targets, by identifying the nature of returns to scale and by identifying a set of efficient peers (representing the current best practice) for each inefficient entity[23]. This is a strength compared to other performance measurement techniques. Note that DEA can be linked to multi-criteria decision analysis techniques (Feng, Lu & Bi, 2004) or used as a multi-criteria decision tool (Yilmaz & Ali Yurdusev, 2011).

---

22 Chapter 2 presents a two-stage DEA model. In the first stage, one could argue that one or several inputs are endogenous. This could be the case. But the fact that the sample contains 90 schools (in other words, it is not a small size sample) and that inefficient entities are precisely the ones which use a high level of inputs let to think that the bias linked to endogeneity (if any) is not pronounced.

23 Interestingly, note that a new journal named 'Data Envelopment Analysis and Decision Science' has been launched in 2012.

The basic models of DEA (Charnes *et al.*, 1978; Banker *et al.*, 1984) can accommodate **continuous variables** and assume that all data are strictly **positive**. But other models, as the Banker and Morey (1986a) one, can also accommodate **categorical variables**. The case of zero and negative data has to be handled carefully. As Thanassoulis, Portela and Despic (2008, p. 309) point out, "the treatment of zero data has not received as much attention perhaps as it should". Basically, **zero data** are replaced by small positive values in DEA. However, it is recommended to first identify whether the zero data is the consequence of an intentional management choice or not. If it is, it probably means that the entities, having chosen not to use a particular input or not to produce a particular output, operate with a different technology. As a result, they should be grouped in a separate set to be compared to other entities with the same technology choice. **Negative data** can be transformed a priori to positive values by adding an arbitrary large positive number to all data. However, this transformation may impact the efficiency results (Seiford & Zhu, 2002) and the identification of returns to scale (Thrall, 1996). As a result, some models have been developed in order to be applied directly to negative data. This is the case of the additive model developed by Charnes, Cooper, Golany, Seiford and Stutz (1985) and the range directional model developed by Portela, Thanassoulis and Simpson (2004).

DEA affects, for each individual entity, a set of input and output weights in order to maximize its efficiency. However, and for different reasons (e.g. the weights assigned to the variables by DEA are considered unrealistic for some entities; the management team may wish to give priority to certain variables; etc.), **preferences** about the relative importance of individual inputs and outputs can be set by the decision maker. This is done by placing weight restrictions onto outputs and inputs (also called multiplier restrictions). Cooper *et al.* (2011) and Thanassoulis *et al.* (2004) provide a review of models regarding the use of weights restrictions. In practice, placing weight restrictions is not an easy task, as decision makers have to agree about the weighting of inputs and outputs.

DEA measures **relative** efficiency. It means that the best efficiency score (1 or 100%) is allocated to the 'best-in-class' entity in the sample set. The advantage of relative efficiency is that the efficient entities are real observed entities (and not virtual ones) which provide such performance in real conditions. Their role of real benchmark cannot be denied by inefficient entities. However, for efficient entities, the basic models of DEA do not measure absolute efficiency. As a result, the entities located on the efficiency frontier have no information about a possible move beyond their 100% score. To address this issue, the concept of **super-efficiency** has been developed. It allows discrimination between efficient entities and allocating efficiency scores higher than 100% (to efficient entities). Andersen and Petersen (1993) provide the first super-efficiency model. Subsequently, other models have been developed. For a review of super-efficiency models, see Zhu (2003, pp. 197-214).

To conclude this critical discussion about the main features of DEA, the fact that efficiency results depend on technical judgements that could be contestable has to be mentioned. Smith (2005) argues that many of the judgements needed to conduct an efficiency analysis are political rather than technical issues (e.g. the choice of variables). An efficiency analysis constitutes an objective basis to hold an open discussion about the way to improve entity efficiency. But it should not be considered as the only criterion for measuring an entity's performance. The user of DEA should therefore recognize the limitations of the efficiency analysis and communicate it clearly to decision makers. And as Jesson, Mayston and Smith (1987, p. 264) point out, one has "to regard the process of performance assessment itself as a learning experience where no final solutions may be available, but some improvements are possible on what has gone before".

1.7 DEA in the education sector

Formally, DEA finds its origin in the doctoral dissertation of Rhodes (1978). It was developed in order to evaluate the 'Program Follow Through', an educational program for disadvantaged pupils in the USA. The work of Charnes *et al.* (1978) echoes the thesis of Rhodes (1978) by formulating the first DEA model, known as the constant returns to scale model.

After Rhodes (1978), Bessent and Bessent (1980) are the first to apply DEA to the education sector[24]. They evaluate the technical efficiency of 55 elementary schools using two inputs and 13 outputs. Charnes, Cooper and Rhodes (1981) publish then a DEA application to the 'Program Follow Through'. They evaluate the technical efficiency of 70 schools, 49 of them participating in the Program, using five inputs and three outputs. Following these pioneering studies, DEA has been widely applied to the education sector. Johnes (2004) provides a (non-exhaustive) review of 55 studies. For each retained study, the authors, the units analyzed, the country of the application, the outputs and the inputs involved as well as comments are provided. The majority of these studies (52) consider aggregate-level units (as universities, elementary schools, etc.) – see Korhonen, Tainio and Wallenius (2001) for an example –; the minority of these studies (3) considers individual-level units (such as graduate students from university departments) – see Johnes (2003) for an example –.

Note that the studies cited in the following part of the current section are not included in the review of Johnes (2004) – with the exception of Jesson, Mayston and Smith (1987) – or in the review presented in Chapter 2 – with the exception of Ouellette and Vierstraete (2005) –.

Three major trends are noticed in the recent use (i.e. the last 15 years) of DEA in the education sector. First, DEA is nowadays applied at all levels of the education sector, including underresearched levels such as vocational education and training. Second, the units analyzed by DEA tend to spread from 'classical' aggregate-level units, such as schools, to underresearched units, covering both aggregate- or individual-level units. Third, innovative DEA applications are

[24] Note that an earlier version of this study has been published by Bessent and Bessent (1979).

implemented. These three major trends are illustrated with various examples hereafter.

In the first trend, one can notice that DEA applications are, nowadays, implemented at all levels of the education sector, with an emphasis on tertiary education. These levels include:

- Primary education, as in Hu, Zhang and Liang (2009) – 58 primary schools in Beijing, China –.
- Lower secondary education, as in Sillah (2012) – 46 secondary schools, Gambia –; Demir and Depren (2010) – 33 secondary schools, Turkey –; Sarrico and Rosa (2009) – 51 Portuguese secondary schools –.
- Obligatory education (primary and lower secondary education), as in Garrett and Kwak (2011) – 447 public school districts in Missouri, USA –; Mizala, Romaguera and Farren (2002) – 2000 obligatory schools, Chile –.
- Upper secondary education, as in Essid, Ouellette and Vigeant (2010; 2013) – 75 high schools, Tunisia –.
- Vocational education, as in Abbott and Doucouliagos (2000) – 23 polytechnics, New Zealand –.
- Tertiary education, as in Hirao (2012) – 50 business schools, USA –; Katharaki and Katharakis (2010) – 20 Greek universities –; Kempkes and Pohl (2010) – 72 German universities –; Johnes and Yu (2008) – 109 Chinese universities –; Agasisti and Salerno (2007) – 52 Italian universities –; Johnes (2006a) – 109 higher education institutions, England –; Martin (2006) – 52 departments within the University of Zaragoza –; Joumady and Ris (2005) – 209 higher education institutions in eight European countries –; Ng and Li (2000) – 84 Chinese higher education institutions –.

In the second trend, the units analyzed by DEA tend to spread from 'classic' aggregate-level units, such as schools, to underresearched units, both at aggregate-level (such as vocational schools, offices of technology transfer, MBA, school boards, local education authorities, etc.) and at individual-level (such as economics graduates) . The implementation of cost efficiency studies (and not mainly technical and scale efficiency studies) are also of note. A few examples are described hereafter:

- Abbott and Doucouliagos (2000) assess the technical and scale efficiency of 23 polytechnics in New Zealand. Note that DEA is rarely implemented in vocational education. The authors use one output (number of full time students) and three inputs (full time equivalent – FTE – teaching staff; FTE non-teaching staff; value of fixed assets). The authors find that the mean technical efficiency and the mean scale efficiency are equal to 0.895 and 0.934 respectively in 1996. These results show that there is a potential of both organizational and scale economies in polytechnics.
- Anderson, Daim and Lavoie (2007) measure the technical efficiency of 54 US university offices of technology transfer. The role of such offices is to manage and transfer research results into other sectors. The authors use five outputs (licensing income (in USD); licenses and options executed; start-up

companies; US patent filed; US patent issued) and one input (total research spending, in USD). Note that weight restrictions are applied on certain outputs in this study. An interesting finding is that the ranking of the technology transfer offices based on DEA differs from the traditional ranking based on licensing income.

- Barnett, Glass, Snowdon and Stringer (2002) assess the cost efficiency of 152 secondary schools in Northern Ireland. They use four outputs (% of students gaining five or more GCSEs at grades A*-C; % of students gaining at most four GCSEs at grades A*-C; % of students gaining five or more GCSEs at grades A*-G; % of students gaining at most four GCSEs at grades A*-G) and five inputs (FTE teachers; average annual salary for these teachers; annual expenditure on teaching staff; annual expenditure on support staff; annual expenditure on other (non-labour) input resources). Unsurprisingly, the results show that cost efficiency is positively related to school size. This finding raises funding issues.
- Colbert, Levary and Shaner (2000) measure the technical efficiency of 24 US MBA and three foreign MBA programs. They use eight outputs (% of alumni who donate money to the program; student satisfaction with teaching; student satisfaction with curriculum; student satisfaction with placement; average salary of graduates; recruiter satisfaction with analytical skills; recruiter satisfaction with team work skills; recruiter satisfaction with graduates' global view) and five inputs (number of faculty; number of students; faculty to student ratio; average GMAT score; number of electives). Several models are run. Based on the results, a new ranking of MBA programs is provided. The authors argue that DEA makes it possible to more fairly compare specific programs.

  Note that Ray and Jeon (2008) conduct another DEA analysis on 65 MBA programs in the USA.
- Although the study of Jesson, Mayston and Smith (1987) is not a recent one, it is of interest as it assesses the technical efficiency of 96 English local education authorities. The authors use two outputs (% of children getting five or more 'O' level (or CSE grade 1) passes; % of children getting three or more graded passes at CSE or 'O' level) and four inputs (% of children in the authority's catchment area whose head is a non-manual worker; % of children not from one-parent families; % of children born in the UK, Ireland, USA or the Old Commonwealth or in households whose parents were born in the UK, Ireland, USA or the Old Commonwealth; secondary school expenditure per pupil). As the results of the study concerning local authorities, they insist on the adequacy of the data used in the DEA model. The authors argue that a framework of analysis has to be established. This framework has to be backed by local authorities[25].

[25] This conclusion echoes the one in the current thesis. As already mentioned, efficiency is one out of three criteria to evaluate the educational system in Switzerland. However, no framework to measure efficiency has yet been provided. Such a framework, such as the one provided in the second essay of this thesis, has to be backed by federal, cantonal and local authorities.

- Johnes (2006b) is a typical individual-level study. It measures teaching efficiency by using the data of 2547 economics graduates. Two outputs represent the graduates' degree results. Three inputs are used: academic ability on arrival at university; gender of graduates; types of schools attended by graduates before entry to university. Several models are run. Johnes (2006b) also assesses the technical efficiency of the 35 departments in which the students are by using aggregate data. She compares the results obtained with the individual-level and the aggregate-level analysis. In doing so, she disentangles the effect of the individual and the effect of the department. The results "suggest that aggregate level DEAs provide efficiency scores which reflect the efforts and characteristics of the students as well as those of the department or institution to which they belong" (p. 453). In other words, a department effect is identified.
- Ouellette and Vierstraete (2005) assess the technical and the allocative efficiency of 142 school boards in Québec, Canada. They use two outputs (FTE pupils in primary and in secondary schools) and 16 inputs (teaching staff (expenditures, quantities, prices); other staff (expenditures, quantities, prices); supplies and materials (expenditures, quantities, prices); energy (expenditures, quantities, prices); other (expenditures, quantities, prices); capital (in square meters). The main finding is that school boards' inefficiency costs 800 million dollars of which 200 million dollars is attributable to unfavorable socio-economic conditions.
- Singh, Rylander and Mims (2012) assess the technical efficiency of online versus offline learning. It is another example of individual-level analysis as it considers 26 students taking an offline course and 44 students taking the online version of the same course. The authors use one input (student's effort level, measured by the number of hours in a week the student spends studying for the course) and three outputs (quantitative scores achieved by the student at the end of the course; the student's viewpoint of how much he/she learned in the course; the student's level of satisfaction with the course). The results show that 56% of the online students are efficient (100%) compared to 38% of the offline students.

The use of DEA appears unlimited in the third trend. It opens the way to innovative and creative applications. Seven examples are presented hereafter:

- De Witte and Van Klaveren (2014) estimate which configuration of teaching activities (divided into homework, lecturing, problems with guidance, problems without guidance, revision, tests and quizzes, classroom management, other activities; all of them being considered as inputs) maximizes the performance of students in mathematics (output). 1790 Dutch students are included in this individual-level analysis. The authors show that high test scores are related to teaching styles that emphasize problem solving (with and without guidance) and homework.
- Fandel (2007) uses DEA to analyze the extent to which the redistribution of funds for teaching and research among the universities of North Rhine-Westphalia (Germany) – mainly based on a negotiation process – is justified by the relative efficiency results of these universities. He argues

that *(1)* universities with higher efficiency scores should no longer lose resources through the redistribution process and *(2)* universities with lower efficiency scores should not receive more than universities with higher efficiency scores through the redistribution process.

- Portela, Camanho and Borges (2011) have implemented a web-based benchmarking platform for the Portuguese secondary schools. This platform integrates DEA. Headteachers have open-access to the platform. They can conduct their own efficiency analysis by selecting the relevant inputs and outputs (for them).
- Ruggiero, Miner and Blanchard (2002) use DEA in the context of school finance equity. They argue that equity analyses based on unadjusted expenditure per pupil fail to recognize that school districts are confronted by different cost environments. As a result, the efficiency with which educational services are provided may not be the same. They apply DEA to adjust expenditures for cost inefficiency.
- In another individual-level units analysis, Vierstraete and Yergeau (2012) assess the technical efficiency of 583 bachelor students. The objective of this study is to identify which method of financing studies (among loans and bursaries from the government, student aid granted directly by universities, scholarships or on-campus jobs, off-campus jobs or parental financial contribution) is linked to the efficient student. The authors use two outputs (accumulated credits, grade point average) and one input (global financial resources). They show that students with a paid job held throughout the year are the most inefficient.
- Chen and Chen (2011) assess the technical efficiency of 99 Taiwanese universities. The originality of this study is the link realized between a total quality performance system called Inno-Qual Performance System – IQPS – used by the universities and DEA. Basically, the authors create five critical indices based *(1)* on the various indicators included in IQPS and *(2)* on qualitative discussion with 29 senior experts. These indices are the outputs: journal articles accepted and published, research patents, financial support from the National Science Council, number of cooperating international universities, promotion and job acquisition for all previous students. Three inputs are used (number of domestic students, number of international members, number of domestic FTE faculty).

The recent trends in the use of DEA do not include the comparison of alternative DEA models, although a few recent studies tend to be interested in this issue (see the second essay in Chapter 3 about it). In this sense, the second essay of this thesis fills a gap and paves the way to further study.

1.8 Applications of DEA in Switzerland

As far as the author is aware, the first use of DEA applied to a Swiss empirical case appears in 2003 (Steinmann & Zweifel, 2003). It follows the use of another non-parametric method, Free Disposal Hull, by Burgat and Jeanrenaud (1990, 1992, 1994). In total, 13 published applications of DEA in Switzerland are identified. Three of them present methodological concerns (Olivares & Schenker-Wicki, 2010; Meunier, 2008; Diagne, 2006). Their results are likely to be invalid. These 13 studies are presented hereafter.

Olivares and Schenker-Wicki (2012)

Olivares and Schenker-Wicki (2012) assess the performance of 10 Swiss and 77 German universities using time series data (2001-2007). This study is the only one using the Malmquist productivity index developed as an extension of DEA to measure efficiency variation over time[26]. It takes into account three inputs (FTE academics; FTE non-academics; operating expenses) and two outputs (students; third-party funds). The main results show that the 'total factor productivity' calculated by the Malmquist index grows by an annual rate of 4% in Switzerland and in Germany. However, the Swiss 'total factor productivity' results entirely from improvements in technical and scale efficiency: a 'catch-up' effect is observed. A slightly negative technology progress is recorded (minus 0.87% on annual rate). This means that the efficiency frontier has shifted downwards in Switzerland. The authors explain this negative 'frontier-shift' by the fact that "the internal organisation of the universities changed substantially, including the skill requirements for management and employees. In other words, such considerable organisational changes do take time and resources in order to reorganise management and the workplace" (p. 32). In Germany, improvements are due to a mix of technical efficiency, scale efficiency and technology progress.

Solaux, Huguenin, Payet and Ramirez (2011)

Solaux *et al.* (2011) use DEA in order to evaluate the scale efficiency of 90 primary schools in the State of Geneva, Switzerland. They consider two inputs (FTE teachers; FTE administrative staff) and four outputs (number of pupils; pupils' performance in French; pupils' performance in German; pupils' performance in mathematics). In order to avoid the ratio form of the pupils' performance outputs, which would invalidate the scale efficiency analysis, Solaux *et al.* (2011) multiply the pupil's performance by the number of pupils. The main results show that the mean scale efficiency is equal to 98%.

Olivares and Schenker-Wicki (2010)

Olivares and Schenker-Wicki (2010) assess the performance of 12 Swiss universities using time series data (1999-2008) and a two-stage model. In the

---

[26] The Malmquist index calculates a 'total factor productivity' which is deconstructed into two components. The first one is called 'catch-up'. This captures the change in technical efficiency over time. The second one is called 'frontier-shift'. This captures the change in technology which occurs over time (i.e. the movement of efficiency frontiers over time). See Annex 1 to learn more about it.

first stage, three inputs (total amount of non-personnel expenditures in CHF; FTE scientific personnel; FTE non-scientific personnel) and three outputs (undergraduate students; postgraduate students; third-party funds in CHF) are used. The authors run a single 'pooled' DEA model containing the data of all universities over time, assuming each year to be an independent observation. In other words, they apply a cross-sectional data model. This choice is questionable, as time series data should be handled with appropriate methods in DEA (windows analysis or Malmquist productivity index). In a second stage, the technical efficiency scores are regressed over 12 non-discretionary variables. Four variables are significant at the 1% level: student-faculty ratio (with a coefficient value of 0); proportion of professor per scientific personnel (with a coefficient value of 1.13); years of studying (with a coefficient value of -0.1); number of students (with a coefficient value of 0).

Schoenenberger, Mack and von Gunten (2009)

Schoenenberger, Mack and von Gunten (2009) apply DEA to 300 Swiss logging companies using time series data (1998-2003). Each year is treated separately. The technical efficiency is then compared over time. Note that it would have been wise to use windows analysis or the Malmquist index. The authors use one output (annual timber in m^3) and four inputs (number of hours worked by staff in the production of wood; number of machine hours performed by all vehicles in the production of wood; administrative costs incurred by the production of wood (in CHF); third-party services (in CHF) in the production of wood). The main result is that 43% of logging companies have technical efficiency scores lower than 50% over time. In a second stage, the technical efficiency scores are regressed over 12 non-discretionary variables[27]. Only one non-discretionary variable is significant at a level of 10, 5 or 1% over at least five out of six years: public subsidies. The authors show that, for instance in 1998, 1000 additional Swiss francs in public subsidies lower the technical efficiency of logging companies by 0.19%.

Soguel and Huguenin (2008)

Soguel and Huguenin (2008) assess the performance of 12 regional social centres in the State of Vaud, Switzerland. They consider two inputs (FTE staff; area in m^2) and two outcomes (number of benefits recipients and average length of recipients' support). The results show that the regional social centres have, on average, a variable return to scale technical efficiency of 94% and a scale efficiency of 96%. Seven out of 12 centres have not yet reached their optimal scale. Two of them operate in a situation of decreasing returns to scale; five of them operate in a situation of increasing returns to scale.

[27] To avoid the problem of technical scores truncated at one, note that the authors actually regress the super-efficiency scores and not the efficiency scores.

Widmer and Zweifel (2008)

Widmer and Zweifel (2008) assess the performance of the 26 Swiss states (or Cantons) using time series data (2000-2004). They construct a 'total public sector performance indicator' which aggregates the provision of eight local government activities. This indicator is used as the output. They considered the real expenditure (in CHF) as the single input. The results, for instance in 2004, show a mean technical efficiency of 85%. A single state is fully efficient: Thurgovia. In a second stage, Widmer and Zweifel (2008) perform a Tobit regression in order to explain the efficiency scores. They use 17 non-discretionary variables. Eight of them are significant at the 90% (or higher) level. The main finding is that the Swiss fiscal equalization program (measured by federal subsidies per capita and the index of financial potential) has a negative effect on the states' efficiency.

Meunier (2008)

Meunier (2008) measures the performance of 156 secondary schools located in 22 states in 2000. She uses two outputs (score for reading in the PISA 2000 test, aggregated by school; inverse standard deviation (by school) of the reading score) and four inputs (number of teachers per pupil; number of hours of supervision per year; number of teachers per pupil having a teaching diploma; number of computers). Meunier (2008) considers both the constant and the variable returns to scale assumptions. Note that the constant returns to scale assumption should not have been applied in this case due to the ratio form of some of the variables (Hollingsworth & Smith, 2003). Note also that the consideration of the inverse standard deviation of the reading score as an output is problematic. For instance, a school with the highest inverse standard deviation (meaning that the pupils' scores are homogenous) and a low reading scores average would appear as fully efficient. This would mean that a school where all pupils fail the reading test could appear as fully efficient. Results show that under the VRS assumption, the mean technical efficiency is equal to 0.8348. 24 schools out of 156 are fully efficient.

Diagne (2006)

Diagne (2006) measures the performance of 27 high schools located in six states in 1999. He considers one output (success rate at the bachelor exams) and four inputs (FTE teachers; % of teachers having more than 10 years of teaching experience; % of teachers with a master or a PhD degree; % of teachers with a permanent working contract). Diagne (2006) performs a constant returns to scale DEA model. Due to the ratio form of several variables, this model is unfortunately inappropriate (Hollingsworth & Smith, 2003). The results are invalid. Diagne (2006) tests several models with different combinations of inputs. In a second stage, the technical efficiency scores are regressed over non-discretionary variables, using both a Tobit and an OLS regression[28]. Keeping in mind that the efficiency scores are flawed, Diagne (2006) shows that, in the first DEA model tested, the % of teachers with a master or a PhD degree is positively related to efficiency (at the 5% level) and that the % of teachers with a permanent working contract is negatively related to efficiency (at the 5% level).

Jeanrenaud and Vuilleumier (2006)

Jeanrenaud and Vuilleumier (2006) assess the performance of 28 Swiss consular posts in Europe, North America and Asia. They consider two inputs (FTE rotational employees; FTE locally-hired employees) and one output (number of visa equivalents). Total efficiency (under the constant returns to scale assumption), pure efficiency (under the variable constant to scale assumption) and scale efficiency are equal to 0.66, 0.76 and 0.88 respectively.

Schenker-Wicki and Hürlimann (2006)

Schenker-Wicki and Hürlimann (2006) measure the performance of 10 Swiss universities using time series data (2000-2003). They consider two outputs (number of diplomas; number of theses) and two inputs (number of students; scientific staff expenses in CHF). The authors run a separate model for each year. The results show that, for instance in 2003, the mean technical efficiency scores is equal to 0.9686, with seven universities (out of 10) being fully efficient. Such a high number of efficient entities in these results leads to assume that the number of variables considered in this study (four) is probably too high for the number of entities assessed (10).

Ferro-Luzzi, Flueckiger, Ramirez and Vassiliev (2006)

Ferro-Luzzi, Flueckiger, Ramirez and Vassiliev (2006) measure the technical efficiency of 132 regional employment offices in 1999. They consider one output (number of hires) and five inputs (number of entries into long-term unemployment; number of unemployment insurance benefit exhaustees; number of re-entries into unemployment four months after having found a job; number of regional employment offices job counsellors; number of registered job-seekers with unemployment insurance benefit entitlement). The mean efficiency score is equal to 0.8459, with 15 offices being fully efficient. In a second stage, the authors identify the determinants of the efficiency of regional employment offices using an OLS regression. Six non-discretionary variables are included in the model. All of them are statistically significant at the 5 or 1% level.

---

[28] Note that the author also runs a one-stage DEA model which includes environmental variables. Unfortunately, as the author does not use the appropriate model, which would have been the Banker and Morey (1986b) model, the results are invalid.

Farsi and Filippini (2005)

Farsi and Filippini (2005) benchmark 52 electricity distribution utilities operating in Switzerland in 1994 using three different methods: corrected ordinary least squares, stochastic frontier analysis and DEA. They consider three inputs (FTE employees; capital stock (installed capacity of the transformers); amount of input energy) and four outputs (annual output in GigaWh; number of customers; load factor (ratio of utility's average load on its peak load); service area in km^2). The mean cost efficiency score estimated by DEA is equal to 0.917. The main results indicate that considerable differences exist in both efficiency scores and ranks across parametric, stochastic and non-parametric methods.

Steinmann and Zweifel (2003)

Steinmann and Zweifel (2003) measure the technical efficiency of 89 Swiss hospitals covering the years 1993-1996. They consider three labour inputs (academic, nursing and administrative staff) and one financial input (non-labour expenses) and five outputs (medical, pediatric, surgical, gynaecological and intensive care discharges). The total of inpatient days is viewed either as an input or an output in two models. The authors do not provide the efficiency scores, but indicate how many hospitals are efficient, or respectively inefficient. About 10% of hospitals are considered as technically efficient. An econometric second stage analysis suggests that subsidies increase inefficiency. The share of junior physicians also enhances efficiency as long as this share does not exceed 12%.

Among the 13 studies reviewed above, six of them focus on school efficiency. Among these six studies, three of them present methodological concerns (Olivares & Schenker-Wicki, 2010; Meunier, 2008; Diagne, 2006). Their results are likely to be invalid. Among the three remaining studies devoted to education, only one (Solaux *et al.*, 2011) is devoted to primary schools. But it only measures scale efficiency, and not technical efficiency. The first essay of the current thesis is, as a result, the first study to measure technical efficiency of primary schools in Switzerland. It is also the fourth (valid) study to use DEA in order to measure efficiency in the Swiss education sector. In this sense, and as the second essay of this thesis, the first essay fills a gap and paves the way to further study.

1.9 Main findings and perspectives

First essay

Determinants of school efficiency: the case of primary schools in the State of Geneva, Switzerland

The public primary school system in the State of Geneva, Switzerland, is characterized by centrally evaluated pupils' performance with the use of standardized tests. As a result, consistent data are collected by the system. The 2010-2011 dataset is used in a two-stage Data Envelopment Analysis (DEA) of school efficiency. At the first stage, DEA is employed to calculate an individual efficiency score for each school[29]. It shows that, on average, each school could reduce its inputs by 7% and still provide the same quality of pupils' performance[30]. At the second stage, efficiency is regressed on school characteristics and environmental variables, none of them being within the control of headteachers[31]. The model is tested for multicollinearity, heteroskedasticity and endogeneity. Four variables are identified as statistically significant.

As school size positively influences efficiency, the proportion of disadvantaged pupils, the provision of special education and the fact of operating on several sites negatively influence efficiency[32]. Although these variables are not under the control of headteachers, it does not mean that nothing can be done to either boost their positive impact or curb their negative impact. Actions can be taken at the State, school and class level.

A likely way to solve the negative influence of multi-site entities would consist of improving the use of ICT in school management and teaching. Selwood and Visscher (2008) advocate the use of school information systems for enhancing school improvement. The use of ICT could also be used for distance learning and distance management. Planning new schools should also consider the advantages of being located on a unique site, which allows reaching a critical size in terms of pupils and teachers.

Positive discrimination is often advocated to correct the negative influence of disadvantaged pupils on school performance. It generally results in allocating more resources to disadvantaged schools. Unfortunately, positive discrimination does not seem to improve pupils' performance neither in Europe (Demeuse, Frandji, Greger & Rochex, 2008) nor in the State of Geneva (Souci & Nidegger, 2010). The impact of positive discrimination on school efficiency is therefore negative: inputs increase without any output improvement. As a result, other actions need to be taken in order to correct the

---

29 Three output variables and three input variables are used in the first stage. The output variables consist of standardized test scores. The input variables include the number of full-time equivalent (FTE) teachers, the number of FTE administrative staff and the school budget in Swiss francs (excluding staff salaries and capital expenditure).

30 Note that DEA measures relative efficiency and not absolute efficiency. A score of 1 (or 100%) means that an entity is the best-in-class. However, it is possible that such an entity could still improve its efficiency. The use of DEA super-efficiency models helps to discriminate among efficient entities by allocating scores higher than 1 (or 100%) to efficient entities. In this sense, the scores generated by super-efficiency models better approximate the notion of absolute efficiency.

31 Ordinay Least Squares (OLS) regression is the method of choice in this study. In addition to OLS, Tobit regression has also been run. Note that the results of the Tobit regression are in line with those of the OLS regression.

32 Unfortunately, no instrumental variable has been identified to test for endogeneity the fact that operations are held on multiple sites. The impact of this latter variable has, therefore, to be taken with cautious. However, it is unlikely that it is endogenous (see Section 2.4.4 for a discussion).

negative influence of disadvantaged socioeconomic status on school performance.

In order to define these actions, one has to identify the social-class differences which explain why disadvantaged children underperform. Through a review of the literature, Rothstein (2010) sums up these differences. Basically, he demonstrates that childrearing and literary practices, health characteristics, housing stability and economic security influence pupils' achievements. Children with low socioeconomic status are disadvantaged in all these areas.

Rather than allocating more resources to schools, policy makers should therefore focus on related social policies. For instance, they could define pre-school, family, health, housing and benefits policies in order to improve disadvantaged children's conditions.

Special education is mainly provided separately, meaning that pupils with special needs are grouped into specific classes. In the State of Geneva, a new law ruling the integration of children and young people with special needs or disability came into force in 2010. It states that integrative solutions are preferred to separative solutions. A move towards pupils with special needs or disabilities integrated into regular classes could increase school efficiency, although this assumption remains to be tested.

Increasing the number of pupils is associated with higher efficiency. Such a finding could suggest that schools are evolving in a situation of increasing returns to scale. The DEA model performed in the first stage does not allow the study to confirm or deny this assumption. This is due to the ratio formulation of the inputs and the outputs which prevents the calculation of scale efficiency (Hollingsworth and Smith, 2003). However, Leithwood and Jantzi (2009) show that primary schools serving socially heterogeneous pupils should be limited in size to not more than 500 pupils in order to maximise efficiency. In the State of Geneva, the average school has 381 pupils, leaving room for improvement (and larger schools). In terms of policy making, the existing – and rigid – class size regulation could be replaced by a more flexible one, allowing headteachers to increase the total number of pupils by increasing class size.

Finally, further analysis could be conducted in order to measure scale efficiency. It should especially determine if there is a size at which school efficiency starts to decline (rather than continuing to increase). Depending on the results of this analysis, a reflection about merging schools facing increasing returns to scale, splitting schools facing decreasing returns to scale, or modeling catchment areas should be undertaken. Further analysis could also consider the possibility of not aggregating the test scores in mathematics, French and German by school year. Instead, the output variables could be disentangled or re-aggregated by topic over time. Finally, follow-up research could investigate the sources of multi-site schools' inefficiency. This research could aim, for instance, to disentangle the pedagogical inefficiency from the organizational inefficiency.

Second essay

DEA does not like positive discrimination: a comparison of alternative models based on empirical data

Due to the existence of free software and pedagogical guides, the use of Data Envelopment Analysis (DEA) has been further democratized in recent years. Nowadays, it is quite usual for practitioners and decision makers with little or no knowledge in operational research to run their own efficiency analyses. Within DEA, several alternative models allow for an environmental adjustment. Five alternative models, each of them easily accessible and achievable by practitioners and decision makers, are performed using the empirical case of the 90 primary schools of the State of Geneva, Switzerland. As the State of Geneva practices an upstream positive discrimination policy towards schools, this empirical case is particularly appropriate for an environment adjustment. The majority of alternative DEA models deliver divergent results.

Applied DEA studies traditionally end with recommendations and policy implications. Most of these studies base their recommendations on the efficiency results produced by one DEA model. This appears to be problematic. As shown in this study, several alternative models measure efficiency within DEA, delivering diverging results. As a result, recommendations and policy implications may differ according to the model used. From a political standpoint, these diverging results could lead to potentially ineffective decisions. From an applied research standpoint, they should represent a serious matter of concern. And from a decision making standpoint, they may lead to opposite managerial options.

Basically, there is no consensus on the best model to use (Cordero-Ferrara *et al.*, 2008). Echoing Smith and Mayston (1987), the choice of models is ultimately a political judgement. Practitioners and decision makers have to select the model which is right for them, in other words, the model which reflects best their own preferences. In this sense, the application of an appropriate multi-criteria decision analysis method to help them select the right model should be investigated in further studies.

Third essay

DEA and non-discretionary inputs: how to select the right model (for you) using multi-criteria decision analysis

Within Data Envelopment Analysis (DEA), several alternative models allow for an environmental adjustment. The majority of them deliver divergent results. From a practical standpoint, but also from a political perspective, decision makers (i.e. top civil servants and ministers) face the difficult task of selecting the most suitable model. This study is performed to overcome this difficulty. By doing so, it fills a research gap.

First, a two-step web-based survey is conducted. In the first step, the survey aims to collect general views from DEA scholars and practitioners to identify the selection criteria. In the second step, the survey aims to prioritize and weight the selection criteria identified in the first step with respect to the goal of selecting the most suitable model. But it also aims to collect the preferences of the respondents about which model is preferable to fulfil the selection criteria.

Second, Analytic Hierarchy Process, a multi-criteria decision analysis method, is used to quantify the preferences expressed in the survey. Results show that the understandability, the applicability and the acceptability of the alternative models are valid selection criteria. When results are aggregated over the respondents, the categorical model developed by Banker and Morey (1986a) emerges as the most suitable model. However, individual results may vary and other models may be identified as the most suitable ones from an individual perspective.

As a weakness of the categorical model is to lessen the discriminating power of DEA, the second most suitable model identified, the two-stage Ray (1991) model, is probably the best option for all situations.

In terms of policy and managerial implications, the results of the current study suggest that:

- The number of selection criteria and alternatives should remain parsimonious in order to avoid the time consuming process of AHP.
- The selection criteria should be backed by the literature or by an expert group. They should not be oriented towards the results in order to avoid opportunistic behavior.

Once the most suitable DEA model is identified, the principles of permanence of methods and of consistency should prevail.

Conclusion and perspectives

As an opening statement in the current thesis, it was mentioned that measuring school efficiency was a challenging task. The same statement still holds as a conclusion statement.

First, a performance measurement technique has to be selected, knowing that alternative techniques could lead to diverging results. Within DEA, one such technique, alternative models have been developed in order to deal with environmental variables. The majority of these models also lead to diverging results. Second, the choice of input and output variables to be included in the efficiency analysis is often dictated by data availability. The choice of the variables remains an issue even when data is available. As a result, the choice of technique, model and variables is probably, and ultimately, a political judgement.

However, conducting an efficiency analysis, even if this analysis is imperfect, allows decision makers to hold an open discussion about the way to improve entities' efficiency. In this way, the results of an efficiency analysis are, in themselves, not the most important part of the process. They represent rather a means by which to reach an objective of continuous improvement within the organizations. Efficiency scores should therefore be interpreted more broadly as orders of magnitude. In all cases, an efficiency analysis represents a step towards evidence-based management or policy. It allows walking away from subjective points of view often expressed by decision makers.

Within DEA, the two-stage model developed by Ray (1991) is probably the most convincing model which allows for an environmental adjustment. In this model, an efficiency analysis is conducted with DEA (first stage) followed by an econometric analysis (second stage) to explain the efficiency scores. An extensive review of the literature in the education sector shows that the econometric models developed in the second stage are seldom tested for multicollinearity, heteroskedasticity and endogeneity. This is a source of concern, as these are classic requirements in such analysis. For instance, the results produced by a model suffering from endogeneity are simply inapplicable if not corrected. Conducting the basic tests of the second-stage model should become the standard in further studies.

An environmental variable of particular interest, tested in this thesis, consists of the fact that operations are held, for certain schools, on multiple sites. Multi-site entities exist in the education sector as well as in other domains, such as healthcare. More multi-site schools are currently being created in Switzerland due to a process of school mergers. Results show that the fact of being located on more than one site has a negative influence on efficiency. More studies are needed to confirm and to interpret this finding. A likely way to solve this negative influence would consist of improving the use of ICT in school management and teaching. Planning new schools should also consider the advantages of being located on a unique site, which allows reaching a critical size in terms of pupils and teachers.

The fact that underprivileged pupils perform worse than privileged pupils has been public knowledge since Coleman *et al.* (1966). As a result, underprivileged pupils have a negative influence on school efficiency. This is confirmed by this thesis for the first time in Switzerland. Several countries have developed priority education policies in order to compensate for the negative impact of disadvantaged socioeconomic status on school performance. These policies have failed. As a result, other actions need to be taken.

In order to define these actions, one has to identify the social-class differences which explain why disadvantaged children underperform. Childrearing and literary practices, health characteristics, housing stability and economic security influence pupil achievement. Rather than allocating more resources to schools, policymakers should therefore focus on related social policies. For instance, they could define pre-school, family, health, housing and benefits policies in order to improve the conditions for disadvantaged children.

Multi-criteria decision analysis methods can help the decision makers to select the most suitable model (i.e. the model which suits best their own preferences). The number of selection criteria should remain parsimonious and not be oriented towards the results of the models in order to avoid opportunistic behaviour. The selection criteria should also be backed by the literature or by an expert group. Once the most suitable model is identified, the principle of permanence of methods should be applied in order to avoid a change of practices over time.

References

Abbott, M. & Doucouliagos, C. (2000). Technical and scale efficiency of vocational education and training institutions: The case of the New Zealand polytechnics. *New Zealand Economic Paper*, *34*(1), 1-23.

Agasisti, T. & Salerno, C. (2007). Assessing the Cost Efficiency of Italian Universities. *Education Economics*, *15*(4), 455-471.

Andersen, P. & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, *39*(10), 1261-1264.

Anderson, T. R., Daim, T. U. & Lavoie, F. F. (2007). Measuring the efficiency of university technology transfer. *Technovation*, *27*(5), 306-318.

Andrews, R., Boyne, G. A. & Enticott, G. (2006). Performance failure in the public sector. Misfortune or mismanagement? *Public Management Review*, *8*(2), 273-296.

Audit Commission (2010). The practice of performance indicators. *Management paper*.

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.

Banker, R. D. & Morey, R. C. (1986a). The Use of Categorical Variables in Data Envelopment Analysis. *Management Science*, *34*(4), 1613-1627.

Banker, R. D. & Morey, R. C. (1986b). Efficiency Analysis for Exogenously Fixed Inputs and Outputs. *Operations Research*, *32*(12), 513-521.

Barnett, R. R., Glass, J. C., Snowdon, R. I. & Stringer, K. S. (2002). Size, Performance and Effectiveness: Cost-Constrained Measures of Best-Practice Performance and Secondary-School Size. *Education Economics*, *10*(3), 291-311.

Barrow, M. M. (1991). Measuring local education authority performance: A frontier approach. *Economics of Education Review*, *10*(1), 19-27.

Bessent, A. M. & Bessent, E. W. (1979). *Determining the Comparative Efficiency of Schools through Data Envelopment Analysis* (Research Report CCS 361). Austin: Center for Cybernetic Studies, University of Texas.

Bessent, A. M. & Bessent, E. W. (1980). Determining the Comparative Efficiency of Schools through Data Envelopment Analysis. *Educational Administration Quarterly*, *16*(2), 57-75.

Bifulco, R. & Bretschneider, S. (2001). Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review*, *20*(5), 417-429.

Blank, J. L. T., van Hulst, B. L., Koot, P. M. & van der Aa, R. (2012). Benchmarking overhead in education: a theoretical and empirical approach. *Benchmarking: An International Journal*, *19*(2), 239-254.

Bouckaert, G. & Halligan, J. (2008). *Managing Performance: International Comparisons*. London: Routledge.

Burgat, P. & Jeanrenaud, C. (1990). Mesure de l'efficacité productive et de l'efficacité-coût : le cas des déchets ménagers en Suisse. *Working Paper No. 9002*. Neuchâtel: Institut de recherches économiques et régionales, Université de Neuchâtel.

Burgat, P. & Jeanrenaud, C. (1992). Measurement of productive efficiency: the example of household waste in Switzerland. *Working Paper No. 9209*. Neuchâtel : Institut de recherches économiques et régionales, Université de Neuchâtel.

Burgat, P. & Jeanrenaud, C. (1994). Technical Efficiency and Institutional Variables. *Swiss Journal of Economics and Statistics*, *130*(4), 709-717.

Chakraborty, K. (1998). *Essays on Scale Economies and Efficiency in Public Education* (Unpublished doctoral dissertation). Utah State University, USA.

Chalos, P. & Cherian, J. (1995). An application of data envelopment analysis to public sector performance measurement and accountability. *Journal of Accounting and Public Policy*, *14*(2), 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. L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429-444.

Charnes, A., Cooper, W. W. & Rhodes, E. L. (1981). Evaluating program and managerial efficiency: An application of DEA to program follow through. *Management Science*, *27*(6), 668-697.

Chen, J.-K. & Chen, I.-S. (2011). Inno-Qual efficiency of higher education: Empirical testing using data envelopment analysis. *Expert Systems with Applications*, *38*(3), 1823-1834.

Coelli, T. J., Prasada Rao, D. S., O'Donnel, C. J. & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis.* New York: Springer.

Colbert, A., Levary, R. R. & Shaner, M. C. (2000). Determining the relative efficiency of MBA programs using DEA. *European Journal of Operational Research*, *125*(3), 656-669.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, M. D & York, R. L. (1966). *Equality of Educational Opportunity.* Washington, D.C.: US Department of Health, Education, and Welfare, Government Printing Office.

Cooper, W. W., Ruiz, J. L. & Sirvent, I. (2011). Choices and Uses of DEA Weights. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 93-126). New York: Springer.

Cooper, W. W., Seiford, L. M. & Zhu, J. (2004). Data Envelopment Analysis: History, Models and Interpretations. In. W. W. Cooper, L. M. Seiford and J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 1-39). Boston: Kluwer Academic Publishers.

Cordero, J. M., Pedraja, F. & Santín, D. (2009). Alternative approaches to include exogenous variables in DEA measures: A comparison using Monte Carlo. *Computers & Operations Research*, *36*(10), 2699-2706.

Cour des comptes (2013). *Gérer les enseignants autrement* (Rapport public thématique). Paris: Cour des comptes.

Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, *64*(8), 672-683.

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

De Bruijn, H. (2002). *Managing Performance in the Public Sector*. London: Routledge.

Demeuse, M., Frandji, D., Greger, D. & Rochex, J.-Y. (2008). *Les politiques d'éducation prioritaire en Europe : Conceptions, mises en œuvre, débats*. Lyon: Institut national de recherche pédagogique.

Demir, I. & Depren, Ö. (2010). Assessing Turkey's secondary schools performance by different region in 2006. *Procedia – Social and Behavioral Sciences*, *2*(1), 2305-2309.

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.

De Witte, K. & Van Klaveren, C. (2014). How are teachers teaching? A nonparametric approach. *Education Economics, 22*(1), 3-23.

Diagne, D. (2006). Mesure de l'efficience technique dans le secteur de l'éducation : une application de la méthode DEA. *Revue suisse d'économie et de statistique*, *142*(2), 231-262.

Dorsch, J. J. & Yasin, M. M. (1998). A framework for benchmarking in the public sector. *International Journal of Public Sector Management*, *11*(2/3), 91-115.

D'Souza, S. (2013). Parent Feedback: A Critical Link in Improving Our Schools. *Public Administration Review*, *73*(3), 413-414.

Emrouznejad, A., Parker, B. R. & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, *42*(3), 151-157.

Essid, H., Ouellette, P. & Vigeant, S. (2010). Measuring efficiency of Tunisian schools in the presence of quasi-fixed inputs: A bootstrap data envelopment analysis approach. *Economics of Education Review*, *29*(4), 589-596.

Essid, H., Ouellette, P. & Vigeant, S. (2013). Small is not that beautiful after all: measuring the scale efficiency of Tunisian high schools using a DEA-bootstrap method. *Applied Economics*, *45*(9), 1109-1120.

Fandel, G. (2007). On the performance of universities in North Rhine-Westphalia, Germany: Government's redistribution of funds using DEA efficiency measures. *European Journal of Operational Research*, *176*(1), 521-533.

Farsi, M. & Filippini, M. (2005). A benchmarking analysis of electricity distribution utilities in Switzerland. *CEPE Working Paper No. 43*. Zurich: Center for Energy Policy and Economics, Swiss Federal Institute of Technology.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of Royal Statistical Society*, *120*(3), 253-281.

Favero, N. & Meier, K. J. (2013). Evaluating Urban Public Schools: Parents, Teachers, and State Assessments. *Public Administration Review*, *73*(3), 401-412.

Feng, Y. J., Lu, H. & Bi, K. (2004). An AHP/DEA method for measurement of the efficiency of R&D management activities in universities. *International Transactions in Operational Research*, *11*(2), 181-191.

Ferro-Luzzi, G., Flueckiger, Y., Ramirez, J. & Vassiliev, A. (2006). Unemployment and employment offices' efficiency: what can be done? *Socio-economic Planning Sciences*, *40*(3), 169-186.

Fuentes, A. (2011). Raising Education Outcomes in Switzerland. *OECD Economics Department Working Papers No. 838*. Paris: OECD Publishing.

Garrett, W. A. & Kwak, N. K. (2011). Performance comparisons of Missouri public schools using data envelopment analysis. In K. D. Lawrence & G. Kleinman (Eds.), *Applications in Multicriteria Decision Making, Data Envelopment Analysis, and Finance (Applications of Management Science, Volume 14)* (pp. 135-155). Bingley: Emerald Group Publishing Limited.

Grin, F. & Hanhart, S. (2003). Modalités de financement de l'éducation: balisage d'une évaluation par les résultats. *Revue Suisse des Sciences de l'Éducation*, *25*(3), 365-372.

Hansen, J. A. (2008). *A comparison of parametric and nonparametric techniques used to estimate school district production functions: analysis of model response to change in sample size and multicollinearity* (Unpublished doctoral dissertation). Indiana University, USA.

Hirao, Y. (2012). Efficiency of the top 50 business schools in the United States. *Applied Economics Letters*, *19*(1), 73-78.

Hollingsworth, B. & Smith, P. (2003). Use of ratios in data envelopment analysis. *Applied Economics Letters*, *10*(11), 733-735.

Hood, C. (1991). A public management for all seasons? *Public Administration*, *69*(1), 3-19.

Hood, C. (2012). Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses. *Public Administration Review*, *72*(S1), 85-92.

Hu, Y., Zhang, Z. & Liang, W. (2009). Efficiency of primary schools in Beijing, China: an evaluation by data envelopment analysis. *International Journal of Educational Management*, *23*(1), 34-50.

Huguenin, J.-M. (2012). Data Envelopment Analysis (DEA): a pedagogical guide for decision makers in the public sector. *Cahier de l'IDHEAP No. 276*. Lausanne: Swiss Graduate School of Public Administration.

Huguenin, J.-M. (2013a). Data Envelopment Analysis (DEA): un guide pédagogique à l'intention des décideurs dans le secteur public. *Cahier de l'IDHEAP No. 278*. Lausanne : Institut de hautes études en administration publique.

Huguenin, J.-M. (2013b). Data Envelopment Analysis (DEA). In A. Ishizaka & P. Nemery (Eds.), *Multi-Criteria Decision Analysis: Methods and Software* (pp. 235-274). Chichester: John Wiley & Sons.

Huguenin, J.-M. (forthcoming). Determinants of school efficiency: the case of primary schools in the State of Geneva, Switzerland. *International Journal of Educational Management*.

International Public Sector Accounting Board (2012). *Handbook of Internantional Public Sector Accounting Pronouncements*. New York: International Federation of Accountants.

Ishizaka, A. & Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. Chichester: John Wiley and Sons.

Jeanrenaud, C. & Vuilleumier, M. (2006). Evaluating Technical Efficiency of Swiss Consulates. *Review of Business and Economics*, *Vol. LI*(3), 266-279.

Jennings, E. T. (2012). Organizational Culture and Effects of Performance Measurement. *Public Administration Review*, *72*(S1), 93-94.

Jesson, D., Mayston, D. & Smith, P. (1987). Performance Assessment in the Education Sector: Educational and Economic Perspectives. *Oxford Review of Education*, *13*(3), 249-266.

Johnes, J. (2003). Measuring teaching efficiency in higher education: an application of data envelopment analysis to graduates from UK universities 1993. *Discussion paper EC7/03*. Lancaster: Lancaster University.

Johnes, J. (2004). Efficiency measurement. In G. Johnes & J. Johnes (Eds.), *International Handbook on the Economics of Education* (pp. 613-742). Cheltenham: Edward Elgar Publishing.

Johnes, J. (2006a). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, *25*(3), 273-288.

Johnes, J. (2006b). Measuring teaching efficiency in higher education: An application of data envelopment analysis to economics graduates from UK Universities 1993. *European Journal of Operational Research*, *174*(1), 443-456.

Johnes, J. & Taylor, J. (1990). *Performance Indicators in Higher Education: UK Universities*. Milton Keynes: Open University Press.

Johnes, J. & Yu, L. (2008). Measuring the research performance of Chinese higher education institutions using data envelopment analysis. *China Economic Review*, *19*(4), 679-696.

Joumady, O. & Ris, C. (2005). Performance in European higher education: A non-parametric production frontier approach. *Education Economics*, *13*(2), 189-205.

Katharaki, M. & Katharakis, G. (2010). A comparative assessment of Greek universities' efficiency using quantitative analysis. *International Journal of Educational Research*, *49*(4), 115-128.

Kempkes, G. & Pohl, C. (2010). The efficiency of German universities – some evidence from nonparametric and parametric methods. *Applied Economics*, *42*(16), 2063-2079.

Knoepfel, P., Larrue, C., Varone, F. & Hill, M. (2011). *Public policy analysis*. Bristol: The Policy Press.

Korhonen, P., Tainio, R. & Wallenius, J. (2001). Value efficiency analysis of academic research. *European Journal of Operational Research*, *130*(1), 121-132.

Kroll, A. (2012, September). *Does Performance Information Use Increase Organizational Performance? Examining an Implicit Assumption*. Paper presented at the Conference of the European Group of Public Administration (EGPA), Bergen, Norway.

Lavado, R. F. & Cabanda, E. C. (2009). The efficiency of health and education expenditures in the Philippines. *Central European Journal of Operations Research*, *17*(3), 275-291.

Lovell, C. A. K. (1993). Production Frontiers and Productive Efficiency. In H. O. Fried, C. A. K Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications* (pp. 3-67). New York: Oxford University Press.

Martin, E. (2006). Efficiency and Quality in the Current Higher Education Context in Europe: an application of the data envelopment analysis methodology to performance assessment of departments within the University of Zaragoza. *Quality in Higher Education*, *12*(1), 57-79.

Mény, Y. & Thoenig, J. C. (1989). *Politiques publiques*. Paris: Presses Universitaires de France.

Meunier, M. (2008). Are Swiss Secondary Schools Efficient? In N. C. Soguel & P. Jaccard (Eds.), *Governance and Performance of Education Systems* (pp. 187-202). Dordrecht: Springer.

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

Neely, A. D., Mills, J. F., Gregory, M. J. & Platts, K. W. (1995). Performance measurement system design – literature review and research agenda. *International Journal of Operations and Production Management*, *15*(4), 80-116.

Nevo, D. (1995). *School-Based Evaluation: A dialogue for school improvement*. Oxford: Pergamon.

Nevo, D. (2007). Evaluation in education. In I. F. Shaw, J. C. Greene and M. M. Mark (Eds.), *The SAGE Handbook of Evaluation* (pp. 441-460). London: SAGE Publications.

Ng, Y. C. & Li, S. K. (2000). Measuring the Research Performance of Chinese Higher Education Institutions: An Application of Data Envelopment Analysis. *Education Economics*, *8*(2), 139-156.

Olivares, M. & Schenker-Wicki, A. (2010). How do Swiss Universities master the reform of the last ten years? Empirical evidence from a data envelopment analysis. *Zurich Open Repository and Archive*. Zurich: University of Zurich.

Olivares, M. & Schenker-Wicki, A. (2012). The Dynamics of Productivity in the Swiss and German University Sector: A Non-Parametric Analysis That Accounts for Heterogeneous Production. *UZH Business Working Paper No. 309*. Zurich: University of Zurich.

O'Neill, J. & West-Burnham, J. (2001). Perspectives on Performance Management. In J. West-Burnham, I. Bradbury & J. O'Neill (Eds.), *Performance Management in Schools: How to Lead and Manage Staff for School Improvement* (pp. 3-14). London: Pearson Education.

OECD (2007). *Education at a Glance 2007: OECD indicators*. Paris: Organisation for Economic Co-operation and development.

Orme, C. & Smith, P. (1996). The Potential for Endogeneity Bias in Data Envelopment Analysis. *Journal of Operational Research Society*, *47*(1), 73-83.

Ouellette, P. & Vierstraete, V. (2005). An evaluation of the efficiency of Québec's school boards using the Data Envelopment Analysis method. *Applied Economics*, *37*(14), 1643-1653.

Portela, M. C. A. S., Camanho, A. S. & Borges, D. N. (2011). BESP – benchmarking of Portuguese secondary schools. *Benchmarking: An International Journal*, *18*(2), 240-260.

Portela, M. C. A. S., Thanassoulis, E. & Simpson, G. P. M. (2004). Negative data in DEA: A directional distance approach applied to bank branches. *Journal of the Operational Research Society*, *55*(10), 1111-1121.

Ray, S. C. & Jeon, Y. (2008). Reputation and efficiency: A non-parametric assessment of America's top-rated MBA programs. *European Journal of Operational Research*, *189*(1), 245-268.

Reschovsky, A. (1994). Fiscal equalization and school finance. *National Tax Journal*, *47*(1), 185-197.

Rhodes, E. L. (1978). *Data envelopment analysis and approaches for measuring the efficiency of decision making units with an application to program follow through in U.S. education* (Unpublished doctoral dissertation). Carnegie-Mellon University, USA.

Rothstein, R. (2010). Family Environment in the Production of Schooling. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of Education* (pp. 148-155). Oxford: Elsevier.

Ruggiero, J. (2004). Performance Evaluation in Education. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 323-346). Dordrecht: Springer.

Ruggiero, J., Miner, J. & Blanchard, L. (2002). Measuring equity of educational outcomes in the presence of inefficiency. *European Journal of Operational Research*, *142*(3), 642-652.

Saaty, T. (1980). *The Analytic Hierarchy Process*. New-York: McGraw-Hill.

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.

Sarrico, C. S., Rosa, M. J. & Coelho, I. P. (2010). The performance of Portuguese secondary schools: an exploratory study. *Quality Assurance in Education*, *18*(4), 268-303.

Schenker-Wicki, A. & Hürlimann, M. (2006). Universités suisses : échec ou succès du financement fondé sur les résultats ? Analyse a posteriori. *Politiques et gestion de l'enseignement supérieur*, *18*(1), 61-78.

Schoenenberger, A, Mack, A. & von Gunten, F. (2009). *Efficacité technique des exploitations forestières publiques en Suisse. Impact des subventions* (Strukturberichterstattung Nr. 42). Berne: Secrétariat d'Etat à l'économie.

Seiford, L. M. & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, *142*(1), 16-20.

Selwood, I. & Visscher, A. J. (2008). The potential of School Information Systems for Enhancing School Improvement. In. N. C. Soguel & P. Jaccard (Eds.), *Governing and Performance of Education Systems* (pp. 269-288). Dordrecht: Springer.

Sheldon, G. (1995). Zur Messung der Effizienz im Bildungsbereich mit Hilfe der Data Envelopment Analysis. *Wirtschaftswissenschaftliches Zentrum der Universität Basel – Studien, Nr. 47*. Basel: University of Basel.

Sillah, B. M. S. (2012). An analysis of efficiency in Senior Secondary Schools in the Gambia 2006-2008: Educational inputs and production of credits in English and Mathematics subjects. *Africa Education Review*, *9*(1), 86-104.

Singh, S., Rylander, D. H. & Mims, T. C. (2012). Efficiency of Online vs. Offline Learning: A Comparison of Inputs and Outcomes. *International Journal of Business, Humanities and Technology*, *2*(1), 93-98.

Smith, P. (2005). *An Introduction to Measuring Efficiency and Productivity in Public Sector Organisations: Course notes*. York: University of York.

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

Smith, P. C. & Street, A. (2006). *Analysis of Secondary School. Efficiency: Final report* (Research Report RR788). London: Department for Education and Skills.

Soguel, N. C. & Huguenin, J.-M. (2008). Comparer l'efficience des prestations financières de l'aide sociale : le cas des centres sociaux régionaux vaudois. In G. Bonoli & F. Bertozzi (Eds.), *Les nouveaux défis de l'Etat social* (pp. 165-184). Lausanne: Presses polytechniques et universitaires romandes.

Solaux, G., Huguenin, J.-M., Payet, J.-P. & Ramirez, J. (2011). Evaluation, concertation, décision : quelle régulation pour le système éducatif ? Le cas de l'enseignement primaire genevois. *Raisons éducatives No. 15*. Bruxelles: De Boeck.

Souci, A. & Nidegger, C. (2010). *Le réseau d'enseignement prioritaire à Genève : quels effets sur les acquis des élèves après quelques années ?* Genève: Service de la recherche en éducation.

Steinmann, L. & Zweifel, P. (2003). On the (in)efficiency of Swiss hospitals. *Applied Economics*, *35*(3), 361-370.

Stevens, P. A. (2001). The determinants of economic efficiency in English and Welsh universities. *Discussion paper No. 185*. London: National Institute of Economic and Social Research.

Stewart, J. & Walsh, K. (1994). Performance Measurement: When Performance can Never be Finally Defined. *Public Money & Management*, *14*(2), 45-49.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman H. O. & Provus, M. M. (1971). *Educational Evaluation and Decision Making*. Itasca: Peacock.

Summermatter, L. & Siegel, J. P. (2009, April). *Defining Performance in Public Management: Variations over time and space*. Paper presented at the Conference of the International Research Society for Public Management (IRSPM), Copenhagen, Denmark.

Sutherland, D., Price, R., Joumard, I. & Nicq, C. (2007). Performance Indicators for Public Spending Efficiency in Primary and Secondary Education. *OECD Economics Department Working Papers No. 546*. Paris: Organisation for Economic Co-operation and Development.

Thanassoulis, E., Portela, M. C. S. & Allen, R. (2004). Incorporating value judgements in DEA. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 99-138). Boston: Kluwer Academic Publishers.

Thanassoulis, E., Portela, M. C. S. & Despic, O. (2008). Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. In H. O. Fried, C. A. Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency and Productivity Growth* (pp. 251-420). New York: Oxford University Press.

Thrall, R. M. (1996). The lack of invariance of optimal dual solutions under translation. *Annals of Operations Research*, *66*(2), 103-108.

Van Helden, G. J. & Tillema, S. (2005). In search of a benchmarking theory for the public sector. *Financial Accountability & Management*, *21*(3), 337-361.

Van der Waldt, G. (2004). *Managing Performance in the Public Sector: Concepts, Considerations and Challenges*. Lansdowne: Juta and Co Ltd.

Varian, H. R. (2010). *Intermediate Microeconomics: a Modern Approach*. New-York: W. W. Norton & Company.

Vierstraete, V. & Yergeau, E. (2011). Performance of the Different Methods of Study Financing: A Measurement through the Data Envelopment Analysis Method. *Managerial and Decision Economics*, *33*(1), 1-9.

Widmer, P. & Zweifel, P. (2008). Provision of Public Goods in a Federalist Country: Tiebout Competition, Fiscal Equalization, and Incentives for Efficiency in Switzerland. *Socioeconomic Institute Working Papers No. 0804*. Zurich: University of Zurich.

Wolter, S. (2010). *Swiss Education Report 2010*. Aarau: Swiss Coordination Centre for Research in Education.

Yilmaz, B. & Ali Yurdusev, M. (2011). Use of data envelopment analysis as a multi criteria decision tool – a case of irrigation management. *Mathematical and Computational Applications*, *16*(3), 669-679.

Zhu, J. (2003). *Quantitative models for performance evaluation and benchmarking*. New York: Springer.

Zikopoulos, P. & Eaton, C. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. New-York: McGraw-Hill Osborne Media.

2
Determinants of school efficiency: the case of primary schools in the State of Geneva, Switzerland[33]

Structured abstract

Purpose

The purpose of this paper is *(1)* to measure school technical efficiency and *(2)* to identify the determinants of primary school performance.

Design/methodology/approach

A two-stage Data Envelopment Analysis (DEA) of school efficiency is conducted. At the first stage, DEA is employed to calculate an individual efficiency score for each school. At the second stage, efficiency is regressed on school characteristics and environmental variables.

Findings

The mean technical efficiency of schools in the State of Geneva is equal to 93%. By improving the operation of schools, 7% (100 – 93) of inputs could be saved, representing 17'744'656 Swiss francs in 2010. School efficiency is negatively influenced by *(1)* operations being held on multiple sites, *(2)* the proportion of disadvantaged pupils enrolled at the school and *(3)* the provision of special education, but positively influenced by school size (captured by the number of pupils).

---

[33] The current version of this essay has been accepted for publication in the International Journal of Educational Management (Emerald Publishing Group).

Practical implications

Technically, the determinants of school efficiency are outside of the control of headteachers. However, it is still possible to either boost their positive impact or curb their negative impact. In the context of the State of Geneva, the policy-related implications of the current study could be summarized as follows. New schools or existing multi-site schools should be concentrated on common sites; if this is not possible, the use of ICT in school management and teaching should be developed and encouraged. In order to correct the negative influence of disadvantaged pupils on school performance, policymakers should focus on related social policies, such as pre-school, health, housing and benefits policies, rather than on allocating additional resources to schools. Finally, with an average of 381 pupils per school, school size could be increased to maximize school efficiency.

Originality/value

Unlike most similar studies, the model in this study is tested for multicollinearity, heteroskedasticity and endogeneity. It is therefore robust. Moreover, one explanatory variable of school efficiency (operations being held on multiple sites) is a truly original variable as it has never been tested so far.

Keywords: school performance; efficiency; two-stage data envelopment analysis; multiple sites.

Article Classification: research paper

JEL classification: I20; I21; I28

2.1 Context and objectives

The measurement of school efficiency is a major concern in Switzerland: improving efficiency in compulsory education is one of four reforms recommended by a recent OECD analysis to raise education outcomes (Fuentes, 2011). Efficiency happens to be one of three criteria selected by the Swiss Conference of Cantonal Ministers of Education to assess the national education system (Wolter, 2010)[34].

Despite this, studies about efficiency of Swiss schools are virtually non-existent. Olivares and Schenker-Wicki (2012, 2010), Meunier (2008), Diagne (2006) and Schenker-Wicki and Hürlimann (2006) represent the only studies to conduct efficiency analysis on this topic. However, none of these studies measure primary school technical efficiency. As a result, the efficiency and the determinants of primary school efficiency in Switzerland are unknown, making difficult to define and to conduct an evidence-based policy. In terms of governance, decision makers still rely on partial productivity ratios (mainly cost per pupil) rather than on more elaborate measures of efficiency. They also lack local empirical evidence about the environmental variables which influence school efficiency.

In order to produce such empirical evidence, this study precisely aims *(1)* to measure school efficiency using an appropriate performance measurement technique and *(2)* to identify the determinants of school efficiency.

The empirical case of the current study covers the full population of public primary schools in the State of Geneva, Switzerland, using cross-sectional data concerning the 2010-2011 school year. These schools are funded by the State government (chiefly for staff salary) and by local authorities – municipalities – (chiefly for school infrastructure). In order to adjust to local environment, partial autonomy in management is granted to schools. For instance, headteachers define job profiles and recruit teachers; they are responsible for school quality and they also chair the school board.

The current paper is organised as follows. The next section (2.2) provides a review of the literature about school efficiency measurement. Section 2.3 presents the retained methodology and data. The results of the efficiency analysis are presented and discussed in Section 2.4. Finally, Section 2.5 contains some concluding remarks.

2.2 Literature and background

2.2.1 Efficiency measurement techniques

Johnes (2004, p. 624) considers two basic approaches to the measurement of efficiency in education: the statistical and the non-statistical approach. The

[34] Effectiveness and equity are the other two criteria.

statistical approach uses econometric techniques, while the non-statistical approach uses linear programming or mathematical algorithms.

Both statistical and non-statistical approaches could be either parametric or non-parametric. However, the statistical approach is often parametric and the non-statistical approach is often non-parametric[35]. In the statistical parametric approach, the production frontier is characterized by the formulation of a function. In the non-statistical non-parametric approach, mathematical algorithms are used to define the production frontier. No function specification has to be formulated.

Both statistical and non-statistical approaches can also be either deterministic or stochastic. Deterministic approach assumes that the differences between the observed outputs and the outputs specified by the production frontier correspond exclusively to inefficiency. Stochastic approach assumes that "deviations from the production function are a consequence no just of inefficiency, but also of measurement errors, random shocks and statistical noise" (Johnes, 2004, p. 625). The aim of stochastic approach is therefore to separate the residual into an inefficiency component and a random component.

Ordinary Least Squares (OLS) and Stochastic Frontier Analysis (SFA) are examples of deterministic and stochastic statistical parametric methods respectively. Applications of these methods in the education sector can be found in Smith and Street (2006) for OLS and SFA or Blank *et al.* (2012) for SFA. Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH) are two deterministic non-statistical non-parametric methods. Applications of these methods in the education sector can be found in Sarrico *et al.* (2010) for DEA or De Witte *et al.* (2010) for FDH.

The main advantage of statistical parametric techniques is that the significance of the frontier's parameters can be tested. But these techniques are unsuitable for applications where there are multiple inputs and multiple outputs. Non-statistical non-parametric techniques are suited to handle such applications. However, these techniques are vulnerable to small sample size.

DEA is the method of choice in this study. It has been retained for its capacity to handle multiple inputs and outputs. Moreover, as pointed out by Agasisti *et al.* (2014, p. 122), "most of the studies of technical efficiency in schools have used DEA as their methodological approach".

DEA finds its origin in Charnes *et al.* (1978) and is first applied to the education sector by Bessent and Bessent (1980). As pointed out by Agasisti *et al.* (2014), a consolidated approach has emerged from the literature when it comes to assessing school efficiency via DEA. This approach is known as the two-stage analysis. It has been developed by Ray (1991) and is recommended, for instance, by Coelli *et al.* (2005, pp. 194-195). At the first stage, DEA is employed to calculate an individual efficiency score for each school. At the second stage, efficiency is regressed on environmental variables (external factors outside of the control of headteachers). Performing a two-stage DEA analysis

---

35 For a detailed taxonomy of approaches, the interested reader will refer to Daraio and Simar (2007, p. 27).

requires the selection of input and output variables to be included in the first stage as well as environmental variables to be included in the second stage.

2.2.2 Input and output variables

A starting point for the choice of input variables consists in the OECD KLEMS model (OECD, 2001), which considers five categories of inputs: capital (K), labor (L), energy (E), materials (M) and services (S). Labor is the most commonly used resource in DEA studies focusing on the education sector. It is measured either in physical terms (number of full-time equivalent staff), such as in Abbott and Doucoulagios (2003) or Avkiran (2001), or in monetary terms (staff salaries), such as in Ahn and Seiford (1993) or Ruggiero (2000). Labor is often expressed per pupil (Mancebón and Mar Molinero, 2000; Mante and O'Brien, 2002).

Expenditure on inputs other than labor and capital usually covers the E, M and S categories. Such inputs are considered by Abbott and Doucouliagos (2003), Arcelus and Coleman (1997) or Beasley (1990, 1995). Finally and due to lack of data, few studies take into account a capital input (Ahn *et al.*, 1989; Ahn and Seiford, 1993; Lovell *et al.*, 1994; Ruggiero, 1996).

On the output side, a large part of the studies focus specifically on standardized test scores as outputs. Among those are Bessent *et al.* (1982), Bradley *et al.* (2001), Chalos (1997), Demir and Depren (2010) or Mizala *et al.* (2002). Agasisti *et al.* (2014, p. 123) note that "such choice represents today the standard for analyzing school efficiency". However, another often used output consists in the number of pupils (i.e. Avkiran, 2001; Coelli *et al.*, 2005; Essid *et al.*, 2013).

2.2.3 Environmental variables

The second stage of the efficiency analysis requires the identification of the environmental variables. Existing two-stage analysis studies cover kindergarten, primary schools, lower and upper secondary schools and universities. Such analyses have been conducted in several countries, including Canada, England, Finland, Greece, Italy, Kuwait, the Netherlands, New Zealand, Norway, Sweden, Switzerland, Thailand and the United States. Environmental variables which emerge from these studies can be grouped into seven categories, presented hereafter.

- **Socioeconomic status of students**

  A higher proportion of disadvantaged students reduces school efficiency. This finding is consistent across studies and appears almost unchallenged (Alexander and Jaforullah, 2004; Alexander *et al.*, 2010; Bradley *et al.*, 2001; Jeon and Shields, 2005; McCarty and Yasawarng, 1993; Ouellette and Vierstraete, 2005; Rassouli-Currier, 2007; Ruggiero and Vitaliano, 1999).

- **Type of school**

  Several studies define various school types (or categories) and test their impact on efficiency (Agasisti, 2013; Alexander and Jaforullah, 2004; Alexander *et al.*, 2010; Bradley *et al.*, 2001; Burney *et al.*, 2011; Duncombe *et al.*, 1997; Lovell *et al.*, 1994; McMillan and Datta, 1998; Mancebón and Mar Molinero, 2000; Ramanathan, 2001). Depending on the study, school type could refer to private versus public school, all-girls versus all-boys school, specialized versus general school, and so on. The only conclusion that can be drawn is that the type of school can matter, sometimes positively, sometimes negatively.

- **School location**

  As for school type, school location can matter when it comes to efficiency (Alexander and Jaforullah, 2004; Alexander *et al.*, 2010; Borge and Naper, 2006; Bradley *et al.*, 2004; Denaux *et al.*, 2011; Kirjavainen and Loikkanen, 1998; Lovell *et al.*, 1994; Ouellette and Vierstraete, 2005; Ramanathan, 2001). For instance, evidence has been found to suggest that the geographical region of a school can either negatively (Agasisti, 2013) or positively (Burney *et al.*, 2011) impact efficiency.

- **Political context**

  Two studies (Borge and Naper, 2006; Waldo, 2007) include the political context as explanatory variables in the second stage. Both of them demonstrate that a higher share of socialists in the local council is associated with lower school efficiency.

- **Competition**

  Bradley *et al.* (2001) find that the number of competitors of a school impacts positively on its efficiency. Duncombe *et al.* (1997) approximate the degree of competition by considering the number of students enrolled in private schools. The authors find a negative association with efficiency. Agasisti (2013) approximates competition by the number of schools in the region. The impact on school efficiency is positive.

- **Teachers characteristics**

  Duncombe *et al.* (1997), Rassouli-Currier (2007) and Ruggiero and Vitaliano (1999) find that the coefficient for teacher salary on staff is negative. Conversely, Burney *et al.* (2011) and Bradley *et al.* (2001) find a positive impact of teacher salary on school efficiency.

  Teacher experience is associated with higher school efficiency in Alexander and Jaforullah (2004) and Alexander *et al.* (2010). Whith regards to teacher qualification, Alexander *et al.* (2010) show that the proportion of teachers who have at least second year university qualifications increases school efficiency. However, it seems that a greater proportion of teachers with formal pedagogical training is associated with lower school efficiency (Waldo, 2007).

- **Size effects (school and class)**

  A clear picture emerges regarding the positive impact of school size, as measured by the number of pupils or students, on school efficiency (Agasisti, 2013; Alexander and Jaforullah, 2004; Alexander *et al.*, 2010; Borge and Naper, 2006; Bradley *et al.*, 2001; Kantabutra and Tang, 2006; Lovell *et al.*, 1994; McMillan and Datta, 1998; Olivares and Schenker-Wicki, 2010; Ramanathan, 2001). This finding is valid across countries and levels of the educational system.

  Class size also positively impacts school efficiency (Kirjavainen and Loikkanen, 1998). This finding is confirmed by Kantabutra and Tang (2006), but only for schools located in urban areas.

2.2.4 School efficiency in Switzerland

Concerning Switzerland, evidence on school efficiency is virtually non-existent. It comprises only five studies, none of them focusing on primary schools. However, two of them apply a two-stage procedure. First, Olivares and Schenker-Wicki (2010) consider a sample of Swiss universities. Results show that the student-faculty ratio, the proportion of professors per scientific staff and the number of students are associated with higher efficiency. Second, Diagne (2006) considers a sample of upper-secondary schools. He shows that the proportion of teachers with qualifications increases efficiency. However, the proportion of teachers with indefinite duration contracts is associated with lower efficiency. Other studies include: Olivares and Schenker-Wicki (2012), who show that improvements in technical efficiency are the most important source of the change in productivity over time of German and Swiss universities; Meunier (2008), who analyzes a sample of lower-secondary schools showing, "that the more the size of the schools increases, the greater is the proportion of efficient schools (…)" (p. 200); and Schenker-Wicki and Hürlimann (2006), who conduct a basic efficiency analysis of a sample of Swiss universities, showing that efficiency has not improved over time.

2.2.5 Originality of the current study

The current study distinguishes itself from the existing literature in four main areas. First, it focuses on primary schools rather than secondary or tertiary schools. As far as the author is aware, only three existing studies specifically cover primary schools, leaving this level of the educational system under-researched in terms of efficiency (Agasisti *et al.*, 2014; Mancebón and Mar Molinero, 2000; Burney *et al.*, 2011). Second, the model used in the second stage of the analysis is tested for multicollinearity, heteroskedasticity and endogeneity. These tests guarantee the robustness of the model. Only five existing studies test for multicollinearity (Agasisti, 2013; Bradley *et al.*, 2010; Burney *et al.*, 2011; Denaux *et al.*, 2011; Ray, 1991), four for heteroskedasticity (Alexander and Jaforullah, 2004; Mancebon and Mar Molinero, 2000; Rassouli-Currier, 2007; Waldo, 2007) and one for endogeneity (Waldo, 2007). Third, the literature review identifies that schools which operate on several sites has never been tested as an explanatory variable of school efficiency. In this study, the number of sites on which schools operate is known. It is therefore a truly original variable to be tested, as headteachers estimate that managing a multi-site school needs more resources than managing a single-site school (Observatory on Primary Education, 2010). In Switzerland, schools operate in a context of school mergers imposed by the State authority. Small schools are grouped into a unique administrative unit, becoming school sites. As a result, the pupils often have to be transported daily from their home town to the appropriate school site; headteachers have to distance manage the different sites; and teachers have to work on several sites, meaning that they sometimes have to move from one site to another during the same day. These elements could alter efficiency. This hypothesis needs to be tested. Fourth, the current study adds new evidence to the situation in Switzerland, which has been under-researched. In doing so, it provides information to decision makers in order to improve school governance.

2.3 Methodology

2.3.1 Two-stage analysis

At the first stage, DEA is employed to calculate an individual efficiency score for each school. As described below in the data section, the inputs and outputs used in this study are formulated as ratios. In such a case, a variable returns to scale (VRS) model, as developed by Banker *et al.* (1984), is required (Hollingsworth and Smith, 2003). As Coelli *et al.* (2005, p. 172) point out, the use of the VRS model permits the calculation of technical efficiency devoid of the scale efficiency effects. The model is input oriented, meaning that it minimizes input for a given level of output.

Following the notation adopted by Johnes (2004, pp. 630-637), it is assumed there are data on s outputs and m inputs for each of n primary schools to be evaluated (n = 90). y_{rk} is the quantity of output r produced by school k. x_{ik} is the quantity of input i consumed by school k. u_r is the weight of output r. v_i is the weight of input i. θ_k represents the measure of VRS efficiency of school k (i.e. 'pure' technical efficiency free from any scale inefficiency). λ_j represents the associated weighting of outputs and inputs of entity j.

The VRS efficiency of the kth school is calculated by solving the following linear problem:

$$\text{Minimize} \quad \theta_k \qquad (1)$$

$$\text{Subject to} \quad y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \le 0 \quad r = 1, \dots, s$$

$$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} \ge 0 \quad i = 1, \dots, m$$

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

$$\lambda_j \ge 0 \quad \forall j = 1, \dots, n$$

The value of θ varies between zero and one ($\theta \le 1$). A value of 1 indicates a school on the best-practice frontier (and hence a technically efficient school).

At the second stage, the efficiency scores are regressed on school characteristics and environmental variables. A Tobit regression, as developed by Tobin (1958), is used in the majority of studies dealing with efficiency in the education sector. However, recent studies have shown that Ordinary Least Squares (OLS) regression is sufficient or even more appropriate to model the efficiency scores (Hoff, 2007; McDonald, 2009). OLS is, therefore, the method of choice in the ensuing study[36].

2.3.2 Data

The empirical case covers the full population (90) of public primary schools in the State of Geneva using cross-sectional data concerning the 2010-2011 school year. The database has been provided by the State of Geneva.

At the first stage, three outputs and three inputs are considered. These variables are all under the control of headteachers.

Outputs include three composite scores (standardized on a scale with a maximum of 100). The first one is composed of pupils' results in French and mathematics standardized tests at the end of the second grade (SCORE2). The second one is composed of pupils' results in French, German and mathematics standardized tests at the end of the fourth grade (SCORE4). Finally, the third one is composed of pupils' results in French, German and mathematics standardized tests at the end of the sixth grade (SCORE6). Several studies focus specifically on standardized test scores as outputs. Among those are Chalos and Cherian (1995), Kirjavainen and Loikkanen (1998), Ruggiero (1996, 2000) and Sengupta (1990).

36 The debate between OLS and Tobit (and even truncated regression) continues. See Simar and Wilson (2011) for discussion.

Inputs include *(1)* the number of full-time equivalent (FTE) teaching staff (TEACHER), *(2)* the number of FTE administrative and technical staff (ADMIN) and *(3)* the school budget in Swiss francs – excluding staff salaries and capital expenditure (BUDGET) –. The three inputs are expressed by pupils to be coherent with the formulation of the outputs. The inputs used in this study are very similar to those used by Arcelus and Coleman (1997). Note that the number of teachers and the number of administrative staff are classical inputs (Abbott and Doucouliagos, 2003; Avkiran, 2001; Grosskopf and Moutray, 2001), as are the overhead expenses (Ahn and Seiford, 1993; Beasley, 1990; Chalos and Cherian, 1995).

Descriptive statistics of outputs and inputs are reported in Table 1.

Table 1
Statistical summary of output and input variables included in the first stage DEA model (sample size = 90 primary schools)

	Mean	SD	Minimum	Maximum
Outputs				
SCORE2 (points/pupil)	78.8082	4.4956	64.9589	91.9591
SCORE4 (points/pupil)	77.2733	3.8718	68.0930	87.3654
SCORE6 (points/pupil)	76.7382	4.5361	64.7010	85.5275
Inputs				
TEACHER (FTE/pupil)	0.0582	0.0043	0.0520	0.0689
ADMIN (FTE/pupil)	0.0035	0.0005	0.0026	0.0052
BUDGET (CHF/pupil)	20.1643	5.8233	8.8186	48.2835

Source: General Direction of Primary Schools, Education Department, State of Geneva.

At the second stage, the data set contains eight explanatory variables divided into two groups: school characteristics and environmental variables. These variables are outside of the control of headteachers.

School characteristics

- SITE: this variable indicates whether a school is located on one site or several. It is set up as a dummy variable, which takes the value of 1 if a school is located on more than one site. The expected sign of SITE is negative, as a greater number of site locations should complicate school organization and alter technical efficiency.
- SPECIAL: this variable indicates whether special education for special needs pupils is available at a particular school. It is set up as a dummy variable, which takes the value of 1 if a school provides special education. The expected sign of SPECIAL is negative (Borge and Naper, 2006; Rassouli-Currier, 2007) as *(1)* school organization with special education is

more restrictive than without it and *(2)* schools with special education mostly admit disadvantaged pupils into special education classes.

- RECEPTION: this variable indicates whether special reception classes for immigrant pupils are available at a particular school. It is set up as a dummy variable, which takes the value of 1 if a school offers special reception classes. The expected sign of RECEPTION is negative because special reception classes are populated by allophone pupils.
- URBAN: this variable indicates whether a school is located in an urban area. It is set up as a dummy variable, which takes the value of 1 if a school is located in an urban area. The expected sign of URBAN is negative, as urban schools tend to be less efficient than rural ones (Alexander *et al.*, 2010; Duncombe *et al.*, 1997).
- CLASS: this variable refers to the number of classes within a school. As the maximum number of pupils per class is regulated by law, this variable is outside of the control of headteachers. The expected sign of CLASS is negative, as a greater number of classes could be due to a smaller number of pupils per class.

Environmental variables

- PUPIL: this variable refers to the number of pupils in a school. The expected sign of PUPIL is positive, as efficiency tends to grow with school size (Alexander *et al.*, 2010; Borge and Naper, 2006; Bradley *et al.*, 2001).
- SOCIO: this variable represents the percentage of pupils (per school) whose parents are blue-collar workers or unqualified workers. It reflects the socioeconomic status of pupils. The expected sign is negative (Alexander *et al.*, 2010; McCarty and Yasawarng, 1993).
- ALLO: this variable represents the percentage of allophone pupils (per school). The expected sign is negative (Ouellette and Vierstraete, 2005).

Descriptive statistics school characteristics and environmental variables are reported in Table 2.

Table 2
Statistical summary of variables included in the second stage DEA model (sample size = 90 primary schools)

	Mean	SD	Minimum	Maximum
School characteristics				
SITE (dummy*)	0.64			
SPECIAL (dummy*)	0.23			
RECEPTION (dummy*)	0.41			
URBAN (dummy*)	0.79			
CLASS	19.69	6.18	9.00	38.00
Environmental variables				
PUPIL	381.38	116.52	157.00	726.00
SOCIO (%)	37.43	13.73	11.00	64.00
ALLO (%)	41.38	14.46	11.08	70.21

* For dummy variables, the mean value gives the proportion of schools in that class. For instance, 64% of schools are located on more than one site.
Source: General Direction of Primary Schools, Education Department, State of Geneva.

2.4 Results

2.4.1 Technical efficiency scores (first stage)

The mean variable returns to scale technical efficiency score (VRSTE) is equal to 0.93 (or 93%). This means that schools could proportionately reduce all their inputs by 7% (100 – 93) whilst maintaining the same quality of pupil performance (outputs). As the calculation of VRSTE is devoid of scale effect, the 7% capacity for improvement resides with school management. VRSTE scores are presented in Table 3.

Table 3
Variable returns to scale technical efficiency scores

School	Technical efficiency	School	Technical efficiency
30	1.00	16	0.95
40	1.00	74	0.95
44	1.00	55	0.94
56	1.00	39	0.94
59	1.00	57	0.94
60	1.00	49	0.94
61	1.00	52	0.94
62	1.00	41	0.93
63	1.00	29	0.93
64	1.00	37	0.93
65	1.00	47	0.93
66	1.00	67	0.92
70	1.00	28	0.92
71	1.00	32	0.92
77	1.00	23	0.91
78	1.00	51	0.91
84	1.00	22	0.91
87	1.00	31	0.91
88	1.00	13	0.91
90	1.00	75	0.91
68	1.00	43	0.90
53	1.00	26	0.90
82	1.00	3	0.89
80	0.99	83	0.89
25	0.99	50	0.89
86	0.99	79	0.89
72	0.98	20	0.89
81	0.98	18	0.88
69	0.98	15	0.88
76	0.98	4	0.88
73	0.98	5	0.87
58	0.98	2	0.86
54	0.97	35	0.85
38	0.97	21	0.85
24	0.96	19	0.83
42	0.96	6	0.82
36	0.96	12	0.81
46	0.96	7	0.81
89	0.96	10	0.81
33	0.95	11	0.80
34	0.95	1	0.79
85	0.95	8	0.78
45	0.95	17	0.76
48	0.95	9	0.76
27	0.95	14	0.76

22.2% of schools have a score of 1. These schools lie on the efficiency or best-practice frontier. All of the other schools are beneath the frontier with respective scores of less than one. 25.6% of schools have a score between 0.999 and 0.95, 24.4% have a score between 0.949 and 0.9, 14.4% have a score between 0.899 and 0.85, 6.7% have a score between 0.849 and 0.8 and 6.7% have a score between 0.799 and 0.75. The lowest score registered is 76%.

2.4.2 Sensitivity analysis (first stage)

Sensitivity analysis aims to identify the impact on school efficiency and ranking when certain parameters are modified in the model. First, the efficiency frontier may be partially modelled with respect to outlier schools. Removing these outliers could result in different efficiency scores and ranks. Second, testing different combinations of inputs and outputs may also provide different efficiency scores and ranks.

A jackknifing procedure is used to deal with potential outlier schools. Such a procedure is used by Borge and Naper (2006), Bradley *et al.* (2001), Hu *et al.* (2009) or Waldo (2007). In this procedure, efficient schools are removed one at a time from the analysis. In this study, 20 schools are 100% efficient. That means that 20 additional models are run, each removing a different efficient school. The similarity of *(1)* school efficiency scores and *(2)* school ranking between the original model and the models where efficient schools are removed one at a time is then tested using Pearson and Spearman rank correlations. Results of this analysis are presented in Table 4.

Table 4
Sensitivity analysis regarding outlier schools

	Mean	Min	Max
VRSTE original model	0.9321	0.7604	1.0000
VRSTE iterated models*	0.9393	0.9286	0.9469
Pearson**	0.9958	0.9553	1.0000
Spearman**	0.9936	0.9497	1.0000

* For each additional model run, a mean is calculated. The mean value indicated in this table refers to the mean of the models' means. The minimum value corresponds to the minimum mean identified within the additional models. The maximum value corresponds to the maximum mean identified within the additional models.

** For each additional model run, a Pearson and a Spearman correlation is calculated with the original model. The mean value indicated in this table refers to the mean of the correlation coefficients observed. The minimum value corresponds to the minimum correlation coefficient observed. The maximum value corresponds to the maximum correlation coefficient observed.

The Pearson and the Spearman mean correlations are positive and considered as perfect (0.9958 for Pearson and 0.9936 for Spearman). The efficiency scores correlation (Pearson) and school ranks correlation (Spearman) range from 0.9553 to 1 and from 0.9497 to 1 respectively. These correlation coefficients are significant at the 1% level. The results show that the efficiency scores and the school ranking are not sensitive to outlier schools.

The efficiency scores and rankings of schools may also vary when different combinations of inputs and outputs are considered and must therefore be tested (Abbott and Doucouliagos, 2000; Burney *et al.*, 2011; Martin, 2006).

Beside the original model containing three inputs and three outputs, six additional models are run. In each of them, a different variable is removed. Table 5 describes these six models. For instance, model 3 contains two inputs (TEACHER and ADMIN) and three outputs (SCORE2, SCORE4 and SCORE6). The variable BUDGET has been removed from this model.

Table 5
Additional DEA models

	Inputs			Outputs		
	TEACHER	ADMIN	BUDGET	SCORE2	SCORE4	SCORE6
Model 1		X	X	X	X	X
Model 2	X		X	X	X	X
Model 3	X	X		X	X	X
Model 4	X	X	X		X	X
Model 5	X	X	X	X		X
Model 6	X	X	X	X	X	

The similarity of *(1)* school efficiency scores and *(2)* school ranking between the original model and the models where input and output variables are removed one at a time is then tested using Pearson and Spearman rank correlations. Results of this analysis are presented in Table 6.

Table 6
Sensitivity analysis regarding input and output variables

	Mean	Min	Max
VRSTE original model	0.9321	0.7604	1.0000
VRSTE iterated models*	0.9127	0.8433	0.9298
Pearson**	0.9413	0.7477	0.9950
Spearman**	0.9414	0.7884	0.9930

* For each additional model run, a mean is calculated. The mean value indicated in this table refers to the mean of the models' means. The minimum value corresponds to the minimum mean observed within the additional models. The maximum value corresponds to the maximum mean observed within the additional models.

** For each additional model run, a Pearson and a Spearman correlation is calculated with the original model. The mean value indicated in this table refers to the mean of the correlation coefficients observed. The minimum value corresponds to the minimum correlation coefficient observed. The maximum value corresponds to the maximum correlation coefficient observed.

The Pearson and the Spearman mean correlations are positive and strong (0.9413 for Pearson and 0.9414 for Spearman). The efficiency scores correlation (Pearson) and school ranking correlation (Spearman) range from 0.7477 to 0.995 and from 0.7884 to 0.993 respectively. These correlation coefficients are significant at the 1% level. The minimum correlation coefficients (0.7477 for Pearson and 0.7884 for Spearman) are observed between the original model and model 1. In model 1, the input variable TEACHER is removed. As teachers represent the most important input variable in a school, it stands to reason that it should be retained within the model. As a result, model 1 can be excluded. These results show that, with the exception of model 1, the efficiency scores and the efficiency rankings are not sensitive to the removal of inputs and outputs.

2.4.3 Regression model (second stage)

To test for potential multicollinearity in the data set, the variance inflation factors (VIF) are assessed. A regression model containing all the explanatory variables mentioned above is run. The mean VIF is equal to 14.51. It is therefore likely that the results are distorted by multicollinearity (Bowerman and O'Connell, 1990; Myers, 1990).

CLASS and PUPIL are the two variables with the highest VIF (54.41 and 45.88 respectively). As an objective of this model is to test the effect of school size, the variable for the number of pupils is kept in the model, as it is a more accurate reflection of school size compared to the number of classes. As a result, CLASS is removed from the model, which is left with seven explanatory variables. The new mean VIF, once the number of classes has been removed from the model, is equal to 2.04. Therefore, it can be concluded that the results of this new model are unlikely to be distorted by multicollinearity.

The OLS model takes the following form:

$$\begin{aligned} TE_k = {} & \alpha_0 + \alpha_1 SITE_k + \alpha_2 SPECIAL_k + \alpha_3 RECEPTION_k + \alpha_4 URBAN_k + \alpha_5 PUPIL_k \\ & + \alpha_6 SOCIO_k + \alpha_7 ALLO_k + e_k \end{aligned}$$

TE_k is the efficiency score, derived from the first stage analysis, of the k_{th} school and e_k is an error term satisfying the usual conditions for ordinary least squares estimation.

In order to identify the functional form of the OLS regression, three Box-Cox models have been run. In the first model, the Box-Cox transformation is applied only to the dependent variable. In the second model, it is applied only to the independent variables. In the third model, it is applied both to the dependent and independent variables. As the variables should only contain strictly positive data, URBAN, SITE, RECEPTION and SPECIAL have been excluded from the second and third model.

In the three models, the three null hypotheses (reciprocal, logarithmic and linear specification respectively) are all rejected at the 1% level, meaning that all possible specifications are rejected.

Results show that the best specification is unclear. The skewness / kurtosis tests are performed on TE. The results show that the hypothesis of a normal distribution is rejected at the 1% level. Several alternative functional forms (cubic, square, square root, 1 / (square root), 1 / square, 1 / cubic) are tested in order to identify a transformation that would convert TE into a normally distributed variable. All of them are rejected at the 5% level.

As a result, the linear form is retained as *(1)* no clear indication points to another specification and *(2)* all Box-Cox models display the lowest chi-square value for the linear specification.

The presence of heteroskedasticity in the second stage is considered. A Breusch-Pagan / Cook-Weisberg test for heteroskedasticity is performed. The null hypothesis is rejected and there is significant evidence of heteroskedasticity. Following this result, the White correction is applied to the model to correct for heteroskedasticity. An OLS regression with robust standard errors is run.

2.4.4 Endogeneity (second stage)

Identifying endogeneity in the second stage appears challenging as *(1)* the efficiency scores themselves are usually unknown from school stakeholders before the DEA analysis is run[37] and *(2)* the efficiency scores are, in fact, built on multiple outputs and multiple inputs. As a result, loops of causality are to be identified between any of the outputs and/or inputs used in the first stage and the independent variables.

In the retained model, it could be argued that simultaneity occurs between the following variables:

- The number of school sites increases where the quantity of teaching and administrative staff increases, and therefore SITE is endogenous to school efficiency. The quantity of staff is used as an input in the first stage. All other things being equal, increasing the number of staff reduces efficiency. In such a case, local authorities decide to increase the number of sites because more staff are working.
- Special education is provided where pupil performance is poor and therefore SPECIAL is endogenous to school efficiency. Pupil performance is used as an output in the first stage. All other things being equal, poor performance reduces efficiency. In such a case, the State authority will provide special education in schools.
- Reception classes are provided where pupil performance is poor, and therefore RECEPTION is endogenous to school efficiency. Pupil performance (measured by standardized tests) is used as an output in the first stage. All other things being equal, poor performance reduces efficiency. In such a case, the State authority will provide reception classes in schools.
- The proportion of disadvantaged pupils increases where pupil performance is poor, and therefore SOCIO is endogenous to school efficiency. Pupil performance is used as an output in the first stage. All other things being equal, poor performance reduces efficiency. In the State of Geneva, school

37 Unknown efficiency scores means that a loop of causality (i.e. efficiency scores explaining the independent variables and not the other way round) is improbable, precisely because efficiency scores are unknown.

catchment areas are defined by the State authority. As a result, the parents' residential address determines which school will be attended by their children. However, it could be argued that some parents develop school catchment area evasion strategies. The objective is to enrol their children into high performance schools, thus parents may strategically move to another catchment area. As the State of Geneva faces a continuous housing crisis, with very limited housing available and high rental rates, only privileged parents can afford to move into these areas. As a result, such moves would increase the proportion of remaining disadvantaged pupils[38].

- The proportion of allophone pupils increases where pupil performance is poor, and therefore ALLO is endogenous to school efficiency. Pupil performance is used as an output in the first stage. All other things being equal, poor performance reduces efficiency. In this case, French-speaking parents move to other neighbourhoods because their childrens' schools have a low performance. This move increases the proportion of allophone pupils.

Endogeneity is solved by using instrumental variables. Instruments are identified following the procedure used by Waldo (2007): first, the instruments have to correlate with the potential endogenous variables; second, they cannot have any explanatory power on efficiency scores if they are to be used as independent variables alongside the potential endogenous variables.

27 variables are tested in order to identify instruments. These variables are all measured at the municipality level in which schools are located. Correlation coefficients between potential endogenous variables and instruments are presented in Table 7. For presentation purposes, only correlation coefficients over |0.5| are listed.

Table 7
Correlation Matrix over |0.5| between potential endogenous variables and instruments

	SOCIO	ALLO
Social assistance rate (%)	0.60	0.63
Agricultural area (%)		-0.60
Habitat and infrastructure area (%)		0.58

Social assistance rate (BENEFIT) is positively correlated with SOCIO and ALLO. The proportion of agricultural area (AGRI) is negatively correlated with

38 For instance, Noreisch (2007) studies the school catchment area evasion in the city of Berlin, Germany. The results show that the higher the percentage of non-German speaking pupils that are enrolled in a school, the more German children avoid it. Although privileged parents do not know the performance of a particular school, they consider the presence of a large proportion of minority pupils "as a hindrance for the cognitive, personal and social development of their children" (van Zanten, 2003, p. 109).

ALLO and the proportion of habitat and infrastructure area (HABIT) is positively correlated with ALLO. To measure the explanatory power of BENEFIT, AGRI and HABIT, two additional models are run. The first one includes BENEFIT and the second one includes BENEFIT, AGRI and HABIT alongside SITE, SPECIAL, RECEPTION, URBAN, PUPIL, SOCIO and ALLO. BENEFIT, AGRI and HABIT are not statistically significant in any of the models. As a result, BENEFIT can be considered as an instrumental variable for SOCIO and BENEFIT, AGRI and HABIT can be considered as instrumental variables for ALLO.

First, the model tests SOCIO as a potential endogenous variable, using BENEFIT as an instrument. A Durbin-Wu-Hausman test is performed. The null hypothesis (Ho) stating that endogeneity is not present in the model is accepted.

Second, the model tests ALLO as a potential endogenous variable, using BENEFIT, AGRI and HABIT as instruments. No endogeneity is found.

Unfortunately, no correlation coefficients over |0.5| were found for SITE, SPECIAL and RECEPTION. Those potential endogenous variables are therefore not tested for endogeneity. However, it is unlikely that these variables are endogenous for the following reasons:

- Considering a principal-agent approach to educational production (Wössmann, 2005), asymmetric information about school data between the principal (i.e. the parents) and the agent (i.e. the headteacher) appears to be strong in the State of Geneva. Information about pupil performance and resource consumption are computed at State level. This is not public knowledge. Efficiency scores have never been measured before this study. As a result, it is unlikely that the variable SITE is endogenous.
- The provision of special education and reception classes does not depend on the State office of compulsory education but on the State office of special education. The presence of special education and reception classes in schools appears to be due to heritage rather than a rational decision based on efficiency analysis. As a result, it is unlikely that the variables SPECIAL and RECEPTION are endogenous.

2.4.5 Determinants of school efficiency (second stage)

SITE, SPECIAL, RECEPTION, URBAN, PUPIL, SOCIO and ALLO explain 68% of technical efficiency scores ($R^2 = 67.89$). Three variables are significant at the 1% level: SITE, PUPIL and SOCIO. One variable is significant at the 5% level: SPECIAL. Detailed results are presented in Table 8.

Table 8
Determinants of school efficiency: results from the OLS regression

	Coefficient	t-statistic	
Constant	1.0205	69.89	**
School characteristics			
SITE	-0.0349	-3.20	**
SPECIAL	-0.0239	-2.08	*
RECEPTION	-0.0081	-0.84	
URBAN	0.0097	0.88	
School environment			
PUPIL	0.0002	5.60	**
SOCIO	-0.0032	-6.29	**
ALLO	-0.0007	-1.30	

** Significant at the 1% level; * Significant at the 5% level

All the variables have the expected sign, with the exception of URBAN which shows a positive sign. However, as URBAN is not statistically significant, it cannot be concluded that this result contradicts Alexander and Jaforullah (2004), Alexander *et al.* (2010) and Duncombe *et al.* (1997).

SITE is negative and significant at the 1% level. Efficiency is negatively influenced by the fact that a school is located on several sites. The movement of SITE from 0 (one site) to 1 (several sites) generates a –0.0349 unit change in the VRSTE score[39].

SPECIAL is negative and significant at the 5% level. Efficiency is negatively influenced by the fact that a school provides special education. The movement of SPECIAL from 0 (no special education) to 1 (with special education) generates a –0.0239 unit change in the VRSTE score.

PUPIL is positive and significant at the 1% level. Efficiency is positively influenced by school size. The value of the coefficient is close to zero. A one unit change in the number of pupils generates a 0.0002 unit change in the VRSTE score.

SOCIO is negative and significant at the 1% level. Efficiency is negatively influenced by the proportion of disadvantaged pupils. A one unit change in the

---

39 Note that an additional regression model was run in which the dummy variable SITE was removed and replaced by a discrete variable (NSITE) accounting for the number of sites on which schools are located (from 1 up to 5 sites, with a mean of 1.87 and a standard deviation of 0.85). In this other model, the coefficient of NSITE is negative and not significant. These results show that being located on one site or on several sites, rather than the number of sites itself, matters.

proportion of disadvantaged pupils generates a 0.0032 unit change in the VRSTE score.

The variables RECEPTION and ALLO have the expected negative sign but are not significant.

The case of school # 14 is emblematic as it embodies a 'worst case' situation. It is a small school (296 pupils) with a high proportion of disadvantaged pupils (64%), located on more than one site and providing special education. All other things being equal, if this school held the mean number of pupils (381), the mean proportion of disadvantaged pupils (37%), was located on one site only and did not provide special education, it would have an efficiency score of 0.8766 instead of 0.7604.

The coefficients of the OLS regression allow the efficiency scores of schools to be adjusted to common levels of non-discretionary variables. In this study, the efficiency scores are adjusted to the following levels of statistically significant non-discretionary variables:

- It is assumed that all schools are considered located on several sites (indeed the majority of schools are located on several sites);
- It is assumed that none of the schools provide special education (indeed the majority of schools do not provide special education);
- It is assumed that all schools have the same number of pupils (the mean value of 381 pupils);
- It is assumed that all schools have the same proportion of disadvantaged pupils (the mean value 37.43%).

Due to the adjustment of the efficiency scores for the statistically significant non-discretionary variables, the maximum value predicted by the OLS model is slightly higher than one. This occurs in the case of relatively efficient schools operating in a relatively unfavourable environment.

Table 9 compares the unadjusted VRSTE and the adjusted VRSTE scores. For instance, school # 21 has an unadjusted score of 0.85. Once this score is corrected to take into account the influence of the significant variables as defined above, school # 21 has an adjusted score of 0.93. The mean efficiency of the unadjusted and the adjusted scores is equal to 0.9321 and 0.9252 respectively. 40 schools out of 90 have an adjusted score higher than their unadjusted score.

Table 9
Adjusted variable returns to scale technical efficiency scores

School	VRSTE	Adjusted VRSTE	School	VRSTE	Adjusted VRSTE
30	1.00	1.04	**16**	0.95	1.00
40	1.00	0.96	**74**	0.95	0.94
44	1.00	0.96	**55**	0.94	0.93
56	1.00	0.97	**39**	0.94	0.94
59	1.00	0.94	**57**	0.94	0.94
60	1.00	0.94	**49**	0.94	0.93
61	1.00	1.00	**52**	0.94	0.91
62	1.00	0.94	**41**	0.93	0.91
63	1.00	0.95	**29**	0.93	0.97
64	1.00	0.90	**37**	0.93	0.94
65	1.00	1.00	**47**	0.93	0.94
66	1.00	0.95	**67**	0.92	0.93
70	1.00	0.94	**28**	0.92	0.93
71	1.00	0.91	**32**	0.92	0.93
77	1.00	0.92	**23**	0.91	0.89
78	1.00	0.91	**51**	0.91	0.90
84	1.00	0.94	**22**	0.91	0.98
87	1.00	0.90	**31**	0.91	0.93
88	1.00	0.93	**13**	0.91	0.94
90	1.00	0.93	**75**	0.91	0.88
68	1.00	0.95	**43**	0.90	0.87
53	1.00	0.97	**26**	0.90	0.94
82	1.00	0.93	**3**	0.89	0.95
80	0.99	0.94	**83**	0.89	0.90
25	0.99	1.01	**50**	0.89	0.90
86	0.99	0.92	**79**	0.89	0.86
72	0.98	0.94	**20**	0.89	0.94
81	0.98	0.92	**18**	0.88	0.91
69	0.98	0.94	**15**	0.88	0.89
76	0.98	0.95	**4**	0.88	0.93
73	0.98	0.93	**5**	0.87	0.89
58	0.98	0.90	**2**	0.86	0.88
54	0.97	0.93	**35**	0.85	0.87
38	0.97	0.96	**21**	0.85	0.93
24	0.96	0.95	**19**	0.83	0.87
42	0.96	0.95	**6**	0.82	0.89
36	0.96	0.98	**12**	0.81	0.85
46	0.96	0.96	**7**	0.81	0.86
89	0.96	0.90	**10**	0.81	0.87
33	0.95	0.94	**11**	0.80	0.83
34	0.95	0.95	**1**	0.79	0.87
85	0.95	0.90	**8**	0.78	0.84
45	0.95	0.97	**17**	0.76	0.86
48	0.95	0.92	**9**	0.76	0.83
27	0.95	0.96	**14**	0.76	0.89
			Mean	0.9321	0.9252

The unadjusted scores are positively correlated with the adjusted scores (Pearson correlation = 0.6963, significant at the 1% level). The unadjusted ranks are also positively correlated with the adjusted ranks (Spearman correlation = 0.5948, significant at the 1% level).

2.5 Discussion and concluding remarks

This study assesses the efficiency of 90 public primary schools located in the State of Geneva, Switzerland. A two-stage procedure is applied. At the first stage, Data Envelopment Analysis (DEA), a performance measurement technique, is used. At the second stage, efficiency scores are regressed on environmental variables.

The results of the first stage show that, in average, the efficiency of primary schools is equal to 93%. By improving the operation of schools, 7% (100 – 93) of inputs could be saved, representing 17'744'656 Swiss francs in 2010[40]. This value constitutes a target for public managers in a context of budget restriction. An individual efficiency score is calculated for each school. In terms of policy implication, this result allows the avoidance of an identical linear cut among schools, as public managers specifically know by how much input must be decreased in each school in order to make it efficient.

The results of the second stage show that four environmental variables influence school efficiency. First, the hypothesis stating that operations being held on multiple sites affect efficiency is confirmed. Multi-site schools are associated with lower efficiency. This latter variable has never been tested so far. Second, the proportion of disadvantaged pupils reduces school efficiency. This finding is in line with previous studies, such as Alexander *et al.* (2010) and McCarty and Yasawarng (1993). Third, efficiency is negatively influenced by the provision of special education. Borge and Naper (2006) and Rassouli-Currier (2007) find similar results. Fourth, school size is associated with higher efficiency, confirming similar results in previous studies (Alexander *et al.*, 2010; Borge and Naper, 2006; Bradley *et al.*, 2001). The identification of these determinants of school efficiency leads to policy and management related implications. These implications are discussed hereafter.

A potential way to tackle the difficulty of managing multi-site schools consists of developing and using information and communication technology (ICT). As advocated by Selwood and Visscher (2008), the use of school information systems enhances school improvement. The use of ICT by headteachers and teachers could help them in their day-to-day management of the classrooms or school sites (distance learning, distance management, etc.). As ICT is actually at an embryonic state in Swiss public schools, an investment effort would first be needed from the State and the local authorities. Another policy implication would consist of concentrating, wherever possible, school buildings of the same

40 Statistics of public operational expenses of the State of Geneva are available at : http://www.bfs.admin.ch/bfs/portal/fr/index/themen/15/02/data/blank/01.html.

administrative entity on a common site. Planning construction of new schools should take into account that it is preferable to reduce the number of sites.

Priority education policy (PEP) is often advocated to correct the negative influence of disadvantaged pupils on school performance. Unfortunately, PEP does not seem to improve pupil performance neither in Europe (Demeuse *et al.*, 2008) nor in the State of Geneva (Souci and Nidegger, 2010). In terms of policy implication, this finding means that other actions need to be taken.

In order to define these actions, one has to identify the social-class differences which explain why disadvantaged children underperform. Rothstein (2010) demonstrates that childrearing and literary practices, health characteristics, housing stability and economic security influence pupil achievement. Children with a low socioeconomic status are disadvantaged in all these areas. For instance, less-educated parents read to young children less often and less consistently; disadvantaged children are in poorer health – mental health, asthma, acute illness, etc. – (see also Currie and Goodman, 2010, for a review about the impact of health on education achievement); they are confronted with housing instability; they suffer from parents confronted with unemployment. Evidence shows that these variables impair skill acquisition. Rather than allocating more resources to schools, policymakers should therefore focus on related social policies. For instance, they could define pre-school, family, health, housing and benefits policies in order to improve the conditions for disadvantaged children.

Special education is intended for pupils with special needs. It is traditionally provided in separate classes. However, a new policy ruling the integration of children and young people with special needs came into force in the State of Geneva in 2010. It states that integrative solutions are preferred to separative solutions. A move towards integrating pupils with special needs into regular classes could increase school efficiency, as underachieving pupils are positively influenced by overachieving pupils. The impacts of this new policy still have to be evaluated.

Finally, increasing the number of pupils is associated with higher efficiency. Such a finding could suggest that schools are evolving in a situation of increasing returns to scale. The DEA model performed in the first stage does not allow the study to confirm or deny this assumption. This is due to the ratio formulation of the inputs and the outputs which prevents the calculation of scale efficiency (Hollingsworth and Smith, 2003). However, Leithwood and Jantzi (2009) show that primary schools serving socially heterogeneous pupils should be limited in size to not more than 500 pupils in order to maximise efficiency. In the State of Geneva, the average school has 381 pupils, leaving room for improvement (and larger schools). In terms of policy making, the existing – and rigid – class size regulation could be replaced by a more flexible one, allowing headteachers to increase the total number of pupils by increasing class size.

In the context of the State of Geneva, the policy-related implications of the current study could be summarized as follows. New schools or existing multi-site schools should be concentrated on common sites; if this is not possible, the use of ICT in school management and teaching should be developed and encouraged. In order to correct the negative influence of disadvantaged pupils on school performance, policymakers should focus on related social policies, such as pre-school, health, housing and benefits policies, rather than on allocating additional resources to schools. Finally, with an average of 381 pupils per school, school size could be increased to maximize school efficiency.

The current study has some shortcomings, some of them calling for further studies:

- Referring to the KLEMS input framework (OCDE, 2001), inputs involved in the first-stage model could also include variables which adequately reflect capital, energy, materials and services used by schools. Such variables are unfortunately either unavailable or only available disaggregated at school level in the State of Geneva. However, the model used in this study includes three inputs, two of which measure the use of labour and correspond to 95% of the public education operating expenses in the State of Geneva. As a result, it is reasonable to consider that the results obtained are robust.
- Outputs involved in the first-stage DEA model of this study reflect quality (pupils' results) and not quantity (such as the number of pupils). As test scores are measured as an average per pupil, information about the size effect is lost. As a result, scale efficiency cannot be measured in the first stage, and no information is produced about the nature of returns to scale. In order to assess scale efficiency, further studies could investigate how the outputs should be formulated. For instance, a possible way to bypass the ratio form of the average test scores could be to multiply the average test scores by the number of pupils.
- Outputs could also include variables reflecting other aspects of human capability (and not only test scores). Unfortunately, in the State of Geneva, such aspects are either not defined or, if defined, not measured.
- This study uses cross-sectional data (school year 2010-2011). As a result, the analysis cannot capture how one or several variables can influence another variable with a time lag. Ideally, time series data would include lagged explanatory variables in the second stage of the DEA model.
- Finally, the current study presents a quantitative analysis. As advocated by Badillo and Paradi (1999), the measurement of efficiency by the use of a quantitative method could advantageously be complemented by a qualitative survey. For instance, Mancebón and Bandrés (1999) interview headteachers of efficient schools in order to identify the best practices that characterize those efficient schools. The current survey could be extended by realizing a qualitative survey of headteachers and eventually of school board members, pupils and parents.

The current study also contains original features. First, the second-stage model is tested for endogeneity. This is rarely done in a two-stage DEA analysis. Models suffering from endogeneity produce biased results. Future studies should therefore systematically test the model for endogeneity. Second, the current study has identified the negative impact of multi-site schools on efficiency. Further studies should confirm this finding in the education sector but also in other fields which are concerned with operations being held on multiple sites, such as the health-care sector.

References

Abbott, M. & Doucouliagos, C. (2000). Technical and scale efficiency of vocational education and training institutions: The case of the New Zealand polytechnics. *New Zealand Economic Papers*, *34*(1), 1-23.

Abbott, M. & Doucouliagos, C. (2003). The efficiency of Australian universities: a data envelopment analysis. *Economics of Education Review*, *22*(1), 89-97.

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., Bonomi, F. & Sibiano, P. (2014). Measuring the "managerial" efficiency of public schools: a case study in Italy. *International Journal of Educational Management*, *28*(2), 120-140.

Ahn, T., Arnold, V., Charnes, A. & Cooper W. W. (1989). DEA and ratio efficiency analyses for public institutions of higher learning in Texas. *Research in Governmental and Nonprofit Accounting*, *5*, 165-185.

Ahn, T. & Seiford, L. M. (1993). Sensitivity of data envelopment analysis to models and variable sets in a hypothesis test setting: the efficiency of university operations. In Y. Ijiri (Eds.), *Creative and Innovative Approaches to the Science of Management* (pp. 191-208). Westport: Quorum Books.

Alexander, W. R. J. & Jaforullah, M. (2004). Explaining efficiency differences of New Zealand secondary schools. *Economics Discussion Papers No. 0403*. Dunedin: University of Otago.

Alexander, W. R. J., Haug, A. A. & Jaforullah, M. (2010). A tow-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools. *Journal of Productivity Analysis*, *34*(2), 99-110.

Arcelus, F. J. & Coleman, D. F. (1997). An efficiency review of university departments. *International Journal of Systems Science*, *28*(7), 721-729.

Avkiran, N. K. (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Economic Planning Science*, *35*(1), 57-80.

Badillo, P.-Y. & Paradi, J. C. (1999). *La méthode DEA: analyse des performances*. Paris: HERMES Science Publications.

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.

Beasley, J. E. (1990). Comparing university departments. *Omega*, *18*(2), 171-183.

Beasley, J. E. (1995). Determining the teaching and research efficiencies. *Journal of the Operational Research Society*, *46*(4), 441-452.

Bessent, A. M. & Bessent, E. W. (1980). Determining the comparative efficiency of schools through data envelopment analysis. *Educational Administration Quarterly*, *16*(2), 57-75.

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

Blank, J. L. T., van Hulst, B. L., Koot, P. M. & van der Aa, R. (2012). Benchmarking overhead in education: a theoretical and empirical approach. *Benchmarking: An International Journal*, *19*(2), 239-254.

Borge, L.-E. & Naper L. R. (2006). Efficiency Potential and Efficiency Variation in Norwegian Lower Secondary Schools. *FinanzArchiv / Public Finance Analysis*, *62*(2), 221-249.

Bowerman, B. L. & O'Connell, R. T. (1990). *Linear Statistical Models: An Applied Approach*. Boston: Duxbury Press.

Bradley, S., Johnes, J. & Little, A. (2010). Measurement and determinants of efficiency and productivity in the further education sector in England. *Bulletin of Economic Research*, *62*(1), 1-30.

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.

Burney, M. A., Johnes, J., Al-Enezi, M. & Al-Mussalam, 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 and 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*, *14*(2), 143-160.

Charnes, A, Cooper, W. W. & Rhodes E. L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429-444.

Coelli, T. J., Prasada Rao, D. S., O'Donnel, C. J. & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York: Springer.

Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data Envelopment Analysis: A comprehensive Text with Models, Applications, References and DEA-Solver Software*. New York: Springer.

Currie, J. & Goodman, J. (2010). Parental Socioeconomic Status, Child Health, and Human Capital. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of Education* (pp. 156-162). Oxford: Elsevier.

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

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.

Demeuse, M., Frandji, D., Greger, D. & Rochex, J.-Y. (2008). *Les politiques d'éducation prioritaire en Europe: Conceptions, mises en œuvre, débats*. Lyon : Institut national de recherche pédagogique.

Demir, I. & Depren, Ö. (2010). Assessing Turkey's secondary schools performance by different region in 2006. *Procedia – Social and Behavioral Sciences*, *2*(1), 2305-2309.

Denaux, Z. S., Lipscomb, C. A. & Plumly, L. W. (2011). Assessing the technical efficiency of public high schools in the state of Georgia. *Review of Business Research*, *11*(5), 46-57.

Diagne, D. (2006). Mesure de l'efficience technique dans le secteur de l'éducation : une application de la méthode DEA. *Revue suisse d'économie et de statistique*, *142*(2), 231-262.

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

Essid, H., Ouellette, P. & Vigeant, S. (2013). Small is not beautiful after all: measuring the scale efficiency of Tunisian high schools using a DEA-bootstrap method. *Applied Economics*, *45*(9), 1109-1120.

Fuentes, A. (2011). Raising Education Outcomes in Switzerland. *OECD Economics Department Working Papers No. 838*. Paris: OECD Publishing.

Grosskopf, S. & Moutray, C. (2001). Evaluating performance in Chicago public high schools in the wake of decentralization. *Economics of Education Review*, *20*(1), 1-14.

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

Hollingsworth, B. & Smith, P. (2003). Use of ratios in data envelopment analysis. *Applied Economics Letters*, *10*(11), 733-735.

Hu, Y., Zhang, Z. & Liang, W. (2009). Efficiency of primary schools in Beijing, China: an evaluation by data envelopment analysis. *International Journal of Educational Management*, *23*(1), 34-50.

Jeon, Y. & Shields, M. P. (2008). Integration and utilization of public education resources in remote and homogenous areas: a case study of the upper peninsula of Michigan. *Contemporary Economics Policy*, *23*(4), 601-614.

Johnes, J. (2004). Efficiency measurement. In G. Johnes & J. Johnes (Eds.), *International Handbook on the Economics of Education* (pp. 613-742). Cheltenham: Edward Elgar Publishing.

Kantabutra, S. & Tang, J. C. S. (2006). Urban-rural and size effects on school efficiency: The case of Northern Thailand. *Leadership and Policy in Schools*, *5*(4), 355-377.

Kirjavainen, T. & Loikkanen, 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.

Leithwood, K. & Jantzi, D. (2009). A Review of Empirical Evidence About School Size Effects: A Policy Perspective. *Review of Educational Research*, *79*(1), 464-490.

Lovell, C. A. K., Walters, L. C. & Wood, L. L. (1994). Stratified models of education production using modified data envelopment analysis and regression analysis. In A. Charnes, W. W. Cooper, A. Y. Lewin & L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology and applications* (pp. 329-351). Dordrecht: Kluwer Academic.

McCarty, T. A. & Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. In H. O. Fried, C. A. K Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications* (pp. 271-287). New York: Oxford University Press.

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

McMillan, M. L. & Datta, D. (1998). The Relative Efficiencies of Canadian Universities: A DEA Perspective. *Canadian Public Policy / Analyse de Politiques*, *24*(4), 485-511.

Mancebón, M. J. & Bandrés, E. (1999). Efficiency Evaluation of Secondary Schools: the key role of model specification and of ex post analysis of results. *Education Economics*, *7*(2), 131-152.

Mancebón, M. J. & Mar Molinero, C. (2000). Performance in Primary Schools. *The Journal of the Operational Research Society*, *51*(7), 843-854.

Mante, B. and O'Brien, G. (2002). Efficiency measurement of Australian public sector organisations: the case of state secondary schools in Victoria. *Journal of Educational Administration*, *40*(3), 274-296.

Martin, M. (2006). Efficiency and Quality in the Current Higher Education Context in Europe: an application of the data envelopment analysis methodology to performance assessment of departments within the University of Zaragoza. *Quality in Higher Education*, *12*(1), 57-79.

Meunier, M. (2008). Are Swiss Secondary Schools Efficient? In N.C. Soguel & P. Jaccard (Eds.), *Governance and Performance of Education Systems* (pp. 187-202). Dordrecht: Springer.

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

Myers, R. (1990). *Classical and Modern Regression with Applications*. Boston: Duxbury Press.

Noreisch, K. (2007). School catchment area evasion: the case of Berlin, Germany. *Journal of Education Policy*, *22*(1), 69-90.

Observatory on Primary Education (2010). *Allocation des ressources aux établissements* (Report of December 2010). Geneva: General Direction of Primary Schools, Education Department, State of Geneva.

Olivares, M. & Schenker-Wicki, A. (2010). How do Swiss Universities master the reform of the last ten years? Empirical evidence from a data envelopment analysis. *Zurich Open Repository and Archive*. Zurich: University of Zurich.

Olivares, M. & Schenker-Wicki, A. (2012). The Dynamics of Productivity in the Swiss and German University Sector: A Non-Parametric Analysis That Accounts for Heterogeneous Production. *UZH Business Working Paper No. 309*. Zurich: University of Zurich.

Organisation for Economic Co-operation and Development (2001). *Measuring Productivity: Measurement of Aggregate and Industry-level Productivity Growth*. Paris: OECD.

Ouellette, P. & Vierstraete, V. (2005). An evaluation of the efficiency of Québec's school boards using the Data Envelopment Analysis method. *Applied Economics*, *37*(14), 1643-1653.

Ramanathan, R. (2001). A Data Envelopment Analysis of Comparative Performance of Schools in the Netherlands. *Opsearch*, *38*(2), 160-182.

Rassouli-Currier, S. (2007). Assessing the Efficiency of Oklahoma Public Schools: A Data Envelopment Analysis. *Southwestern Economic Review*, *34*(1), 131-144.

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

Rothstein, R. (2010). Family Environment in the Production of Schooling. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of Education* (pp. 148-155). Oxford: Elsevier.

Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, *90*(3), 553-565.

Ruggiero, J. (2000). Nonparametric estimation of returns to scale in the public sector with an application to the provision of educational services. *Journal of the Operational Research Society*, *51*(8), 906-912.

Ruggiero, J. & Vitaliano, D. F. (1999). Assessing the efficiency of public schools using data envelopment analysis and frontier regression. *Contemporary Economic Policy*, *17*(3), 321-331.

Sarrico, C. S., Rosa, M. J. & Coelho, I. P. (2010). The performance of Portuguese secondary schools: an exploratory study. *Quality Assurance in Education*, *18*(4), 268-303.

Schenker-Wicki, A. & Hürlimann, M. (2006). Universités suisses : échec ou succès du financement fondé sur les résultats ? Analyse a posteriori. *Politiques et gestion de l'enseignement supérieur*, *18*(1), 61-78.

Selwood, I. & Visscher, A. J. (2008). The potential of School Information Systems for Enhancing School Improvement. In. N. C. Soguel & P. Jaccard (Eds.), *Governing and Performance of Education Systems* (pp. 269-288). Dordrecht: Springer.

Sengupta, J. K. (1990). Tests of efficiency in DEA. *Computers Operations Research*, *17*(2), 123-132.

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

Smith, P. C. & Street, A. (2006). *Analysis of Secondary School. Efficiency: Final report* (Research Report RR788). London: Department for Education and Skills.

Souci, A. & Nidegger, C. (2010). *Le réseau d'enseignement prioritaire à Genève : quels effets sur les acquis des élèves après quelques années ?* Genève : Service de la recherche en éducation.

Tobin, J. (1958). Estimation for relationships with limited dependent variables. *Econometrica*, *26*(1), 24-36.

Van Zanten, A. (2003). Middle-class Parents and Social Mix in French Urban Schools: reproduction and transformation of class relations in education. *International Studies in Sociology of Education*, *13*(2), 107-123.

Waldo, S. (2007). Efficiency in Swedish Public Education: Competition and Voter Monitoring. *Education Economics*, *15*(2), 231-251.

Wössmann, L. (2005). The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, *13*(2), 143-169.

Wolter, S. (2010). *Swiss Education Report 2010*. Aarau: Swiss Coordination Centre for Research in Education.

3

DEA does not like positive discrimination: a comparison of alternative models based on empirical data

Structured abstract

Purpose

Within Data Envelopment Analysis (DEA), several alternative models allow for an environmental adjustment. The aim of this study is to test how these models, each of which measure efficiency, potentially lead to diverging results.

Design/methodology/approach

Five alternative models, each user-friendly and easily accessible to practitioners and decision makers, are performed using empirical data of 90 primary schools in the State of Geneva, Switzerland. The models are compared with one another on the basis of several indicators: mean efficiency, median efficiency, minimum efficiency, maximum efficiency, number of efficient schools, Pearson correlation and Spearman rank correlation. A Wilcoxon signed rank sum test is also performed.

Findings

The models are compared on the basis of a pairwise comparison. Except for two pairs of models (out of 15) whose results seem to converge, each and every other pair of models provides diverging results. In other words, the majority of alternative models deliver divergent results. This finding is valid for the specific empirical dataset used in the current study. For this reason, it cannot be generalized to other datasets. However, the fact that the efficiency scores

diverge in the current study may suggest that the results obtained from several alternative models may diverge in other cases too.

Practical implications

Applied DEA studies traditionally end with recommendations and policy implications. Most of these studies base their recommendations on the efficiency results produced by a particular DEA model. This appears to be problematic. As shown in this study, several alternative models to measure efficiency, within DEA, deliver diverging results. Consequently, recommendations and policy implications may differ according to the model used. From a political standpoint, these diverging results could potentially lead to ineffective decisions. From an applied research standpoint, they should represent a serious matter of concern. And from a decision making standpoint, they may lead to opposing managerial choices.

Originality/value

Unlike studies using simulated data, the current study intentionally uses empirical data in order to address the issue faced by practitioners and decision makers who perform their own efficiency analysis.

Five alternative models are compared. With the exception of one existing study, no other existing study tests so many models. The models are all user-friendly and easily accessible to practitioners.

A new model has been developed. It is specifically customized to handle cases of additional funding allocated to disadvantaged schools.

Keywords: data envelopment analysis; environmental variables; comparison.

Article Classification: research paper

JEL classification: C61; D24

3.1 Context

The use of Data Envelopment Analysis (DEA) is experiencing rapid and continuous growth. In 2002, Tavares (2002) identified 3203 DEA publications (journal articles, research articles, event articles, books and dissertations). In 2008, Emrouznejad, Parker and Tavares (2008) inventoried more than 7000 publications. This growth reflects the need for user-friendly performance measurement methods. In recent years, the use of DEA has been further democratized due to *(1)* the existence of free software, such as Win4DEAP, Efficiency Measurement System or DEA Solver, *(2)* the publication of pedagogical guides (Coelli, 1996; Coelli, Prasada Rao, O'Donnell & Battese, 2005, pp. 161-206; Huguenin, 2012; Huguenin, 2013a; Huguenin, 2013b) and *(3)* the teaching of DEA in under- and postgraduate programs[41]. Nowadays, it is quite usual for practitioners and decision makers with little or no background in operational research and economics to run their own efficiency analysis[42]. For instance, a web-based platform integrating DEA has been developed in Portugal for secondary schools' headteachers (Portela, Camanho & Borges, 2011). Users are able to perform their own efficiency analysis by selecting the schools to be included in the dataset and the variables to be included in the analysis[43].

The external environment could influence the ability of management to convert inputs into outputs and, as a result, impact entities' technical efficiency. Following Coelli *et al.* (2005, p. 190), an environmental variable is defined as a factor that could influence the efficiency of an entity, where such a factor is not a traditional input and is assumed to be outside of the manager's control. Because it is not under the control of managers, such a factor is also called a non-discretionary variable[44]. It cannot be varied at the discretion of an individual manager but nevertheless needs to be taken into account to measure efficiency (Cooper, Seiford & Tone, 2007, p. 215). This paper considers traditional inputs as those covered by the OECD KLEMS model, which considers five categories of inputs: capital (K) labour (L), energy (E), materials (M) and services (S) (OECD, 2001).

Examples of environmental variables include ownership differences (such as public versus private), location characteristics, labour relations (such as

---

41 For instance, DEA is taught at the University of Lausanne, Switzerland, in three different courses: *(1)* Public Sector Performance Measurement (Master of Science in Public Policy and Management), *(2)* Public Sector Financial Management (Master of Advanced Studies in Public Administration) and *(3)* Benchmarking (Certificate of Advanced Studies in Administration and Management of Educational Establishments). About 90 decision makers in the public sector are trained annually in the use of DEA.

42 The author of this study regularly meets Swiss headteachers who use DEA to assess individual teachers, classes or schools.

43 Note that this plateform represents an example of 'ascending' benchmarking (Viger, 2007), where the starting point of the analysis comes from the base (i.e. the headteachers).

44 Non-controllable variables and exogenous variables are used as synonyms to non-discretionary variables in the DEA literature.

conflictual versus peaceful relationships between trade unions and employers' organizations) and government regulations (Fried, Schmidt & Yaisawarng, 1999). Location characteristics consist of the environmental variables which are specific to the location of an entity, such as a supermarket influenced by population density.

In the education sector, three main generic drivers can be considered as environmental variables. They influence pupil performance but are outside of the control of headteachers (Soteriou, Karahanna, Papanastasiou & Diakourakis, 1998, p. 68, based on Thanassoulis, 1996, p. 883). They consist of *(1)* pupil characteristics, such as intelligence, willingness or effort propensity, *(2)* family and the external environment, such as the socioeconomic status of pupils and *(3)* school related factors (which are outside of the control of headteachers). In this latter category, school size (as measured by the number of pupils) is, for instance, outside of the control of headteachers in Switzerland, as they have to register every single pupil residing in the catchment area defined by school authorities.

Environmental variables in school efficiency measurement using Data Envelopment Analysis (DEA) include (non exhaustive list):

- school location in a particular region (Agasisti, 2013; Burney, Johnes, Al-Enezi & Al-Musallam, 2013);
- types of school, such as private schools (Agasisti, 2013; Kirjavainen & Loikkanen, 1998; Lovell, Walters and Wood, 1994; Ramanathan, 2001), all-girls schools (Alexander & Jaforullah, 2004; Alexander, Haug & Jaforullah, 2010; Bradley, Johnes & Millington, 2001), urban and rural schools (Agasisti, 2013; Alexander & Jaforullah, 2004; Alexander, Haug & Jaforullah, 2010; Denaux, Lipscomb & Plumly, 2011; Kantabutra & Tang, 2006; Kirjavainen & Loikkanen, 1998);
- socioeconomic status of pupils (Alexander & Jaforullah, 2004; Alexander, Haug & Jaforullah, 2010; Borge & Naper, 2006; Bradley, Johnes & Little, 2010; Denaux *et al.*, 2011; McCarty & Yaisawarng, 1993; Ouellette & Vierstraete, 2005; Rassouli-Currier, 2007);
- school size (Agasisti, 2013; Borge & Naper, 2006; Bradley *et al.*, 2001; Duncombe, Miner & Ruggiero, 1997; Kantabutra & Tang, 2006; Kirjavainen & Loikkanen, 1998);
- political factors (Borge & Naper, 2006; Waldo, 2007);
- teacher characteristics (Alexander & Jaforullah, 2004; Alexander, Haug & Jaforullah, 2010; Bradley, *et al.*, 2001; Burney, Johnes, Al-Enezi & Al-Musallam, 2013; Diagne, 2006; Duncombe, Miner & Ruggiero, 1997; Lovell, Walters & Wood, 1994; Ruggiero & Vitaliano, 1999).

Positive discrimination policies are implemented by public authorities to adjust for the environment[45]. They aim to compensate the negative impact of

[45] Some of these policies are essentially built on an ideological basis (Demeuse & Friant, 2012).

environmental variables (mainly socioeconomic status of pupils) on school performance. In Europe, these priority education policies are defined as

> policies designed to have an effect on educational disadvantaged through systems or programs of focused action (whether the focus be determined according to socioeconomic, ethnic, linguistic, religious, geographic, or educational criteria) by offering something more ('better' or 'different') to designated populations (Frandji, 2008, p. 12).

Within DEA, several models allow for an environmental adjustment. Following Muñiz (2002), they can be grouped in three categories: *(1)* one-stage models (Banker & Morey, 1986a; Banker & Morey, 1986b; Ruggiero, 1996[46]; Yang and Paradi in Muñiz, Ruggiero, Paradi and Yang, 2006), *(2)* multi-stage models including two-stage (Ray, 1988; Ray, 1991), three-stage (Ruggiero, 1998; Fried, Lovell, Schmidt & Yaisawarng, 2002; Muñiz, 2002) and four-stage models (Fried, Schmidt & Yaisawarng, 1999) and *(3)* program analysis models (Charnes, Cooper & Rhodes, 1981)[47]. There are few published studies which compare these models with one another.

The empirical field of this study considers the case of public primary schools in the State of Geneva, Switzerland. As the State of Geneva has implemented upstream positive discrimination measures since 2008, this empirical case is particularly appropriate for an environmental adjustment. The Geneva public school system is described in the next section.

3.2 Geneva public school system

In the State of Geneva, education is compulsory at early childhood (corresponding to the international standard classification of education ISCED # 0) for a duration of 2 years, primary (ISCED # 1) for a duration of 6 years and lower secondary education (ISCED # 2) for a duration of 3 years.

In 2010-2011, the State of Geneva registered 90 public primary schools. These schools are funded by the State government (chiefly for staff salary) and by local authorities – municipalities – (chiefly for school infrastructure). Pupil competences are assessed with the use of standardized tests at three different times in two or three subjects. At the end of the second grade, French (mother tongue) and mathematics are assessed; at the end of the fourth and sixth grade, French, German (first foreign language) and mathematics are assessed.

---

46 Ruggiero (1996) develops an additional one-stage model. However, this model seems to be rather an extension of the Banker and Morey (1986a) model that allows for categorical variables. As it allows continous environmental variables, it is comparable to the Banker and Morey (1986b) model (Ruggiero, 1996, p. 555).

47 Note that Yang and Pollitt (2009) propose the following categories: separative models (in which Charnes, Cooper & Rhodes (1981) and Banker & Morey (1986a) would be classified), one-stage models, two-stage models, three-stage models and four-stage models.

Primary schools are managed by headteachers assisted by one or several teachers working part time as headteachers' assistants. Staff consists of teachers, secretaries and schoolkeepers (maintenance). In some schools, educators are also active.

In order to adjust to local environment, partial autonomy in management is granted to schools. For instance, headteachers define job profiles and recruit teachers; they are responsible for school quality (and hence pupil performance); and they also chair the school board.

Every school has a board composed by representatives of the school staff, parents and city civil-servants and is chaired by the headteacher. The board demonstrates instances of democracy where stakeholders are informed and consulted. Whilst they only have limited authority about school management, they can make propositions about day-to-day school life. School boards aim to develop better relationships between school, families and local communities.

The main characteristics of primary schools are as follows:

- A school can be located on one or several sites (up to five); which implies that school buildings can be spread over several locations (or sites);
- Special education is only available in a limited number of schools (21 schools out of 90); which means that pupils with special needs are grouped in the schools where special education is available;
- Special reception classes for immigrant pupils are only available in a limited number of schools (35 schools out of 90).

The State of Geneva practices a policy of positive discrimination towards schools. Additional teaching resources are allocated to disadvantaged schools. Five school categories (A to E) are defined according to the percentage of pupils (per school) whose parents are blue-collar workers or unqualified workers – category # 9 of the International Standard Classification of Occupations – (Observatory on Primary Education, 2010). This variable, SOCIO, reflects the socioeconomic status of pupils. For instance, schools with a SOCIO proportion of more than 50% are considered as the most disadvantaged schools and are classified in the E category. Table 10 describes the quantity of additional teaching staff per pupil that schools receive.

Table 10
Positive discrimination in Geneva: more teaching staff for disadvantaged schools

Category (# of schools)	Pupils in the lowest socioeconomic category (%)	Pupil/teacher target ratio	Teacher/pupil target ratio	Additional teaching staff per pupil (%)
A (15)	0.00-19.99	18.55	0.0539	0.00
B (20)	20.00-29.99	18.15	0.0551	2.20
C (20)	30.00-39.99	17.45	0.0573	6.30
D (15)	40.00-49.99	16.65	0.0601	11.41
E (20)	50.00-100.00	15.25	0.0656	21.64

Source: General Direction of Primary Schools, Education Department, State of Geneva.

A school in category A has a target teacher/pupil ratio of 0.054 (i.e. 18.55 pupils per teacher). This target is defined by the State authority. Such a school does not receive any additional resources as it is in the most advantaged category. A school in category B has a target teacher/pupil ratio of 0.0551 (i.e. 18.15 pupils per teacher). It receives 2.2% additional teaching staff (i.e. 0.0539 + (0.022 x 0.0539)). A school in category C has a target teacher/pupil ratio of 0.0573 (i.e. 17.45 pupils per teacher). It receives 6.3% additional teaching staff (i.e. 0.0539 + (0.063 x 0.0539)). A school in category D has a target teacher/pupil ratio of 0.0601 (i.e. 16.65 pupils per teacher). It receives 11.41% additional teaching staff (i.e. 0.0539 + (0.1141 x 0.0539)). Finally, a school in category E has a target teacher/pupil ratio of 0.0656 (i.e. 15.25 pupils per teacher). It receives 21.64% additional teaching staff (i.e. 0.0539 + (0.2164 x 0.0539))[48].

3.3 Objective

The aim of this study is to test how several alternative DEA models, each of which measure efficiency, can deliver diverging results. Unlike studies using simulated data, this study intentionally uses empirical data. As a result, the comparison is made between the estimates of the alternative models, without knowing whether these estimates approximate the 'true' efficiency measure (which could be estimated with a simulation analysis)[49]. By using empirical data, this study addresses the issue faced by practitioners and decision makers who perform their own efficiency analysis. If the alternative models produce divergent results, the choice of model becomes a strategic issue.

---

48 As the detrimental condition of the environment is compensated by additional resources, the relevance to actually use a model which allows for an environmental adjustment is open to debate. Consider two schools with one pupil each. Both of them obtain a test's results of 6. The first school faces a detrimental environment and receives 20% additional teaching staff. Instead of having one teacher, it thus has 1.2 teachers. The second school faces a favourable environment. It does not receive additional resources, and stays with one teacher. With a classical DEA model, with no environmental adjustment, the first school obtains an efficiency score of 83.3% and the second one a score of 100%. The first school is penalized for having received additional resources. In order to be 100% efficient, its pupils should obtain a test's result 20% higher than the pupil attending the second school. However, one cannot expect from the disadvantaged pupil to become 20% better than the advantaged pupil. One can probably only expect that the disadvantaged pupil becomes as good as the advantaged pupil. Another point to take into consideration is that the test's results are bounded to a maximum number of points. Consider that 6 is the best grade possible. If both pupils obtain a 6, the first school will always be less efficient, because it is not possible for its pupil to score higher than 6. As a result, it seems appropriate to use a model which allows for an environmental adjustment. With such a model, both schools obtain an efficiency score of 100% in the above mentioned example.

49 Another research question, not treated in this study, would be to determine whether the estimates of alternative models converge or diverge with the 'true' efficiency. This question cannot be answered by using empirical data, as the 'true' efficiency is unknown. The only way to calculate the 'true' efficiency would consist of *(1)* defining a production function, *(2)* generating inputs from a random distribution and *(3)* deriving outputs. Note that existing studies using simulated data provide mixed results about the convergence of alternative models with the 'true' efficiency (see Section 3.5 about it).

The alternative models tested in this study are all user-friendly and easily accessible to practitioners and decision makers. The empirical case is the 90 primary schools of the State of Geneva, Switzerland. It is particularly well suited to test several alternative models, as *(1)* the State of Geneva practices positive discrimination towards disadvantaged schools and *(2)* schools are grouped in five categories defined by one continuous variable (percentage of pupils whose parents are blue-collar workers or unqualified workers). According to their respective category, schools receive additional teaching staff.

3.4 Adjusting for the environment in DEA

Within DEA, several models allow for an environmental adjustment. Following Muñiz (2002), they can be grouped into three categories: *(1)* one-stage models (Banker & Morey, 1986a; Banker & Morey, 1986b; Ruggiero, 1996[50]; Yang & Paradi model in Muñiz, Ruggiero, Paradi & Yang, 2006, p. 1176), *(2)* multi-stage models including two-stage (Ray, 1991), three-stage (Ruggiero, 1998; Fried, Lovell, Schmidt & Yaisawarng, 2002; Muñiz, 2002) and four-stage models (Fried, Schmidt & Yaisawarng, 1999) and *(3)* program analysis models (Charnes, Cooper & Rhodes, 1981).

The models which allow for an environmental adjustment are shortly introduced hereafter, alongside their main advantages and drawbacks (Thanassoulis, Portela & Despic, 2008). The basic variable returns to scale DEA model (VRS) is first recalled (Banker, Charnes & Cooper, 1984). This basic model does not allow for an environmental adjustment.

Banker, Charnes and Cooper (1984) – No environmental adjustment

The basic VRS model measures entities' technical efficiency under the assumption of variable returns to scale.

Following the notation adopted by Johnes (2004, pp. 630-637), there are data on s outputs and m inputs for each of n primary schools to be evaluated (n = 90 in the current study). y_{rk} is the quantity of output r produced by school k. x_{ik} is the quantity of input i consumed by school k. θ_k represents the VRS efficiency of school k (i.e. 'pure' technical efficiency free from any scale inefficiency). λ_j represents the associated weighting of outputs and inputs of entity j.

---

[50] Ruggiero (1996) develops an additional one-stage model. However, this model seems to be an extension of the Banker and Morey (1986a) model that allows for categorical variables. As it allows continous environmental variables, it is comparable to the Banker and Morey (1986b) model (Ruggiero, 1996, p. 555).

The VRS efficiency of the k^{th} school is calculated by solving the following linear problem:

$$\text{Minimize} \quad \theta_k \tag{1}$$

$$\text{Subject to} \quad y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \qquad r = 1, \ldots, s$$

$$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} \geq 0 \qquad i = 1, \ldots, m$$

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

$$\lambda_j \geq 0 \quad \forall j = 1, \ldots, n$$

Banker and Morey (1986a) – One-stage model

The Banker and Morey (1986a) model, also called the categorical model, can be applied when:

- DMUs are grouped into different categories according to the condition of the environment;
- And the environmental variable can be ordered from the least to the most detrimental upon efficiency.

For instance, the 90 primary schools in the State of Geneva are divided into five hierarchical categories (A to E). Schools in category A face the most advantageous environment. Schools in category E face the most detrimental environment. If the measurement of efficiency did not take into account the fact that schools face different environments (i.e. it considered each school to be in the same category), the evaluation would be unfair on the schools facing a difficult environment and too indulgent on the schools facing an advantageous environment.

In the Banker and Morey (1986a) model, 'E' schools are classified as category 1, 'D' schools as category 2, 'C' schools as category 3, 'B' schools as category 4 and 'A' schools as category 5. School efficiency is then evaluated in the following way, using the basic VRS (or constant returns to scale) model:

- Schools in category 1 are only evaluated against schools within this group;
- Schools in category 2 are evaluated with reference to schools in category 1 and 2;
- Schools in category 3 are evaluated with reference to schools in category 1, 2 and 3;
- Schools in category 4 are evaluated with reference to schools in category 1, 2, 3 and 4;

- Finally, schools in category 5 are evaluated with reference to schools in category 1, 2, 3, 4 and 5.

The Banker and Morey (1986a) model evaluates schools under operating handicaps which take into account their particular environments. This ensures that no school is compared to another with a more favourable environment. The VRS formulation of the categorical model is presented in Appendix 1.

Garrett and Kwak (2011) apply the Banker and Morey (1986a) model in the case of 447 school districts in the State of Missouri, USA. They use relative district wealth as the categorical variable with three categorical levels (rich, average and poor).

The main advantage of the Banker and Morey (1986a) model is that it is appropriate for dealing with non-discretionary variables that are qualitative or categorical. Moreover, it is easy to calculate. The method is simple and therefore transparent. There are at least two disadvantages to this approach. First, the various categories have to be ordered hierarchically (from the least to the most favourable). This ordering is not always natural. Second, the Banker and Morey (1986a) model reduces the discriminating power of DEA which depends on the number of entities relative to the number of variables included in the model. As the Banker and Morey (1986a) model considers various sub-samples according to the number of categories, the smaller the sub-sample, the lower the discriminating power between entities that is achieved by DEA (all other things being equal).

Banker and Morey (1986b) – One-stage model

The Banker and Morey (1986b) model directly includes environmental variable(s) as continuous non-discretionary input or output variables in the linear programming formulation. This model takes into account the fact that environmental variables are outside of the control of management and cannot be treated as discretionary factors. As a result, the constraints on the environmental variable are modified. Assuming an input-orientation with variable returns to scale, the inputs are divided into discretionary (x^D) and non-discretionary (x^{ND}) sets. The VRS formulation of the categorical model is presented in Appendix 1.

The environmental variable has to be included as a non-discretionary input or output variable. This implies it is first necessary to decide upon the direction of influence of the environmental variable. Following Coelli *et al.* (2005):

> If the variable is believed to have a positive effect upon efficiency then it should be included in the linear program in the same way as a non-discretionary input would be included. (…). On the other hand, if instead we have a set of 'negative-effect' environmental variables to add to the model then they should be included in the linear program in the same way as a non-discretionary output would be included (p. 192).

Muñiz (2002) strictly applies the Banker and Morey (1986b) model in the context of the education sector. He tests several models considering different non-discretionary variables: percentage of students who usually study more than 10 hours a week; percentage of students who believe that both their parents and teachers have high prospects with regard to their academic future; percentage of students whose annual family income exceeds two and a half million pesetas; percentage of students who did not change teaching centres in that academic year or in the previous; percentage of students who are only child.

Several studies include non-discretionary variables as inputs or outputs but perform a standard DEA model (i.e. a constant returns to scale or a variable returns to scale model) instead of a Banker and Morey (1986b) model. This leads to biased (if not invalid) results. Examples are found in Garrett and Kwak (2011) or in Diagne (2006).

The main advantage of the Banker and Morey (1986b) model is that it is able to accommodate multiple and continuous non-discretionary variables. However, this approach presents various disadvantages:

- Ruggiero (1996) shows that Banker and Morey's (1986b) model formulation leads to referent points that are not feasible. See Ruggiero (2004, pp. 330-331) for a numerical example.
- The Banker and Morey (1986b) model requires a prior understanding and specification of the direction of influence of the non-discretionary variables.

Assuming that the direction of influence of the non-discretionary variables is understood, the Banker and Morey (1986b) model is easy to calculate. The method is simple and therefore transparent.

Ruggiero (1996) – One-stage model

The Banker and Morey (1986b) model defines efficiency with respect to discretionary variables only. Ruggiero (1996) shows that it leads to referent points that are not feasible. He provides a one-stage model to correct this problem by excluding all entities with a more favorable environment from the evaluation of each entity. The Ruggiero (1996) model is quite similar to the Banker and Morey (1986a) model, with the difference that it allows for continuous environmental variables[51].

As the Ruggiero (1996) model is not applied in this study, its formulation is not presented. See Ruggiero (1996, pp.559-560) for the model specification.

Ruggiero (1996) provides an application of the model to the case of school districts in the State of New York. He uses the percentage of adults with college education as an environmental input.

The main advantages of the Ruggiero (1996) model are that *(1)* it is able to accommodate multiple and continuous non-discretionary variables and that *(2)*

---

[51] The Ruggiero (1996) model is an extension of the Banker and Morey (1986a) model that allows for categorical variables. As it allows continous environmental variables, it is comparable to the Banker and Morey (1986b) model (Ruggiero, 1996, p. 555).

it does not lead to non-feasible referent points. However, this approach suffers from various drawbacks:

- Similar to the Banker and Morey (1986b) model, the Ruggiero (1996) model requires a prior understanding and specification of the direction of influence of the non-discretionary variables.
- The Ruggiero (1996) model is not able to consistently handle many non-discretionary variables. As Ruggiero (2004, p. 332) points out

> A potentially more serious problem is the inability to handle many non-discretionary factors. As the number of non-discretionary inputs increases, the probability of overestimating efficiency increases. As a result, inefficient DMUs could be identified as efficient by default. This model does not recognize tradeoffs that exist between the non-discretionary variables; a given DMU under analysis could have a favourable environment because it has favourable levels of most non-discretionary factors but have a limited referent set only because it has an unfavourable level of at least one non-discretionary input (p. 332).

As the Ruggiero (1996) model is not included in a DEA software package, it is not easy to calculate, although the method appears simple and therefore transparent.

Yang and Paradi in Muñiz, Ruggiero, Paradi and Yang (2006, p. 1176) **– One-stage model**

The Yang and Paradi model applies a handicapping measure based on the levels of the non-discretionary variables. Entities with a favourable environment are penalized by the handicapping measure. In such a case, inputs are adjusted with a higher handicap (i.e. they are augmented) and/or outputs are adjusted with a lower handicap (i.e. they are reduced). As a result, adjusted inputs have a higher value than original inputs and adjusted outputs have a lower value than original outputs. The VRS formulation of the Yang and Paradi model is presented in Appendix 1.

Muñiz *et al.* (2006) provide an application of the Yang and Paradi model using simulated data. The decision to adjust data before running a DEA model is supported by Barnum and Gleason (2008).

The main advantage of the Yang and Paradi model is that it does not lessen the discriminating power of DEA, as it does not categorize the entities. The use of handicapping measures presents two disadvantages. First, the direction of influence has to be understood prior to the variables' adjustment. Second, the values of the handicapping measures have to be defined. In most cases, the extent to which the variables have to be augmented or lowered is unclear. In the context of this study, it makes sense to apply the Yang and Paradi model as the handicapping values are known.

Assuming that the handicapping measures h_j and $\hat{h}_j$ have been defined, the Yang and Paradi model is moderately easy to calculate[52]. The method is simple and therefore transparent.

Ray (1991) – Two-stage model

The two-stage model is first introduced by Ray (1988) and further developed by Ray (1991). In the first stage, a basic DEA model (1) is performed using only discretionary variables. After obtaining the technical efficiency scores (TE) from the first stage, Ray (1991) uses an OLS model to regress these scores upon non-discretionary variables in the second stage. The second stage regression is specified as follows:

$$TE_k = \alpha_0 + \beta_1 X_1 + \ldots \beta_v X_v + e_k \qquad (2)$$

The error term represents the efficiency. Since Ray (1991), other types of regression have been used in the second stage. For instance, McCarty and Yaisawarng (1993) are the first to use a Tobit regression.

Applications of the two-stage models in the education sector include Agasisti (2013), Borge and Naper (2006), Burney, Johnes, Al-Enezy and Al-Musallam (2013), Denaux *et al.* (2011), Rassouli-Currier (2007) or Waldo (2007).

According to Coelli (2005, pp. 194-195), the two-stage model presents the advantages of being able to accommodate *(1)* more than one variable and *(2)* both categorical and continuous variables. Moreover, it does not require a prior understanding of the direction of influence of the non-discretionary variables. It is also easy to calculate. The method is simple and therefore transparent. As the second stage introduces a regression analysis, the Ray (1991) model presents the disadvantages inherent to such techniques. Mainly, it requires the specification of a functional form to the regression model. Any misspecification may distort the results. Cordero *et al.* (2009) also point out that the adjustment of efficiency scores takes into account (only) the radial component of inefficiency and not the potential inefficiency derived from slacks[53].

---

[52] Priority education policies or "PEPs" (also known as positive discrimination policies) aim to compensate for the negative impact of environmental variables (mainly socioeconomic status of pupils) on school performance. Such policies have been introduced in the US, England, Belgium, France, Greece, Portugal, Czech Republic, Romania or Sweden (Demeuse, Frandji, Greger & Rochex, 2012). Additional funding allocated to disadvantaged schools is an example of PEPs focused on the institutions. When the additional funding is known, the value of the handicapping measure of the Yang and Paradi model can easily be calculated. This is the case in the context of this study.

[53] Efficiency scores generated by DEA are similar with or without the calculation of slacks. In the two-stage method, the coefficients of the regression are calculated towards the efficiency scores as a dependent variable. Their values will be identical whether these scores belong to entities whose inefficiency is composed by only a radial factor or a radial and a slack factor.

Ruggiero (1998) – Three-stage model

The first two stages of the Ruggiero (1998) model are identical to those used in Ray (1991). In the third stage, the parameters estimated from the second stage regression are used to construct an index for the non-discretionary variables. The following index x^{ND} is considered: $x^{ND} = \sum_{u=1}^{v} \beta_u x_u^{ND}$, where v is the number of non-discretionary variables. The DEA model is run again in the third stage by using the index for the non-discretionary variables to exclude all entities with a more favourable environment from the evaluation of each entity[54].

As the Ruggiero (1998) model is not applied in this study; its formulation is not presented. See Ruggiero (2004, pp. 333-334) for the model specification.

Ruggiero (1998; 2004) provides an application of his three-stage model using simulated data.

The advantages and disadvantages of the Ray (1991) model apply to stage one and two of the Ruggiero (1998) model. An additional disadvantage arises in the third stage, as the efficient entities (on the frontier) are the same as those which would be computed by using a DEA model in which all variables were discretionary. This is the case because the efficiency frontier is the same in both situations. As a result, only the scores of the inefficient entities are modified by the Ruggiero (1998) model. This approach is difficult to calculate. It is sophisticated and therefore not transparent.

Muñiz (2002) – Three-stage model

The first stage of the Muñiz (2002) model uses model # 1 (with only discretionary variables) to compute technical efficiency scores. Muñiz's (2002) following approach focuses on the slacks, which are added in model # 3 hereafter. Considering output slacks, s_r, and input slacks, s_i, the model can be described:

Minimize $\quad \theta_k - \varepsilon \sum_{r=1}^{s} s_r - \varepsilon \sum_{i=1}^{m} s_i \qquad (3)$

Subject to $\quad y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} + s_r = 0 \qquad r = 1, \ldots, s$

$$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} - s_i = 0 \qquad i = 1, \ldots, m$$

54 The specification of the Ruggiero (1998) three-stage model is therefore similar to the specification of the Ruggiero (1996) one-stage model. It only replaces the original values of the environmental variables in the one-stage model by the index of environmental variables in the third-stage model.

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

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

Here, ε is a non-Archimedean value defined to be smaller than any positive number, but greater than 0.

In the Muñiz (2002) model, the total slack values in each variable, defined as the sum of the radial and the non-radial movements, are used[55]. By taking into account the total slack, the model avoids losing information from the non-radial movement.

The slacks computed by model # 3 are confounded by the influence of the non-discretionary variables (i.e. the non-discretionary inputs in the input-oriented model), since they have not been included in the first stage.

The objective of the second stage is to distinguish between the slacks associated with *(1)* the real technical inefficiency of the entity and *(2)* the non-discretionary variables. A separate DEA analysis is performed for the slacks of each (discretionary) variable. The model must therefore be run once for each discretionary variable. The slack detected for every entity in a specific variable is used as a variable itself (to be minimized) in the respective DEA models. The objective of this second stage is to minimize the slacks in a discretionary variable subject to the non-discretionary variables. In other words, the second stage determines the minimum amount of slacks achievable by an entity for a specific variable subject to the value of the non-discretionary variables.

To perform the third stage, original data of each entity are adjusted by removing the slack values associated with the non-discretionary variables. The third stage consists of a DEA model which uses the adjusted data values of the discretionary variables. The technical efficiency scores are not confounded by the influence of non-discretionary variables anymore, as the slacks calculated in the third stage are due exclusively to the inefficient performance of the entity.

As the Muñiz (2002) model is not applied in this study; its formulation is not presented. See Muñiz (2002, pp. 628-631) for the model specification.

Muñiz (2002) applies his model to 62 high schools in the State of Asturias, Spain. In the second stage, he uses the following variables as non-discretionary inputs: percentage of students who usually study more than 10 hours a week; percentage of students who believe that both their parents and teachers have high prospects with regard to their academic future; percentage of students whose annual family income exceeds two and a half million pesetas; percentage of students who did not change teaching centres in that academic year or in the previous; percentage of students who are only child.

---

[55] By contrast, the Ruggiero (1998) model only takes the radial movement into account in the second stage.

The main advantage of the Muñiz (2002) model is the use of non-parametric techniques in every stage[56]. As a result, no functional form has to be specified. This is useful when the productive process of entities under analysis is unknown. The Muñiz (2002) model also captures information included in the slacks. High cost of time and calculation are the main disadvantages of this approach, which is sophisticated and therefore not transparent.

Fried, Lovell, Schmidt and Yaisawarng (2002) – Three-stage model

The first stage of the Fried *et al.* (2002) model uses model # 3 (with only discretionary variables) to compute technical efficiency scores. The slacks are broadly interpreted as being composed of three effects: the influence of the environment (first effect), inefficiencies due to management (second effect) and statistical noise arising from measurement errors (third effect). The second stage aims to decompose the slacks into these three effects using stochastic frontier analysis (SFA).

In the second stage, the dependent variables are the total input slacks (radial and non-radial movements). They are regressed against the non-discretionary variables (first effect). SFA separates residual into two parts: managerial inefficiencies (second effect) and statistical noise (third effect).

In the third stage, discretionary variables data are adjusted in a manner that accounts for the influence of the environment and statistical noise. The first stage is then repeated using the adjusted data, providing technical efficiency scores devoid of environmental effects and statistical noise.

As the Fried *et al.* (2002) model is not applied in this study; its formulation is not presented. See Fried *et al.* (2002, pp. 160-164) for the model specification.

As far as the author is aware, the Fried *et al.* (2002) model has not been applied to the education sector. Applications of this model are found in Yanyan (2012) with respect to commercial banks; Shang, Hung, Lo and Wang (2008) with respect to hotels or Lee (2008) with respect to paper companies.

The Fried *et al.* (2002) model presents the advantages of being able to accommodate the following into the second stage: *(1)* more than one variable and *(2)* both categorical and continuous variables. Moreover, it does not require a prior understanding of the direction of influence of the non-discretionary variables and it captures the information included in the slacks. As the second stage introduces a SFA, the Fried *et al.* (2002) model presents the disadvantages inherent to such technique. As the residual is separated into an error component and an inefficiency component in SFA, it requires specification of the distributional form of the efficiency component. Any misspecification may distort the results. The Fried *et al.* (2002) model is

[56] The use of non-parametric techniques in every stage has also a drawback, as it is sensitive to outliers.

difficult to calculate and time-consuming. The method is sophisticated and therefore not transparent.

Fried, Schmidt and Yaisawarng (1999) – Four-stage model

The first stage uses model # 3 (with only discretionary variables) to compute technical efficiency scores. In the second stage, the total slacks are regressed upon the environmental variables. In the third stage, the parameters estimated from the second stage regression are used to predict the total input slacks (if the model is input-oriented) or the total output surplus (if the model is output-oriented). These predicted values are used to calculate adjusted values of the original inputs or outputs. In the fourth stage, the DEA model is run again using the adjusted data. It provides technical inefficiency scores devoid of environmental influence.

As the Fried *et al.* (1999) model is not applied in this study; its formulation is not presented. See Fried *et al.* (1999, pp. 252-255) for the model specification.

Sav (2013) and Cordero-Ferrara, Pedraja-Chaparro and Salinas-Jiménez (2008) provide the only two existing applications of the Fried *et al.* (1999) model in the education sector. Sav (2013) measures technical efficiency of 227 universities. Three environmental variables are used in the second stage: the state and local government contribution to public university operating expenses per full-time equivalent student; the number of high school students per 1000 that score at the 80th percentile and above on either the SAT or ACT tests; the number of college freshmen imported from other states relative to the number of resident freshman attending college out-of-state. Cordero-Ferrara *et al.* (2008) measure the efficiency of 80 high-schools in the State of Extramaduria, Spain. They use three non-discretionary components in the second stage. These components are derived from eleven non-discretionary variables using Principal Component Analysis.

The Fried *et al.* (1999) four-stage model presents the advantages of being able to accommodate in the second stage *(1)* more than one variable and *(2)* both categorical and continuous variables. Moreover, it does not require a prior understanding of the direction of influence of the non-discretionary variables. It captures the information included in the slacks. As the second stage introduces a regression analysis, the Fried *et al.* (1999) model presents the disadvantages inherent to such techniques. Mainly, it requires the specification of a functional form to the regression model. Moreover, a significant relationship between the slacks and the environmental variable has to be identified in order to apply this approach. The Fried *et al.* (1999) model is moderately complicated to calculate. The method is sophisticated and therefore not transparent.

Charnes, Cooper and Rhodes (1981) – Program analysis model

The program analysis model developed by Charnes *et al.* (1981) is an alternative approach to the previous ones. Its objective is not to adjust the efficiency scores to the environment but to reveal potential efficiency differences between several 'programs'. The Charnes *et al.* (1981) model consists of three steps.

In the first step, the entire sample is divided into sub-samples of entities facing the same environment (or operating the same 'program'). DEA models are solved for each sub-sample separately. In the second step, all observed data points are projected onto their respective frontiers to 'artificially' eliminate inefficiency attributed to management. Finally, a single DEA model is run using the data projected values. Note that remaining technical inefficiency can be attributed, in this model, to environmental variables[57].

The first application of the program analysis model in the education sector was produced by Charnes *et al.* (1981). Schools running under the 'Program Follow Through' are compared to schools not running under this program[58]. Other applications include Portela and Thanassoulis (2001), Soteriou, Karahanna, Papanastasiou and Diakourakis (1998) or Diamond and Medewitz (1990). Portela and Thanassoulis (2001) use the program analysis model to assess pupils within schools of the same type and within schools of all types. Soteriou *et al.* (1998) assess the efficiency of secondary schools in Cyprus. They separate schools into two groups operating in an urban or a rural environment. Diamond and Medewitz (1990) assess the efficiency of high-school classes. They consider two categories of classes: in the first one, the Developmental Economic Education Program is applied; in the second one, it is not.

The main advantage of the Charnes *et al.* (1981) model is that it is appropriate for dealing with non-discretionary variables that are qualitative or categorical. Moreover, it can be applied even when there is no natural ordering of the environmental variable. This means that the direction of influence does not need to be specified. The model is easy to calculate. It is simple and therefore transparent. The main disadvantage of the Charnes *et al.* (1981) model is that it lessens the discriminating power of DEA, which depends on the number of entities relative to the number of variables included in the model. As the Charnes *et al.* (1981) model considers various sub-samples, the smaller the sub-sample, the lower the discriminating power between entities that is achieved by DEA (all other things being equal).

---

57 This is a major difference between the program analysis model and other models. The remaining technical inefficiency in all other models can be attributed to managerial inefficiency.

58 The 'Program Follow Through' was launched in 1968 for a period of ten years in the United States as a federally sponsored program providing health, educational and social services to disadvantaged early primary school pupils and their family.

3.5 Comparing the models: a literature review

Various studies have conducted benchmark analysis of alternative methods to measure efficiency (such as COLS, SFA, DEA or Free Disposal Hull). Evidence suggests that the choice of technique affects efficiency scores and rankings of entities. See Johnes (2004, pp. 661-662) for a short review. For instance, Farsi and Filippini (2005) assess the electricity distribution utilities in Switzerland. They study the sensitivity of three benchmarking methods, one being non-parametric and two being parametric: DEA, COLS and SFA. Their results indicate that both efficiency scores and rankings of entities are significantly different across methods. Another example is provided by Badillo and Paradi (1999, p 76-100), who show that diverging results are observed when only non-parametric methods are used, such as DEA and Free Disposal Hull (FDH).

Alternative models to measure efficiency, within DEA, can also lead to diverging results but this has been far less investigated. Whilst few studies address this issue, interest seems to have been growing in recent years.

Some studies (Cordero, Pedraja & Santin, 2009; Estelle, Johnson & Ruggiero, 2010; Harrison, Rouse & Armstrong, 2012; Muñiz *et al.*, 2006; Ruggiero, 1996; Ruggiero, 1998; Ruggiero, 2004) use simulated data to compare alternative DEA models to the 'true' efficiency estimates performed by the simulation. However, the objective of these studies is to allow for comparisons between efficiency estimates performed by the alternative models and 'true' efficiency estimates. The objective of these studies is not, therefore, to determine if the efficiency estimates performed by the alternative models are convergent or divergent.

Very few studies (namely Cordero-Ferrara *et al.*, 2008; Muñiz, 2002; Yang and Pollitt, 2009) use empirical data in order to specifically benchmark alternative DEA models[59]. In these studies, comparisons are made between the efficiency estimates of the alternative models.

As practitioners and decision makers tend to perform their own efficiency analysis, the potential issue of diverging results is a matter of concern. If the alternative models produce divergent results, the choice of model becomes a strategic issue.

Studies using simulated and empirical data are presented hereafter.

[59] Other studies, such as Diagne (2006), McCarty and Yaisawarng (1993) or Garrett and Kwak (2011), perform alternative DEA models using empirical data. However, the objective of these studies is not to compare the models. Moreover, in the three studies cited above, a standard DEA model is performed instead of the Banker and Morey (1986b) model, although non-discretionary variables are included. The results are therefore flawed, rendering invalid comparisons.

Simulated data

Cordero, Pedraja and Santín (2009) consider the following models: one-stage by Banker and Morey (1986b), two-stage by Ray (1991) with a Tobit regression, three-stage by Muñiz (2002) and four-stage by Fried *et al.* (1999). Technical efficiency scores of these four methods are compared to a 'true' efficiency measure[60]. The four-stage model obtains the best results, although its Spearman rank's correlation with the 'true' efficiency is moderate (lower than 0.8). Note that the other models have very weak or weak Spearman rank's when the sample of DMUs is small (50). Estelle *et al.* (2010) show that the methodology used for comparison in Cordero *et al.* (2009) is flawed. Results are therefore called into question. Ultimately, Cordero-Ferrara, Pedraja-Chaparro and Salinas-Jiménez (2008) conclude that there is no consensus on the best model to use.

Estelle *et al.* (2010) consider the one-stage Banker and Morey (1986b) model and three variants of the three-stage Ruggiero (1998) model (alternatively using ordinary least squares, fractional logit and non-parametric regression in the second stage). Using simulated data, results are compared to the 'true' efficiency estimates. The three-stage model performs better than the one-stage model according to three indicators: correlation, rank correlation and mean absolute deviation between 'true' and estimated efficiency. The three variants of the Ruggiero (1998) model are very close one to one another. They have a strong correlation and rank correlation with the 'true' efficiency (higher than 0.8).

Harrison, Rouse and Armstrong (2012) use simulated data to compare the standard variable returns to scale (VRS) model (without non-discretionary variables), the one-stage Banker and Morey (1986a) model and the one-stage Banker and Morey (1986b) model with the 'true' efficiency estimates performed by the simulation. Discussing, first, all alternative models which allow for an environmental adjustment, Harrison *et al.* (2012) note that "there is no DEA model that is clearly superior in controlling for non-discretionary inputs (…)" (p. 263). Considering then the objective of their study, they conclude that *(1)* the Banker and Morey models perform equally well and *(2)* the Banker and Morey models should be used in preference to the standard VRS model when the influence of the environment is moderate to high.

Muñiz *et al.* (2006) use simulated data to compare the one-stage Banker and Morey (1986b) model, the three-stage Muñiz (2002) model, the three-stage Ruggiero (1998) model and the one-stage Yang and Paradi model (in Muñiz, Ruggiero, Paradi and Yang (2006, p. 1176)) with the 'true' efficiency estimates performed by the simulation. Three indicators are used to assess the models' performance: the rank correlation between 'true' and estimated efficiency, the mean absolute deviation and the percentage of entities for which inefficiency is overestimated. The Banker and Morey (1986b) model, using the variable returns to scale assumption, provides a rank correlation close to the Muñiz (2002) model. It does not perform as well as the other models. The Muñiz

60 'True' efficiency is determined by a known artificial set of data as the production function, used to simulate data.

(2002) model is the second best performer when the number of variables is small, but the results worsen when the number of variables increases. The Ruggiero (1998) model is the best performer in all the cases analyzed except one scenario. Finally, the Yang and Paradi model tends to overestimate inefficiency. It is also negatively affected by an increase in the number of variables.

Ruggiero (1996) uses simulated data to compare the one-stage Ruggiero (1996) model and the one-stage Banker and Morey (1986b) model with the 'true' efficiency estimates performed by the simulation. Based on several indicators (Ruggiero, 1996, p. 561), he shows that the Ruggiero (1996) model performs better than the Banker and Morey (1986b) model. This latter model tends to underestimate efficiency scores. Ruggiero (1996) applies his model to an empirical case of 556 school districts in the State of New York, USA. Unfortunately, he does not run a Banker and Morey (1986b) model with the same data in order to compare the results.

Ruggiero (1998) uses simulated data to compare a standard DEA model (without non-discretionary variables), the one-stage Banker and Morey (1986a) model, the one-stage Ruggiero (1996) model, the one-stage Banker and Morey (1986b) model, the two-stage Ray (1991) model (with two variants in the second stage regression analysis – linear and log-linear –) and the three-stage Ruggiero (1998) model (with two variants in the second stage regression – linear and log-linear –) with the 'true' efficiency estimates performed by the simulation. Four indicators are used to assess the models' performance: the correlation and the rank correlation between 'true' and estimated efficiency, the mean absolute deviation and the percentage of entities for which efficiency is inferior to the 'true' efficiency. The two- and three-stage models perform better than the one-stage models (including the standard DEA). The Ray (1991) model (both linear and log-linear variants) performs better than all other models based on the correlation and rank correlation criteria. These main results are confirmed by Ruggiero (2004).

Empirical data

Cordero-Ferrara *et al.* (2008) discuss various models including Banker and Morey (1986b), Ruggiero (1996)[61], Ray (1991), Fried *et al.* (2002) and Fried *et al.* (1999). They conclude that "an analysis of the different options does not allow us to conclude that any one is better than the others, that is, none of them is free of constraint" (p. 1329). Cordero-Ferrara *et al.* (2008) also apply a second-stage model, using Tobit regression, and the fourth-stage Fried *et al.* (1999) model to an empirical case (80 high-schools in the State of Extramaduria, Spain). They compare the fourth-stage model with a standard DEA model containing only discretionary variables. The rank correlation (Spearman) between the two models is 0.714. The number of efficient schools and the mean efficiency increase in the four-stage model.

[61] This model is wrongly cited as Ruggiero (1998) in Cordero-Ferrara *et al.* (2008, p. 1326).

Muñiz (2002) uses empirical data on 62 high-schools in the State of Asturias (Spain). He tests a standard DEA model without non-discretionary variables, the Banker and Morey (1986b) model and the three-stage Muñiz (2002) model. The most important finding lies in the number of efficient schools: 5 in the standard model, 12 in the three-stage model and 30 in the one-stage model. Based on the comparison between the one- and the three-stage models, Muñiz (2002) also shows that the majority of schools (75%) present less than 10% difference in efficiency scores. Schools facing a difference of more than 10% are usually efficient in the one-stage model, but not in the third-stage model. These results provide support for the Banker and Morey (1986b) model, as "it can be stated that both classifications present similar results except in the case when part of the units are considered efficient in the one-stage model" (Muñiz, 2002, p. 637). Unfortunately, no Pearson and Spearman correlations are run by Muñiz (2002).

Yang and Pollitt (2009) use empirical data on 221 Chinese coal-fired power-plants. They test a standard DEA model (without non-discretionary variables), the one-stage Banker and Morey (1986b) model, the two-stage Ray (1991) model (with two variants in the second stage regression – Tobit and logistic –), the three-stage Fried *et al.* (2002) model and the four-stage Fried *et al.* (1999) model. The Yang and Pollitt (2009) study distinguishes itself from other studies comparing models in the sense that they include undesirable outputs. The fact that the number of non-discretionary variables included in the alternative models is not the same (two in the one-stage model, seven in the other models) must also be noted[62]. Based on the correlations and the rank correlations between the efficiency scores of the alternative models, the following comments can be made:

- The standard DEA model and the third-stage model have a perfect correlation (0.98) and a perfect rank correlation (0.988);
- The standard DEA model and the fourth-stage model have a strong correlation (0.885) and a strong rank correlation (0.889);
- The three- and four-stage models have, in general, a higher correlation with the other models; on this basis, Yang and Pollitt (2009) suggest that "it indicates that these two models are able to explain most of the features of the other models, thus suggesting their superiority" (p. 1104).
- Correlations between the models vary from 0.311 to 0.98; rank correlations vary from 0.441 to 0.988. This suggests that alternative models lead to diverging results.

To sum up

The best available evidence, synthetized in Table 11, suggests that:

- There is no consensus on the best model to use (Cordero-Ferrara *et al.*, 2008);

[62] As the one-stage Banker and Morey (1986b) model cannot accommodate dummy variables, only two remaining non-discretionary variables were included in it.

- The one-stage Banker and Morey (1986a; 1986b) models perform equally well (Harrison *et al.*, 2012);
- The one-stage Banker and Morey (1986b) and the three-stage Muñiz (2002) model present similar results for a majority of entities under assessment (Muñiz, 2002);
- The one-stage Ruggiero (1996) model performs better than the Banker and Morey (1986b) model (Ruggiero, 1996);
- The three-stage Ruggiero (1998) model perform better than the Banker and Morey (1986b) model, the Yang and Paradi model and the Muñiz (2002) model (Muñiz *et al.*, 2006);
- Based on the correlation and the rank correlation criteria, the two-stage Ray (1991) model performs better than the one-stage Banker and Morey (1986b) model, the one-stage Ruggiero (1996) model and the three-stage Ruggiero (1998) model (Ruggiero, 1998); this evidence is confirmed in Ruggiero (2004);
- The three-stage Fried *et al.* (2002) and the four-stage Fried *et al.* (1999) are only compared to other models in Yang and Pollitt (2009). As the models include undesirable outputs and as the number of non-discretionary variables varies across the models, the results of this study cannot be generalized.

Table 11
Models' performance

Best available evidence	Reference
Banker and Morey (1986a) = Banker and Morey (1986b)	Harrison *et al.* (2012)
Banker and Morey (1986b) = Muñiz (2002)	Muñiz (2002)
Ray (1991) > Banker and Morey (1986b)	Ruggiero (1998); Ruggiero (2004)
Ray (1991) > Ruggiero (1996)	Ruggiero (1998); Ruggiero (2004)
Ray (1991) > Ruggiero (1998)	Ruggiero (1998); Ruggiero (2004)
Ruggiero (1996) > Banker and Morey (1986b)	Ruggiero (1996)
Ruggiero (1998) > Banker and Morey (1986b)	Ruggiero (1998)
Ruggiero (1998) > Yang and Paradi (2006)	Ruggiero (1998)
Ruggiero (1998) > Muñiz (2002)	Muñiz *et al.* (2006)

Although a formal rule of transitivity cannot be applied, the best evidence available suggests that the Ray (1991) model seems superior to alternative models.

The following critical comments conclude this literature review:

- The indicators used to assess the models' performance in the above mentioned studies (with simulated and empirical data) are probably not sufficient. It would have been wise to add statistical hypothesis tests in order to refine the analyses. Yang and Pollitt (2009) are the only ones to perform a Wilcoxon-Mann-Whitney test. However, this test does not seem appropriate in their context, and should probably have been substituted by a Wilcoxon signed rank sum test (see Section 3.8 about it).

- Some studies (Muñiz *et al.*, 2006; Ruggiero, 1996; Ruggiero, 1998; Yang & Pollitt, 2009) do not indicate the level of significance of the correlation coefficients (Pearson and/or Spearman) between the results of the alternative models and the 'true' efficiency measure. As a result, it is difficult to validly take into account their conclusion with respect to these indicators.
- Studies using simulated data merely indicate which is the model whose results are the closest to the 'true' efficiency measure. But they do not indicate whether the convergence between the alternative models (even the 'best' one) and the 'true' efficiency measure is acceptable or not. As a result, it is difficult to draw a general conclusion stating that the alternative models produce convergent or divergent results with the 'true' efficiency measure.

3.6 Methodology

The choice of alternative models later used in this study is made from a practitioner's standpoint according to three criteria: the degree of sophistication of the models, the level of computational skills needed to perform the models and the inclusion of models in DEA software[63]. Three commercial (*PIM-DEA* ®, *DEA-Solver PRO* ® and *DEAFrontier* ®) and two free (*Win4DEAP* and *Efficiency Measurement System* – EMS –) software packages are considered.

The degree of sophistication is considered as:

- Low for one-stage models which can be performed in existing software;
- Moderate for one-stage models which are not included in existing software and which need, as a result, coding from the practitioners;
- Moderate for two-stage models;
- High for three- and four-stage models.

The level of computational skills is considered as:

- Low if the model can be performed using an existing software;
- Moderate if it requires two different software packages but can easily be performed;
- High if it requires coding or two different software packages and a good command of these packages.

To be retained, a model has to show a low or moderate degree of sophistication, a low or moderate level of computational skills and be included in existing software. Table 12 presents the alternative models according to the three criteria.

---

[63] Note that Yang and Pollitt (2009, p. 1098) consider the easiness to understand, to apply and to interpret the models as advantages. The simplicity of calculation is considered by Muñiz (2002, p. 632) as another advantage.

Table 12
Ten models are assessed according to their sophistication, the computational skills needed to perform them and their inclusion in existing software

Model	Degree of sophistication	Level of computational skills	Software	Comments
One-stage				
Banker and Morey (1986a)	Low	Low	PIM-DEA ®, DEA-Solver PRO ® Win4DEAP, EMS	The categorical model can be performed in Win4DEAP and EMS by running a separate model for each category
Banker and Morey (1986b)	Low	Low	PIM-DEA ®, DEA-Solver PRO ®, DEAFrontier ®, EMS	
Ruggiero (1996)	Moderate	High		
Yang and Paradi in Muniz *et al.* (2006)	Moderate	Moderate	PIM-DEA ®, DEA-Solver PRO ®, DEAFrontier ®, Win4DEAP, EMS	The data needs first to be adjusted
Two-stage				
Ray (1991)	Moderate	Moderate	First stage in all software packages	The second stage can easily be performed in software like STATA ®, SPSS ® or R; simple and multiple regression can also be run in Excel ®
Three-stage				
Ruggiero (1998)	High	High	First stage in all software packages	The second stage can easily be performed in software like STATA ®, SPSS ® or R; the construction of the index to be used in the third stage requires a good command of these software packages; the third stage cannot be performed in existing software packages
Fried *et al.* (2002)	High	High	First and third stage in all software packages	The second stage can be performed in software like STATA ®, SPSS ® or R; it requires a very good command of these software packages; data need to be adjusted in order to run the third stage
Muniz (2002)	High	High	PIM-DEA ®, DEA-Solver PRO ®, DEAFrontier ®, Win4DEAP, EMS	The DEA model can be run separately at all stages in all existing software; however, the number of models to be run and the necessary adjustement to the data in the second stage means that the Fried *et al.* (2002) model not user-friendly
Four-stage				
Fried et al. (1999)	High	High	First and fourth stage in all software packages	The second and third stages can be performed in software like STATA ®, SPSS © or R; it requires a good command of these sofware packages
Program analysis				
Charnes *et al.* (1981)	Low	Low	PIM-DEA ®, DEA-Solver PRO ®, DEAFrontier ®, Win4DEAP, EMS	A DEA model for each subset has to be run separately; data need to be adjusted before running the last model containing all firms

Five models are retained: one-stage Banker and Morey (1986a) – BM1986a –; one-stage Banker and Morey (1986b) – BM1986b –; one-stage Yang and Paradi – YP2006 –; two-stage Ray (1991) – R1991 –; program analysis Charnes *et al.* (1981) – C1981 –. With the exception of YP2006, these models coincidentally correspond to those recommended by Coelli *et al.* (2005, pp. 191-194).

According to the best evidence available, the two-stage model performs better than the one-stage Ruggiero (1996) model and the three-stage Ruggiero (1998) model (Ruggiero, 1998, 2004). As the three-stage Ruggiero (1998) model performs better than the three-stage Muñiz (2002) model, it appears logical to retain the two-stage model.

Although they have been criticized, the one-stage Banker and Morey (1986a) and Banker and Morey (1986b) models are supported by Harrison *et al.* (2012) who note that these models are widely used by researchers. They have generated at least 239 different publications (Löber & Staat, 2010, p. 810). Harrison *et al.* (2010, p. 263) stress that it suggests that many researchers have found these models appropriate for their particular context. They also mention that "given there is no DEA model that is clearly superior in controlling for non-discretionary inputs, researchers continue to refer to the work of Banker and Morey (1986a, b)" (p. 263).

The three-stage Fried *et al.* (2002) model and the four-stage Fried *et al.* (1999) model suffer from a lack of comparison with other models. Yang and Pollitt (2009, p. 1097) clearly considered the three-stage model as the most sophisticated. As the comparison performed by Yang and Pollitt (2009) includes undesirable outputs, it is not clear if the conclusion of their study would have remained the same had the undesirable outputs been excluded. Further comparative studies featuring theses models are therefore needed.

The Charnes *et al.* (1981) program analysis model is retained. Unlike the other models, it estimates a technical efficiency devoid of managerial efficiency. As a result, it cannot be directly compared to the other models. However, this model is retained to test if its results are, somehow unexpectedly, convergent to the estimates of other models.

Finally, a standard VRS model (without non-discretionary variables) is also retained as a base case – VRS –.

The models are compared with one another on the basis of several indicators: mean efficiency, median efficiency, minimum efficiency, maximum efficiency, number of efficient schools, Pearson correlation and Spearman rank correlation. These are standard indicators commonly used in studies comparing models, such Cordero *et al.* (2009), Estelle *et al.* (2010), Harrison *et al.* (2012), Muñiz *et al.* (2006), Ruggiero (1996), Ruggiero (1998), Cordero-Ferrara *et al.* (2008), Muñiz (2002) and Yang and Pollitt (2009). In addition to these indicators, Yang and Pollitt (2009) perform the Wilcoxon-Mann-Whitney test. However, as developed in Section 3.8, the Wilcoxon-Mann-Whitney test does not seem appropriate in the case described in Yang and Pollitt (2009). As a result, the Wilcoxon signed rank sum test is preferred and retained in this study.

3.7 Data and models

Database

At the State of Geneva level, information about school input and output are atomized into various databases belonging to different administrative units. Public access to these databases is denied, making information about school production process unknown and opaque. However, cross-sectional data concerning the 2010-2011 school year and the 90 public primary schools has been secured for this study[64]. It includes pupils' results at standardized tests (aggregated at schools level), the number of full-time equivalent staff and various environmental variables. Useful data had to first be gathered from the different administrative units and second be organized into a workable order.

Discretionary and non-discretionary variables

Three discretionary outputs and three discretionary inputs are considered. These variables are all under the control of headteachers and are aggregated over schools.

Discretionary outputs include three composite scores (on a standardized scale with a maximum of 100) purely reflecting the quality of the education process. The first one is composed of pupils' results in French and mathematics standardized tests at the end of the second grade (SCORE2). The second one is composed of pupils' results in French, German and mathematics standardized tests at the end of the fourth grade (SCORE4). Finally, the third one is composed of pupils' results in French, German and mathematics standardized tests at the end of the sixth grade (SCORE6).

A large part of the studies focus specifically on standardized test scores as outputs[65]. Among those are Bessent and Bessent (1980), Bessent, Bessent, Kennington and Reagan (1982), Bradley *et al.* (2001), Chalos and Cherian (1995), Chalos (1997), Demir and Depren (2010), Kirjavainen and Loikkanen (1998), Mizala, Romaguera and Farren (2002), Ray (1991), Ruggiero (1996, 2000) or Sengupta (1990). Agasisti *et al.* (2014, p. 123) note that "such choice represents today the standard for analyzing school efficiency".

Discretionary inputs include *(1)* the number of full-time equivalent (FTE) teaching staff (TEACHER), *(2)* the number of FTE administrative and

[64] The 2010-2011 school year is a typical school year. Nothing makes the author think that the results of this study would have been different if another school year was used.

[65] The fact to include variables reflecting other aspects of human capability (and not only test scores) is open to debate. For instance, David Broddy, chairman of the Society of Heads, made the following statement at the Society of Heads' annual meeting in 2013 (Paton, 2013): "What part have we played in allowing that only academic success is a measure of human capability? That a definition of a "good" school is one that rises to the top of exam league tables and the definition of a "bright" pupil is one that gets A* grades?"
Unfortunately, in the State of Geneva, such other aspects are either not defined or, if defined, not measured.

technical staff (ADMIN) and *(3)* the school budget in Swiss francs – excluding staff salaries and capital expenditure (BUDGET) –[66]. The three inputs are expressed by pupils to be coherent with the formulation of the outputs. Note that BUDGET consists of a (relatively) small financial amount received by schools according to the number and the types of classes it runs. It can be used to finance teachers conducting supplementary tasks (i.e. tasks which do not appear in their contracts) or to buy school materials, support cultural activities, etc.

In 2010, according to the Swiss Federal Statistical Office, the first two inputs (TEACHER and ADMIN) correspond to 94.9% of the public education operating expenses of the State of Geneva (State and local authorities – municipalities –)[67]. They are formulated in FTE as opposed to monetary terms given that schools are not responsible for the age pyramid of their teachers and other staff. Taking into account the wages of the employees (which automatically grow higher alongside seniority) would unfairly alter the efficiency of a school with a greater proportion of senior staff[68].

The inputs are very similar to those used by Arcelus and Coleman (1997) – FTE teachers, FTE support staff, operating expenses and library expenses – although BUDGET is a feature in this study. The number of teachers and the number of administrative staff are classical inputs (Abbott & Doucouliagos, 2003; Avkiran, 2001; Grosskopf & Moutray, 2001) as are the overhead expenses (Ahn & Seiford, 1993; Beasley, 1990; Chalos & Cherian, 1995; Engert, 1996).

In the State of Geneva, schools are grouped into five categories according to a single non-discretionary variable: the percentage of pupils (per school) whose parents are blue-collar workers or unqualified workers (SOCIO). Note that the positive discrimination policy impacts only TEACHER (see Table 10 in Section 3.2) and not ADMIN or BUDGET.

Descriptive statistics of the variables are reported in Table 13. For instance, schools in category C have 0.0566 teachers per pupil, 0.0034 administrative staff per pupil and CHF 18.8496 per pupil. Pupils in category C schools obtain 77.9059 points at the end of grade 2, 77.751 at the end of grade 4 and 76.8070 points at the end of grade 6. The SOCIO variable has an average value of 38.15% in category C.

Note that 34 schools out of 90 (37.8%) are not grouped according to their theoretical category. For instance, 34% of pupils at school # 74 are classified as

---

[66] Note that test scores from previous school year could be used as an input when longitudinal data is available.

[67] These statistics are available at : http://www.bfs.admin.ch/bfs/portal/fr/index/themen/15/02/data/blank/01.html.

[68] The question to include wages as an input instead of FTE is open to debate. It would probably be appropriate in a context where schools can freely set teachers' salary. But in a context where teachers' salary is set by public authority and grow automatically alongside seniority, higher wages are not a good proxy of teaching quality. For instance, Woessmann (2003) shows that the teachers' age influences negatively pupil's performance.

disadvantaged. This school should be in category C, but is actually categorized in B. Several assumptions can explain this observation:

- The State authority has the discretionary power to group schools in other categories despite the value of SOCIO. Out of the 34 schools which are not grouped according to their theoretical category, 26 are grouped in a more advantaged category than the one in which they should be included[69]. For example, 23% of pupils at school # 79 are classified as disadvantaged, indicating that it should be in category B, but is actually categorized in A.
- Headteachers use their negotiation power in order to move their school to a more disadvantaged category than the one in which they should be included. Out of the 34 schools which are not grouped according to their theoretical category, 8 are grouped in a more disadvantaged category[70].

The fact that some schools are not grouped in their theoretical category has an impact on the Banker and Morey (1986a) model and the Charnes *et al.* (1981) model, as these two models are based on entities' categories. In this study, two alternatives are therefore considered:

- In the first one, the Banker and Morey (1986a) model and the Charnes *et al.* (1981) model are based on the observed schools' categorization (BM1986a-O and C1981-O);
- In the second one, the Banker and Morey (1986a) model and the Charnes *et al.* (1981) model are based on the theoretical schools' categorization (BM1986a-T and C1981-T).

---

[69] As TEACHER depends on the category, this could reflect the State's willingness to minimize expenses. But considerations other than financial could also explain the fact that some schools are not grouped in their theoretical category. For instance, the State authority may have considered that, for other reasons than the socioeconomic statu, some particular schools should be moved to another category.

[70] As TEACHER depends on the category, headteachers have an interest to be in a more disadvantaged category in order to receive more resources.

Table 13
Statistical summary of output and input variables included in the first stage DEA model - Observed category - (sample size = 90 primary schools)

	OBSERVED CATEGORY					
	A	B	C	D	E	Total
Number of schools	15	20	20	15	20	90
OUTPUTS						
SCORE2 (points/pupil)						
Mean	81.1284	80.6277	77.9059	77.9345	76.8063	78.8082
SD	2.3604	4.2687	4.0426	4.6632	5.1538	4.4956
Minimum	76.1504	71.9674	68.9075	69.0868	64.9589	64.9589
Maximum	84.0542	91.9591	83.2465	85.6571	88.8975	91.9591
SCORE4 (points/pupil)						
Mean	80.0865	79.1950	77.7510	75.9859	73.7298	77.2733
SD	2.2735	3.6067	2.8951	3.9605	2.9263	3.8718
Minimum	75.8127	68.9830	72.7422	68.0930	68.9577	68.0930
Maximum	83.4049	87.3654	81.5557	81.3806	78.5661	87.3654
SCORE6 (points/pupil)						
Mean	80.2470	78.6407	76.8070	76.2867	72.4740	76.7382
SD	2.6391	3.8218	3.6879	4.8185	3.6197	4.5361
Minimum	75.6189	70.2255	66.1693	66.2378	64.7010	64.7010
Maximum	85.1323	84.5935	81.4771	85.5275	78.5470	85.5275
INPUTS						
TEACHER (FTE/pupil)						
Mean	0.0558	0.0550	0.0566	0.0581	0.0648	0.0582
SD	0.0018	0.0017	0.0013	0.0018	0.0035	0.0043
Minimum	0.0532	0.0520	0.0546	0.0559	0.0572	0.0520
Maximum	0.0599	0.0583	0.0596	0.0618	0.0689	0.0689
ADMIN (FTE/pupil)						
Mean	0.0035	0.0034	0.0034	0.0035	0.0037	0.0035
SD	0.0005	0.0007	0.0005	0.0004	0.0005	0.0005
Minimum	0.0027	0.0026	0.0026	0.0029	0.0032	0.0026
Maximum	0.0045	0.0052	0.0044	0.0041	0.0050	0.0052
BUDGET (CHF/pupil)						
Mean	22.3694	19.8652	18.8496	19.1546	20.8817	20.1643
SD	6.2819	7.5281	4.0493	5.1942	5.4515	5.8233
Minimum	13.8019	13.2040	8.8186	13.3897	13.6034	8.8186
Maximum	32.1989	48.2835	27.6211	31.3439	33.3575	48.2835
NON-DISC. VARIABLE						
SOCIO						
Mean	19.6000	26.0500	38.1500	46.2000	54.9000	37.4333
SD	3.6801	6.9998	3.7455	3.7455	5.4086	13.7253
Minimum	15.0000	11.0000	29.0000	39.0000	45.0000	11.0000
Maximum	28.0000	37.0000	46.0000	54.0000	64.0000	64.0000

Source: General Direction of Primary Schools, Education Department, State of Geneva.

An unexpected observation emerges from Table 13. The average teacher/pupil ratio is lower in category B (0.055) than in category A (0.0558). Theoretically, it should be higher. This is partially explained by the fact that:

- Eight schools grouped in category A should actually belong to category B as they present a value of SOCIO higher than 19.99%. This implies that the teacher/pupil ratio of category A is pushed upwards.

- Three schools grouped in category B should actually belong to category A as they present a value of SOCIO lower than 20%. This implies that the teacher/pupil ratio of category B is pushed downwards.

Descriptive statistics of the variables based on their theoretical category are reported in Table 14. The average teacher/pupil ratio in category B (0.0554) is still lower than in category A (0.0557). This means that even when the categories are theoretically (re)composed, other unknown factors influence the allocated quantity of teaching staff[71].

[71] This could simply be due to the fact that the number of teachers cannot be easily adjusted – up- or down – from one school year to the next. For instance, assume that the ratio of teachers to pupils has to be reduced in a school. As the number of pupils is non-discretionary, the State authority has to reduce the number of teachers in this school. However, teachers benefit from the guarantee of employment. Except under exceptional circumstances, they cannot be fired. Neither can they be forced to move to another school. Consequently, if teachers refuse to relocate to another school, the ratio of teachers to pupils cannot be reduced and would thus remain 'artificially high'. This could be the case in category A schools.

Table 14
Statistical summary of output and input variables included in the first stage DEA model - Theoretical category - (sample size = 90 primary schools)

	THEORETICAL CATEGORY					
	A	B	C	D	E	Total
Number of schools	10	18	20	23	19	90
OUTPUTS						
SCORE2 (points/pupil)						
Mean	81.3409	79.8970	79.9912	77.9780	76.2036	78.8082
SD	2.5378	2.9255	4.2877	4.6426	5.3413	4.4956
Minimum	75.7948	74.4785	71.9674	68.9075	64.9589	64.9589
Maximum	84.0542	85.0372	91.9591	85.6571	88.8975	91.9591
SCORE4 (points/pupil)						
Mean	81.5598	79.8333	77.8690	76.0102	73.4941	77.2733
SD	3.2380	1.6527	2.7511	3.2162	3.3763	3.8718
Minimum	76.0190	75.8127	68.9830	69.5880	68.0930	68.0930
Maximum	87.3654	81.8875	80.8669	81.3806	80.2393	87.3654
SCORE6 (points/pupil)						
Mean	81.1046	78.6428	78.0296	76.5137	71.5482	76.7382
SD	2.8699	3.7172	2.4987	4.2160	3.4189	4.5361
Minimum	75.6189	70.2255	73.5402	66.1693	64.7010	64.7010
Maximum	85.1323	82.5055	84.5935	85.5275	78.3456	85.5275
INPUTS						
TEACHER (FTE/pupil)						
Mean	0.0557	0.0554	0.0563	0.0579	0.0646	0.0582
SD	0.0011	0.0020	0.0023	0.0023	0.0039	0.0043
Minimum	0.0543	0.0520	0.0526	0.0551	0.0562	0.0520
Maximum	0.0583	0.0599	0.0616	0.0661	0.0689	0.0689
ADMIN (FTE/pupil)						
Mean	0.0035	0.0036	0.0032	0.0035	0.0036	0.0035
SD	0.0003	0.0008	0.0004	0.0005	0.0004	0.0005
Minimum	0.0031	0.0026	0.0026	0.0027	0.0032	0.0026
Maximum	0.0040	0.0052	0.0041	0.0044	0.0050	0.0052
BUDGET (CHF/pupil)						
Mean	22.6474	20.9626	18.0419	19.7753	20.8063	20.1643
SD	6.6455	8.1456	3.6295	4.7841	5.5989	5.8233
Minimum	13.8019	13.7500	8.8186	13.3897	13.6034	8.8186
Maximum	32.1989	48.2835	23.1899	31.3439	33.3575	48.2835
NON-DISC. VARIABLE						
SOCIO						
Mean	16.1000	23.3889	35.0500	44.4348	56.0000	37.4333
SD	2.3310	3.1086	2.8741	2.8095	4.2817	13.7253
Minimum	11.0000	20.0000	30.0000	40.0000	50.0000	11.0000
Maximum	19.0000	29.0000	39.0000	49.0000	64.0000	64.0000

Source: General Direction of Primary Schools, Education Department, State of Geneva, and own calculation.

Table 15 compares the teacher/pupil target ratio (second column) with the teacher/pupil effective ratio (fourth column) in the two alternatives considered: observed (upper part of the table) versus theoretical (lower part of the table) categorization. The third column of the table recalls the percentage of targeted additional teaching staff that each category should have when compared with category A. For instance, schools in category D should have 11.41% more teaching staff than schools in category A. As these values are target values, they are the same in both alternatives. Finally, the fifth column displays the percentage of real additional teaching staff that each category gets when compared with category A. For instance, schools in category C have 1.49% more teaching staff than schools in category A when the categorization is observed, but only 1.08% when the categorization is theoretically-based.

Table 15
Teacher/pupil ratio in the observed and in the theoretical categorization

Observed category (# of schools)	Teacher/pupil target ratio	Additional teaching staff per pupil (%) (target)	Teacher/pupil effective ratio	Additional teaching staff per pupil (%) (effective)
A (15)	0.0539	0.00	0.0558	0.00
B (20)	0.0551	2.20	0.0550	-1.34
C (20)	0.0573	6.30	0.0566	1.49
D (15)	0.0601	11.41	0.0581	4.18
E (20)	0.0656	21.64	0.0648	16.15
Theoritical category (# of schools)	**Teacher/pupil target ratio**	**Additional teaching staff per pupil (%) (target)**	**Teacher/pupil effective ratio**	**Additional teaching staff per pupil (%) (effective)**
A (10)	0.0539	0.00	0.0557	0.00
B (18)	0.0551	2.20	0.0554	-0.54
C (20)	0.0573	6.30	0.0563	1.08
D (23)	0.0601	11.41	0.0579	4.01
E (19)	0.0656	21.64	0.0646	15.99

Source: General Direction of Primary Schools, Education Department, State of Geneva, and own calculation.

Except for schools in category B, the observed value of additional teaching staff is higher than the theoretically-based categorization. As a result, the values of the observed categorization are closer to the targeted additional teaching staff values than the theoretical categorization. This could also explain why the observed categorization of schools differs from the theoretical one (i.e. it has been adjusted from the theoretical categorization in order to better reduce the gap towards the targeted values)[72].

The correlation matrix of the input and output variables is presented in Table 16.

Table 16
Correlation Matrix for the variables

	TEACHER	ADMIN	BUDGET	SCORE2	SCORE4	SCORE6	SOCIO
TEACHER	1.00						
ADMIN	0.29 **	1.00					
BUDGET	0.08	-0.10	1.00				
SCORE2	-0.22 *	-0.09	0.07	1.00			
SCORE4	-0.46 **	-0.01	-0.07	0.33 **	1.00		
SCORE6	-0.49 **	-0.09	0.05	0.30 **	0.49 **	1.00	
SOCIO	0.75 **	0.07	-0.04	-0.36 **	-0.67 **	-0.61 **	1.00

** Significant at the 1% level; * Significant at the 5% level

72 To sum up, the separation between categories is not as complete as might be desired. In addition to the human resources factor, other possible contaminating effects could emerge from pupils' mobility from one school to another during the school year.

Statistically significant correlations are discussed hereafter. On the input side, the correlation between TEACHER and ADMIN is positive but very weak[73]. On the output side, correlations are positive but very weak between SCORE2 and SCORE 4 (0.33) and between SCORE2 and SCORE6 (0.3) and weak between SCORE4 and SCORE6 (0.49).

Correlations between TEACHER and the discretionary output variables are negative and very weak (TEACHER and SCORE2) or weak (TEACHER and SCORE4, TEACHER and SCORE6). This finding is coherent with Hanushek (2006). Based on a meta-analysis, he shows that school resources are weakly associated with school performance. The fact that the value of the correlation is increasing (or worsening) between TEACHER and SCORE2 (–0.22), SCORE4 (–0.46) and SCORE6 (–0.49) is intriguing. A possible interpretation of this result is that the number of teachers matters more in the early grades than in the later grades.

The correlation between the non-discretionary variable SOCIO and TEACHER is positive but only moderate (0.75). This reflects the fact that, despite the positive discrimination policy, the State of Geneva retains discretionary power in the allocation of resources, or that the rigidity in terms of human resource management does not always allow the State authority to increase or reduce the teacher/pupil ratio as desired. Unsurprisingly, correlations between SOCIO and SCORE2, 4 and 6 is negative.

Models

All DEA models are run using a variable returns to scale (VRS) assumption and an input orientation. The free software package *Win4DEAP* is used to perform all models except the Banker and Morey (1986b) model[74]. For this model, the free package *EMS* is used[75]. The software *STATA* ® is used to perform the second stage of the two-stage model.

The standard VRS model is performed without SOCIO. The Banker and Morey (1986a) model, the Charnes *et al.* (1981) model and the Yang and Paradi model are performed according to *(1)* the five observed school categories and *(2)* the five theoretical school categories. SOCIO is included as a continuous non-discretionary variable in the Banker and Morey (1986b)

---

73 Correlation coefficients are considered as perfect between 1 and 0.98 (or -1 and-0.98), strong between 0.97 and 0.8 (or -0.97 and -0.8), moderate between 0.79 and 0.6 (or -0.79 and -0.6), weak between 0.59 and 0.35 (or -0.59 and -0.35) and very weak between 0.34 and 0 (or -0.34 and 0).

74 As DEAP is a DOS program, a user friendly Windows interface has been developed for it (Win4DEAP). These 'twin' software packages have to be both downloaded and extracted to the same folder. Win4DEAP cannot work without DEAP.
DEAP Version 2.1: http://www.uq.edu.au/economics/cepa/deap.htm
Win4DEAP Version 1.1.3: http://www8.umoncton.ca/umcm-deslierres_michel/dea/install.html

75 The Banker and Morey (1986b) model is not included in Win4DEAP, but is included in EMS or commercial software packages such as PIM-DEA or DEA-Solver PRO.
EMS: http://www.holger-scheel.de/ems/

model. In order to allow a coherent comparison, SOCIO is also the only environmental variable considered in the two-stage Ray (1991) model[76]. Finally, note that no bootstrapping procedure is applied[77].

Two alternative variants of the Yang and Paradi model are performed. The first variant applies the values of h_j and $\hat{h}_j$ displayed in Table 17 to all inputs and outputs (YP2006-I&O). For instance, schools' inputs in category C are multiplied by a factor of 0.9407; offsetting the additional the 6.3% of resources received by schools C according to the teacher/pupil target ratio (see Table 15); outputs in category C are multiplied by a factor of 1.063; allowing for the 6.3% augmenting of outputs. The second variant applies the values of h_j to all inputs but does not adjust the outputs (YP2006-I)[78].

Table 17
Inputs are multiplied by the h_j factor and outputs by the $\hat{h}_j$ factor

Category	h_j	$\hat{h}_j$
A	1	1
B	0.9784	1.0220
C	0.9407	1.0630
D	0.8976	1.1141
E	0.8221	1.2164

As the positive discrimination policy of the State of Geneva concerns only the number of teaching staff, the handicapping measure in the Yang and Paradi model could be modified in order to be exclusively oriented towards it. This modified model corresponds to an extension of YP2006 and is customized for cases of positive discrimination concerning specific variables in the model. As the original YP2006 model adjusts all discretionary inputs and outputs, the modified model restricts the adjustment exclusively to the variables impacted

76 Other environmental variables than SOCIO could have been added in the BM1986a, BM1986b and R1991 models. The decision to include only SOCIO is justified by the fact that the State of Geneva uses SOCIO as the only variable in order to model its positive discrimination policy. The results of the models are therefore influenced by a single environmental variable. They could have been different if additional environmental variables had been considered.

77 Applying a bootstrapping procedure could make sense in the case of the Banker and Morey (1986a) and the Charnes *et al.* (1981) models, as the E category contains a limited number of schools. This option has not been retained because it introduces a supplementary difficulty and sophistication for practitioners and decision makers. Bootstrapping procedures are not included in the basic version of existing software packages, and therefore need coding skills from the practitioners.

78 The Yang and Paradi model formulation specifies that the handicapping measure is applied to all inputs and/or outputs. In this study, the handicap measure is applied *(1)* to all inputs and outputs and *(2)* to all inputs only, as the positive discrimination policy in the State of Geneva is oriented towards inputs.

by the positive discrimination policy (one in the Geneva case). Other discretionary inputs and outputs are not adjusted. As a result, inputs are divided into two categories: inputs impacted by the positive discrimination policy (x_{uj}^{WithPD}) and inputs not impacted by the positive discrimination policy (x_{ij}^{NoPD}). Assume h_j is the handicapping measure to adjust input variables impacted by the positive discrimination policy. The adjusted inputs are $h_j x_{uj}^{WithPD}$. There are data on s outputs, m inputs not impacted by the positive discrimination policy and v inputs impacted by the positive discrimination policy for each of n primary schools to be evaluated. y_{rk} is the quantity of output r produced by school k. x_{uj}^{WithPD} is the quantity of input u consumed by school k. x_{ij}^{NoPD} is the quantity of input i consumed by school k. λ_j represents the associated weighting of outputs and inputs of entity j. θ_k represents the VRS efficiency of school k (i.e. 'pure' technical efficiency free from any scale inefficiency).

This modified model, named Huguenin (H2014), is specified as follows:

$$\text{Minimize} \quad \theta_k \tag{4}$$

$$\text{Subject to} \quad y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \qquad r = 1, \ldots, s$$

$$\theta_k x^{NoPD}{}_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij}^{NoPD} \geq 0 \qquad i = 1, \ldots, m$$

$$\theta_k h_k x^{WithPD}{}_{uk} - \sum_{j=1}^{n} \lambda_j h_j x_{uj}^{WithPD} \geq 0 \qquad u = 1, \ldots, v$$

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

$$\lambda_j \geq 0 \quad \forall j = 1, \ldots, n$$

The handicapping measure h_j takes the values displayed in Table 17 (second column). These values allow TEACHER to be adjusted for the additional staff allocated under the positive discrimination policy. For instance, schools in category D obtain 11.41% additional teaching staff. The handicapping measure of 0.8976 applied for schools in category D allows for the actual number of teaching staff that these schools would have obtained if the positive discrimination policy had not been implemented[79]. As BM1986a, C1981 and YP2006, the Huguenin model is performed according to *(1)* the five observed school categories and *(2)* the five theoretical school categories.

Two-stage model

As only one environmental variable is used as an explanatory variable in the second stage of the Ray (1991) model, no risk of multicollinearity arises. The OLS model takes the following form[80]:

$TE_k = \alpha_0 + \alpha_1 SOCIO + e_k$

TE_k is the gross efficiency score, derived from the first stage analysis, of the k^{th} school and e_k is an error term satisfying the usual conditions for ordinary least squares estimation.

The potential presence of heteroskedasticity in the second stage is considered. A Breusch-Pagan / Cook-Weisberg test for heteroskedasticity is performed. It tests the null hypothesis (Ho) that the error variances are all equal versus the alternative that the error variances are a multiplicative function of one or more variables. If Ho is accepted, it indicates homoskedasticity; if it is rejected, it indicates heteroskedasticity.

The χ^2 of the Breusch-Pagan / Cook-Weisberg test is equal to 7.83 with a p-value of 0.0051. As the p-value is smaller than 0.05, the null hypothesis is rejected indicating that there is significant evidence of heteroskedasticity. Following this result, the model is corrected for heteroskedasticity by running an OLS regression with robust standard errors.

In the second-stage regression, it could be argued that the proportion of disadvantaged pupils increases where pupil performance is poor, and therefore SOCIO is endogenous to school efficiency. Pupil performance (measured by standardized tests) is used as an output in the first stage. All other things being equal, poor performance reduces efficiency. Privileged parents will move to other neighbourhoods in order to remove their children from low performing schools. As the State of Geneva is facing a housing crisis, with limited housing available and high rental rates, only privileged parents could afford to move into these areas. This move consequently increases the proportion of remaining disadvantaged pupils. As a result, a risk of simultaneity could occur between SOCIO and SCORE2, 4 and 6.

---

79 It could be criticised that *(1)* the handicapping measure is only applied to inputs impacted by the positive discrimination policy and *(2)* the choice of using the target additional teaching staff values as handicapping measures are questionable. However, these decisions are justified by the fact that the State of Geneva estimates that its positive discrimination policy only impacts one discretionary input (TEACHER) and that the targeted values adequately reflect the environmental influence. It is not the aim of this study to assess the relevancy of these political decisions.

80 Recent studies have shown that Ordinary Least Squares (OLS) regression is sufficient or even more appropriate to model the efficiency scores (Hoff, 2007; McDonald, 2009). OLS is, therefore, the method of choice in the ensuing study.

Endogeneity is solved by using instrumental variables. Instruments are identified following the procedure used by Waldo (2007): first, the instruments have to correlate with the potential endogenous variables; second, they must not have any explanatory power on efficiency scores if they are to be used as independent variables alongside the potential endogenous variables.

27 variables are tested in order to identify instruments. These variables are all measured at the municipality level in which schools are located[81]. The potentially endogenous SOCIO variable presents a correlation coefficient above |0.5| with only one variable: social assistance rate (BENEFIT), with a positive correlation of 0.6.

To measure the explanatory power of BENEFIT, an additional model is run. It includes BENEFIT alongside SOCIO. BENEFIT has a coefficient value of minus 0.0002792 (t value of –0.08) and is not statistically significant. As a result, BENEFIT can be considered as an instrumental variable.

The model tests SOCIO as a potentially endogenous variable, using BENEFIT as an instrument. A Durbin-Wu-Hausman test is performed. The null hypothesis (Ho) states that endogeneity is not present in the model. If Ho is accepted, it indicates the absence of endogeneity; if it is rejected, it indicates that endogeneity exists within the model. The χ^2 of the Durbin-Wu-Hausman test is equal to 0.00638 with a p-value of 0.93635. As the p-value is larger than 0.05, the null hypothesis is accepted. No endogeneity is found.

This is not surprising. In this study, SOCIO is assumed to be the cause of SCORE2, 4 and 6. If information about pupil performance (measured by standardized tests) was public knowledge, it could potentially encourage parents to move into catchment areas of better schools. However, in a principal-agent approach of educational production (Wössmann, 2005), asymmetric information about school data between the principal (i.e. the parents) and the agent (i.e. the headteacher) appears to be strong in the State of Geneva. Information about school quality (pupil performance) and resource consumption are computed at the State level and is unknown by parents. Therefore, parents cannot base their move on rational data and it is unlikely that SOCIO is endogenous.

---

81 The 27 variables are as follows: population (2011), population density per km^2,(2011), proportions of the population (2011) between *(1)* 0 and 19 years old, *(2)* 20 and 64 years old, *(3)* over 64 years old, area in km^2 (1992/1997, latest data available), habitat and infrastructure area (%), agricultural area (%), wooded area (%), unproductive area (%), total number of jobs (2008, latest data available), number of jobs in the primary sector, number of jobs in the secondary sector, number of jobs in the tertiary sector, total number of companies (2008, latest data available), number of companies in the primary sector, number of companies in the secondary sector, number of companies in the tertiary sector, number of newly built apartments (2010), social assistance rate (2011), share of votes in the last federal election for left parties (2011), tax burden for married people with two children and an annual revenue of 100'000 CHF (State, municipal and religious tax, in % of gross labour income) (2011), budget surplus (excess revenue) (2011), gross debt (2011), taxable wealth of natural persons (2008, latest data available), taxable income of natural persons (2008, latest data available), taxable profit of corporations (2009, latest data available).

The procedure of Gasparini and Ramos (2003), applied in De Witte and Moesen (2010) or in Agasisti, Bonomi and Sibiano (2014), is used to derive adjusted net efficiency scores for each school:

$$\theta_k^{Net} = \mathrm{e_k} + (1 - \max_{i=1,\ldots,n} \mathrm{e_i})$$

where θ_k^{Net} is the adjusted net efficiency score of the k[th] school and $\mathrm{e_k}$ is the residual for each school obtained from the OLS estimation[82].

3.8 Results

Descriptive statistics

Table 18 displays the descriptive statistics of:

- The standard VRS model;
- The five models which allow for an environmental adjustment (BM1986a; BM1986b; R1991; YP2006; H2014);
- The C1981 model; noting that this model shows efficiency scores devoid of managerial inefficiency but does not adjust for the environment.

The upper part of Table 18 displays results for the observed categorization; the lower part displays results for the theoretical categorization.

For instance, the YP2006-I&O model has, considering the theoretical categorization, a mean efficiency score of 0.9345 with a standard deviation of 0.056. The median efficiency score is 0.9452. This means that half the schools have a score higher than 0.9452 and half the schools have a score lower than 0.9452. In this model, the minimum efficiency score obtained by a school is 0.7976 and the maximum score is 1. 19 schools (row 'Number of efficient schools') are fully efficient.

[82] As the efficiency scores are adjusted with the maximum observed residual, the procedure of Gasparini and Ramos (2003) results in the identification of a single efficient entity.

Table 18
Descriptive statistics of the five models which allow for an environmental adjustment and the two models without environmental adjustment (VRS and C1981)

	VRS	BM1986a	BM1986b	R1991	YP2006 I&O	YP2006 I	H2014	C1981
Observed category								
Mean	0.9321	0.9787	0.9793	0.9009	0.9340	0.9654	0.9650	0.9517
SD	0.0671	0.0342	0.0339	0.0450	0.0516	0.0392	0.0401	0.0537
Min.	0.7604	0.8572	0.8415	0.7939	0.7981	0.8556	0.8544	0.7976
Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Median	0.9481	1.0000	1.0000	0.9041	0.9344	0.9763	0.9779	0.9657
Number of efficient schools	20.0000	51.0000	46.0000	1.0000	17.0000	31.0000	31.0000	25.0000
Theoretical category								
Mean	0.9321	0.9751	0.9793	0.9009	0.9345	0.9604	0.9587	0.9553
SD	0.0671	0.0394	0.0339	0.0450	0.0560	0.0455	0.0463	0.0537
Min.	0.7604	0.8482	0.8415	0.7939	0.7976	0.8344	0.8338	0.7813
Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Median	0.9481	1.0000	1.0000	0.9041	0.9452	0.9769	0.9747	0.9774
Number of efficient schools	20.0000	50.0000	46.0000	1.0000	19.0000	29.0000	29.0000	28.0000

Based on Table 18, the following facts are established:

- No obvious difference emerges from the descriptive statistics between results of the observed and the theoretical categorizations.
- BM1986a and BM1986b have a lower discriminating power than the other models; more than half of the schools are efficient in these models. They have the highest mean efficiency scores amongst all models[83].
- VRS and YP2006-I&O have close mean efficiency scores.
- YP-I and H2014 have close mean efficiency scores and a similar number of efficients schools.
- H2014 and C1981 have close mean efficiency scores.

Comparison between the observed and the theoretical categorizations

Table 19 displays the Pearson and Spearman correlation coefficients between the same models in the observed and the theoretical categorizations. For instance, the Pearson correlation between the efficiency scores of YP2006-I&O in the observed categorization and the efficiency scores of YP2006-I&O in the theoretical categorization is equal to 0.8022.

[83] The use of super-efficiency models could have been imagined in order to allow discrimination between efficient entities. These models allocate scores higher than 1 (or 100%) to efficient entities. However, super-efficiency models are sophisticated and not easily achievable by practitioners. Moreover, they may lead to infeasible solutions (Thrall, 1996; Zhu, 1996).

Table 19
Pearson and Spearman correlation coefficients between the same models in the observed and the theoretical categorizations

	BM1986a	BM1986b	R1991	YP2006 I&O	YP2006 I	H2014	VRS	C1981
Pearson	0.7696**	1.0000**	1.0000**	0.8022**	0.8343**	0.8736**	1.0000**	0.8834**
Spearman	0.6872**	1.0000**	1.0000**	0.7810**	0.8192**	0.8606**	1.0000**	0.8955**

** Significant at the 1% level; * Significant at the 5% level

As the efficiency scores of BM1986b, R1991 and VRS do not differ in the two categorizations, the Pearson and Spearman correlations are perfect. The efficiency scores of the other models are impacted by the type of categorization. YP2006-I, H2014 and C1981 present strong correlations in both Pearson and Spearman. YP2006-I&O has a strong Pearson correlation and a moderate Spearman correlation. BM1986a have moderate correlations in both Pearson and Spearman.

The difference between the two categorizations is further tested in order to determine whether differences occur by chance or are statistically significant.

Cooper, Seiford and Tone (2007) show that "since the theoretical distribution of the efficiency score in DEA is usually unknown, we are forced to deal with nonparametric statistics for which the distribution of the DEA scores are statistically independent" (p. 233). They use a Wilcoxon-Mann-Whitney test in order to identify whether the differences between two different groups (for instance entities located in an urban environment versus entities located in a rural environment) are significant[84].

Yang and Pollitt (2009, p. 1103) use the Wilcoxon-Mann-Whitney test in order to identify whether the difference between the efficiency scores of different models containing the same entities is significant. As the Wilcoxon-Mann-Whitney test seems appropriate in the case described in Cooper *et al.* (2007), it does not seem appropriate in the case described in Yang and Pollitt (2009). For this latter case, the Wilcoxon signed rank sum test seems better suited and is therefore appropriate to test repeated measurements on a single sample (or two related samples or matched samples) in order to assess whether their population mean ranks differ[85]. It therefore seems appropriate to compare efficiency scores of different models containing the same entities. A Wilcoxon signed rank sum test is thus performed between each model in the two categorizations. For instance, the efficiency scores of BM1986a in the observed categorization are tested against the efficiency scores of BM1986a in the theoretical categorization.

Table 20 displays results of the Wilcoxon signed rank sum test for the three models which allow for an environmental adjustment and are impacted by the

84 The Wilcoxon-Mann-Whitney test is the non-parametric version of the independent samples t-test.

85 The Wilcoxon signed rank sum test is the non-parametric version of the paired samples t-test.

type of categorization (BM1986a, YP2006 and H2014) and for C1981 which is also impacted by the type of categorization but does not allow for an environmental adjustment[86]. The null hypothesis states that there is no statistically significant difference between the efficiency scores of the same model in the two categorization alternatives. The null hypothesis is accepted for BM1986a, YP2006-I&O and C1981 but is rejected for YP2006-I and H2014 at the 1% level[87]. For YP2006-I and H2014, there is a statistically significant difference between the efficiency scores in the observed categorization and the efficiency scores in the theoretical categorization[88]. These two models are probably more sensitive to the type of categorizations because they cumulate two adjustments. First (like the other models), the entities are re-organized according to the two types of categorizations. Note that the difference between the two categorizations is marginal, in the sense that one entity could be moved from school category A to school category B, for example but never from school category A to school category C, D or E. Second (unlike the other models), the inputs of these two models are adjusted within the categories before performing the efficiency analysis.

Table 20
Wilcoxon signed rank sum test

		YP2006			
	BM1986a	**I&O**	**I**	**H2014**	**C1981**
z-statistic	0.8370	0.4520	2.8200	3.5140	-1.5710
p-value	0.4024	0.6515	0.0048	0.0004	0.1161

Based on Table 19 and Table 20, the following facts are established:

- All correlations between the same models in the two types of categorization are positive and strong, with the exception of BM1986a (Pearson and Spearman correlations) and YP2006-I&O (Spearman correlations), which are also positive but only moderate.
- The null hypothesis of the Wilcoxon signed rank sum test cannot be accepted for all models which allow for an environmental adjustment. As a

86 The BM1986b and the R1991 models allow for an environmental adjustement but are not impacted by the type of categorization.

87 A Wilcoxon-Mann-Whitney test has also been performed. The null hypothesis is accepted for all models.

88 Note that the Kolmogorov-Smirnov test is another non-parametric hypothesis test used in DEA (Banker, Zheng and Natarajan, 2010). As it compares the distribution of two independent samples (and not repeated measurements on a single sample), it does not seem appropriate in the context of this study. The null hypothesis of the Kolmogorov-Smirnov test states that there is no statistically significant difference between the distribution functions of the same model in the two categorizations alternatives. For the record, a Kolmogorov-Smirnov test has been performed between each model in the two categorizations. The null hypothesis is accepted for all models.

result, the distinction between the two types of categorization will be kept in the upcoming analysis.

Pearson correlation

Table 21 displays the Pearson correlation coefficients between each pair of models in the **observed** categorization alternative.

Focusing on the five models which allow for an environmental adjustment (the cells in the first six rows and columns shaded in light grey), the correlation coefficients are positive and vary from 0.3301 (BM1986b and YP2006-I&O) to 0.9499 (YP2006-I and H2014). Every single correlation is significant at the 1% level. Three correlations are higher than 0.8 and can be described as strong (BM1986a and H2014; BM1986a and YP2006-I; YP2006-I and H2014). Seven correlations are moderate (0.6787 between BM1986a and R1991; 0.7007 between BM1986b and R1991; 0.6572 between BM1986b and YP2006-I; 0.6434 between BM1986b and H2014; 0.7454 between R1991 and YP2006-I; 0.7837 between R1991 and H2014; 0.6002 between YP2006-I&O and YP2006-I). Finally, five correlations are weak (0.5811 between BM1986a and BM1986b; 0.5762 between BM1986a and YP2006-I&O; 0.3301 between BM1986b and YP2006-I&O; 0.5158 between R1991 and YP2006-I&O; 0.4931 between YP2006-I&O and H2014).

Table 21
Pearson correlation coefficients (observed categorization)

	BM1986a	BM1986b	R1991	YP2006 I&O	YP2006 I	H2014	VRS	C1981
BM1986a	1.0000							
BM1986b	0.5811**	1.0000						
R1991	0.6787**	0.7007**	1.0000					
YP2006-I&O	0.5762**	0.3301**	0.5158**	1.0000				
YP2006-I	0.9201**	0.6572**	0.7454**	0.6002**	1.0000			
H2014	0.8983**	0.6434**	0.7837**	0.4931**	0.9499**	1.0000		
VRS	0.6144**	0.4864**	0.6682**	-0.0003	0.6162**	0.7529**	1.0000	
C1981	0.1679	0.2627*	0.4388**	-0.3321**	0.2214*	0.4036**	0.8778**	1.0000

** Significant at the 1% level; * Significant at the 5% level

The Pearson correlation coefficients between the standard VRS model and the five models which allow for an environmental adjustment are positive and weak (BM1986b), positive and moderate (BM1986a; R1991; YP2006-I; H2014) or negative and very weak (YP2006-I&O). Note that this latter correlation is not statistically significant. VRS and C1981 have a strong positive correlation.

The Pearson correlation coefficients between the C1981 model and the five models which allow for an environmental adjustment are positive and very weak (BM1986a; BM1986b; YP2006-I), positive and weak (R1991; H2014) or negative and very weak (YP2006-I&O). Note that the correlation between C1981 and BM1986a is not statistically significant.

Table 22 displays the Pearson correlation coefficients between each pair of models in the **theoretical** categorization alternative.

Focusing on the five models which allow for an environmental adjustment (the cells in the first six rows and columns shaded in light grey), the correlation coefficients are positive and vary from 0.5050 (YP2006 and H2014) to 0.9652 (YP2006-I and H2014). Every single correlation is significant at the 1% level. Six correlations are higher than 0.8 and can be described as strong (BM1986a and BM1986b; BM1986a and YP2006-I; BM1986a and H2014; R1991 and YP2006-I; R1991 and H2014; YP2006-I and H2014). Seven correlations are moderate (0.7218 between BM1986a and R1991; 0.6238 between BM1986a and YP2006-IO&O; 0.7007 between BM1986b and R1991; 0.7641 between BM1986b and YP2006-I; 0.7343 between BM1986b and H2014; 0.6329 between R1991 and YP2006-I&O; 0.6108 between YP2006-I&O and YP2006-I). Finally, two correlations are weak (0.5414 between BM1986b and YP2006-I&O; 0.5050 between YP2006-I&O and H2014).

Table 22
Pearson correlation coefficients (theoretical categorization)

				YP2006				
	BM1986a	**BM1986b**	**R1991**	**I&O**	**I**	**H2014**	**VRS**	**C1981**
BM1986a								
BM1986b	0.8145**	1.0000						
R1991	0.7218**	0.7007**	1.0000					
YP2006-I&O	0.6238**	0.5414**	0.6329**	1.0000				
YP2006-I	0.9071**	0.7641**	0.8394**	0.6108**	1.0000			
H2014	0.8693**	0.7343**	0.8306**	0.5050**	0.9652**	1.0000		
VRS	0.5958**	0.4864**	0.6682**	0.0490	0.6935**	0.7914**	1.0000	
C1981	0.0421	0.0392	0.3328**	-0.3584**	0.2303*	0.3765**	0.8238**	1.0000

** Significant at the 1% level; * Significant at the 5% level

The Pearson correlation coefficients between the standard VRS model and the five models which allow for an environmental adjustment are positive and moderate (R1991; YP2006-I; H2014), positive and weak (BM1986a; BM1986b) or positive and very weak (YP2006-I&O). Note that this latter correlation is not statistically significant. VRS and C1981 have a strong positive correlation.

The Pearson correlation coefficients between the C1981 model and the five models which allow for environmental adjustment are positive and very weak (BM1986a; BM1986b; R1991; YP2006-I), positive and weak (H2014) or negative and weak (YP2006-I&O). Note that the correlations between C1981 and BM1986a or BM1986b are not statistically significant.

Based on Table 21 and Table 22, the following facts are established:

- The Pearson correlation between C1981 and the five models which allow for an environmental adjustment is either weak or very weak. This is not a surprise, as C1981 does not adjust for the environment.

- In some cases, the Pearson correlation between VRS and models which allow for an environmental adjustment is moderate and statistically significant. This was not expected, as VRS does not adjust for the environment.
- The Pearson correlations among the five models which allow for an environmental adjustment are positive. However, they are mainly moderate.
- Overall, nine correlations are strong (30%), fourteen are moderate (47%) and seven are weak (23%). The nine strong correlations link the following models: BM1986a and H2014, BM1986a and YP2006-I, YP2006-I and H2014 in the observed categorization; BM1986a and BM1986b, BM1986a and YP2006-I, BM1986a and H2014, R1991 and YP2006-I, R1991 and H2014, YP2006-I and H2014 in the theoretical categorization. Note that BM1986a appears five times in these nine strong correlations.
- The Pearson correlation coefficient analysis is the first indication that the results of a majority of models which allow for an environmental adjustment are divergent. To be considered as convergent, a strong or a perfect correlation would be needed.

Spearman correlation

Table 23 displays the Spearman's rank correlation coefficients between each pair of models in the **observed** categorization alternative.

Focusing on the five models which allow for an environmental adjustment (the cells in the first six rows and columns shaded in light grey), the correlation coefficients are positive and vary from 0.3710 (BM1986b and YP2006-I&O) to 0.9072 (YP2006-I and H2014). Every single correlation is significant at the 1% level. Three correlations are higher than 0.8 and can be described as strong (BM1986a and YP2006-I; BM1986a and H2014; YP2006-I and H2014)[89]. Seven correlations are moderate (0.6022 between BM1986a and R1991; 0.6081 between BM1986a and YP2006-I&O; 0.6010 between BM1986b and R1991; 0.6482 between BM1986b and YP2006-I; 0.6163 between BM1986b and H2014; 0.6638 between R1991 and YP2006-I; 0.7008 between R1991 and H2014). Finally, five correlations are weak (0.5964 between BM1986a and BM1986b; 0.3710 between BM1986b and YP2006-I&O; 0.5178 between R1991 and YP2006-I&O; 0.5848 between YP2006-I&O and YP2006-I; 0.4704 between YP2006-I&O and H2014).

[89] These three pairs of models are also associated with a strong Pearson correlation.

Table 23
Spearman correlation coefficients (observed categorization)

	BM1986a	BM1986b	R1991	YP2006 I&O	YP2006 I	H2014	VRS	C1981
BM1986a	1.0000							
BM1986b	0.5964**	1.0000						
R1991	0.6022**	0.6010**	1.0000					
YP2006-I&O	0.6081**	0.3710**	0.5178**	1.0000				
YP2006-I	0.8508**	0.6482**	0.6638**	0.5848**	1.0000			
H2014	0.8092**	0.6163**	0.7008**	0.4704**	0.9072**	1.0000		
VRS	0.5033**	0.4363**	0.5580**	-0.0183	0.5467	0.7192**	1.0000	
C1981	0.1498	0.1925	0.3377**	-0.2587*	0.2587*	0.4493**	0.8756**	1.0000

** Significant at the 1% level; * Significant at the 5% level

The Spearman's rank correlation coefficients between the standard VRS model and the five models which allow for an environmental adjustment are positive and weak (BM1986a; BM1986b; R1991; YP2006-I), positive and moderate (H2014) or negative and very weak (YP2006-I&O). VRS and C1981 have a strong positive correlation. Note that the correlations between VRS and YP2006-I&O or YP2006-I are not statistically significant.

The Spearman correlation coefficients between the C1981 model and the five models which allow for an environmental adjustment are positive and very weak (BM1986a; BM1986b; R1991; YP2006-I), positive and weak (H2014) or negative and very weak (YP2006-I&O). Note that the correlations between C1981 and BM1986a or BM1986b are not statistically significant.

Table 24 displays the Spearman's rank correlation coefficients between each pair of models in the **theoretical** categorization alternative.

Focusing on the five models which allow for an environmental adjustment (the cells in the first six rows and columns shaded in light grey), the correlation coefficients are positive and vary from 0.4594 (YP2006 and H2014) to 0.9179 (YP2006-I and H2014). Every single correlation is significant at the 1% level. Two correlations are higher than 0.8 and can be described as strong (BM1986a and YP2006-I; YP2006-I and H2014). Eleven correlations are moderate (0.7207 between BM1986a and BM1986b; 0.6576 between BM1986a and R1991; 0.6649 between BM1986a and YP2006-I&O; 0.7986 between BM1986a and H2014; 0.6010 between BM1986b and R1991; 0.6804 between BM1986b and H2014; 0.6353 between R1991 and YP2006-I&O; 0.7752 between R1991 and YP2006-I; 0.7643 between R1991 and H2014; 0.6035 between YP2006-I&O and YP2006-I). Finally, two correlations are weak (0.5257 between BM1986b and YP2006-I&O; 0.4594 between YP2006-I&O and H2014)[90].

[90] These two pairs of models are also associated with a weak Pearson correlation.

Table 24
Spearman correlation coefficients (theoretical categorization)

	BM1986a	BM1986b	R1991	YP2006 I&O	YP2006 I	H2014	VRS	C1981
BM1986a	1.0000							
BM1986b	0.7207**	1.0000						
R1991	0.6576**	0.6010**	1.0000					
YP2006-I&O	0.6649**	0.5257**	0.6353**	1.0000				
YP2006-I	0.8449**	0.7000**	0.7752**	0.6035**	1.0000			
H2014	0.7986**	0.6804**	0.7643**	0.4594**	0.9179**	1.0000		
VRS	0.5169**	0.4363**	0.5580**	0.0244	0.6326**	0.7872**	1.0000	
C1981	0.0817	0.0970	0.2842**	-0.2640*	0.2843**	0.4586**	0.8250**	1.0000

** Significant at the 1% level; * Significant at the 5% level

The Spearman's rank correlation coefficients between the standard VRS model and the five models which allow for an environmental adjustment are positive and moderate (YP2006-I; H2014), positive and weak (BM1986a; BM1986b, R1991) or positive and very weak (YP2006-I&O). Note that this latter correlation is not statistically significant. VRS and C1981 have a strong positive correlation.

The Spearman correlation coefficients between the C1981 model and the five models which allow for an environmental adjustment are positive and very weak (BM1986a; BM1986b; R1991; YP2006-I; H2014) or negative and very weak (YP2006-I&O). Note that the correlations between C1981 and BM1986a or BM1986b are not statistically significant.

Based on Table 23 and Table 24, the following facts are established:

- The Spearman correlation between C1981 and each of the five models which allow for an environmental adjustment is either weak or very weak. This is not a surprise, as C1981 does not adjust for the environment.
- The Spearman correlation between VRS and models which allow for an environmental adjustment is either weak or very weak. It is moderate in only two cases (YP2006-I; H2014).
- The Spearman correlations among the five models which allow for an environmental adjustment are positive. However, they are mainly moderate.
- Overall, eighteen correlations are moderate (60%), seven are weak (23%) and five are strong (17%). The five strong correlations link the following models: BM1986a and YP2006-I, BM1986a and H2014, YP2006-I and H2014 in the observed categorization; BM1986a and YP2006-I, YP2006-I and H2014 in the theoretical categorization. Note that BM1986a appears three times in these five strong correlations.
- After the Pearson correlation analysis, the Spearman's rank correlation coefficient analysis is the second indication that the results of the majority of models which allow for an environmental adjustment are divergent. To be considered convergent, a strong or a perfect correlation would be needed.

Comparison between the models in the observed categorization

The Wilcoxon signed rank sum test is used to assess the difference between each pair of models in the **observed** categorization alternative. For example, the test is performed between the efficiency scores of BM1986a and the efficiency scores of R1991. Results are displayed in Table 25. The first number appearing in a given cell is the z-statistic and the second number is the p-value. For instance, the Wilcoxon signed rank sum test between BM1986b and H2014 has a z-statistic of 5.124 and a p-value of 0.

Table 25
Wilcoxon signed rank sum test between each pair of models in the observed categorization

	BM1986a	BM1986b	R1991	YP2006 I&O	YP2006 I	H2014	VRS	C1981
BM1986a		-0.1330	-8.2370	-7.8620	-7.3980	-7.3980	-7.2450	-4.2670
		0.8944	0.0000	0.0000	0.000	0.0000	0.0000	0.0000
BM1986b	0.1330		-8.2370	-6.7670	-5.3430	-5.1240	-7.4470	-4.4870
	0.8944		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R1991	8.2370	8.2370		6.1420	8.2370	8.2370	4.8730	6.8490
	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000
YP2006 I&O	7.8620	6.7670	-6.1420		6.1400	5.5100	0.2800	2.2530
	0.0000	0.0000	0.0000		0.0000	0.0000	0.7797	0.0242
YP2006 I	7.3980	5.3430	-8.2370	-6.1400		1.6220	-4.5950	-1.5930
	0.0000	0.0000	0.0000	0.0000		0.1047	0.0000	0.1111
H2014	7.3980	5.1240	-8.2370	-5.5100	-1.6220		-5.5420	-2.3880
	0.0000	0.0000	0.0000	0.0000	0.1047		0.0000	0.0169
VRS	7.2450	7.4470	-4.8730	-0.2800	4.5950	5.5420		5.6140
	0.0000	0.0000	0.0000	0.7797	0.0000	0.0000		0.0000
C1981	4.2670	4.4870	-6.8490	-2.2530	1.5930	2.3880	-5.6140	
	0.0000	0.0000	0.0000	0.0242	0.1111	0.0169	0.0000	

The null hypothesis is rejected at the 5% level for all but four pairs of models (BM1986a and BM1986b; YP2006-I&O and VRS; YP2006-I and H2014; YP2006-I and C1981)[91]. These four pairs of models appear in light grey cells. For the following pairs of models, there is a statistically significant difference between the efficiency scores in the first model mentioned and the efficiency scores in the second model mentioned: BM1986a and R1991, BM1986a and YP2006-I&O, BM1986a and YP2006-I, BM1986a and H2014, BM1986a and VRS, BM1986a and C1981, BM1986b and R1991, BM1986b and YP2006-I&O, BM1986b and YP2006-I, BM1986b and H2014, BM1986b and VRS, BM1986b and C1981, R1991 and YP2006-I&O, R1991 and YP2006-I, R1991 and H2014, R1991 and VRS, R1991 and C1981, YP2006-I&O and

91 Two additional tests have also been performed (a Wilcoxon-Mann-Whitney test and a Kolgomorov-Smirnov test). The results are similar to the Wilcoxon signed rank sum test, except that the null hypothesis is accepted by two additional pairs of models (H2014 and C1981; VRS and C1981).

YP2006-I, YP2006-I&O and H2014, YP2006-I&O and C1981, YP2006-I and VRS, H2014 and VRS, H2014 and C1981, VRS and C1981.

Among the four pairs for which the null hypothesis is accepted, only two pairs concern models which exclusively allow for an environmental adjustment (BM1986a and BM1986b; YP2006-I and H2014).

Based on Table 25, the following facts are established:

- Two of the pairs of models which allow for an environmental adjustment do not have a statistically significant difference between their efficiency scores (BM1986a and BM1986b; YP2006-I and H2014). The other pairs of models which allow for an environmental adjustment have a statistically significant difference between their efficiency scores (BM1986a and R1991; BM1986a and YP2006-I&O; BM1986a and YP2006-I; BM1986a and H2014; BM1986b and R1991; BM1986b and YP2006-I&O; BM1986b and YP2006-I; BM1986b and H2014; R1991 and YP2006-I&O; R1991 and YP2006-I; R1991 and H2014; YP2006-I&O and H2014).
- The Wilcoxon signed rank sum test performed on every pair of models in the observed categorization is the third indication that the results for the majority of models which allow for an environmental adjustment are divergent.

Comparison between the models in the theoretical categorization

The Wilcoxon signed rank sum test is used to assess the difference between each pair of models in the **theoretical** categorization alternative. For example, the test is performed between the efficiency scores of BM1986b and the efficiency scores of R1991. Results are displayed in Table 26. The first number appearing in a given cell is the z-statistic and the second number is the p-value. For instance, the Wilcoxon signed rank sum test between R1991 and YP2006-I&O has a z-statistic of –6.307 and a p-value of 0.

Table 26
Wilcoxon signed rank sum test between every pair of models in the theoretical categorization

	BM1986a	BM1986b	R1991	YP2006 I&O	YP2006 I	H2014	VRS	C1981
BM1986a		0.4070	-8.2370	-7.8260	-7.4940	-7.3630	-7.1290	-2.7510
		0.6838	0.0000	0.0000	0.0000	0.0000	0.0000	0.0059
BM1986b	-0.4070		-8.2370	-7.1400	-6.4570	-6.4010	-7.3130	-3.3610
	0.6838		0.0000	0.0000	0.0000	0.0000	0.0000	0.0008
R1991	8.2370	8.2370		6.307	8.2370	8.2370	4.885	6.7220
	0.0000	0.0000		0.000	0.0000	0.0000	0.000	0.0000
YP2006 I&O	7.8260	7.1400	-6.307		5.4660	4.5340	0.0720	2.3000
	0.0000	0.0000	0.000		0.0000	0.0000	0.9423	0.0215
YP2006 I	7.4940	6.4570	-8.2370	-5.4660		-0.5250	-4.2400	-0.6920
	0.0000	0.0000	0.0000	0.0000		0.5996	0.0000	0.4888
H2014	7.3630	6.4010	-8.2370	-4.5340	0.5250		-4.9300	-0.8720
	0.0000	0.0000	0.0000	0.0000	0.5996		0.0000	0.3831
VRS	7.1290	7.3130	-4.885	-0.0720	4.2400	4.9300		5.0910
	0.0000	0.0000	0.000	0.9423	0.0000	0.0000		0.0000
C1981	2.7510	3.3610	-6.7220	-2.3000	0.6920	0.8720	-5.0910	
	0.0059	0.0008	0.0000	0.0215	0.4888	0.3831	0.0000	

The null hypothesis is rejected at the 5% level for all but five pairs of models (BM1986a and BM1986b; YP2006-I&O and VRS; YP2006-I and H2014; YP2006-I and C1981; H2014 and C1981)[92]. These five pairs of models appear in light grey cells. Compared to the observed categorization, the null hypothesis is accepted for an additional pair (H2014 and C1981). For the following pairs of models, there is a statistically significant difference between the efficiency scores in the first model mentioned and the efficiency scores in the second model mentioned: BM1986a and R1991, BM1986a and YP2006-I&O, BM1986a and YP2006-I, BM1986a and H2014, BM1986a and VRS, BM1986a and C1981, BM1986b and R1991, BM1986b and YP2006-I&O, BM1986b and YP2006-I, BM1986b and H2014, BM1986b and VRS, BM1986b and C1981, R1991 and YP2006-I&O, R1991 and YP2006-I, R1991 and H2014, R1991 and VRS, R1991 and C1981, YP2006-I&O and YP2006-I, YP2006-I&O and H2014, YP2006-I&O and C1981, YP2006-I and VRS, H2014 and VRS, VRS and C1981.

Among the five pairs for which the null hypothesis is accepted, only two pairs concern models which exclusively allow for an environmental adjustment (BM1986a and BM1986b; YP2006-I and H2014). These pairs are the same identified by the Wilcoxon signed rank sum test in the observed categorization.

Based on Table 26, the following facts are established:

92 Two additional tests have also been performed (a Wilcoxon-Mann-Whitney test and a Kolgomorov-Smirnov test). The results of the Wilcoxon-Mann-Whitney test are similar to the results of the Wilcoxon signed rank sum test. The results of the Kolgomorov-Smirnov test are similar to the results of the Wilcoxon signed rank sum test, except that the null hypothesis is accepted by two additional pairs of models (H2014 and VRS; VRS and C1981).

- Two of the pairs of models which allow for an environmental adjustment do not have a statistically significant difference between their efficiency scores (BM1986a and BM1986b; YP2006-I and H2014).
- The other pairs of models which allow for an environmental adjustment have a statistically significant difference between their efficiency scores (BM1986a and R1991; BM1986a and YP2006-I&O; BM1986a and YP2006-I&O; BM1986a and H2014; BM1986b and R1991; BM1986b and YP2006-I&O; BM1986b and YP2006-I; BM1986b and H2014; R1991 and YP2006-I&O; R1991 and YP2006-I; R1991 and H2014; YP2006-I&O and YP2006-I; YP2006-I&O and H2014).
- The Wilcoxon signed rank sum test performed on each pair of models in the theoretical categorization is the fourth indication that the results for the majority of models which allow for an environmental adjustment are divergent.

To sum up

Table 27 sums up the Pearson, Spearman and Wilcoxon signed rank sum analysis. Among the five models which allow for an environmental adjustment (BM1986a, BM1986b, R1991, YP2006 and H2014):

- The results of BM1986a seem to diverge with R1991, YP2006 and H2014 based on the Wilcoxon test. Consequently, the choice of the model (made by politicians or decision makers) impacts school management in terms of schools' input targets and rankings. According to the model selected, the technical efficiency score and the ranking of a particular school are divergent.

  The results of BM1986a and BM1986b seem to converge based on the Wilcoxon test. This finding is in line with Harrison *et al.* (2012) who conclude that both models perform equally well with small or medium sample sizes. However, the Pearson and the Spearman correlations are weak in the observed categorization. From a managerial perspective, the choice of the model is therefore not meaningless in terms of schools' efficiency scores and rankings.

  Figure 4 shows the efficiency scores (in the observed categorization) of BM1986a and BM1986b for each school[93]. Eight schools out of 90 have an efficiency score which differs by more than 5% between the two models. These schools are assigned by their respective numbers on the figure. For instance, school # 11 has an efficiency score of 1 and is equally ranked # 1 ex aequo with the other efficient schools in the BM1986a model[94]; however, it has an efficiency score of 0.8415 and is ranked # 90

93 In order to facilitate comparison, schools are arranged in the figure according to the efficiency scores obtained by the BM1986a model. Note that the Y-axis is truncated at the value of 0.7.

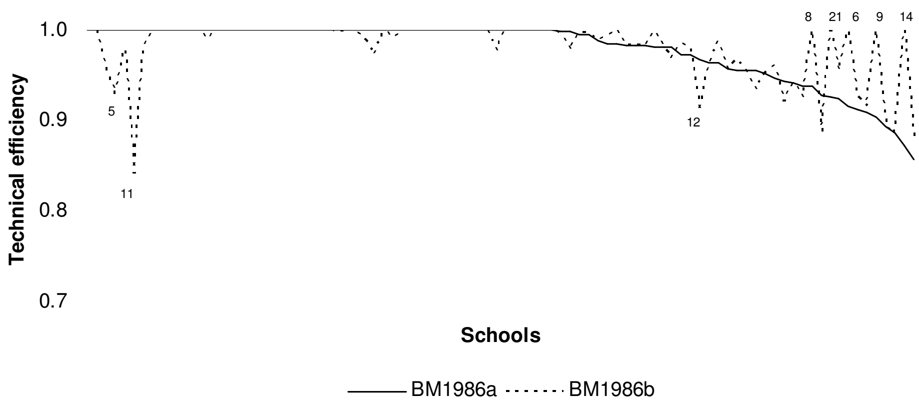
94 According to the Spearman method of calculating the ranks, school # 11 is ranked # 26 (compared to all the other efficient schools).

in the BM1986b model. For such a school, the choice of the model implies serious managerial consequences. In the BM1986a model, school # 11 is considered efficient and is top-ranked. In the BM1986b model, school # 11 should reduce its inputs by 15.85% in order to become efficient and is ranked last.

Among these eight schools, seven are in category E and one in category D (school # 21). Schools # 5, 11 and 12 have a SOCIO value of under 50%. The other five schools have a SOCIO value higher than 60%. Two interpretations can be made. First, it seems that the difference of efficiency scores between BM1986a and BM1986b grows alongside the value of SOCIO. Second, it seems that, among the eight schools, BM1986b tends to allocate a smaller efficiency score, compared to BM1986a, to schools with a relatively small value of SOCIO, and a higher efficiency score to schools with a relatively high value of SOCIO. This is not surprising, as BM1986a does not discriminate among schools in the same category, as opposed to BM1986b, which actually takes into consideration the individual value of SOCIO for each school.

Note that when the eight schools mentioned above are taken out of the sample, the Pearson and the Spearman correlations of the 82 remaining schools have a value of 0.9261 and of 0.8191 respectively. Both correlations are considered as strong and are significant at the 1% level.

Figure 4
Efficiency scores provided by BM1986a and BM1986b for each school

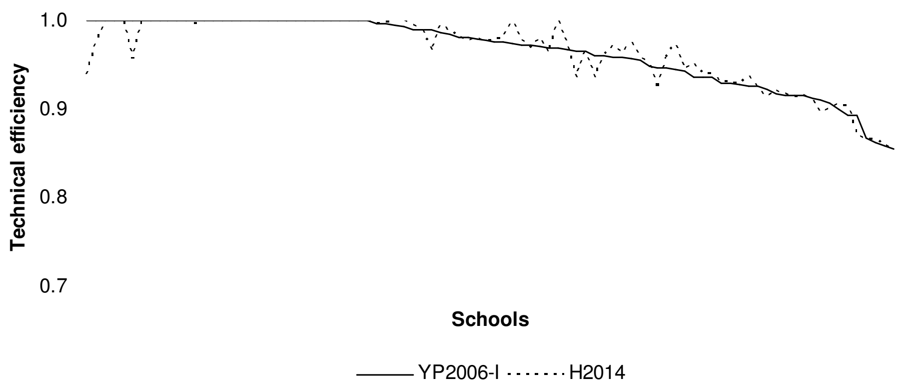


- The results of BM1986b seem to diverge with R1991, YP2006-I&O, YP2006-I and H2014 based on the Wilcoxon test.
- The results of R1991 seem to diverge with BM1986a, BM1986b, YP2006-I&O, YP2006-I and H2014 based on based on the Wilcoxon test.
- The results of YP2006-I&O seem to diverge with BM1986a, BM1986b, R1991, YP2006-I and H2014 based on the Wilcoxon test. The fact that

YP2006-I&O and YP2006-I diverge is problematic. It shows that, within the same model, the choice of adjusting inputs and/or outputs lead to different results.

- The results of YP2006-I seem to diverge with BM1986a, BM1986b, R1991 and YP2006-I&O based on the Wilcoxon test. However, they seem to converge with H2014 (see Figure 5[95]). The converging results between YP2006-I and H2014 are easily understandable, as these two models are very similar. Recall that when H2014 adjusts only the input impacted by the positive discrimination policy (TEACHER), YP2006-I adjusts all inputs, impacted or not by the above mentioned policy (TEACHER, ADMIN and BUDGET).

Figure 5
Efficiency scores provided by YP2006-I and H2014 for each school



- The results of H2014 seem to diverge with BM1986a, BM1986b, R1991 and YP2006-I&O based on the Wilcoxon test. However, they seem to converge with YP2006-I.
- The results of the Wilcoxon signed rank sum test in the observed and in the theoretical categorizations are convergent for the pairs of models composed exclusively by those allowing for an environmental adjustment.

In cases of divergence, the choice of the model (made by politicians or decision makers) impacts school management in terms of schools' input targets and rankings. According to the model selected, the technical efficiency score and the ranking of a particular school are divergent.

When the models allowing for an environmental adjustment are compared with the VRS and the C1981 models, the following conclusions are made:

- The results of VRS seem to diverge with BM1986a, BM1986b, R1991, YP2006-I and H2014 based on the Wilcoxon test.

---

95 In order to facilitate comparison, schools are arranged in the figure according to the efficiency scores obtained by the YP2006-I model. Note that the Y-axis is truncated at the value of 0.7.

The VRS results seem to converge with YP2006-I&O based on the Wilcoxon test. As YP2006-I&O adjusts the efficiency scores for the environmental influence and VRS does not, the fact that these two models provide convergent efficiency scores is a counterintuitive result. However, Muñiz *et al.* (2006) show that the YP2006 model tends to overestimate inefficiency (in other words, to underestimate efficiency) when compared to the 'true' efficiency. As a matter of fact, the mean efficiency of the VRS and the YP2006 models is 0.9321 and 0.934 respectively. Among all of the models which allow for an environmental adjustment (except for the R1991 model), the YP2006-I&O model produces the lowest mean efficiency. This could explain why its results are convergent with the VRS results.

- The results of C1981 seem to diverge with BM1986a, BM1986b, R1991 and YP2006-I&O based on the Wilcoxon test. However, they converge with YP2006-I.

  The picture between C1981 and H2014 is not clear. Based on the Wilcoxon test, the results of these two models seem to diverge when the observed categorization is considered; but the results seem to converge when the theoretical categorization is considered. In both cases, the Pearson and Spearman correlations are weak. The convergence in the case of the theoretical categorization is surprising, as H2014 adjusts for the environment and C1981 adjusts for managerial inefficiency. In H2014, efficiency scores are devoid of environmental effects and reveal managerial inefficiency. In C1981, efficiency scores are devoid of managerial inefficiency and reveal the impact of the environment.

Interested readers will find a graphical representation of every pair of models in Appendix 2.

Table 27
A diagnostic per model

Model	Pearson	Spearman	Wilcoxon	Diagnostic
BM1986a	**Observed categorization (O)** Perfect: - Strong: YP2006-I, H2014 Moderate: R1991, VRS Weak: BM1986b, YP2006-I&O Very weak: C1981 **Theoretical categorization (T)** Perfect: - Strong: BM1986b, YP2006-I, H2014 Moderate: R1991, YP2006-I&O Weak: VRS Very weak: C1981	**Observed categorization (O)** Perfect: - Strong: YP2006-I, H2014 Moderate: R1991, YP2006-I&O Weak: BM1986b, VRS Very weak: C1981 **Theoretical categorization (T)** Perfect: - Strong: YP2006-I Moderate: BM1986b, R1991, YP2006-I&O, H2014 Weak: VRS Very weak: C1981	**No difference** BM1986b (O+T) **Difference** R1991 (O+T) YP2006-I&O (O+T) YP2006-I (O+T) H2014 (O+T) VRS (O+T) C1981 (O+T)	**Comments** No difference between BM1986a and BM1986b Weak Pearson and Spearman correlations with BM1986b (O) Strong Pearson and moderate Spearman correlations with BM1986b (T) **Conclusion** The results of BM1986a and BM1986b seem to converge The results of BM1986a and the other models seem to diverge
BM1986b	**Observed categorization (O)** Perfect: - Strong: - Moderate: R1991, YP2006-I, H2014 Weak: BM1986a, VRS Very weak: YP2006-I&O, C1981 **Theoretical categorization (T)** Perfect: - Strong: BM1986a Moderate: R1991, YP2006-I, H2014 Weak: YP2006-I&O, VRS Very weak: C1981	**Observed categorization (O)** Perfect: - Strong: - Moderate: R1991, YP2006-I, H2014 Weak: BM1986a, YP2006-I&O, VRS Very weak: C1981 **Theoretical categorization (T)** Perfect: - Strong: - Moderate: BM1986a, R1991, YP2006-I, H2014 Weak: YP2006-I&O, VRS Very weak: C1981	**No difference** BM1986a (O+T) **Difference** R1991 (O+T) YP2006-I&O (O+T) YP2006-I (O+T) H2014 (O+T) VRS (O+T) C1981 (O+T)	**Comments** No difference between BM1986b and BM1986a Weak Pearson and Spearman correlations with BM1986a (O) Strong Pearson and moderate Spearman correlations with BM1986a (T) **Conclusion** The results of BM1986b and BM1986a seem to converge The results of BM1986b and the other models seem to diverge
R1991	**Observed categorization (O)** Perfect: - Strong: - Moderate: BM1986a, BM1986b, H2014, VRS, YP2006-I Weak: YP2006-I&O, C1981 Very weak: - **Theoretical categorization (T)** Perfect: - Strong: YP2006-I, H2014 Moderate: BM1986a, BM1986b, YP2006-I&O, VRS Weak: - Very weak: C1981	**Observed categorization (O)** Perfect: - Strong: - Moderate: BM1986a, BM1986b, YP2006-I, H2014 Weak: YP2006-I&O, VRS Very weak: C1981 **Theoretical categorization (T)** Perfect: - Strong: - Moderate: BM1986a, BM1986b, YP2006-I&O, YP2006-I, H2014 Weak: VRS Very weak: C1981	**Difference** BM1986a (O+T) BM1986b (O+T) YP2006-I&O (O+T) YP2006-I (O+T) H2014 (O+T) VRS (O+T) C1981 (O+T)	**Comments** Difference between R1991 and the other models **Conclusion** The results of R1991 and the other models seem to diverge

Table 27
A diagnostic per model (continued)

Model	Pearson	Spearman	Wilcoxon	Diagnostic
YP2006-I&O	**Observed categorization (O)** Perfect: - Strong: - Moderate: YP2006-I Weak: BM1986a, R1991, H2014 Very weak: BM1986b, VRS, C1981 **Theoretical categorization (T)** Perfect: - Strong: - Moderate: BM1986a, R1991, YP2006-I Weak: BM1986b, H2014, C1981 Very weak: VRS	**Observed categorization (O)** Perfect: - Strong: - Moderate: BM1986a Weak: BM1986b, R1991, YP2006-I, H2014 Very weak: VRS, C1981 **Theoretical categorization (T)** Perfect: - Strong: - Moderate: BM1986a, R1991, YP2006-I Weak: BM1986b, H2014 Very weak: VRS, C1981	**No difference** VRS (O+T) **Difference** BM1986a (O+T) BM1986b (O+T) R1991 (O+T) YP2006-I (O+T) H2014 (O+T) C1981 (O+T)	**Comments** No difference between YP2006-I&O and VRS Very weak Pearson and Spearman correlations with VRS (O+T) **Conclusion** The results of YP2006-I&O and VRS seem to converge The results of YP2006-I&O and the other models seem to diverge
YP2006-I	**Observed categorization (O)** Perfect: - Strong: BM1986a, H2014 Moderate: BM1986b, R1991, YP2006-I&O, VRS Weak: - Very weak: C1981 **Theoretical categorization (T)** Perfect: - Strong: BM1986a, R1991, H2014 Moderate: BM1986b, YP2006-I&O, VRS Weak: - Very weak: C1981	**Observed categorization (O)** Perfect: - Strong: BM1986a, H2014 Moderate: BM1986b, R1991 Weak: YP2006-I&O, VRS Very weak: C1981 **Theoretical categorization (T)** Perfect: - Strong: BM1986a, H2014 Moderate: BM1986b, R1991, YP2006-I&O, VRS Weak: - Very weak: C1981	**No difference** H2014 (O+T) C1981 (O+T) **Difference** BM1986a (O+T) BM1986b (O+T) R1991 (O+T) YP2006-I&O (O+T) C1981 (O+T)	**Comments** No difference between YP2006-I and H2014 Strong Pearson and Spearman correlations with H2014 (O+T) No difference between YP2006-I and C1981 Very weak Pearson and Spearman correlations with C1981 (O+T) **Conclusion** The results of YP2006-I and H2014 seem to converge The results of YP2006-I and C1981 seem to converge The results of YP2006-I and the other models seem to diverge
H2014	**Observed categorization (O)** Perfect: - Strong: BM1986a, YP2006-I Moderate: BM1986b, R1991, VRS Weak: YP2006-I&O, C1981 Very weak: - **Theoretical categorization (T)** Perfect: - Strong: BM1986a, R1991, YP2006-I Moderate: BM1986b, VRS Weak: YP2006-I&O, C1981 Very weak: -	**Observed categorization (O)** Perfect: - Strong: BM1986a, YP2006-I Moderate: BM1986b, R1991, VRS Weak: YP2006-I&O, C1981 Very weak: - **Theoretical categorization (T)** Perfect: - Strong: YP2006-I Moderate: BM1986a, BM1986b, R1991, VRS Weak: YP2006-I&O, C1981 Very weak: -	**No difference** YP2006-I (O+T) C1981 (T) **Difference** BM1986a (O+T) BM1986b (O+T) R1991 (O+T) YP2006-I&O (O+T) VRS (O+T) C1981 (O)	**Comments** No difference between H2014 and YP2006-I (O+T) Strong Pearson and Spearman correlations with H2013 (O+T) No difference between H2014 and C1981 (T) Weak Pearson and Spearman correlations with C1981 (O+T) **Conclusion** The results of H2014 and YP2006-I seem to converge The results of H2014 and C1981 seem to converge (T) The results of H2014 and C1981 seem to diverge (O) The results of H2014 and the other models seem to diverge

Table 27
A diagnostic per model (continued)

Model	Pearson	Spearman	Wilcoxon	Diagnostic
VRS	**Observed categorization (O)** Perfect: - Strong: C1981 Moderate: BM1986a, R1991, YP2006-I, H2014 Weak: BM1986b Very weak: YP2006-I&O **Theoretical categorization (T)** Perfect: - Strong: C1981 Moderate: R1991, YP2006-I, H2014 Weak: BM1986a, BM1986b Very weak: YP2006-I&O	**Observed categorization (O)** Perfect: - Strong: C1981 Moderate: H2014 Weak: BM1986a, BM1986b, R1991, YP2006-I Very weak: YP2006-I&O **Theoretical categorization (T)** Perfect: - Strong: C1981 Moderate: YP2006-I, H2014 Weak: BM1986a, BM1986b, R1991 Very weak: YP2006-I&O	**No difference** YP2006-I&O (O+T) **Difference** BM1986a (O+T) BM1986b (O+T) R1991 (O+T) YP2006-I (O+T) H2014 (O+T) C1981 (O+T)	**Comments** No difference between VRS and YP2006-I&O Very weak Pearson and Spearman correlations with YP2006-I&O (O+T) **Conclusion** The results of VRS and YP2006-I&O seem to converge The results of VRS and the other models seem to diverge
C1981	**Observed categorization (O)** Perfect: - Strong: VRS Moderate: - Weak: R1991, H2014 Very weak: BM1986a, BM1986b, YP2006-I&O, YP2006-I **Theoretical categorization (T)** Perfect: - Strong: VRS Moderate: - Weak: YP2006-I&O, H2014 Very weak: BM1986a, BM1986b, R1991, YP2006-I	**Observed categorization (O)** Perfect: - Strong: VRS Moderate: - Weak: H2014 Very weak: BM1986a, BM1986b, R1991, YP2006-I, YP2006-I&O **Theoretical categorization (T)** Perfect: - Strong: VRS Moderate: - Weak: H2014 Very weak: BM1986a, BM1986b, R1991, YP2006-I&O, YP2006-I	**No difference** YP2006-I (O+T) H2014 (T) **Difference** BM1986a (O+T) BM1986b (O+T) R1991 (O+T) YP2006-I&O (O+T) H2014 (O) VRS (O+T)	**Comments** No difference between C1981 and YP2006-I (O+T) Very weak Pearson and Spearman correlations with YP2006-I (O+T) No difference between C1981 and H2014 (T) Weak Pearson and Spearman correlations with H2014 (O+T) **Conclusion** The results of C1981 and YP2006-I seem to converge The results of C1981 and H2014 seem to converge (T) The results of C1981 and H2014 seem to diverge (O) The results of C1981 and the other models seem to diverge

3.9 Further analysis

This study could be prolonged by several means which are discussed hereafter.

- When dealing with empirical data, the quality of data is a serious concern, especially when a particular variable is used to group entities into different categories. Even when the quality of data seems appropriate, the discretionary power of decision makers could potentially bias the categories. Using different or additional variables to group entities into categories could also potentially modify the results. In the current study, two alternative categorizations have been considered (and tested). 37.8% of schools have been moved from the first categorization (observed) to the second categorization (theoretical). It has been concluded that the results of the models which allow for an environmental adjustment are unaffected by the categorization. Further studies should confirm this conclusion.
- Additional models (three- and four-stage models) could be performed and compared with the models included in the current study. However, as models are compared in pairs, the results of the pairs of models performed in this study would remain the same. It must be noted that this study has positioned itself from the standpoint of practitioners and decision makers. As a result, it has voluntarily omitted some models.
- The Pearson and the Spearman correlations might be influenced by the fact that many schools have efficiency scores equal to one. Table 18 displays the number of efficient schools by model. The BM1986a model identifies the highest number of efficient schools: 51 out of 90 (56.67%) in the observed categorization. The R1991 model identifies the lowest number of efficient schools: 1 out of 90 (1.11%). Table 21 and Table 23 display the Pearson and the Spearman correlations across models in the observed categorization. The variations in correlation coefficients between the models do not seem to be influenced by the number of efficient schools. For instance, the Pearson correlations between BM1986a (51 efficient schools) and the other models are as follows: 0.5811 (BM1986b, 46 efficient schools); 0.6787 (R1991, one efficient school); 0.5762 (YP2006, 17 efficient schools); 0.9201 (YP2006-I, 31 efficient schools); 0.8983 (H2014, 31 efficient schools); 0.6144 (VRS, 20 efficient schools); 0.1679 (C1981, 25 efficient schools)[96].
- In relation to variables, Smith and Mayston (1987) argue the following:

  > The choice and relative importance of outputs is ultimately a political judgement, and no amount of mathematical analysis can reconcile the diversity of views concerning priorities in the public sector. The user of DEA has to recognise this limitation, and at the very least it would seem sensible to test the implications of a variety of output sets (p.188).

[96] In this example, the Pearson correlation between the number of efficient schools and the associated Pearson coefficients is equal to 0.3365.

The main findings of the current study indicate that results diverge according to the model performed (with the exception of the BM1986a and BM1986b models and of the YP2006-I and H2014 models). Ultimately, there is no consensus on the best model to use (Cordero-Ferrara *et al.*, 2008). Echoing Smith and Mayston (1987), the choice of model is ultimately a political judgement. Practitioners and decision makers have to select the model which is right *for them*, in other words, the model which best suits their own criteria (not to say the model which best serves their own interests). In this sense, the application of an appropriate multi-criteria decision analysis method to help decision makers select the right model should be investigated in further studies.

3.10 Conclusion

This study tests how several alternative models, within DEA, potentially lead to divergent results. Five models which allow for an environmental adjustment are retained based on their degree of sophistication, their inclusion in existing software and the level of computational skills that they require. These models are the following: Banker and Morey (1986a), Banker and Morey (1986b), Ray (1991), Yang and Paradi in Muñiz *et al.* (2006, p. 1176) and a new model developed in this study called Huguenin (2014). Unlike studies using simulated data to compare efficiency scores from several models, this study uses empirical data concerning 90 primary schools in the State of Geneva, Switzerland. With the exception of Ruggiero (1998), no existing study tests so many models.

The results of the five models are compared on the basis of *(1)* a Pearson and a Spearman correlation analysis and *(2)* a Wilcoxon signed rank sum test analysis. Except for BM1986a and BM1986b and for YP2006-I and H2014, whose results seem to converge, each and every other pair of models (for instance R1991 and BM1986b) provide diverging results. In other words, the efficiency scores generated by the models forming each pair are significantly different. This finding is valid for the specific empirical dataset used in the current study. For this reason, it cannot be generalized to other datasets. However, the fact that the efficiency scores diverge in the current study may suggest that the results obtained from several alternative models may diverge in other cases too.

Applied DEA studies traditionally end with recommendations and policy implications. See for instance McCarty and Yaisawarng (1993, pp. 285-286), Kantabutra and Tang (2006, pp. 370-372) or Jeon and Shields (2008, p. 611). Most of these studies base their recommendations on the efficiency results produced by a particular DEA model. This appears to be problematic. As shown in this study, several alternative models to measure efficiency, within DEA, deliver diverging results. Consequently, recommendations and policy implications may differ according to the model used. From a political standpoint, these diverging results could potentially lead to ineffective decisions. From an applied research standpoint, they should represent a serious matter of concern. And from a decision making standpoint, they may lead to opposing managerial choices.

As no consensus emerges on the best model to use, practitioners and decision makers may be tempted to select the model which is right *for them*, in other words, the model which best suits their own criteria and preferences. The choice of model thus becomes a strategic issue. Further studies should identify and validate such criteria. Once these criteria are known, the application of an appropriate multi-criteria decision analysis method to help decision makers select the right model should also be investigated.

References

Abbott, M. & Doucouliagos, C. (2003). The efficiency of Australian universities: a data envelopment analysis. *Economics of Education Review*, *22*(1), 89-97.

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., Bonomi, F. & Sibiano, P. (2014). Measuring the "managerial" efficiency of public schools: a case study in Italy. *International Journal of Educational Management, 28*(2), 120-140.

Ahn, T. & Seiford, L. M. (1993). Sensitivity of data envelopment analysis to models and variable sets in a hypothesis test setting: the efficiency of university operations. In Y. Ijiri (Eds.), *Creative and Innovative Approaches to the Science of Management* (pp. 191-208). Westport: Quorum Books.

Alexander, W. R. J. & Jaforullah, M. (2004). Explaining efficiency differences of New Zealand secondary schools. *Economics Discussion Papers No. 0403*. Dunedin: University of Otago.

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*(2), 99-110.

Arcelus, F. J. & Coleman, D. F. (1997). An efficiency review of university departments. *International Journal of Systems Science*, *28*(7), 721-729.

Avkiran, N. K. (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Economic Planning Science*, *35*(1), 57-80.

Badillo, P.-Y. & Paradi, J. C. (1999). *La méthode DEA: analyse des performances*. Paris: HERMES Science Publications.

Banker, R. D., Zheng, Z. & Natarajan R. (2010). DEA-based hypothesis tests for comparing two groups of decision making units. *European Journal of Operational Science*, *206*(1), 231-238.

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.

Banker, R. D. & Morey, R. C. (1986a). The Use of Categorical Variables in Data Envelopment Analysis. *Management Science*, *34*(4), 1613-1627.

Banker, R. D. & Morey, R. C. (1986b). Efficiency Analysis for Exogenously Fixed Inputs and Outputs. *Operations Research*, *32*(12), 513-521.

Barnum, D. T. & Gleason, J. M. (2008). Bias and precision in the DEA two-stage method. *Applied Economics*, *40*(18), 2305-2311.

Beasley, J. E. (1990). Comparing university departments. *Omega*, *18*(2), 171-183.

Bessent, A. M. & Bessent, E. W. (1980). Determining the comparative efficiency of schools through data envelopment analysis. *Educational Administration Quarterly*, *16*(2), 57-75.

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

Bifulco, R. & Bretschneider, S. (2001). Estimating school efficiency. A comparison of methods using simulated data. *Economics of Education Review*, *20*(5), 417-422.

Borge, L.-E. & Naper L. R. (2006). Efficiency Potential and Efficiency Variation in Norwegian Lower Secondary Schools. *FinanzArchiv / Public Finance Analysis*, *62*(2), 221-249.

Bradley, S., Johnes, J. & Little, A. (2010). Measurement and determinants of efficiency and productivity in the further education sector in England. *Bulletin of Economic Research*, *62*(1), 1-30.

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.

Burney, M. A., Johnes, J., Al-Enezi, M. & Al-Mussalam, 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 and 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*, *14*(2), 143-160.

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.

Coelli, T. J. (1996). A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program. *CEPA Working Paper 96/08*. Brisbane: Centre for Efficiency and Productivity Analysis, University of Queensland.

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

Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data Envelopment Analysis: A comprehensive Text with Models, Applications, References and DEA-Solver Software*. New York: Springer.

Cordero-Ferrara, J. M., Pedraja-Chaparro, F. & Salinas-Jiménez, J. (2008). Measuring efficiency in education: an analysis of different approaches for incorporating non-discretionary inputs. *Applied Economics*, *40*(10), 1323-1339.

Cordero, J. M., Pedraja, F. & Santín, D. (2009). Alternative approaches to include exogenous variables in DEA measures: A comparison using Monte Carlo. *Computers & Operations Research*, *36*(10), 2699-2706.

De Witte, K. & Moesen, W. (2010). Sizing the government. *Public Choice, 145*(1), 39-55.

Demeuse, M., Frandji, D., Greger, D. & Rochex, J.-Y. (2012). *Education Policies and Inequalities in Europe*. New York: Palgrage Macmillan.

Demeuse, M. & Friant, N. (2012). Evaluer les politiques d'éducation prioritaire en Europe : un défi méthodologique. *Revue Suisse des Sciences de l'Education*, *34*(1), 39-55.

Demir, I. & Depren, Ö. (2010). Assessing Turkey's secondary schools performance by different region in 2006. *Procedia – Social and Behavioral Sciences*, *2*(1), 2305-2309.

Denaux, Z. S., Lipscomb, C. A. & Plumly, L. W. (2011). Assessing the technical efficiency of public high schools in the state of Georgia. *Review of Business Research*, *11*(5), 46-57

Diagne, D. (2006). Mesure de l'efficience technique dans le secteur de l'éducation : une application de la méthode DEA. *Revue suisse d'économie et de statistique*, *142*(2), 231-262.

Diamond, A. & Medewitz, J. N. (1990). Use of data envelopment analysis in an evaluation of the efficiency of the DEEP program for economic education. *Journal of Economic Education*, *21*(3), 337-354.

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

Emrouznejad, A., Parker, B. R. & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, *42*(3), 151-157.

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

Estelle, S. M., Johnson, A. L. & Ruggiero, J. (2010). Three-stage DEA models for incorporating exogenous inputs. *Computers & Operations Research*, *37*(6), 1087-1090.

Farsi, M. & Filippini, M. (2005). A benchmarking analysis of electricity distribution utilities in Switzerland. *CEPE Working Paper No. 43*. Zurich: Center for Energy Policy and Economics, Swiss Federal Institute of Technology.

Frandji, D. (2008). Pour une comparaison des politiques d'éducation prioritaire en Europe. In M. Demeuse, D. Frandji, D. Greger & J.-Y. Rochex (Eds.), *Les politiques d'éducation prioritaire en Europe: Conceptions, mises en oeuvre, débats* (pp. 9-34). Lyon : Institut national de Recherche pédagogique.

Fried, H. O., Lovell, C. A. K., Schmidt, S. S. & Yaisawarng, S. (2002). Accounting for environmental effects and statistical noise in data envelopment analysis. *Journal of Productivity Analysis*, *17*(1/2), 157-174.

Fried, H. O., Schmidt, S. S. & Yaisawarng, S. (1999). Incoporating the Operating Environment Into a Nonparametric Measure of Technical Efficiency. *Journal of Productivity Analysis*, *12*(3), 249-267.

Garrett, W. A. & Kwak, N. K. (2011). Performance comparisons of Missouri public schools using data envelopment analysis. In K. D. Lawrence & G. Kleinman (Eds.), *Applications in Multicriteria Decision Making, Data Envelopment Analysis, and Finance (Applications of Management Science, Volume 14)* (pp. 135-155). Bingley: Emerald Group Publishing Limited.

Gasparini, C. & Ramos, F. (2003). Efetividade e eficiencia no ensino medio brasileiro. *Economia Aplicada, 7*(2), 389-411.

Grosskopf, S. & Moutray, C. (2001). Evaluating performance in Chicago public high schools in the wake of decentralization. *Economics of Education Review*, *20*(1), 1-14.

Hanushek, E. A. (2006). School Resources. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education, Volume 2* (pp. 865-903). Amsterdam: North-Holland.

Harrison, J., Rouse, P. & Armstrong, J. (2012). Categorical and continuous non-discretionary variables in data envelopment analysis: a comparison of two single-stage models. *Journal of Productivity Analysis*, *37*(3), 261-276.

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

Huguenin, J.-M. (2012). Data Envelopment Analysis (DEA): a pedagogical guide for decision makers in the public sector. *Cahier de l'IDHEAP No. 276*. Lausanne: Swiss Graduate School of Public Administration.

Huguenin, J.-M. (2013a). Data Envelopment Analysis (DEA): un guide pédagogique à l'intention des décideurs dans le secteur public. *Cahier de l'IDHEAP No. 278*. Lausanne : Institut de hautes études en administration publique.

Huguenin, J.-M. (2013b). Data Envelopment Analysis (DEA). In A. Ishizaka & P. Nemery (Eds.), *Multi-Criteria Decision Analysis: Methods and Software* (pp. 235-274). Chichester: John Wiley & Sons.

Jeon, Y. & Shields, M. P. (2008). Integration and utilization of public education resources in remote and homogenous areas: a case study of the upper peninsula of Michigan. *Contemporary Economics Policy*, *23*(4), 601-614.

Johnes, J. (2004). Efficiency measurement. In G. Johnes & J. Johnes (Eds.), *International Handbook on the Economics of Education* (pp. 613-742). Cheltenham: Edward Elgar Publishing.

Kantabutra, S. & Tang, J. C. S. (2006). Urban-rural and size effects on school efficiency: The case of Northern Thailand. *Leadership and Policy in Schools*, *5*(4), 355-377.

Kirjavainen, T. & Loikkanen, 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.

Lee, J.-Y. (2008). Application of the three-stage DEA in measuring efficiency – an empirical evidence. *Applied Economics Letters*, *15*(1), 49-52.

Löber, G. & Staat, M. (2010). Integrating categorical variables in Data Envelopment Analysis models: A simple solution technique. *European Journal of Operational Research*, *202*(3), 810-818.

Lovell, C. A. K., Walters, L. C. & Wood, L. L. (1994). Stratified models of education production using modified data envelopment analysis and regression analysis. In A. Charnes, W. W. Cooper, A. Y. Lewin & L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology and applications* (pp. 329-351). Dordrecht: Kluwer Academic.

McCarty, T. A. & Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. In H. O. Fried, C. A. K Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications* (pp. 271-287). New York: Oxford University Press.

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

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

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

Muñiz, M., Paradi, J., Ruggiero, J. & Yang, Z. (2006). Evaluating alternative DEA models used to control for non-discretionary inputs. *Computers & Operations Research*, *33*(5), 1173-1183.

Observatory on Primary Education (2010). *Allocation des ressources aux établissements* (Report of December 2010). Geneva: General Direction of Primary Schools, Education Department, State of Geneva.

Ouellette, P. & Vierstraete, V. (2005). An evaluation of the efficiency of Québec's school boards using the Data Envelopment Analysis method. *Applied Economics*, *37*(14), 1643-1653.

Organisation for Economic Co-operation and Development (2001). *Measuring Productivity: Measurement of Aggregate and Industry-level Productivity Growth*. Paris: OECD.

Paton, G. (2013, March 4). Schoolchildren losing the power to concentrate in class. *The Daily Telegraph*.

Portela, M. C. A. S., Camanho, A. S. & Borges, D. N. (2011). BESP – benchmarking of Portuguese secondary schools. *Benchmarking: An International Journal*, *18*(2), 240-260.

Portela, M. C. A. S. & Thanassoulis, E. (2001). Decomposing school and school type efficiency. *European Journal of Operational Research*, *132*(2), 114-130.

Ramanathan, R. (2001). A Data Envelopment Analysis of Comparative Performance of Schools in the Netherlands. *Opsearch*, *38*(2), 160-182.

Rassouli-Currier, S. (2007). Assessing the Efficiency of Oklahoma Public Schools: A Data Envelopment Analysis. *Southwestern Economic Review*, *34*(1), 131-144.

Ray, S. C. (1988). Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. *Socio-economic Planning Sciences*, *22*(4), 167-176.

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

Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, *90*(3), 553-565.

Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, *111*(3), 461-469.

Ruggiero, J. (2000). Nonparametric estimation of returns to scale in the public sector with an application to the provision of educational services. *Journal of the Operational Research Society*, *51*(8), 906-912.

Ruggiero, J. (2004). Performance Evaluation in Education. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 323-346). Dordrecht: Springer.

Ruggiero, J. & Vitaliano, D. F. (1999). Assessing the efficiency of public schools using data envelopment analysis and frontier regression. *Contemporary Economic Policy*, *17*(3), 321-331.

Sav, G. T. (2013). Four-Stage DEA Efficiency Evaluations: Financial Reforms in Public University Funding. *International Journal of Economics and Finance*, *5*(1), 24-33.

Sengupta, J. K. (1990). Tests of efficiency in DEA. *Computers Operations Research*, *17*(2), 123-132.

Shang, J.-K., Hung, W.-T., Lo, C.-F. & Wang, F.-C. (2008). Ecommerce and hotel performance: three-stage DEA analysis. *The Service Industries Journal*, *28*(4), 529-540.

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

Soteriou, A. C., Karahanna, E., Papanastasiou, C. & Diakourakis, M. S. (1998). Using DEA to evaluate the efficiency of secondary schools: the case of Cyprus. *International Journal of Educational Management*, *12*(2), 65-73.

Syrjänen, M. J. (2004). Non-discretionary and discretionary factors and scale in data envelopment analysis. *European Journal of Operational Research*, *158*(1), 20-33.

Tavares, G. (2002). A bibliography of Data Envelopment Analysis (1978-2001). *Rutcor Research Report No. 01-02*. Piscataway: Rutgers University.

Thanassoulis, E. (1996). Altering the Bias in Differential School Effectiveness Using Data Envelopment Analysis. *The Journal of the Operational Research Society*, *47*(7), 882-894.

Thanassoulis, E., Portela, M. C. S. & Despic, O. (2008). Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. In H. O. Fried, C. A. K. Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency and Productivity Growth* (pp. 251-420). Oxford: Oxford University Press.

Thrall, R. M. (1996). Duality, classification and slacks in DEA. *Annals of Operations Research*, *66*(2), 109-138.

Viger, G. (2007, June). *L'analyse comparative au service de l'amélioration et de la performance.* Note de cadrage presented at the third meeting of the Contrôle de Gestion des Programmes.

Waldo, S. (2007). Efficiency in Swedish Public Education: Competition and Voter Monitoring. *Education Economics*, *15*(2), 231-251.

Woessmann, L. (2003). Schooling Resources, Educational Institutions and Student Performance: the International Evidece. *Oxford Bulletin of Economics and Statistics*, *65*(2), 117-170.

Wössmann, L. (2005). The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, *13*(2), 143-169.

Yang, H. & Pollitt, M. (2009). Incorporating both undesirable outputs and uncontrollable variables into DEA: The performance of Chinese coal-fired power plants. *European Journal of Operational Research*, *197*(3), 1095-1105.

Yanyan, P. (2012). Commercial Bank Branch Efficiencies Based on Three-Stage DEA Model. In L. Zhang & C. Zhang (Eds.), *Engineering Education and Management* (pp. 465-470). Berlin: Springer.

Zhu, J. (1996). Robustness of the efficient DMUs in data envelopment analysis. *European Journal of Operational Research*, *90*(3), 451-460.

Appendix 1

Banker and Morey (1986a) – One-stage model

The VRS formulation of the categorical model is specified as follows:

Minimize θ_k (5)

Subject to $y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \qquad r = 1, \ldots, s$

$$\theta_k x_{ik}^D - \sum_{j=1}^{n} \lambda_j x_{ij}^D \geq 0 \qquad i = 1, \ldots, m$$

$$\sum_{j=1}^{n} \lambda_j d_{rj}^{(Cr)} \leq d_{rk}^{(Cr)} \qquad r = 1, \ldots, R$$

$$C\,r = 1, \ldots, C-1$$

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

$$\lambda_j \geq 0 \quad \forall j = 1, \ldots, n$$

The third set of constraints corresponds to an index of dummy variables d_{rk}^{Cr} representing categories of the environment. C represents the category level (e.g. school category C) and r represents the category variable (where there is more than one category variable). In the case of the State of Geneva, there is only one category variable (SOCIO). For example, if there are five category levels (A to E), this can be coded using four dummy variables where:

- $d^{(1)}$ equals zero for schools in category E and one for schools in category D, C, B and A;
- $d^{(2)}$ equals zero for schools in category E and D and one for schools in category C, B and A;
- $d^{(3)}$ equals zero for schools in category E, D and C and one for schools in category B and A;

$d^{(4)}$ equals zero for schools in category E, D, C and B and one for schools in category A.

Banker and Morey (1986b) – One-stage model

The VRS formulation of the Banker and Morey (1986b) model is specified as follows:

$$\text{Minimize} \quad \theta_k \tag{6}$$

$$\text{Subject to} \quad y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \le 0 \qquad r = 1, \ldots, s$$

$$\theta_k x_{ik}^{D} - \sum_{j=1}^{n} \lambda_j x_{ij}^{D} \ge 0 \qquad i = 1, \ldots, m$$

$$x_{uk}^{ND} - \sum_{j=1}^{n} \lambda_j x_{uj}^{ND} \ge 0 \qquad u = 1, \ldots, v$$

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

$$\lambda_j \ge 0 \quad \forall j = 1, \ldots, n$$

In the above model, an additional constraint is included for each non-discretionary input (x^{ND}). These constraints are similar to the constraints for the discretionary inputs (x^{D}) with the exception that the efficiency component is not included. As a result, the efficiency is defined with respect to the discretionary inputs only.

Yang and Paradi model in Muñiz, Ruggiero, Paradi and Yang (2006, p. 1176) – **One-stage model**

Assume h_j is the handicapping measure to adjust input variables and $\hat{h}_j$ the handicapping measure to adjust output variables. The adjusted input is $h_j x_{ij}$ and the adjusted output is $\hat{h}_j y_{rj}$. The model is specified as follows:

$$\text{Minimize} \quad \theta_k \tag{7}$$

$$\text{Subject to} \quad \hat{h}_k y_{rk} - \sum_{j=1}^{n} \lambda_j \hat{h}_j y_{rj} \le 0 \qquad r = 1, \ldots, s$$

$$\theta_k h_k x_{ik} - \sum_{j=1}^{n} \lambda_j h_j x_{ij} \ge 0 \qquad i = 1, \ldots, m$$

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

$$\lambda_j \ge 0 \quad \forall j = 1, \ldots, n$$

Charnes, Cooper and Rhodes (1981) – Program analysis model

Charnes *et al.* (1981) use a constant returns to scale model to assess efficiency. This model is defined as follows:

$$\text{Minimize} \quad \theta_k \tag{8}$$

$$\text{Subject to} \quad y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \le 0 \qquad r = 1, \ldots, s$$

$$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} \ge 0 \qquad i = 1, \ldots, m$$

$$\lambda_j \ge 0 \quad \forall j = 1, \ldots, n$$

Appendix 2

This appendix presents two graphical representations for each pair of models in the observed categorization. Both graphs are built with the same data. Note that the Y-axis of all graphs is truncated at the value of 0.7.

The first graph arranges schools in the figure according to the five school categories (category E to category A from left to right). Among a category (for instance school category E), schools are listed by alphabetical order. This graphical representation allows identifying visually where the divergence is mostly concentrated. For instance, Figure 6 displays the efficiency scores of BM1986a and BM1986b. The gap between the two curves is more important on the left of the graph, meaning that the divergence occurs mostly in the disadvantaged schools.

The second graph arranges schools in the figure according to the efficiency scores obtained by one of the two models contained in the graph. For instance, Figure 7 arranges schools in the figure according to the efficiency scores obtained by the BM1986a model.

Figure 6
Efficiency scores provided by BM1986a and BM1986b for each school – first graph

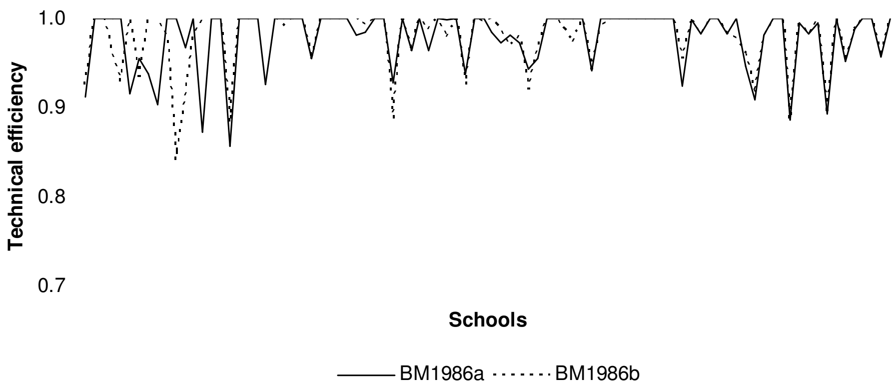


Figure 7
Efficiency scores provided by BM1986a and BM1986b for each school – second graph

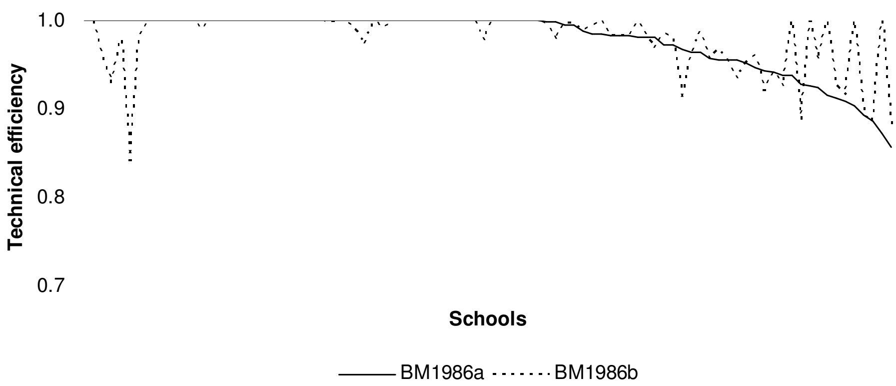


Figure 8
Efficiency scores provided by BM1986a and R1991 for each school – first graph

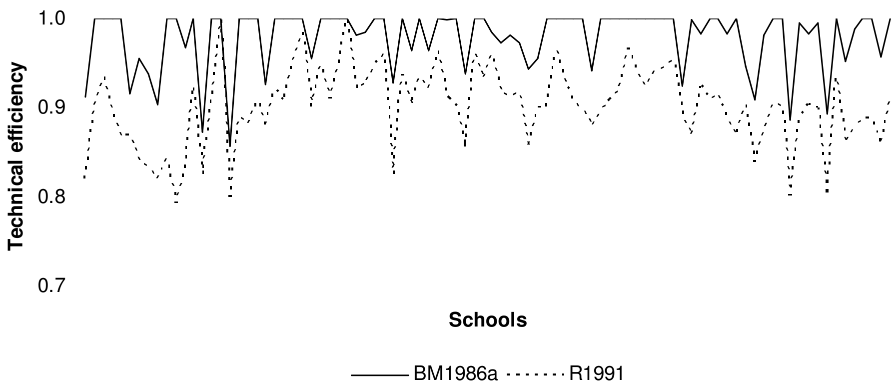


Figure 9
Efficiency scores provided by BM1986a and R1991 for each school – second graph

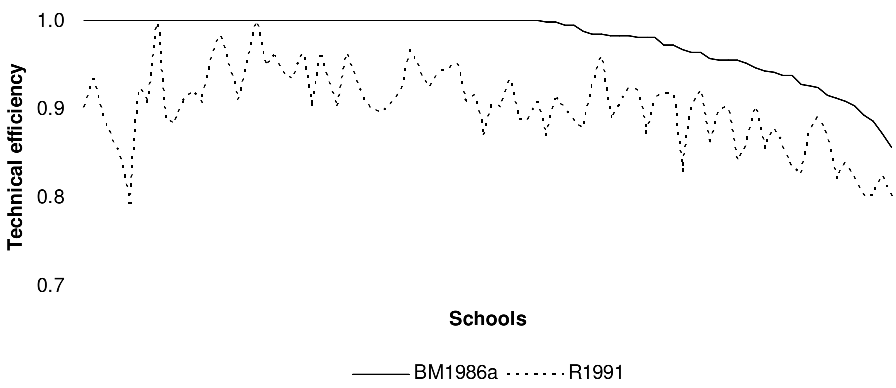


Figure 10
Efficiency scores provided by BM1986a and YP2006-I&O for each school – first graph

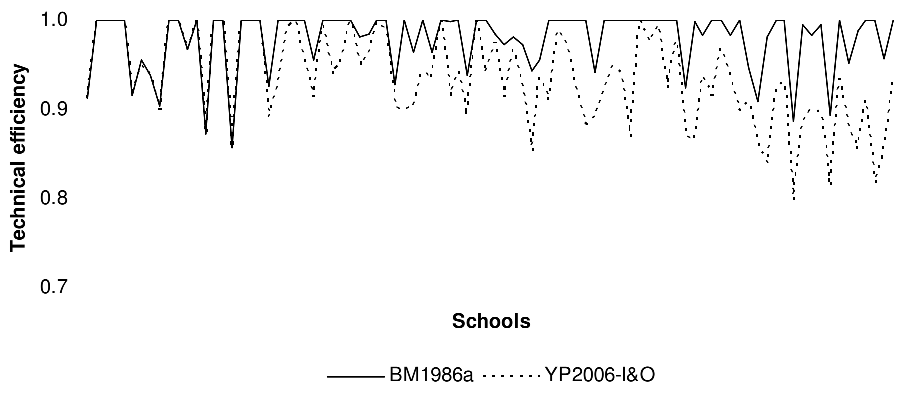


Figure 11
Efficiency scores provided by BM1986a and YP2006-I&O for each school – second graph

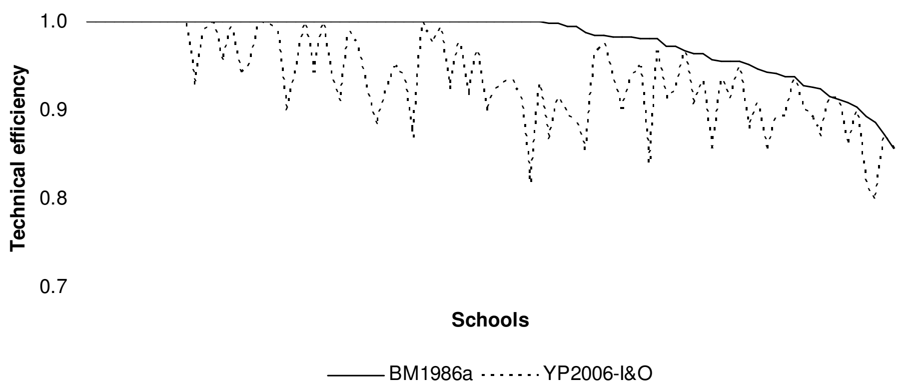


Figure 12
Efficiency scores provided by BM1986a and YP2006-I for each school – first graph

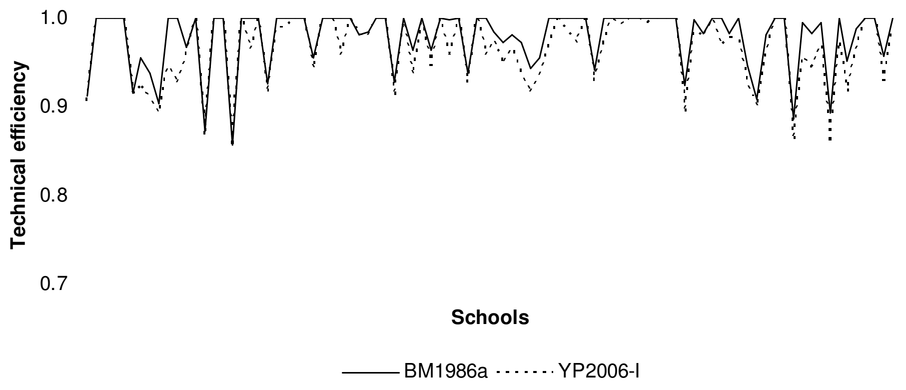


Figure 13
Efficiency scores provided by BM1986a and YP2006-I for each school – second graph

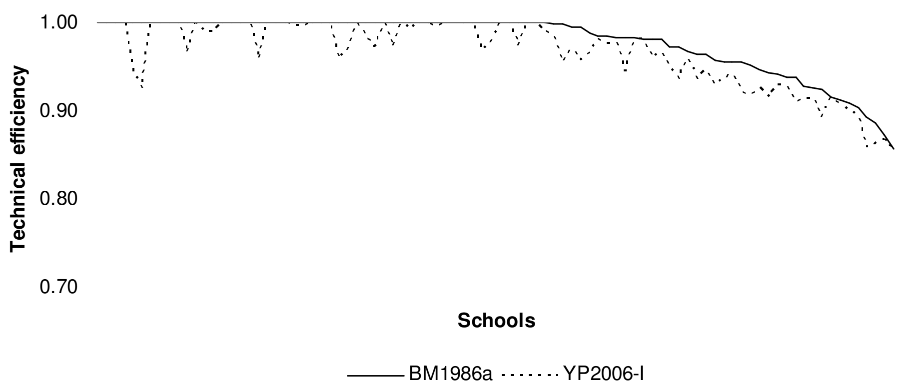


Figure 14
Efficiency scores provided by BM1986a and H2014 for each school – first graph

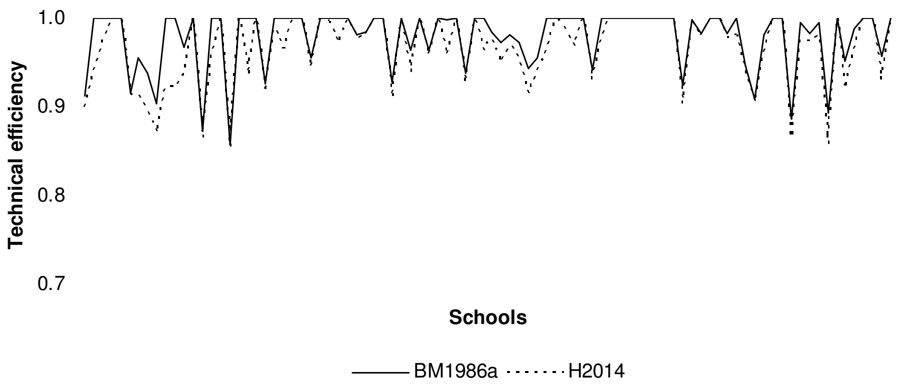


Figure 15
Efficiency scores provided by BM1986a and H2014 for each school – second graph

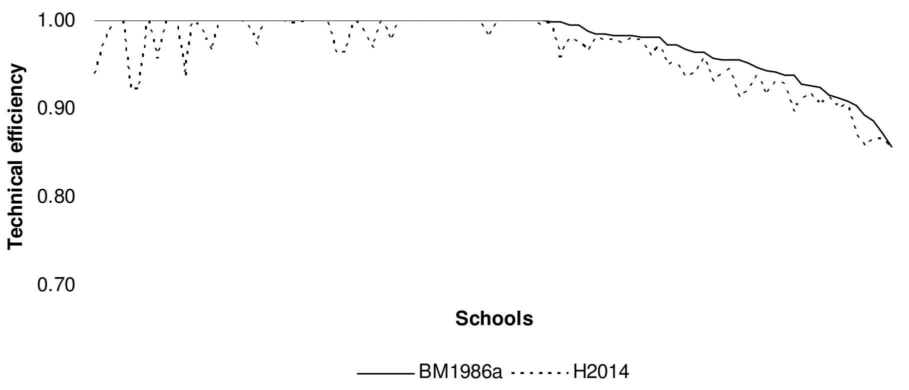


Figure 16
Efficiency scores provided by BM1986a and VRS for each school – first graph

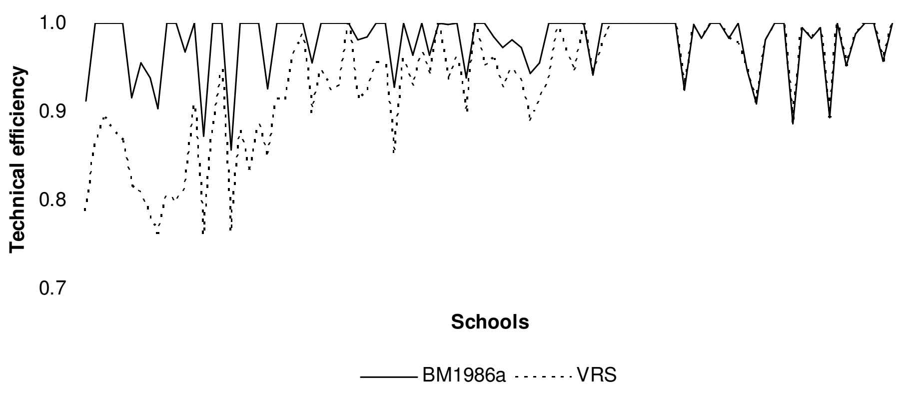


Figure 17
Efficiency scores provided by BM1986a and VRS for each school – second graph

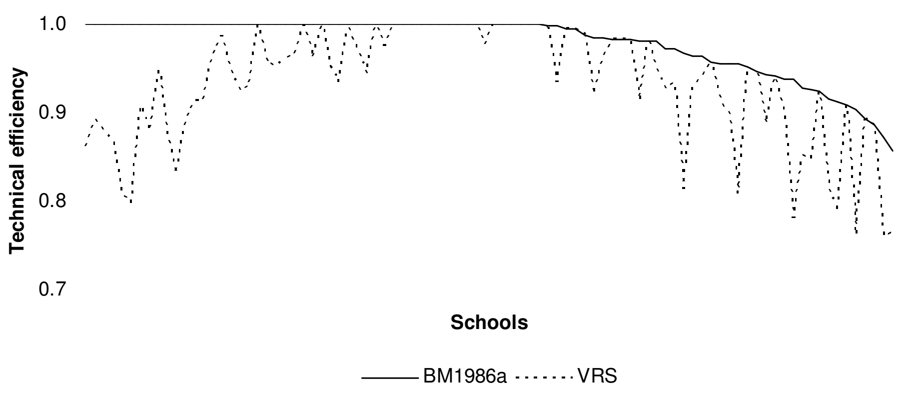


Figure 18
Efficiency scores provided by BM1986a and C1981 for each school – first graph

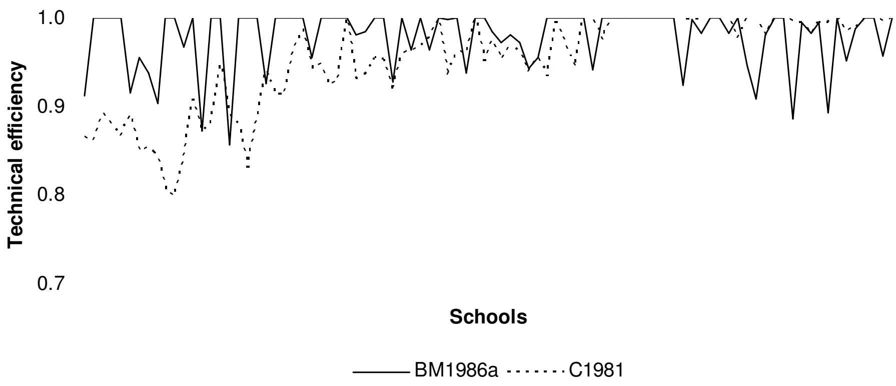


Figure 19
Efficiency scores provided by BM1986a and C1981 for each school – second graph

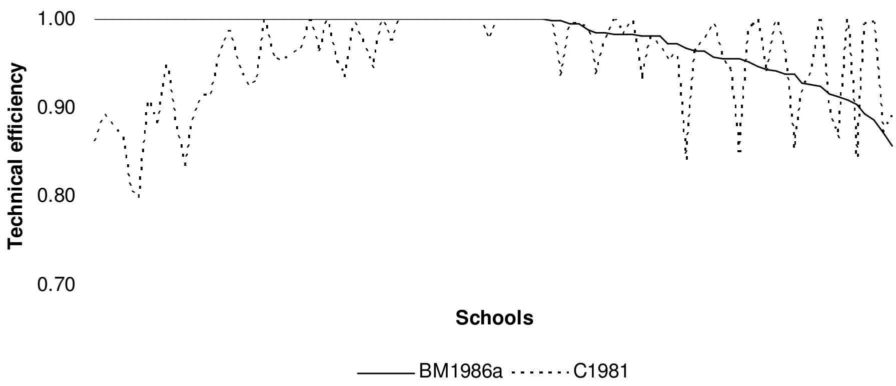


Figure 20
Efficiency scores provided by BM1986b and R1991 for each school – first graph

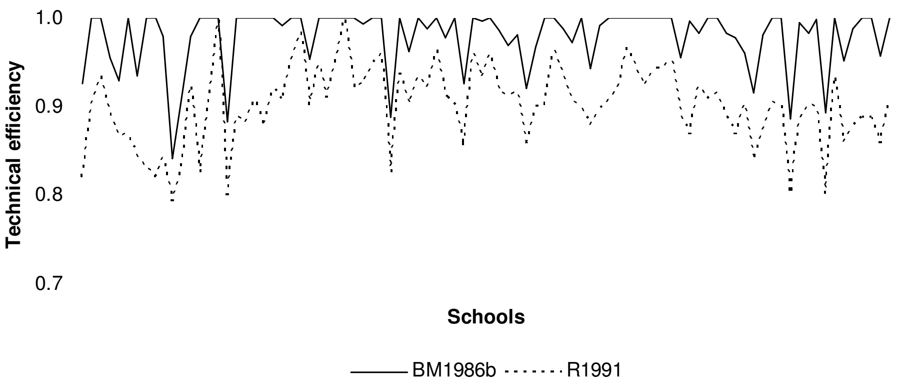


Figure 21
Efficiency scores provided by BM1986b and R1991 for each school – second graph

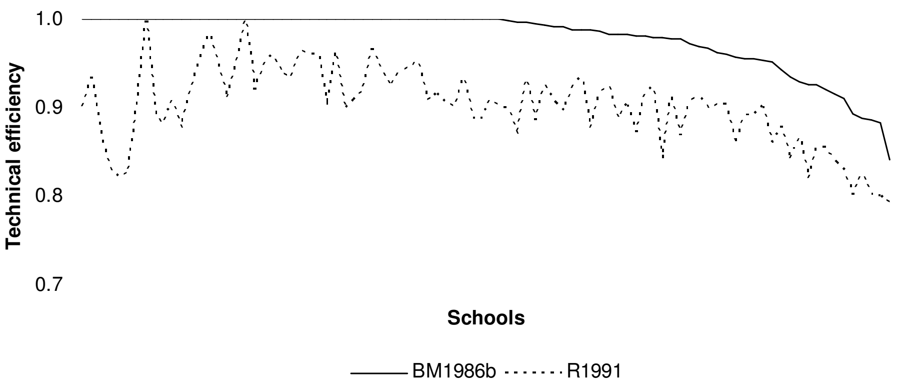


Figure 22
Efficiency scores provided by BM1986b and YP2006-I&O for each school – first graph

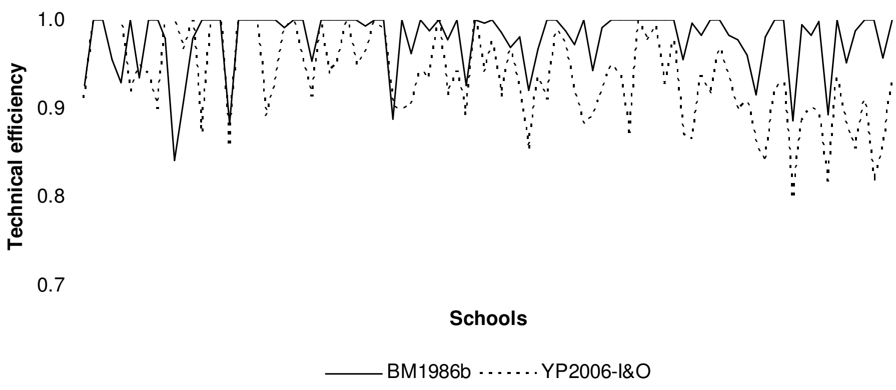


Figure 23
Efficiency scores provided by BM1986b and YP2006-I&O for each school – second graph

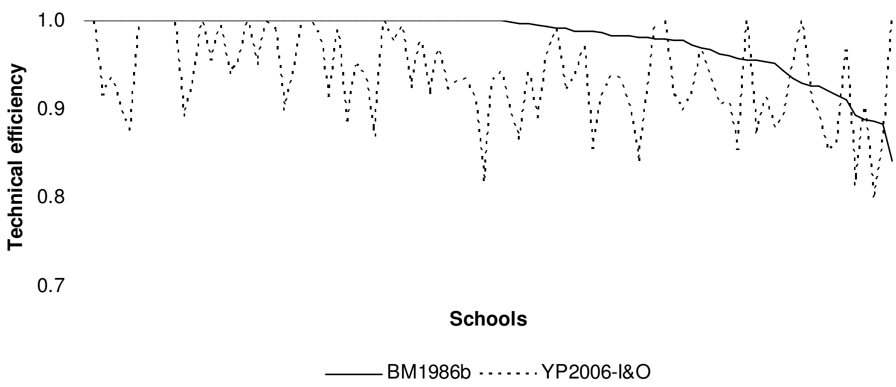


Figure 24
Efficiency scores provided by BM1986b and YP2006-I for each school – first graph

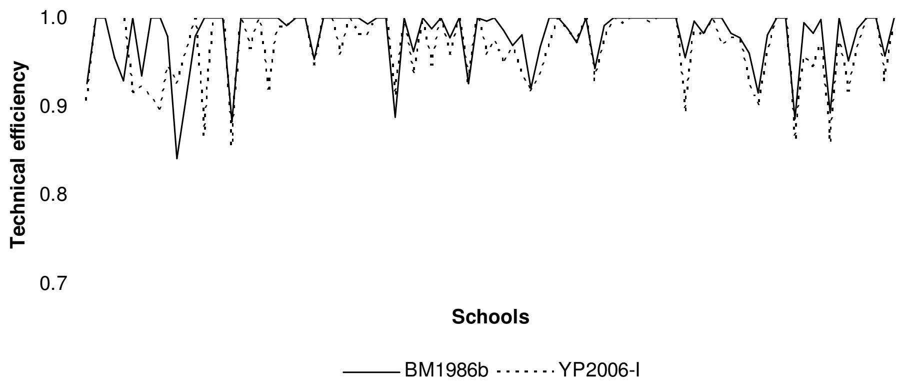


Figure 25
Efficiency scores provided by BM1986b and YP2006-I for each school – second graph

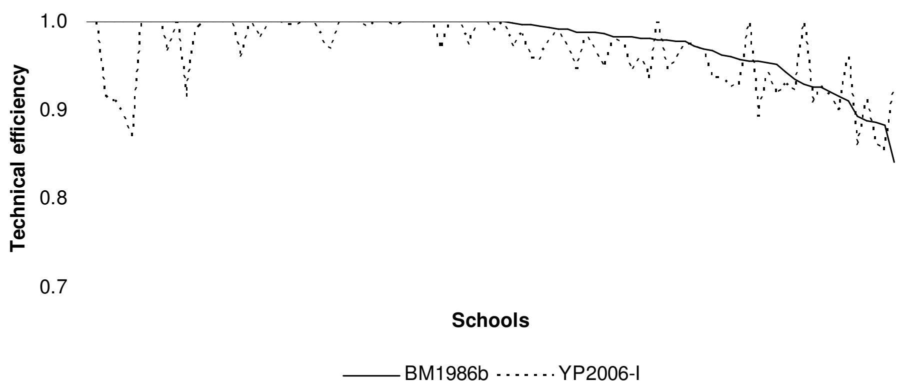


Figure 26
Efficiency scores provided by BM1986b and H2014 for each school – first graph

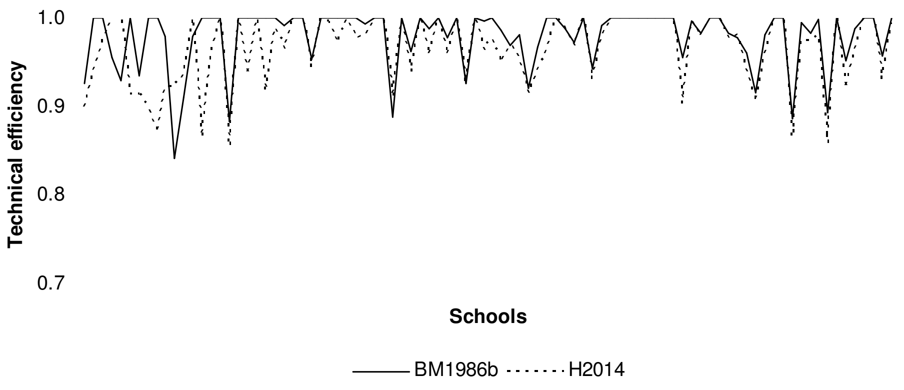


Figure 27
Efficiency scores provided by BM1986b and H2014 for each school – second graph

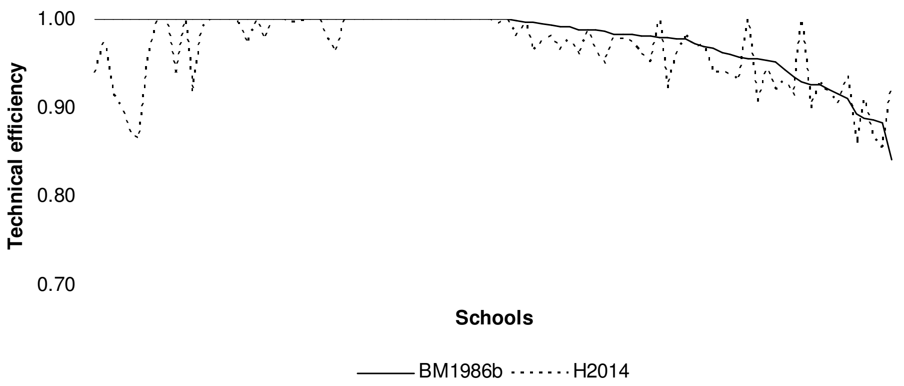


Figure 28
Efficiency scores provided by BM1986b and VRS for each school – first graph

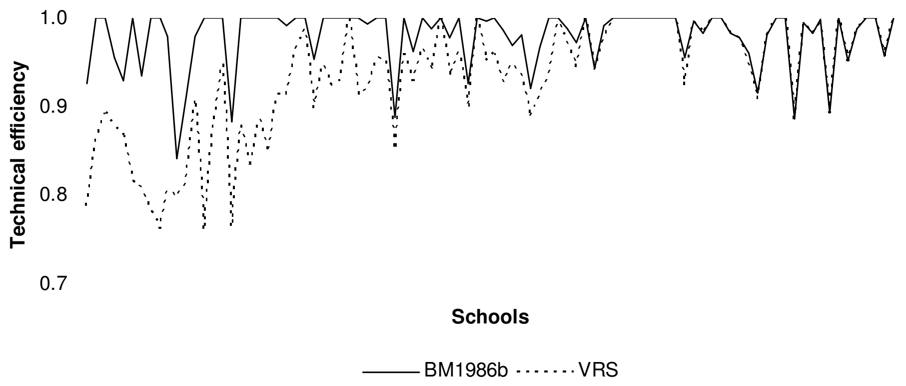


Figure 29
Efficiency scores provided by BM1986b and VRS for each school – second graph

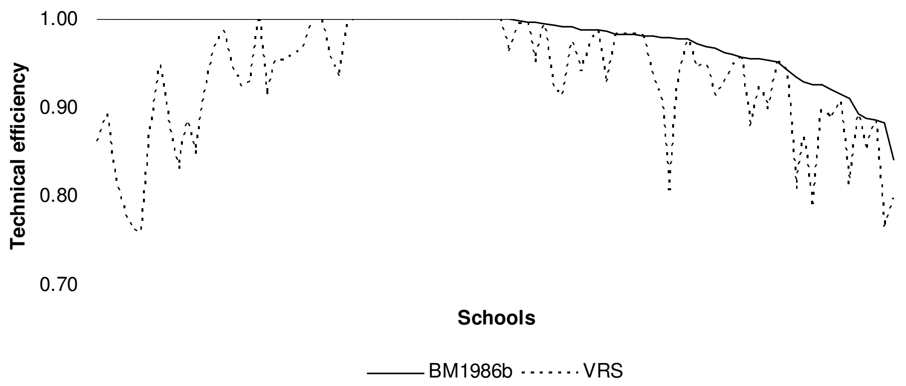


Figure 30
Efficiency scores provided by BM1986b and C1981 for each school – first graph

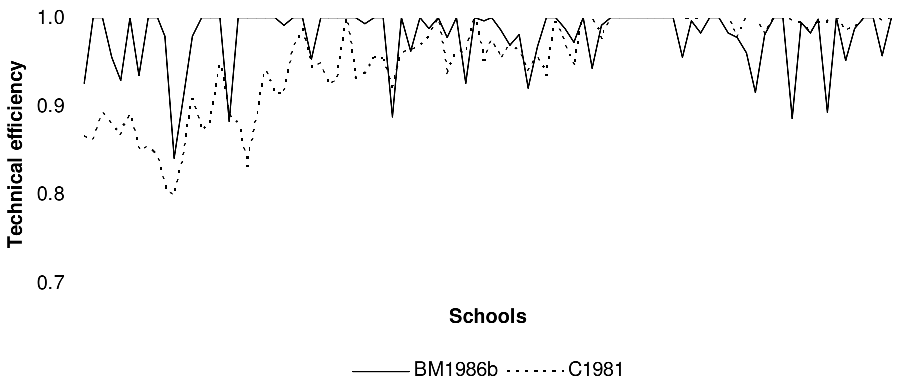


Figure 31
Efficiency scores provided by BM1986b and C1981 for each school – second graph

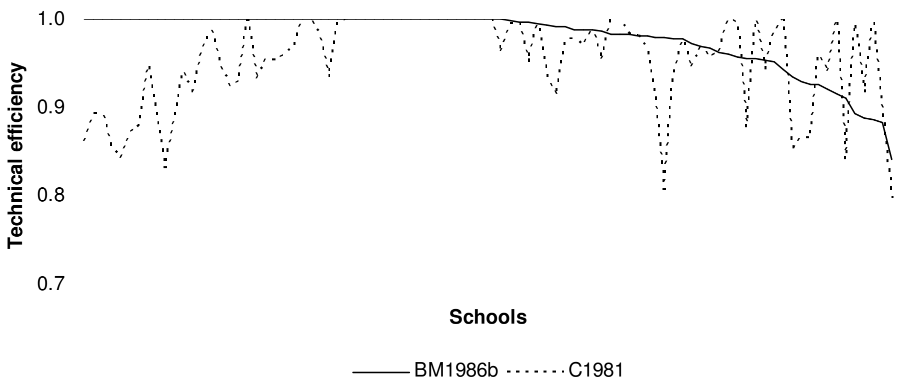


Figure 32
Efficiency scores provided by R1991 and YP2006-I&O for each school – first graph

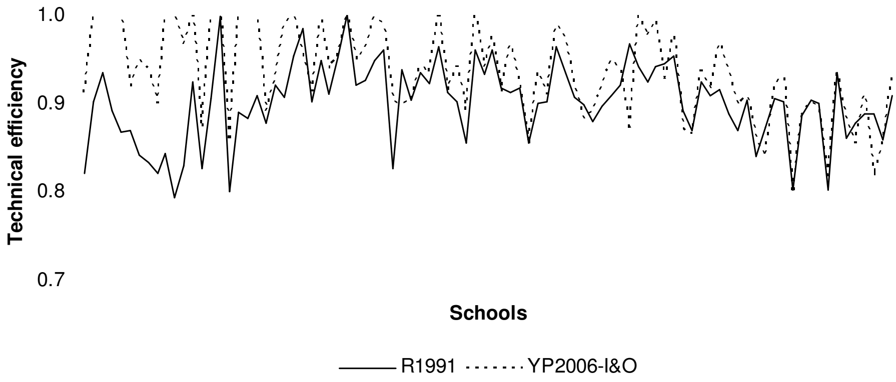


Figure 33
Efficiency scores provided by R1991 and YP2006-I&O for each school – second graph

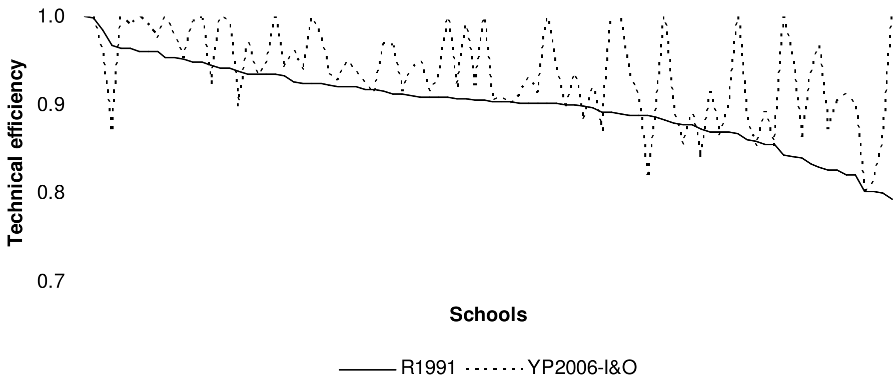


Figure 34
Efficiency scores provided by R1991 and YP2006-I for each school – first graph

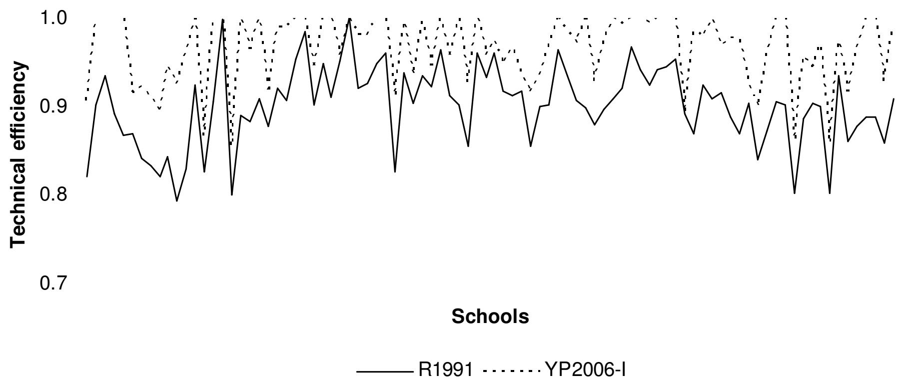


Figure 35
Efficiency scores provided by R1991 and YP2006-I for each school – second graph

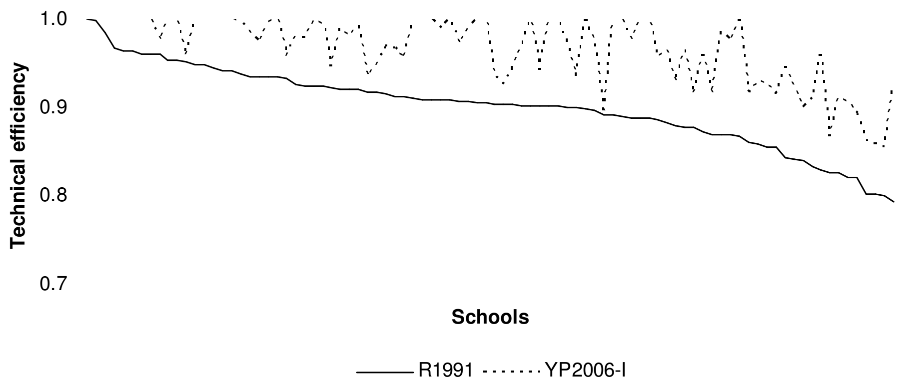


Figure 36
Efficiency scores provided by R1991 and H2014 for each school – first graph

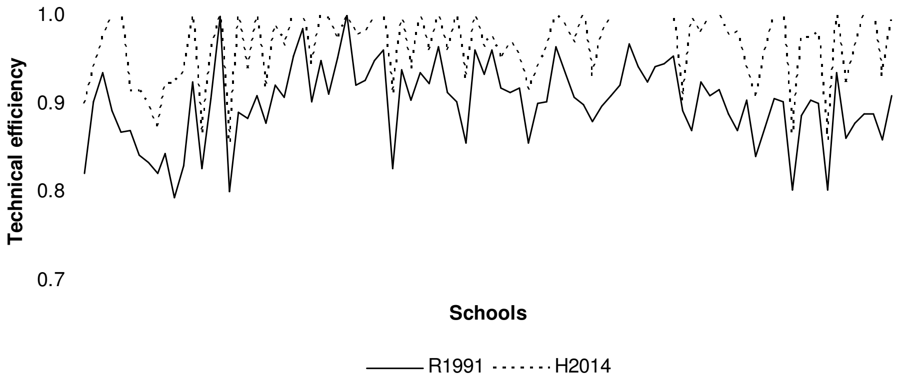


Figure 37
Efficiency scores provided by R1991 and H2014 for each school – second graph

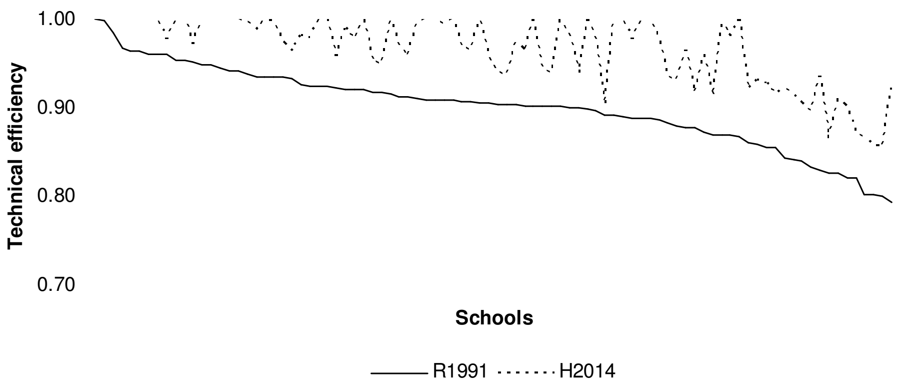


Figure 38
Efficiency scores provided by R1991 and VRS for each school – first graph

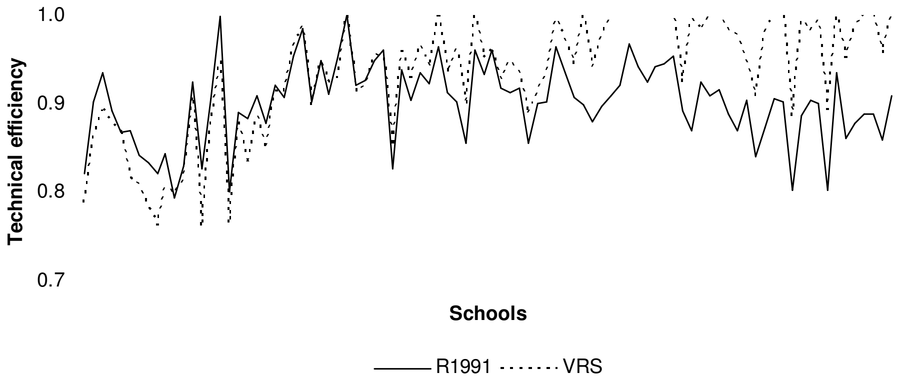


Figure 39
Efficiency scores provided by R1991 and VRS for each school – second graph

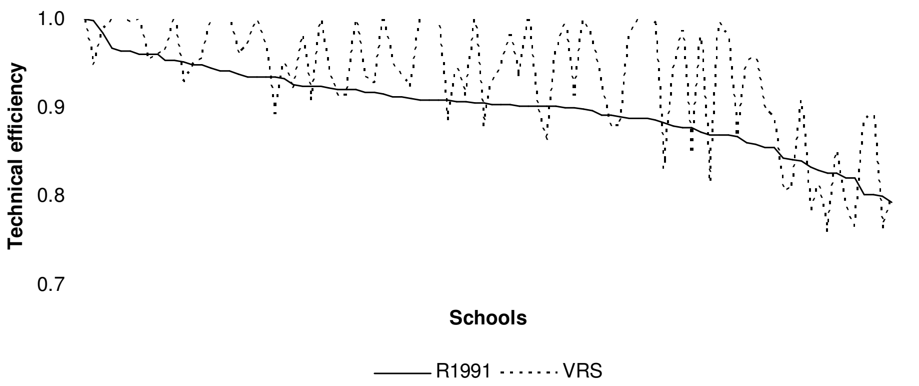


Figure 40
Efficiency scores provided by R1991 and C1981 for each school – first graph

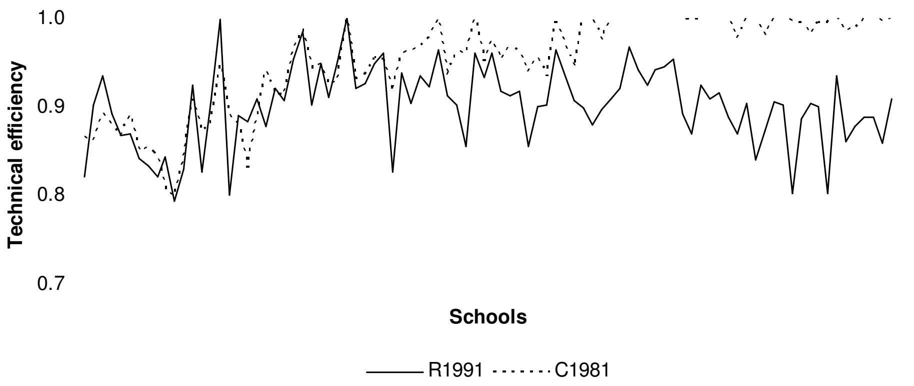


Figure 41
Efficiency scores provided by R1991 and C1981 for each school – second graph

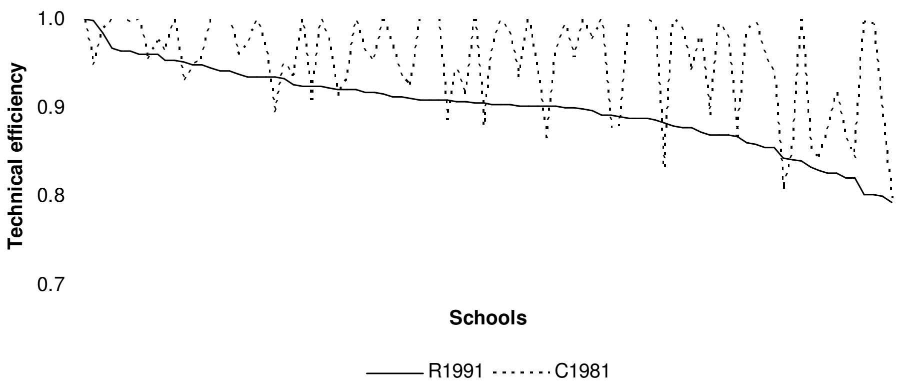


Figure 42
Efficiency scores provided by YP2006-I&O and YP2006-I for each school – first graph

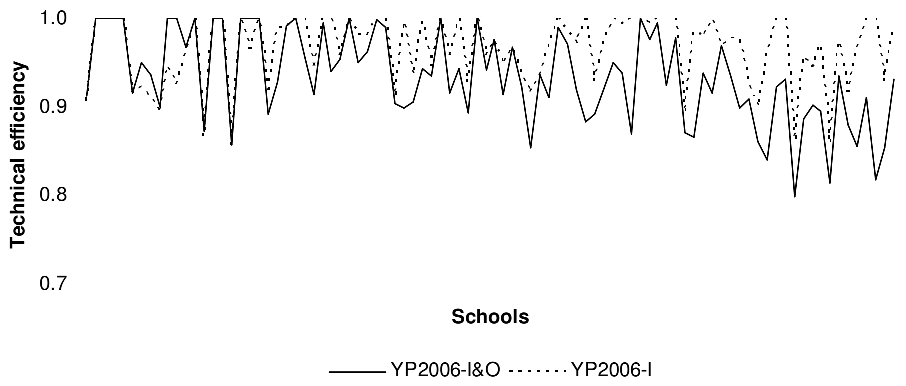


Figure 43
Efficiency scores provided by YP2006-I&O and YP2006-I for each school – second graph

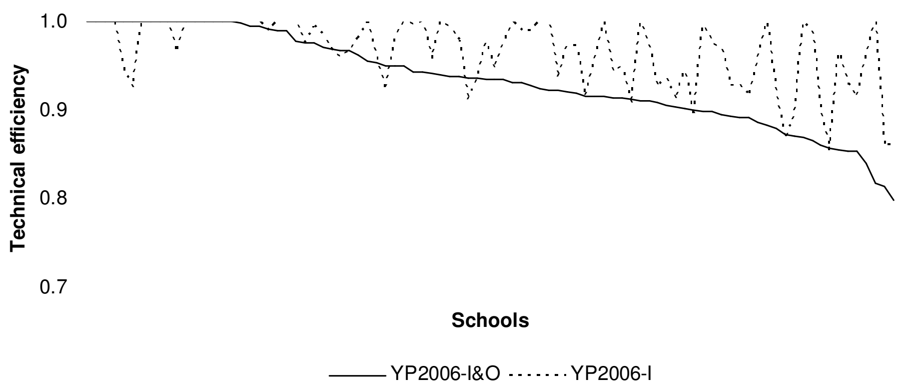


Figure 44
Efficiency scores provided by YP2006-I&O and H2014 for each school – first graph

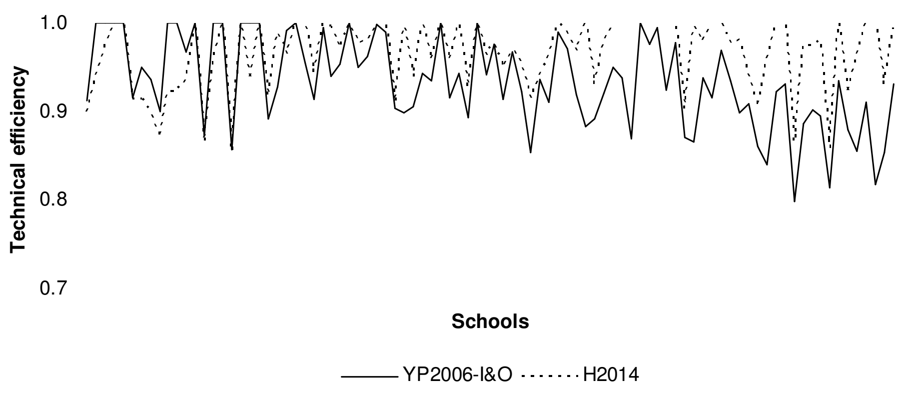


Figure 45
Efficiency scores provided by YP2006-I&O and H2014 for each school – second graph

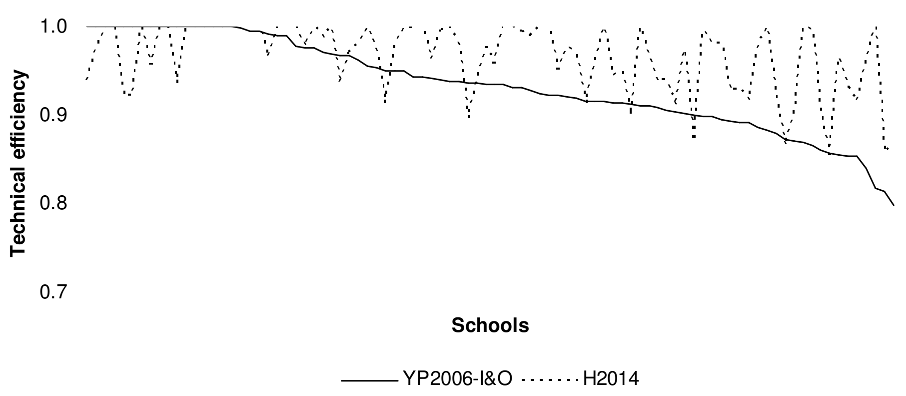


Figure 46
Efficiency scores provided by YP2006-I&O and VRS for each school – first graph

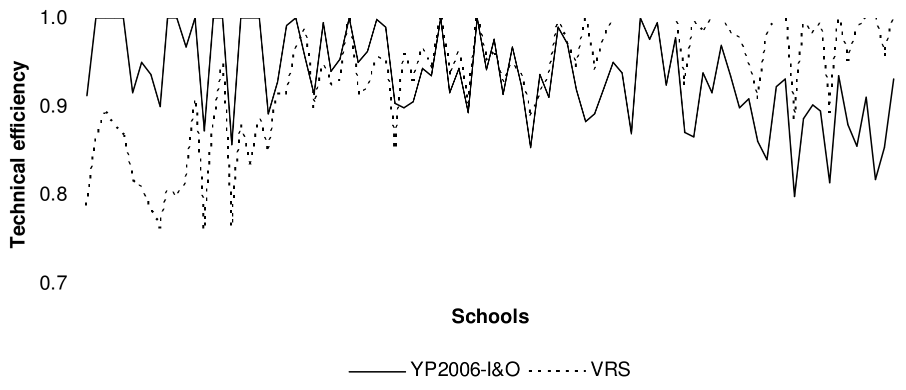


Figure 47
Efficiency scores provided by YP2006-I&O and VRS for each school – second graph

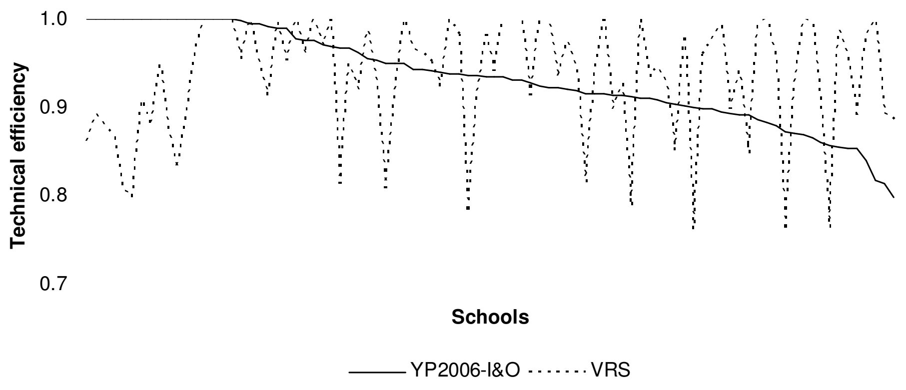


Figure 48
Efficiency scores provided by YP2006-I&O and C1981 for each school – first graph

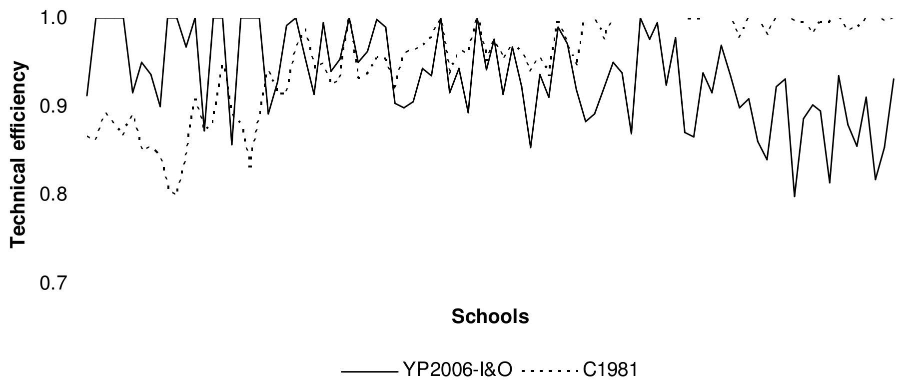


Figure 49
Efficiency scores provided by YP2006-I&O and C1981 for each school – second graph

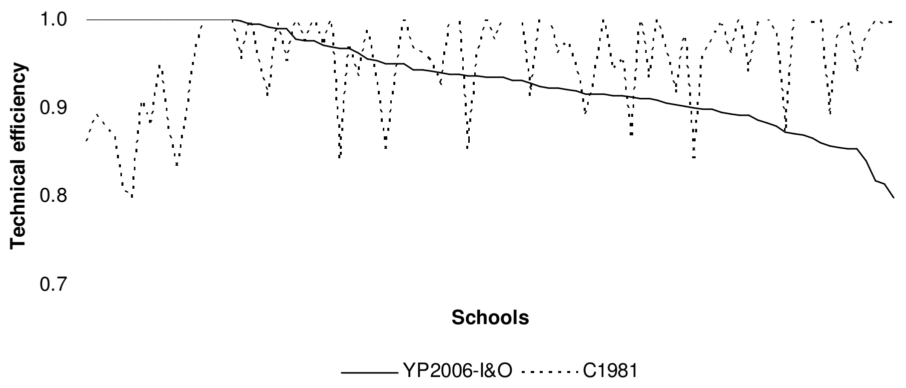


Figure 50
Efficiency scores provided by YP2006-I and H2014 for each school – first graph

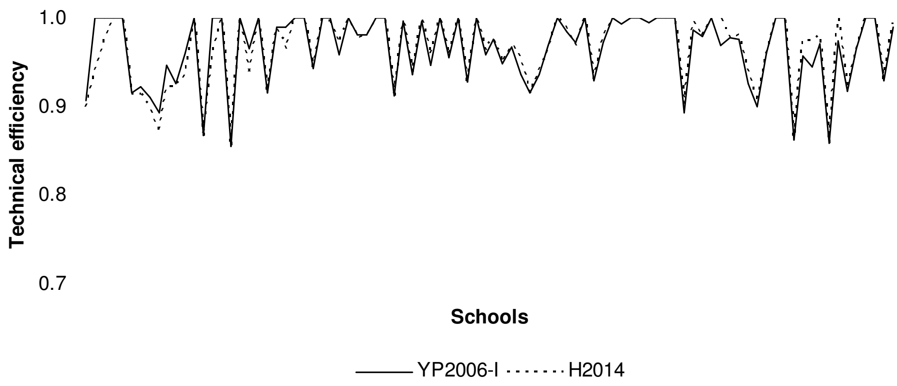


Figure 51
Efficiency scores provided by YP2006-I and H2014 for each school – second graph

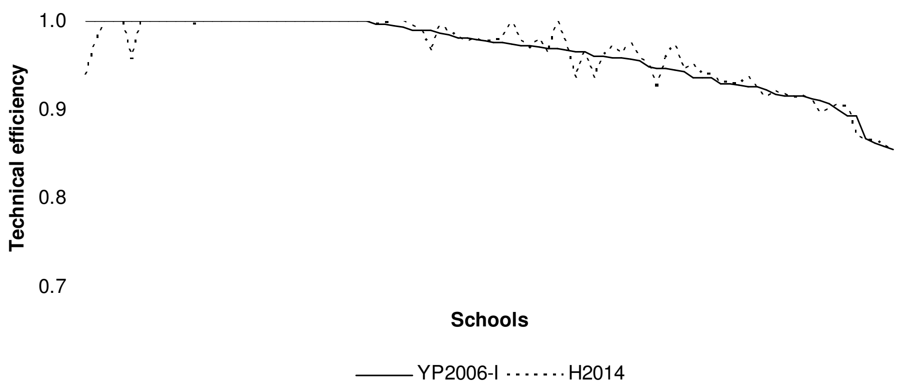


Figure 52
Efficiency scores provided by YP2006-I and VRS for each school – first graph

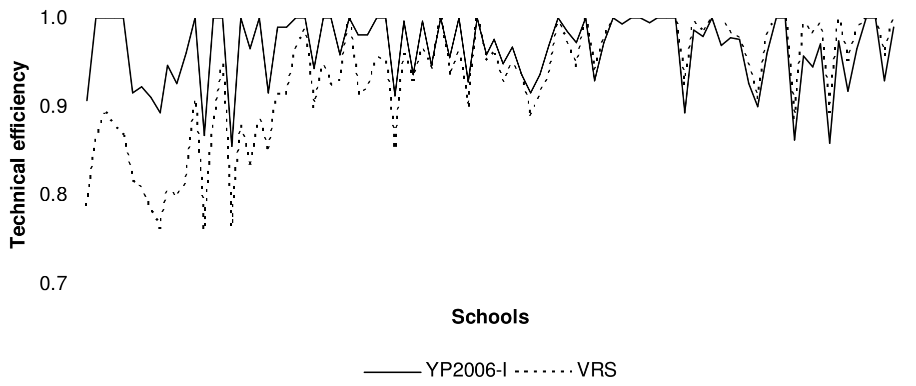


Figure 53
Efficiency scores provided by YP2006-I and VRS for each school – second graph

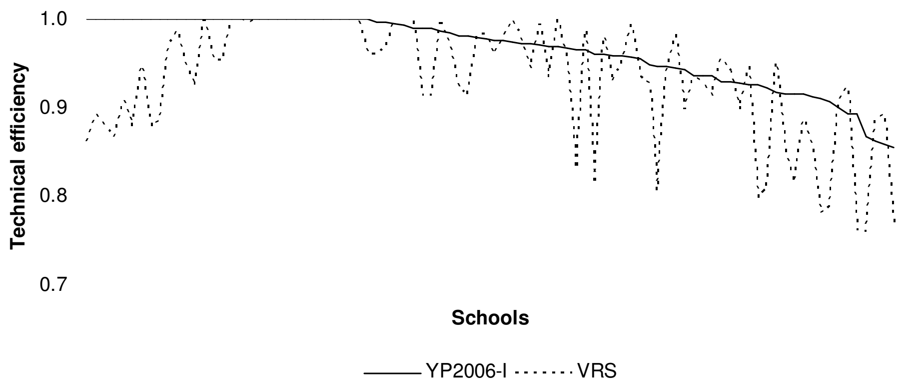


Figure 54
Efficiency scores provided by YP2006-I and C1981 for each school – first graph

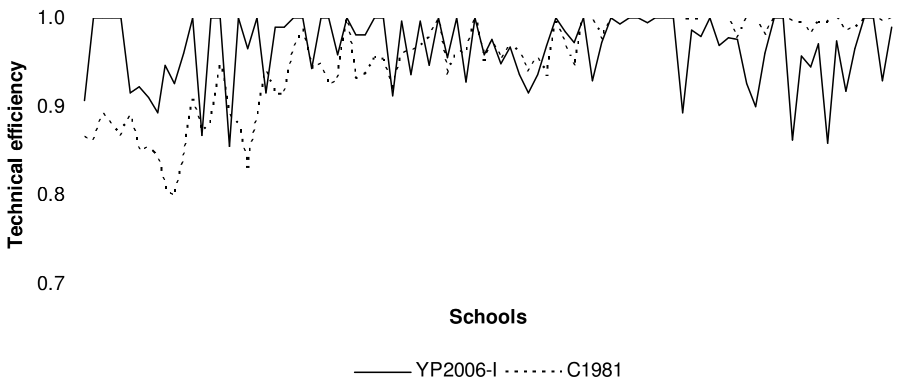


Figure 55
Efficiency scores provided by YP2006-I and C1981 for each school – second graph

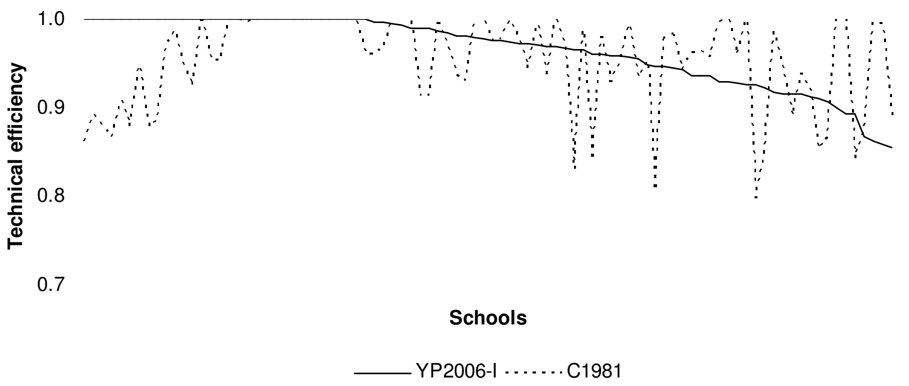


Figure 56
Efficiency scores provided by H2014 and VRS for each school – first graph

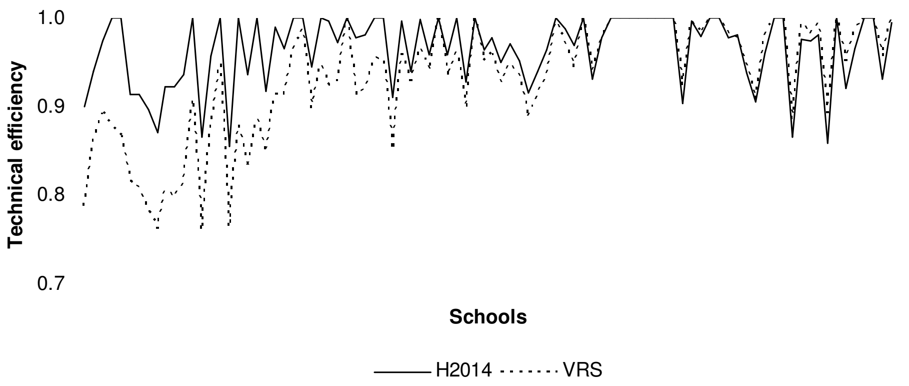


Figure 57
Efficiency scores provided by H2014 and VRS for each school – second graph

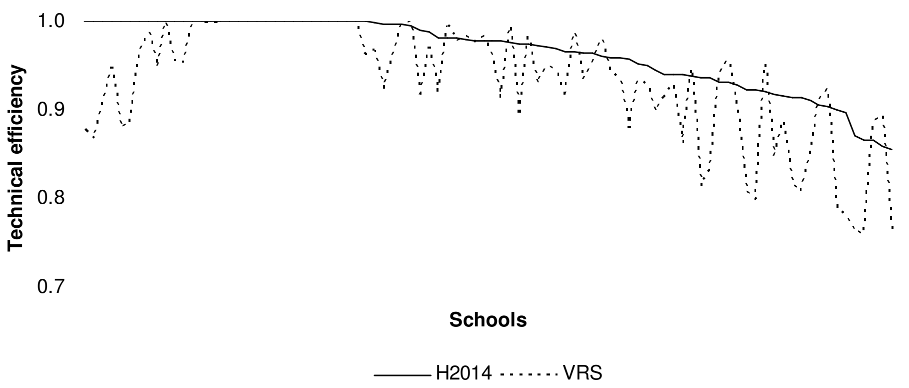


Figure 58
Efficiency scores provided by H2014 and C1981 for each school – first graph

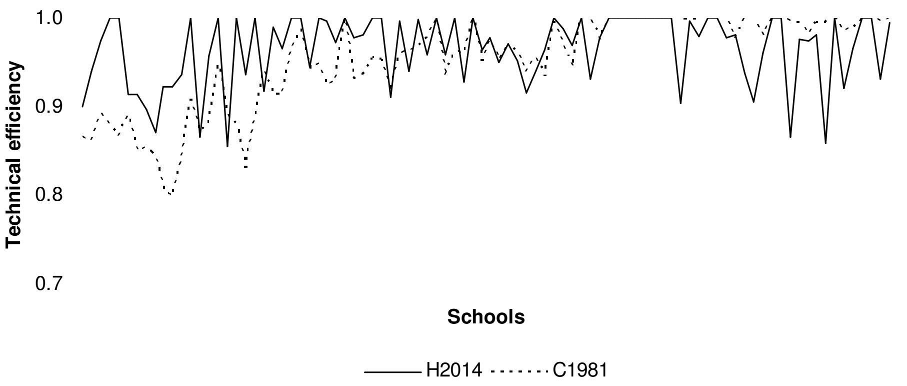


Figure 59
Efficiency scores provided by H2014 and C1981 for each school – second graph

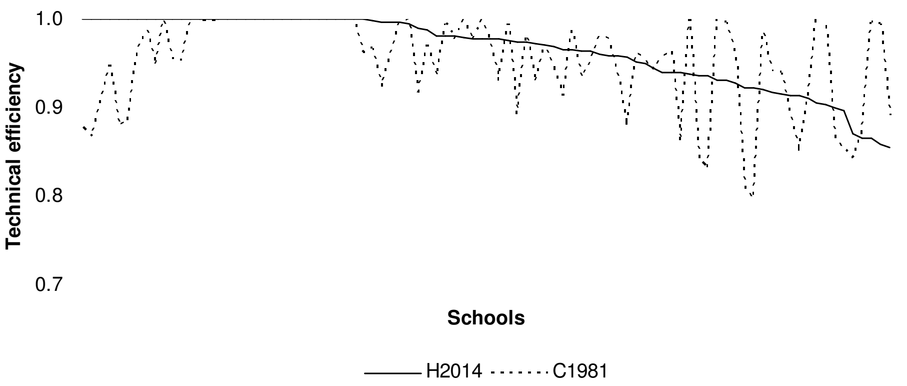


Figure 60
Efficiency scores provided by VRS and C1981 for each school – first graph

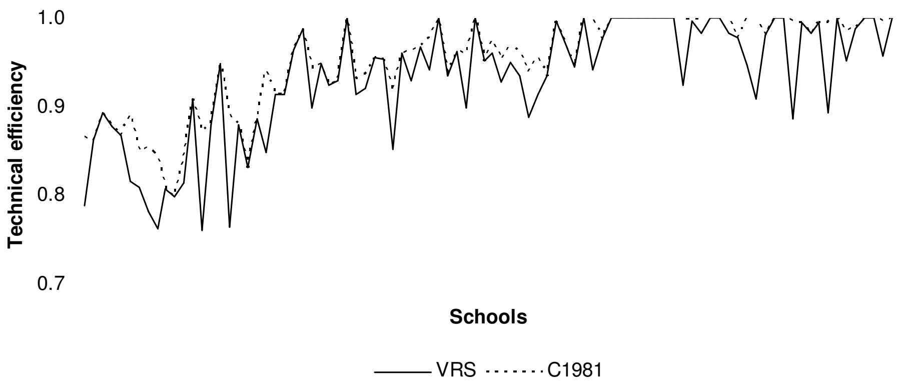
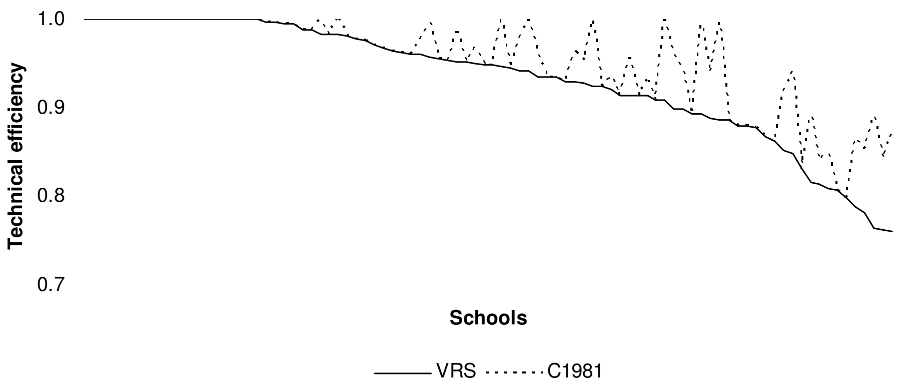


Figure 61
Efficiency scores provided by VRS and C1981 for each school – second graph



4

DEA and non-discretionary inputs: how to select the most suitable model (for you) using multi-criteria decision analysis

Structured abstract

Purpose

Performance measurement techniques include several methods, such as econometric or linear programming methods. They possibly deliver divergent results. Within the same method, such as Data Envelopment Analysis (DEA), different models also deliver divergent results. The aim of this study is to illustrate how a multi-criteria decision analysis method could be applied in order to help select the most suitable DEA model among alternative models.

Design/methodology/approach

First, a two-step web-based survey is conducted. In the first step, the survey aims to collect general views from DEA scholars and practitioners to identify the selection criteria. In the second step, the survey aims to prioritize and weight the selection criteria identified in the first step with respect to the goal of selecting the most suitable model. But it also aims to collect the preferences of the respondents about which model is preferable to fulfil each selection criterion. Second, Analytic Hierarchy Process, a multi-criteria decision analysis method, is used to quantify the preferences expressed in the survey.

Findings

Results show that the understandability, the applicability and the acceptability of the alternative models are valid selection criteria. When results are aggregated

over the respondents, the categorical model developed by Banker and Morey (1986a) emerges as the most suitable model. However, individual results may vary and other models may be identified as the most suitable ones from an individual perspective.

Practical implications

In terms of policy and managerial implications, the results of the current study suggest that:

- The number of selection criteria and alternatives (i.e. models) should remain parsimonious in order to avoid the time consuming process of AHP.
- The selection criteria should be backed by the literature or by an expert group. They should not be oriented towards the results of the models in order to avoid a biased model selection and potential opportunistic behaviour from decision makers.

Once the most suitable DEA model is identified, the following principles should prevail:

- The principle of permanence of methods: This principle states that the retained methods (in the current case, the DEA model retained) are not modified from one period of time to the other. It allows the coherence and comparison of the efficiency results produced.
- The principle of consistency: This principle requires the decision maker to be consistent from one period of time to another in applying the same DEA model.

Originality/value

Multi-criteria decision analysis methods have never been applied with the objective of selecting the most suitable efficiency measurement technique or, within a particular technique, the most suitable model. This study is performed to overcome this difficulty. By doing so, it fills a research gap.

Keywords: data envelopment analysis; alternative models; analytic hierarchy process.

Article Classification: research paper

JEL classification: C6; D24; D70

4.1 Context

The external environment could influence the ability of management to convert inputs into outputs and, as a result, impact entities' technical efficiency. Following Coelli, Prasada Rao, O'Donnel and Battese (2005, p. 190), an environmental variable is defined as a factor that could influence the efficiency of an entity, where such a factor is not a traditional input and is assumed to be outside of the manager's control. Because it is not under the control of managers, such a factor is also called a non-discretionary variable. It cannot be varied at the discretion of an individual manager but nevertheless needs to be taken into account to measure efficiency (Cooper, Seiford & Tone, 2007, p. 215).

Examples of environmental variables include ownership differences (such as public versus private), location characteristics, labor relations (such as conflicting versus peaceful relationships between trade unions and employers' organizations) and government regulations (Fried, Schmidt & Yaisawarng, 1999). In the education sector, for instance, three main generic drivers can be considered as environmental variables. They influence pupil performance but are outside of the control of headteachers (Soteriou, Karahanna, Papanastasiou & Diakourakis, 1998, p. 68, based on Thanassoulis, 1996, p. 883). They consist of *(1)* pupil characteristics, such as intelligence, willingness or effort propensity, *(2)* family and the external environment, such as the socioeconomic status of pupils and *(3)* school related factors (which are outside of the control of headteachers).

Data Envelopment Analysis (DEA) is a commonly used approach to the measurement of efficiency. Within DEA, several models allow for an environmental adjustment. Following Muñiz (2002), they can be grouped in three categories: *(1)* one-stage models (Banker & Morey, 1986a; Banker & Morey, 1986b; Ruggiero, 1996; Yang and Paradi in Muñiz, Ruggiero, Paradi and Yang, 2006), *(2)* multi-stage models including two-stage (Ray, 1988; Ray, 1991), three-stage (Ruggiero, 1998; Fried, Lovell, Schmidt & Yaisawarng, 2002; Muñiz, 2002) and four-stage models (Fried, Schmidt & Yaisawarng, 1999) and *(3)* program analysis models (Charnes, Cooper & Rhodes, 1981)[97]. See Huguenin (2014) for a presentation of these models.

There are few published studies which compare these models with one another. Some studies (Cordero, Pedraja & Santin, 2009; Estelle, Johnson & Ruggiero, 2010; Harrison, Rouse & Armstrong, 2012; Muñiz *et al.*, 2006; Ruggiero, 1996; Ruggiero, 1998; Ruggiero, 2004) use simulated data to compare alternative DEA models' results to the 'true' efficiency estimates performed by

97 Note that Yang and Pollitt (2009) propose the following categories: separative models (in which Charnes, Cooper & Rhodes (1981) and Banker & Morey (1986a) would be classified), one-stage models, two-stage models, three-stage models and four-stage models.

the simulation[98]. These studies provide mixed results about the convergence of alternative models with the 'true' efficiency.

Other studies (Cordero-Ferrara, Pedraja-Chaparro & Salinas-Jiménez, 2008; Huguenin, 2014; Muñiz, 2002; Yang and Pollitt, 2009) use empirical data in order to specifically benchmark alternative DEA models. In these studies, comparisons are made between the efficiency estimates of the alternative models. The best available evidence suggests that there is no consensus on the best model to use (Cordero-Ferrara *et al.*, 2008). It also suggests that the majority of models deliver diverging results (Huguenin, forthcoming)[99]. In other words, the efficiency scores generated by the models are significantly different. Consequently, recommendations and policy implications may differ according to the model used. From a political standpoint, these diverging results could potentially lead to ineffective decisions. From an applied research standpoint, they should represent a serious matter of concern. And from a decision making standpoint, they may lead to opposing managerial choices.

As no consensus emerges on the best model to use, practitioners and decision makers face the difficulty of selecting the model which is right *for them*, in other words, the model which best reflects their own preferences. Some authors, such as Wong and Li (2006), qualify this difficulty as the selection 'dilemma'. The choice of model thus becomes a strategic issue.

4.2 Objectives

The aim of this study is to illustrate how a multi-criteria decision analysis method could be applied in order to help select the most suitable DEA model among a choice of alternative models[100]. As far as the author is aware, this has never been done before.

To reach this objective, the following methodology has been developed:

- First, a two-step web-based survey is conducted. In the first step, the survey aims to collect general views from DEA scholars and practitioners, based on their judgement and experience, in order to identify the selection criteria. In the second step, the survey aims to prioritize and weight the selection criteria identified in the first step with respect to the goal of selecting the most suitable model. But it also aims to collect the preferences of the

98 'True' efficiency is determined by an artificial set of data as the production function, used to simulate data, is known.

99 Only two models seem to produce converging results according to Huguenin (forthcoming-a): Banker and Morey (1986a) and Banker and Morey (1986b). This finding is coherent with Harrison *et al.* (2012).

100 Note that the issue of this essay is not specific to DEA models but concerns all efficiency measurement techniques (Ordinary Least Squares, Stochastic Frontier Analysis, Free Disposal Hull, etc.). The issue of this essay has, therefore, a broader scope than just selecting between alternative DEA models.

respondents about which model is preferable to fulfil each selection criterion.

- Second, the selection of the most suitable model is then conducted using Analytic Hierarchy Process, a multi-criteria decision analysis method, with a limited number of individual cases for illustration purposes. Every individual case can potentially lead to the selection of a different model. Aggregated results can however be derived from individual results.

The current study thus focuses on the process of selecting the most suitable model, rather than on the result itself generated by this process. As a result, it does not aim to identify the most suitable model which is representative of a particular population, for instance the DEA community. The preferences expressed by the sample of respondents are, as a result, used for illustrative purpose.

4.3 DEA and AHP: a literature review

Data Envelopment Analysis

DEA is a performance measurement technique. It finds its origin in Charnes, Cooper and Rhodes (1978). See Huguenin (2013) for a synthetized presentation of DEA or Cooper, Seiford and Tone (2007) for a comprehensive treatment of the methodology. Charnes *et al.* (1978) develop a first basic model under the assumption of constant returns to scale. Banker, Charnes and Cooper (1984) develop a second basic model under the assumption of variable returns to scale (see Section 4.4 for the model specification). The DEA approach is:

- Non-statistical, as it uses linear programming;
- Non-parametric, as no function specification of the production frontier has to be formulated;
- Deterministic, as it considers that the differences between the observed outputs and the outputs specified by the production frontier correspond exclusively to inefficiency.

The use of DEA is experiencing rapid and continuous growth. In 2002, Tavares (2002) identified 3203 DEA publications (journal articles, research articles, event articles, books and dissertations). In 2008, Emrouznejad, Parker and Tavares (2008) inventoried more than 7000 publications.

DEA has been applied to various areas, both in the private and in the public sectors, such as banking (Nguyen, Roca & Sharma, 2014; Holod & Lewis, 2011), insurance companies (Eling & Huang, 2010; Borges, Nektarios & Barros, 2008), retailing stores (Vaz & Camanho, 2012; Malhotra, Malhotra & Lafond, 2010), hotels (Manasakis, Apostolakis & Datseris, 2013; Barros & Santos, 2006), airlines (Lee & Wothington, 2014; Fethi, Jackson & Weyman-Jones, 2000), airports (Suzuki, Nijkamp, Pels & Rietveld, 2014; Adler & Berechman, 2001), ports (Lee, Yeo & Thai, 2014; Tongzon, 2001), cement industry (Oggioni, Riccardi & Toninelli, 2011; Sharma, 2008), petroleum industry (Sueyoshi & Goto, 2012; Al-Najjar & Al-Jaybajy, 2012), power plants (Liu, Lin & Lewis, 2010; Azizi, Lofti, Saati & Vahidi, 2007; Park & Lesourd, 2000), employment offices (Andersson, Manson & Sund, 2014; Sheldon, 2003), health care (Thanassoulis, Portela Silva & Graveney, 2014; Gautam, Hicks, Johnson & Mishra, 2013), transportation (Mallikarjun, Lewis & Sexton, 2014; Caulfield, Bailey & Mullarkey, 2013), farming (Kelly, Shalloo, Geary, Kinsella & Wallace, 2012; Picazo-Tadeo, Gómez-Límon & Reig-Martínez, 2011), education (Huguenin, forthcoming; Harrison & Rouse, 2014), police forces (Verma & Gavirneni, 2006; Sun, 2002), fire stations (Friebelová, Friebel & Marková, 2009; Lan, Chuang & Chen, 2009), prisons (Marques & Simões, 2009; Butler & Johnson, 1997), waste collection (Rogge & De Jaeger, 2013; Ichinose, Yamamoto & Yochida, 2013) regions (Rabar, 2013; De Witte & Moesen, 2010) and many others.

Unlike statistical approaches, DEA can accommodate both multiple inputs and multiple outputs. It is a strength in the context of the public sector where multiple non-monetary outputs are generally provided. However, the specification (i.e. the choice and/or the quantity) of inputs and outputs to be included in the analysis impacts efficiency results. Moreover, the number of inputs and outputs which can be included in the model depends on the entities' sample size (Cooper, 2006). DEA is sensitive to small sample size, as it lessens its discriminating power. This is probably its main drawback. Unlike statistical parametric approaches, DEA does not need the production function and the distribution of inefficiencies to be specified. As a result, it avoids the potential problems of mis-specification. But it also means that "there are no familiar parametric tests with which to check the validity of the model" (Johnes, 2004, p. 643). Given the strengths and weaknesses of DEA (but also probably given its ease of use), most studies in various fields have opted for DEA as their methodological approach, as in the education sector (Agasisti, Bonomi & Sibiano, 2014).

Analytic Hierarchy Process

AHP is a Multi-Criteria Decision Analysis (MCDA) method. MCDA methods have been developed to help the decision maker in the personal decision process. These methods take into account the preferences of the decision maker, which is subjective information. As an ideal solution suiting all the criteria usually does not exist, MCDA methods identify a compromise solution.

Roy (1981) defines four main types of problems which require decision making: choice, sorting, ranking and description. The current essay aims to select the most suitable DEA model (i.e. the single best model). This is an example of a choice problem. Several MCDA methods are appropriate for a choice problem, such as AHP, Analytic Network Process, Multi-Attribute Utility Theory, MACBETH, PROMETHEE, ELECTRE or TOPSIS. The choice of the appropriate method is difficult as "there has been no possibility of deciding whether one method makes more sense than another in a specific problem situation" (Roy & Bouyssou, 1993, cited by Hishizaka & Nemery, 2013, p. 6).

AHP and MACBETH (Bana e Costa & Vansnick, 1999; Bana e Costa, De Corte & Vansnick, 2003, 2005) are two methods which:

- Allow a compensable score, meaning that a bad score for one criterion is compensated by a good score on another;
- Adopt a pairwise comparison process; this process allows comparing pairs of criteria with respect to the goal of the decision making process (for instance, to select the right model) and pairs of alternatives (for instances, models) with respect to every single criterion[101].

AHP and MACBETH share many similarities. However, AHP compares the pairs of criteria or alternatives on a ratio scale whereas MACBETH compares them on an interval scale. The drawback of an interval scale is that results are modified when a change on the scale is adopted, for instance when the measurement of a distance is in meters or in kilometres. A ratio scale avoids this drawback. In the current study, AHP, a method developed by Saaty (1977, 1980), is thus retained as the method of choice.

As pointed out by Vaidya and Kumar (2006, p. 1), AHP "(…) is one of the most widely used multiple criteria decision-making tools". The same authors show that AHP is predominantly used in the areas of selection and evaluation[102]. Ishizaka and Labib (2011), Ho (2008) and Vaidya and Kumar (2006) provide reviews of AHP applications.

In the area of selection, AHP has been applied to various fields such as software packages (Lai, Wong & Cheung, 2002), contractors (Al-Harbi, 2001), site locations (Korpela & Tuominen, 1996), delivery methods (Al Khalil, 2002), vendors (Tam & Tummala, 2001), suppliers (Tahriri, Osman, Ali, Yusuff & Esfandiary, 2008), manufacturing systems (Bayazit, 2005), drugs (Vidal, Sahin, Martelli, Berhoune & Bonan, 2010), staff (Celik, Kandakoglu & Er, 2009) and many others.

In the area of evaluation, AHP has been applied to various fields such as scientific journals (Forgionne, Kohli & Jennings, 2002), Enterprise Resource Planning systems (Al-Rawashdeh, Al'azzeh & Al-Qatawneh, 2014), banking data (Yin, Pu, Liu & Zhou, 2014), urban parks (Wang & Zhang, 2014), teaching (Yin, 2013), websites (Lin, 2010), weapons (Cheng, Yang & Hwang, 1999), environmental impact of industrial alternatives (Sólnes, 2003), universities (Lee, 2010), hospitals (Chen, 2006), bank mergers and acquisitions (Arbel & Orgler, 1990) and many others.

Few AHP papers focus on method selection (as in this current research). Applications have been found in the areas of selecting the most suitable underground mining method (Gupta & Kumar, 2012) or the most suitable bridge construction method (Pan, 2008). However, as far as the author is aware, AHP has never been used in order to select a performance measurement

---

101 For a taxonomy of MCDA methods, see Ishizaka and Nemery (2013, pp. 1–9).

102 According to Forman and Gass (1991), AHP can also be applied to the areas of ranking, prioritization, resource allocation, benchmarking and conflict resolution. See Saaty (2012) for numerous examples of applications.

technique. Neither has it been used in order to select, among a particular performance technique, the most suitable model.

The main advantage of AHP is its ability to rank alternatives according to their effectiveness in meeting potential conflicting criteria. The approach is flexible as it is tailored to reflect the individual preferences of decision makers, but also to reflect the consensual preferences of a group of decision makers (Ramanathan, 2001). The fact that AHP is able to deconstruct a complex decision problem to its simpler constituent parts is also appreciated (Macharis, Springael, De Brucker & Verbeke, 2004).

However, AHP has several shortcomings (Lin, Lee and Ho, 2011). Among those, the use of a linear scale has been debated and other types of scales have been proposed, such as the geometric scale (Lootsma, 1989), the inverse linear scale (Ma & Zheng, 1991) or the balanced scale (Salo & Hämäläinen, 1997). Saaty (1980, 1991a) maintains that the linear scale is the most appropriate one. From a practical point of view, the linear scale is also the only one implemented in the two leading AHP software packages, *Expert Choice* ® and *MakeItRational* ®. The fact that the linear scale is limited by a 1-9 point scale is also criticized, as decision makers could find it difficult to discriminate between two points on the scale, such as between point 7 and point 8 for example. Some authors have proposed an alternative grading scale. For instance, Hajkowicz, Young, Wheeler, MacDonald and Young (2000) propose a simple 2-point scale. With such a scale, the decision makers only indicate if an alternative is equally, more, or less important than another. Another drawback of AHP is that the number of pairwise comparisons to be conducted may become very large, and thus become time consuming (Macharis *et al.*, 2004). Finally, many researchers have observed that ranking irregularities can occur when a copy (Belton & Gear, 1983) or a near-copy (Dyer, 1990) of an alternative is added or removed from the original problem to solve[103]. This phenomenon is called the rank reversal of the alternatives and has initiated a fierce debate among the AHP community. This debate is illustrated by the exchange of views of Holder (1990, 1991) and Saaty (1991a, 1991b). Several authors have proposed approaches to avoid the rank reversal problem, such as Wang and Elhag (2006) and Millet and Saaty (2000).

Combining DEA and AHP

AHP is often used in combination with other techniques (Ho, 2008). The interaction between AHP and DEA is quite recent[104]. In 2008, Ho (2008) identified in a partial literature review only four journal articles combining

---

103 Note that the phenomenon of rank reversal is not limited to AHP but concerns other MCDA methods (Wang & Luo, 2009).

104 Index numbers are an alternative to DEA for the measurement of changes in total factor productivity. Several formulae for price and quantity index numbers have been developed (see Coelli *et al.*, 2005, for a review). In general, note that AHP has also been used in interaction with such index numbers. For instance, Frei and Harker (1999) use AHP in order to aggregate various measures of productivity.

these two techniques (Takamura & Tone, 2003; Yang & Kuo, 2003; Saen, Memariani & Lofti, 2005; Ertay, Ruan & Tuzkaya, 2006). The combination of the two methods has been growing rapidly since the review realized by Ho (2008). Among the recent applications, Pakkar (2012) develops, for instance, an integrated approach to the DEA and AHP methodologies which defines a domain of efficiency based on two sets of weights.

Several ways of combining DEA and AHP have been identified in the literature. They are synthetized hereafter:

- AHP is used to convert qualitative data (i.e. input and/or output) into quantitative data. Quantitative data are then used in DEA models (Kong & Fu, 2012; Pakkar, 2012; Lin, Lee & Ho, 2011; Azadeh, Ghaderi & Izadbakhsh, 2008; Korpela, Lehmusvaara & Nisonen, 2007; Ertay, Ruan & Tuzkaya, 2006; Feng, Lu & Bi, 2004; Yang & Kuo, 2003; Shang & Sueyoshi, 1995). This approach has the advantage of achieving the quantification of qualitative data. But, as pointed out by Pakkar (2012), AHP is a subjective data-oriented procedure while DEA is an objective data-oriented approach. As a result, subjective data are introduced alongside objective data in DEA models. Results may vary greatly depending on the values of the subjective data, which themselves depend on individual preferences of decision makers who convert qualitative data into (subjective) quantitative data.
- AHP is used to aggregate (and thus to reduce) the number of inputs or outputs (Cai and Wu, 2001; Korhonen, Tainio & Wallenius, 2001). This approach is used when inputs and outputs can be grouped into common categories. It has the advantage of allowing the use of DEA when, in an initial situation, there are too many variables according to the number of entities to assess. However, this approach has the drawback to lose information about 'sub-inputs' which are aggregated into categories. The authors fail to consider alternative methods within the DEA framework, such as the window analysis, to deal with such situations.
- AHP is used to restrict the input and output weights to be used in DEA models (Takamura & Tone, 2003; Seifert & Zhu, 1998). This approach allows the decision makers to express their preferences about the weight of the variables. Again, these preferences are subjective and influence DEA results. From a general point of view, the issue of placing weight restrictions onto inputs and outputs is not obvious and should be justified. See for instance Stewart (1996) for a discussion.
- AHP is used to estimate the missing data for slightly non-homogeneous entities in order to make it possible to perform DEA models (Saen, Memariani & Lofti, 2005). This approach has the advantage to allow the use of DEA which would be useless otherwise because of missing data. Note that the algorithm developed by the authors is based on the AHP technique but does not need the expression of subjective preferences, as it considers the mean values of inputs and outputs.
- AHP is used as an alternative method (i.e. instead of DEA) to rank efficient units (Jablonsky, 2007). This approach is an alternative of the super-

efficiency models developed within the framework of DEA which also allow ranking the efficient units (Zhu, 2003). However, the super-efficiency models could lead to infeasible solutions. This pitfall is avoided with the use of AHP. It is not clear how convergent the rankings produced by super-efficiency models and the AHP approach are. Moreover, AHP ranking may vary greatly according to the subjective preferences expressed by the decision makers.

- AHP is used to identify relevant inputs and outputs to be included in DEA models (Yoo, 2003). This approach allows the decision makers to prioritize the inputs and the outputs to be included in the DEA analysis. By doing so, they (voluntarily) restrict the number of variables. It is not clear in Yoo (2003) why they should impose such a restriction. The author also fails to provide a comparison with a DEA model which would include all the variables available.
- AHP is used to weight the amount of change in initial inputs and outputs of entities in target setting (Lozano & Villa, 2009). Decision makers are asked which inputs and outputs they wish to improve, allow to worsen or want to keep constant. This approach is an interactive one. Decision makers are able to visualize the projected points of the entities by varying the inputs and the outputs. The main drawback of this approach is that the decision makers are able to 'influence' the results as they wish.
- DEA is used as an alternative method (i.e. instead of AHP) in order to calculate the local weights of alternatives with respect to the selection criteria. The global weights of alternatives are then calculated with AHP (Nachiappan & Ramanathan, 2008; Sevkli, Lenny Koh, Zaim, Demirbag & Tatoglu, 2007; Ramanathan, 2006). This approach, known as DEAHP, is used when the decision makers are not able to decide whether a criterion or an alternative is better than another. It is more objective, but, as a result, does not take into account the preferences of the decision makers. Some authors have also shown that DEAHP may produce irrational results in certain situations (Wang, Parkan & Luo, 2007; Wang & Chin, 2009).
- DEA is performed and the pairwise comparison matrix is identified. This matrix is used in a single hierarchical level AHP model in order to generate an alternative ranking of entities (Guo, Liu & Qiu, 2006; Sinuany-Stern, Mehrez & Hadad, 2000). The added value of this approach is not evident as the ranking produced is not necessarily compatible with the ranking generated by DEA alone.

4.4 Methodology

The current section introduces the models' specifications of DEA and AHP. It also presents the web-based survey addressed to DEA scholars and practitioners in order *(1)* to identify the selection criteria and *(2)* to collect the preferences of the members of the DEA community. Finally, the current section describes the models retained as alternatives.

Data Envelopment Analysis

The objective of the current study is to investigate how AHP can be applied in order to select the most suitable DEA model. As a result, no DEA model is empirically applied. However, the specification of the model developed by Banker *et al.* (1984) is mentioned hereafter.

Following the notation adopted by Johnes (2004, pp. 630-637), there are data on s outputs and m inputs for each of n entities to be evaluated. y_{rk} is the quantity of output r produced by entity k. x_{ik} is the quantity of input i consumed by entity k. θ_k represents the VRS efficiency of entity k (i.e. 'pure' technical efficiency free from any scale inefficiency). λ_j represents the associated weighting of outputs and inputs of entity j.

The VRS efficiency of the k^th^ entity is calculated by solving the following linear problem:

$$\text{Minimize} \quad \theta_k \tag{1}$$

$$\text{Subject to} \quad y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \qquad r = 1, \dots, s$$

$$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} \geq 0 \qquad i = 1, \dots, m$$

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

$$\lambda_j \geq 0 \quad \forall j = 1, \dots, n$$

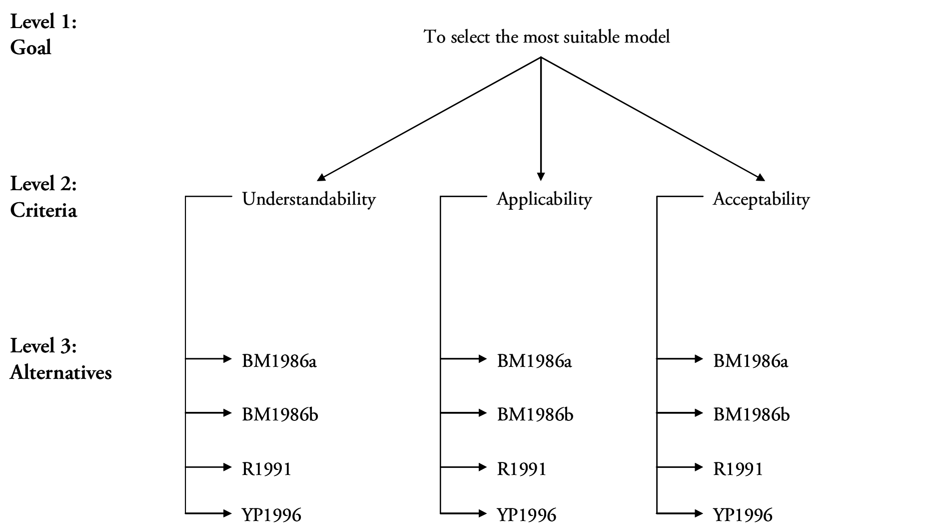
Analytic Hierarchy Process

In AHP, the problem is structured as a hierarchy including three levels (see Figure 62 for an illustration). The first level is the goal of the complex problem to be solved, the second one represents the criteria and the third one represents the alternatives. In the context of the current study:

- The goal of the complex problem consists of selecting the most suitable model.
- The selection criteria retained at the end of the web-based survey process are the understandability, the applicability and the acceptability of the alternative models. These selection criteria have been first identified on the basis of a literature review and second confirmed by a web-based survey addressed to Data Envelopment Analysis scholars and practitioners (see sub-section 'Web-based survey' later in this Section about the methodology and Section 4.5 about the results of the survey).

- The alternatives are the DEA models. Four alternative models are considered[105]. These models are the Banker and Morey (1986a) model (BM1986a), the Banker and Morey (1986b) model (BM1986b), the two-stage Ray (1991) model (R1991) and the Yang and Paradi model (in Muñiz, Ruggiero, Paradi and Yang (2006, p. 1176) (YP2006). Note that other models could have been considered. The choice of these four models is discussed in the sub-section 'Models retained as alternatives' later in this section.

Figure 62
To select the most suitable model among four alternatives according to three criteria



At each level of the hierarchy (with the exception of the top level), the pairwise comparisons of the decision maker are collected in a matrix. At the second level, the decision maker has to judge which criterion is most important for selecting the model. For instance, he compares the understandability versus the applicability criteria – as the first pairwise of this level – with respect to the goal of selecting the most suitable model. At the third level, the decision maker has to judge which model is preferable to fulfil the given criteria. For instance, he compares the BM1986a model versus the BM1986b model – as the first pairwise of this level – with respect to the understandability criterion. The process stands for all levels of the hierarchy and for all pairwise comparisons on criteria and alternatives (models).

105 A number of four alternatives implies six pairwise comparisons per criterion. The number of pairwise comparisons to be conducted is obtained by the following formula: $(n^2 - n) / 2$, where n is the number of alternatives.

The scale for comparing two alternatives (or criteria) is defined by Saaty (1977) (see Table 28). It is a linear scale with the integers 1 to 9[106]. As Saaty (2012, p. 73) points out, “experience has confirmed that a scale of nine units is reasonable and reflects the degree to which we can discriminate the intensity of relationships between elements”. For instance, a score of 3 means that the first criterion – for example understandability – is three times as important as the second criterion – for example applicability – with respect to the most suitable model. It also means that it is moderately more important. A score of 1 means that the two criteria are of equal importance.

Table 28
The AHP 1-9 scale

	Degree of importance	
Numerical scale	Verbal scale	Explanation
1	Equal importance	Two criteria contribute equally to the goal Two alternatives are equal with respect to one criterion
2	Weak or slight	(For compromise between equal importance -1- and moderate importance -3-)
3	Moderate importance	Experience and judgement slightly favor one criterion/alternative over another
4	Moderate plus	(For compromise between moderate importance -3- and strong importance -5-)
5	Strong importance	Experience and judgement strongly favor one criterion/alternative over another
6	Strong plus	(For compromise between strong importance -5- and very strong or demonstrated importance -7-)
7	Very strong or demonstrated importance	A criterion/alternative is favored very strongly over another; its dominance is demonstrated in practice
8	Very, very strong	(For compromise between very strong or demonstrated -7- and extreme importance -9-)
9	Extreme importance	The evidence favoring one criterion/alternative over another is of the highest possible order of affirmation

Source: adapted from Saaty (2008)

AHP calculates three types of scores – called priorities –:

- The criteria priorities are scores (or weights) that reflect the importance of each criterion with respect to the goal. These priorities are calculated with pairwise comparisons collected in a matrix.
- The local alternative priorities are scores that reflect the importance of an alternative with respect to one specific criterion. These priorities are calculated with pairwise comparisons collected in a matrix.
- The criteria and the local alternative priorities are then used to calculate the global alternative priorities. These priorities reflect the importance of alternatives across all criteria.

The pairwise matrix A is filled with the pairwise comparisons a_{ij}, where a_{ij} is the comparison of alternative (or criterion) i with alternative (or criterion) j.

[106] Note that the linear scale with the integers 1 to 9 is based on psychological observations (Stevens, 1957).

a_{ij} is defined as the ratio $\frac{p_i}{p_j}$, where p_i is the priority of the alternative (or criterion) i and p_j is the priority of the alternative (or criterion) j.

The matrix A is considered as consistent if it respects the transitivity and the reciprocity rules. The transitivity rule is respected when $a_{ij} = a_{ik} \times a_{kj}$, where i, j, and k are alternatives (or criteria) of the matrix. The reciprocity rule is respected when $a_{ij} = \frac{1}{a_{ji}}$.

The matrix A is used to calculate the criteria priorities and the local alternative priorities. Several methods have been developed to do so. AHP usually uses the eigenvalue method. In this method, the vector of the priorities p is calculated by solving the equation

$$Ap = np \qquad (2)$$

where n is the dimension of the matrix A and $p = p_1,\ldots,p_n$ are the priorities.

In AHP, the consistency of the matrix has to be checked, as priorities make sense only if calculated from consistent or near consistent matrices (Ishizaka & Labib, 2009). Saaty (1977) uses a consistency ratio (CR). If the ratio is inferior to 0.1, meaning that the inconsistency is inferior to 10% of 500 randomly filled matrices, the matrix is considered to be of an acceptable consistency. If the ratio is superior to 0.1, the values of the matrix have to be adjusted in order to make it consistent. As a result, the decision makers have to revise their preferences.

Finally, the global priorities for each alternative have to be calculated from the local alternative priorities across all criteria. Two aggregation approaches are possible: the distributive mode and the ideal mode. In an open system, where alternatives can be added or removed and the preferences are allowed for alternatives to be dependent on other alternatives, the distributive mode is indicated (Millet & Saaty, 2000). This is the case in the current analysis, as one may want to add or to remove DEA models as alternatives. The distributive mode, also called the additive aggregation, is defined by

$$P_i = \sum_j w_j p_{ij} \qquad (3)$$

where P_i is the global alternative priority of alternative i, w_j is the weight of criterion j and p_{ij} is the local alternative priority i with regard to criterion j.

AHP has been adapted in order to be applied to group decisions. As the survey used in this analysis links a collection of individuals in different location, it is difficult, not to say impossible, to reach a consensus among the decision makers. As a result, the geometrical means of the individual evaluations is used to fit the matrix (Saaty & Vargas, 2005).

Web-based survey

Following the approach adopted by Wong and Li (2008) – applied to intelligent building experts and practitioners –, a two-step web-based survey was conducted among the DEA community in April and May 2014. The parent population is composed of 164 DEA scholars and practioners who participated in the 9th international conference on DEA[107]. This conference took place in August 2011 at the University of Macedonia, Thessaloniki, Greece. The parent population constitutes a sample of convenience because of its accessibility and proximity to the current research (and to its author, who also participated to the above mentioned conference), but also because its members have a good knowledge and command of DEA. In this sense, except for a brief reminder of the DEA models included in the survey, it was not necessary to extensively explain the models to them[108].

The goal of the survey is *(1)* to identify the selection criteria to be used in the multi-criteria decision analysis and *(2)* to collect the preferences of members of the DEA community in order to calculate the criteria, local alternative and global alternative priorities (scores).

The survey was administered to the members of the DEA community through a web-based survey in two steps[109]. As part of the process, the anonymity of the respondents was guaranteed. Original screen captures of the web-based survey are presented in Appendix 1.

In the first step (April 2014), a preliminary list of four pre-determined selection criteria (understandability, applicability, acceptability and cost-benefit) was provided to the members of the DEA community (see the following sub-section about it). They were asked to complete the list by adding criteria if necessary. They were also asked to remove one or all the preliminary criteria if they considered them as inappropriate. The survey document stressed the fact that these criteria would be used to select the most suitable model among alternative models. Two assumptions were communicated. The first one indicated that the 'true' efficiency was unknown[110]. As a result, the deviation

---

107 Note that the e-mail addresses of the participants were obtained on the basis of the list of participants through an internet search. As a result, the e-mail addresses were neither asked nor obtained through the International Data Envelopment Analysis Society (iDEAs), the organizer of the conference. On the 242 participants to the conference, 164 valid e-mail addresses were identified (67.8%).

108 The results in terms of criteria and preferences are likely to be influenced by the characteristics of the sample of convenience. Another sample would probably produce different results. As the goal of this study is to apply the process of AHP in order to select the most suitable model, rather than the result itself generated by this process, this is not a cause of concern. Note that decision makers are often novice in the use of DEA. As a result, they are not familiar with DEA models. Tutoring about these models would constitute a pre-requisite should they compose the sample of convenience.

109 The web-based survey was designed and administered with *SurveyMonkey* ® (www.surveymonkey.com), an online web-survey tool.

110 'True' efficiency is determined by an artificial set of data as the production function, used to simulate data, is known.

between the 'true' efficiency and the estimated efficiency could not be considered as a criterion. The second one indicated that sufficient information about discretionary and non-discretionary variables was available in order to perform all models. In particular, the influence direction of the non-discretionary variables was known and the non-discretionary variables were available in categorical and continuous terms.

In the second step (May 2014), the members of the DEA community were asked, on a pairwise comparison basis, to express their preferences[111]. These preferences concerned:

- First the importance of each criterion with respect to the goal (i.e. to select the most suitable model);
- Second the importance of each model with respect to each criterion.

The survey listed all the possible pairwise comparisons. For each pairwise comparison, the respondents had to select, using a drop-down menu, one option among nine options, each of them corresponding to one of the nine degrees of importance included in the AHP scale. For instance, option # 3 corresponds to the third degree of the scale, stating that a criterion (respectively an alternative) is moderately preferred to another criterion (respectively another alternative). The conversion from verbal to numerical scale was realized by the author of the current study when reporting the respondents' preferences into an AHP software package.

Note that some researchers, instead of performing a second-step web-based survey, form an evaluation team of a limited number of members in order to collect the preferences (Tam & Tummala, 2001).

Preliminary criteria

None of the alternative DEA models is devoid of drawbacks nor can they always be applied to all empirical cases. But when some alternative models can be applied to a similar empirical case, the DEA literature does not provide guidance about the task of selecting the most suitable model. This is not surprising, as there is no consensus on the best model (Cordero-Ferrara *et al.*, 2008). Harrison, Rouse and Armstrong (2012, p. 263) also conclude that "there is no DEA model that is clearly superior in controlling for non-discretionary inputs (…)".

Very few authors apply alternative models to a similar empirical case in order to compare their results (Huguenin, 2014; Yang & Pollitt, 2009; Cordero-Ferrara *et al.* 2008; Muñiz, 2002). Among them, Yang and Pollitt (2009, p. 1098) compare the models not only in terms of their results but also in terms of their general advantages and disadvantages. But they failed to provide a clear list of criteria to structure this comparison. They consider the ease of interpretation, the ease of application and various technical characteristics as potential

[111] Note that the second step of the survey was sent to the full parent population, inclusive of members of the DEA community who did not participate in the first stage.

advantages or disadvantages. Muñiz (2002, p. 632), meanwhile, provides a list of criteria to compare a one-stage model with a three-stage model. With the exception of the simplicity in calculation and the possibility to include simultaneously all variables in the same stage, the other criteria refer to the results which are generated (total slack use and discriminating power).

As a result, no clear list of criteria destined to select the most suitable model emerges from the literature. Assuming *(1)* that the appropriate types of variables are available to perform the alternative models (for instance, the same variable is available in continous and categorical terms) and *(2)* that no technical characteristics prevent the use of a particular model, then the criteria should focus on general non-technical (or qualitative) characteristics, such as the ease of interpretation or the ease of application.

One could argue that some of the non-technical criteria should be oriented toward the results of the alternative models. For instance, the results of alternative models performed on a sample of schools whose resources are linked to an environmental variable will inform the decision maker about the model which provides the most favorable results for the less favored schools. If the decision maker supports the implemented priority education policy, he will probably select the model whose results show evidence to support such policy. But a decision maker who does not support such policy will reject this model and select another one.

This example shows that the inclusion of results-oriented criteria is likely to trigger opportunistic behaviour from the decision maker. A model selection based on results-oriented criteria presents at least two additional drawbacks:

- The selected model will probably not remain the same over time, as the selection process will follow the production of more or less favorable results.
- If the selected model happens to be a sophisticated one, it could be difficult to understand as it lacks transparency. It could also be difficult to communicate. Supporters of the selected model may have a hard time convincing opponents of its appropriateness and relevance.

For these reasons, the current study supports the use of general non-technical criteria which are not results-oriented[112]. A preliminary list of such criteria is proposed in the survey. These criteria alongside working definitions are presented hereafter:

1. Understandability

   The model is simple and transparent. It is easy to understand. As a result, it is easy to communicate. This criterion includes the notions of ease of interpretation highlighted by Yang and Pollitt (2009). Note that the criterion of understandability is also considered as a qualitative

[112] In the business world, the selection of the most suitable model has to be done by decision makers who are informed users of DEA but usually not DEA specialists. This advocates further in favor of general non-technical criteria.

criterion in other fields, such as in accounting (International Public Sector Accounting Standards Board, 2013).

2. Applicability

   The model is easy to apply. It is easy to run (or perform). Results are easy to calculate. The existence of a user-friendly software packages to run the model facilitates its application. The fact that a model contains various stages, where each stage needs potentially a different software package, makes the model less applicable. The criterion of applicability includes the notions of ease of application and simplicity in calculation highlighted by Yang and Pollitt (2009) and Muñiz (2002).

3. Acceptability

   The model is acceptable to the various stakeholders. For instance, the model used to benchmark hospitals should be accepted by surgeons, physicians, nurses, patients, and so on. The intrinsic characteristics of the model make it acceptable or not. A model could be easily understandable but not acceptable as its caracteristics do not make sense. For instance, a stakeholder could argue that a model which incorporates non-discretionary variables alongside discretionary variables during the same stage is not acceptable. The criterion of acceptability is not evoked in the DEA literature. However, it is likely that the most suitable model should gain a consensus regarding its acceptability among stakeholders.

4. Cost-benefit

   Running a model imposes costs. The benefits of a model should justify its costs. Assessing whether the benefits justify the costs is a matter of judgement. The costs of a model include the costs of collecting and processing the data. The benefits of a model include the added-value in terms of results provided by the model. The criterion of cost-benefit is not evoked in the DEA literature. However, it is used in other fields, such as in accounting (International Public Sector Accounting Standards Board, 2013).

Note that in AHP studies, three alternative approaches are used in order to identify selection criteria:

- Selection criteria are defined by the researchers, without justification. They are considered as given. See for instance Pak (2013), Ishizaka, Balkenborg and Kaplan (2011), Bertonlini, Braglia and Carmignani (2006) or Al Kahlil (2002). This approach may be perceived as lacking rigor. However, it fits well in the spirit of AHP, which considers the individual preferences of decision makers. Nothing prevents the selection criteria being chosen by decision makers in order to reflect their individual preferences.
- Selection criteria are based on a literature review. These criteria are usually grouped and re-organized by the researchers. See for instance Huang, Chu and Chiang (2008) or Al-Harbi (2001). A drawback of this approach is that the criteria may vary from one study to another. This is partly due to

the conflicting interests of multiple stakeholders considered in the studies (Park & Min, 2011).

- Selection criteria are based on surveys. See for instance Park and Min (2011) or Tam and Tummala (2001). Usually, the surveys used contain a list of pre-determined criteria based on a literature review. A drawback of this approach is that the criteria retained may vary according to the sample of respondents. Usually, the representativeness of the sample of respondents is not tested in these studies. Another drawback is that the surveys are often rigid. They do not permit respondents to add or remove criteria.

Models retained as alternatives

The alternative models considered in the web-based survey are all, to some extent, user-friendly and easily accessible to practitioners and decision makers (Huguenin, 2014). A limited number of four models is included in the survey, as expressing preferences on a pairwise comparison basis is time consuming[113]. The models retained are:

- **The Banker and Morey (1986a) model (BM1986a)**

  In this model, the entities are grouped into homogenous categories defined by the level of the environmental variables. In order to measure efficiency, entities are compared only with other entities with similar or worse environmental variables.

  Following the notation adopted by Johnes (2004, pp. 630-637), there are data on s outputs and m inputs for each of n entities to be evaluated. y_{rk} is the quantity of output r produced by entity k. x_{ik}^{D} is the quantity of discretionary input i consumed by entity k. θ_k represents the measure of the variable return to scale (VRS) technical efficiency of entity k (i.e. 'pure' technical efficiency free from any scale inefficiency). λ_j represents the associated weighting of outputs and inputs of entity j.

  The formulation of the categorical model is specified as follows:

$$\text{Minimize} \quad \theta_k \tag{4}$$

$$\text{Subject to} \quad y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \le 0 \qquad r = 1, \ldots, s$$

[113] Four models to judge with respect to, for instance, three criteria corresponds to 18 pairs of models (6 per criterion). Five models would have meant a total of 30 pairs (10 per criterion). As the perceived burden of responding to a survey is tied to its length (Handwerk, Carson & Blackwell, 2000), it is preferable to keep the questionnaire as short as possible. Note also that Saaty (1980) considers that the number of alternatives should be lower than nine in order to keep the evaluation simple enough.

$$\theta_k x_{ik}^D - \sum_{j=1}^{n} \lambda_j x_{ij}^D \geq 0 \qquad i = 1,\ldots,m$$

$$\sum_{j=1}^{n} \lambda_j d_{rj}^{(Cr)} \leq d_{rk}^{(Cr)} \qquad r = 1,\ldots,R$$

$$C\, r = 1,\ldots,C-1$$

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

$$\lambda_j \geq 0 \quad \forall j = 1,\ldots,n$$

The third set of constraints corresponds to an index of dummy variables d_{rk}^{Cr} representing categories of the environment. C represents the category level (e.g. school category C) and r represents the category variable (where there are more than one category variable).

- **The Banker and Morey (1986b) model (BM1986b)**

  In this model, the environmental variables are included directly into the model as non-discretionary variables. This model takes into account the fact that environmental variables are not under the control of management and cannot be treated as discretionary factors. As a result, the constraints on the environmental variable are modified. Assuming an input-orientation with variable returns to scale, the inputs are divided into discretionary (x^D) and non-discretionary sets (x^{ND}). The model is specified as follows:

Minimize θ_k (5)

Subject to
$$y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \qquad r = 1,\ldots,s$$

$$\theta_k x_{ik}^D - \sum_{j=1}^{n} \lambda_j x_{ij}^D \geq 0 \qquad i = 1,\ldots,m$$

$$x_{uk}^{ND} - \sum_{j=1}^{n} \lambda_j x_{uj}^{ND} \geq 0 \qquad u = 1,\ldots,v$$

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

$$\lambda_j \geq 0 \quad \forall j = 1,\ldots,n$$

In the above model, an additional constraint is included for each non-discretionary input. These constraints are similar to the constraints for the discretionary inputs with the exception that the efficiency component is not included. As a result, the efficiency is defined with respect to the discretionary inputs only.

Although they have been criticized, BM1986a and BM1986b are supported by Harrison *et al.* (2012), who note that these models are widely used by researchers. They have generated at least 239 different publications (Löber & Staat, 2010, p. 810). Harrison *et al.* (2012, p. 263) stress that it suggests that many researchers have found these models appropriate for their particular context. They also mention that "given there is no DEA model that is clearly superior in controlling for non-discretionary inputs, researchers continue to refer to the work of Banker and Morey (1986a, b)" (p. 263).

- **The Ray (1991) model (R1991)**

  This model contains two stages. In the first stage, a basic DEA model is performed using only discretionary variables. After obtaining the technical efficiency scores (TE) from the first stage, Ray (1991) uses an OLS model to regress these scores upon non-discretionary variables in the second stage. The second stage regression is specified as follows:

$$TE_k = \alpha_0 + \beta_1 X_1 + \ldots \beta_v X_v + e_k \qquad (6)$$

  The error term represents the efficiency. Since Ray (1991), other types of regression have been used in the second stage. For instance, McCarty and Yaisawarng (1993) are the first to use a Tobit regression.

  R1991 is recommended by Coelli *et al.* (2005) in most cases. It has demonstrated its superiority to other models which allow for an environmental adjustment (Ruggiero, 1998, 2004).

- **The Yang and Paradi model** (in Muñiz, Ruggiero, Paradi and Yang, 2006, p. 1176) (**YP2006**).

  This model applies a handicapping measure based on the levels of the non-discretionary variables. Entities with a favorable environment are penalized by the handicapping measure. Non-discretionary inputs are adjusted with a higher handicap and non-discretionary outputs are adjusted with a lower handicap. As a result, adjusted inputs have a higher value than original inputs and adjusted outputs have a lower value than original outputs. Assume h_j is the handicapping measure to adjust input variables and $\hat{h}_j$ the handicapping measure to adjust outputs variables. The adjusted input is $h_j x_{ij}$ and the adjusted output is $\hat{h}_j y_{rj}$.The model is specified as follows:

$$\text{Minimize} \quad \theta_k \tag{7}$$

$$\text{Subject to} \quad \hat{h}_k y_{rk} - \sum_{j=1}^{n} \lambda_j \hat{h}_j y_{rj} \leq 0 \qquad r = 1, \ldots, s$$

$$\theta_k h_k x_{ik} - \sum_{j=1}^{n} \lambda_j h_j x_{ij} \geq 0 \qquad i = 1, \ldots, m$$

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

$$\lambda_j \geq 0 \quad \forall j = 1, \ldots, n$$

YP2006 is relatively little known and used. Compared to BM1986a, it does not lessen the discriminating power of DEA, as it does not categorize the entities. YP2006 is particularly suited when discretionary inputs and/or outputs are augmented or diminished according to the condition of the environment.

Note that other models than BM1986a, BM1986b, R1991 and YP2006 could have been considered. The inclusion or the removal of one or several models is likely to modify the results in terms of local and global alternative scores. However, as the current study focuses on the process of selecting the most suitable model rather on the results itself generated by this process, the choice of models included in the survey is, in itself, not determinant.

4.5 Results

Web-based survey – First step

16 respondents (9.8%) participated to the first step of the survey[114]. The criteria of understandability, applicability and acceptability are backed by a majority of respondents. They are kept in the second step of the survey. One respondent notes that "understandability and acceptability are the most important criteria when it comes to environmental implications". The criterion of applicability is critically discussed by two respondents. The first one notes that "the word sounds fine, but the definition gives the impression that if it is easy to calculate, it is valuable. I disagree with this premise". However, the second one estimates that "applicability is an important issue".

When asked to remove one criterion if they were forced to do so, eight respondents out of 16 (50%) answer that they would remove the criterion of

114 Shih and Fan (2008) show that web-based surveys produce an average response rate of 34%. A general concern of web-based surveys is their low response rates compared to other types of surveys (Monroe & Adams, 2012). In the context of the current study, the low response rate (9.8%) is not relevant as it does not bear on representativeness (Cook, Heath & Thomson, 2000).

cost-benefit from the list. This is probably due to the fact that this criterion partly overlaps the criterion of applicability. As a result, the respondents possibly consider that cost-benefit and applicability are somehow correlated. If a model is easy to apply, then performing it triggers lower costs. Obviously, the respondents favour applicability over cost-benefit, as only four respondents (25%) would remove applicability. Finally, two respondents would remove understandability (12.5%) and two would remove acceptability (12.5%). Following this result and to keep the number of selection criteria as low as possible, the criterion of cost-benefit has not been kept in the second step of the survey.

Five respondents suggest additional criteria alongside working definitions:

- "Suitability: all models are not suitable to a particular case study".
- "Capability: the models' assumptions should be as close as possible to the reality".
- "Statistical properties: the models should reflect the physical processes of production".
- "Does the model work properly"?
- "Adaptivity – Flexibility – Time appropriateness: A model should be easily adaptable to incorporate parameter and context variability. In addition, it should provide results in a timely manner".

The above mentioned criteria are all relevant. However, the survey mentions that sufficient information is available in order to perform all models. As a result, the criterion of suitability does not need to be added in the second step. The survey also mentions that the 'true' efficiency is unknown, meaning that the production function is not identified. As a result, the criteria of capability and statistical properties are not retained in the second step. Finally, the survey makes the implicit assumption that all models work properly, making the criterion 'Does the model work properly' unnecessary. The suggestion to include the criteria of adaptivity, flexibility and time appropriateness remains open. As they are mentioned by only one respondent, and not confirmed by a second one, the decision has been made not to retain them in the second step of the survey.

Web-based survey – Second step

10 respondents (6.1%) participated to the second step of the survey. Among them, only six were complete (3.7%). The response rate of the second step is lower than the response rate of the first step. It could probably be explained by the length of the questionnaire. Four out of 10 respondents dropped-off after the first three questions. It could also be explained by the degree of difficulty of thinking to lead, even for respondents used to DEA, and by the heterogeneity of the parent population, both geographically and professionally (a mix of scholars and practitioners). Again, in the context of the current study, the low response rate (3.7%) is not relevant as it does not bear on representativeness (Cook, Heath & Thomson, 2000). Moreover, AHP does not need to involve a

large sample. It is particularly appropriate for research focusing on a specific issue with a small sample (Lam & Zao, 1998).

Respondents' preferences are converted from a verbal to a numerical scale by the author of the current study. They are entered into *MakeItRational* ®, a decision-making software based on AHP. *MakeItRational* ® is used to calculate the criteria, local alternative and global alternative priorities (scores). When inconsistency appears, pairwise comparisons are revised following the software suggestion. The local and global alternative scores are calculated according to the distributive mode[115].

Criteria priorities

The criteria priorities reflect the importance of each criterion with respect to the goal. They are displayed in Table 29. For each of the six respondents and for the group as a whole (columns '1' to 'Group'), the weights of each criterion are presented. For instance, respondent # 3 clearly prefers the criterion of acceptability with a score of 57.14 (column '3'; line 'Acceptability'). The same respondent considers then that the criterion of applicability is more important than the criterion of understandability, with scores of 28.57 and 14.29 respectively. Note that the sum of the weights of the three criteria equals 100.

The preferences on the criteria vary among the respondents. For instance, respondent # 4 favors the criterion of applicability (69.12) while respondent # 6 prefers the criterion of understandability (64.41). The methodology clearly allows every single decision maker to value the criteria according to his own preferences.

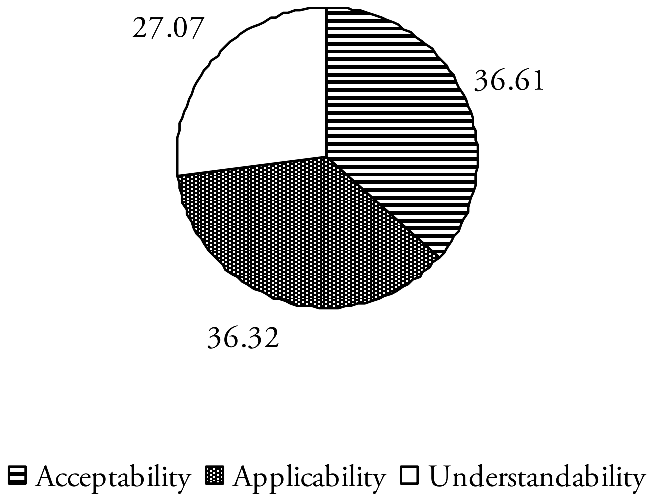
Table 29
Criteria priorities

	Weight (%)						
Criterion	**1**	**2**	**3**	**4**	**5**	**6**	**Group**
Understandability	26.84	33.33	14.29	16.01	9.89	64.41	27.07
Applicability	11.72	33.33	28.57	69.12	36.43	27.06	36.32
Acceptability	61.44	33.33	57.14	14.86	53.68	8.53	36.61

When the individual results are aggregated over the group (last column of the table), results show that the weights of the three criteria are balanced (see Figure 63). However, the criteria of acceptability and applicability, which present close scores, are preferred to the criteria of understandability.

[115] Note that the local and global alternative scores have also been calculated according to the ideal mode. Results in terms of ranking are similar to those calculated according to the distributive mode.

Figure 63
Criteria priorities aggregated over the group



Local alternative priorities

The local alternative priorities reflect the importance of an alternative with respect to one specific criterion. They are unweighted by the criteria weights. In other words, a score of 100 is allocated to each criterion, and is spread over the four alternatives. The local alternative priorities are displayed in Table 30.

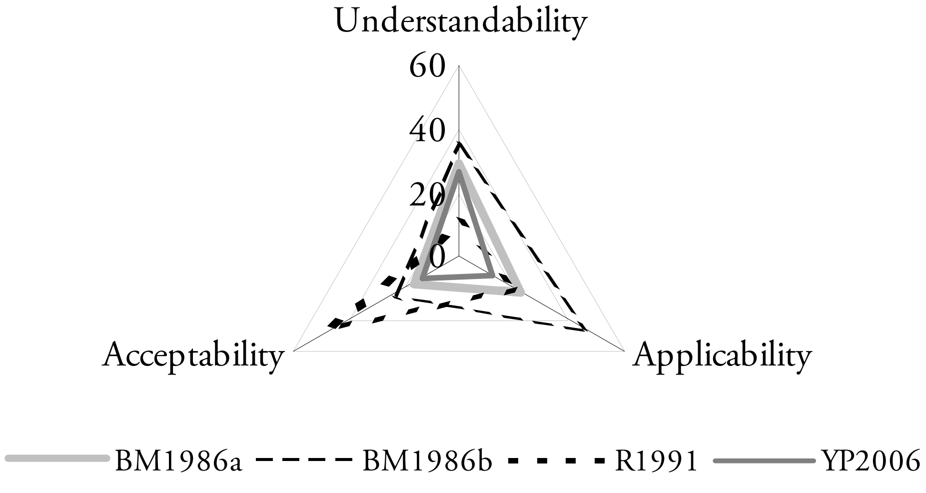
The mathematical aggregation of the individual results is presented in the last part of the table (section 'Group'). For instance, BM1986a is the preferred alternative when it comes to understandability (score of 32.32; column 'Understandability', line 'BM1986a' in the 'Group' section of the table). It is followed by R1991 (26.82), YP2006 (23.24) and BM1986b (17.62). Note that the sum of the four weights associated with the four alternatives for each criterion equals 100. When it comes to applicability, R1991 is the favored model (31.6). And when it comes to acceptability, BM1986a is the preferred model (28.06).

Table 30
Local alternative priorities

Respondent	Alternative	Weight (%) Understandability	Applicability	Acceptability
1	BM1986a	27.65	49.74	25.71
	BM1986b	6.00	8.72	7.04
	R1991	14.21	8.31	59.37
	YP2006	52.15	33.23	7.88
2	BM1986a	7.69	8.96	12.36
	BM1986b	30.77	22.08	29.94
	R1991	30.77	28.65	18.82
	YP2006	30.77	40.31	38.87
3	BM1986a	54.16	11.91	26.00
	BM1986b	19.02	11.91	60.17
	R1991	16.59	68.62	5.92
	YP2006	10.23	7.55	7.91
4	BM1986a	39.92	10.00	42.36
	BM1986b	10.81	23.50	14.67
	R1991	39.79	54.86	7.97
	YP2006	9.48	11.65	35.00
5	BM1986a	29.29	21.99	16.82
	BM1986b	34.07	45.98	23.62
	R1991	9.75	20.09	46.10
	YP2006	26.90	11.94	13.46
6	BM1986a	29.93	66.81	19.90
	BM1986b	8.64	16.70	8.38
	R1991	47.38	11.90	23.83
	YP2006	14.05	4.59	47.89
Group	BM1986a	32.32	26.76	28.06
	BM1986b	17.62	24.06	22.65
	R1991	26.82	31.60	24.74
	YP2006	23.24	17.58	24.55

Table 30 also displays the individual results for each respondent (sections '1' to '6' of the table). For every respondent, it is possible to display the results in a spider diagram. This type of representation highlights the strengths and weaknesses of each model from the point of view of the respondent. For instance, Figure 64 illustrates the results of respondent # 5. In this case, R1991 is strong on the criterion of acceptability but weak on the criteria of understandability and applicability. BM1986b is strong on the criterion of applicability, medium on the criterion of understandability and weak on the criterion of acceptability. BM1986a and YP2006 are balanced on the three criteria.

Figure 64
Local alternative priorities of respondent # 5



One may have expected more consistent results among the respondents, as they belong to the same parent population of DEA specialists. For instance, it could be considered surprising, at first glance, that respondent # 1 weights the understandability of BM1986a at 27.65% when respondent # 2 weights the same criteria for the same model at 7.69%. However, it has to be kept in mind that the preferences expressed by the respondents are subjective. They are based on their knowledge of the models and their own experience. Given this, it is conceivable that these preferences could differ. Moreover, it has to be kept in mind that the respondents had no opportunity to harmonize their answers as they were located in different locations. This prevented the building of a consensus.

Global alternative priorities

The global alternative priorities, displayed in Table 31, reflect the importance of alternatives across all criteria. They rank the alternative models from the most to the less suitable.

The mathematical aggregation of the individual results is presented in the last part of the table (section 'Group'). BM1986a is the preferred model, with a total score of 28.74 (column 'Total', line 'BM1986a' in the 'Group' section). Understandability, applicability and acceptability contribute 8.75, 9.72 and 10.27 respectively towards the total score of 28.74 (columns 'Understandability', 'Applicability' and 'Accessibility', line 'BM1986a' in the 'Group' section). Note that the sum of these three weights equals the total score of BM1986a of 28.74. The total score of R1991 (27.79) is close to the score of BM1986a. BM1986b and YP2006 follow with total scores of 21.8 and 21.66 respectively. Note that the sum of the total scores of the four alternatives equals 100.

Individual results of the six respondents are separately presented in the table (sections '1' to '6'). Results vary according to the preferences of each respondent. For instance, R1991 is the most suitable model for respondents # 1 (41.27), # 4 (45.47) and # 5 (33.03). Respondents # 1 and # 5 prefer this model especially because of its acceptability (36.48 and 24.75 respectively). Respondent # 4 particularly appreciates the applicability of this model (37.92).

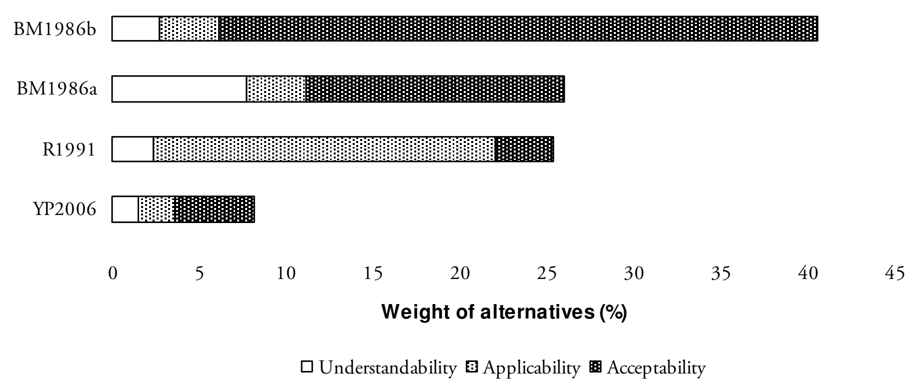
YP2006 is the most suitable model for respondent # 2 (36.65). The contribution to this total score by the three criteria is balanced (13.44 for applicability, 12.96 for acceptability and 10.26 for understandability). BM1986b is the most suitable model for respondent # 2, who appreciates its acceptability (34.38). Finally, respondent # 6 favors BM1986a, especially because of its understandability (19.28) and its applicability (18.08).

Table 31
Global alternative priorities

Respondent	Alternative	Weight (%)			
		Total	Understandability	Applicability	Acceptability
1	BM1986a	29.05	7.42	5.83	15.80
	BM1986b	6.95	1.61	1.02	4.32
	R1991	41.27	3.81	0.97	36.48
	YP2006	22.73	14.00	3.90	4.84
2	BM1986a	9.67	2.56	2.99	4.12
	BM1986b	27.60	10.26	7.36	9.98
	R1991	26.08	10.26	9.55	6.27
	YP2006	36.65	10.26	13.44	12.96
3	BM1986a	26.00	7.74	3.40	14.86
	BM1986b	40.50	2.72	3.40	34.38
	R1991	25.36	2.37	19.61	3.38
	YP2006	8.14	1.46	2.16	4.52
4	BM1986a	19.60	6.39	6.91	6.39
	BM1986b	20.16	1.73	16.24	2.18
	R1991	45.47	6.37	37.92	1.18
	YP2006	14.77	1.52	8.05	5.20
5	BM1986a	19.94	2.90	8.01	9.03
	BM1986b	32.80	3.37	16.75	12.68
	R1991	33.03	0.96	7.32	24.75
	YP2006	14.23	2.66	4.35	7.22
6	BM1986a	39.06	19.28	18.08	1.70
	BM1986b	10.80	5.57	4.52	0.71
	R1991	35.77	30.52	3.22	2.03
	YP2006	14.37	9.05	1.24	4.08
Group	BM1986a	28.74	8.75	9.72	10.27
	BM1986b	21.80	4.77	8.74	8.29
	R1991	27.79	7.26	11.48	9.05
	YP2006	21.66	6.29	6.39	8.99

The results diplayed in Table 31 can, for each respondent, be illustrated in a stacked bar diagram for a better visualization. For instance, the results of respondent # 3 are presented in Figure 65. The bars represent the global alternative priorities of the models. BM1986b is the preferred model of respondent # 3, especially because of its high acceptability. BM1986a and R1991 have very close scores. But they are not appreciated for the same reasons. While BM1986a is appreciated for its good acceptability, R1991 is appreciated for its good applicability.

Figure 65
Global alternative priorities of respondent # 3



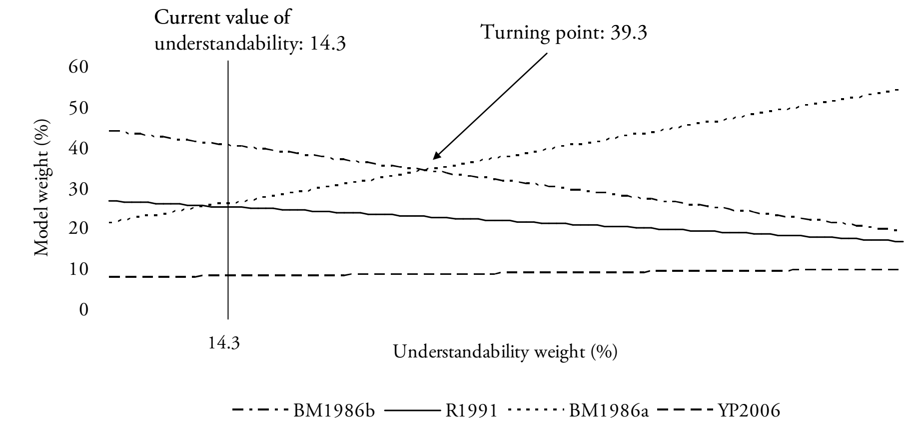
Sensitivity analysis

A sensitivity analysis can be performed in the AHP framework. It is conducted in *MakeItRational* ®. It allows assessment of the impact of changes to one criterion weight over the global alternative priority. Figure 66, Figure 67 and Figure 68 illustrate the sensitivity analysis performed on respondent # 3. This respondent prefers BM1986b (see Table 31). He gives the following scores to the criteria of understandability, applicability and acceptability: 14.3, 28.6 and 57.1 (see Table 29).

Figure 66 focuses on the criterion of understandability. The global alternative priority of each model (y-axis) is represented by a linear curve. The x-axis indicates the weight of understandability. The preferred model in the current situation (i.e. with a value of understandability of 14.3) is BM1986b. It can be visualized by the upper curve on the graph with an understandability weight of 14.3. If the current understandability weight is reduced, then the preferred model remains BM1986b. If the current understandability weight is increased to over 39.3 (turning point on the graph), then the preferred model is no longer BM1986b but BM1986a.

When the weight of understandability is modified from 14.3% to 39.3%, it leaves 60.7% (100 – 39.3) for the other criteria (applicability and acceptability) while keeping the proportionality between them.

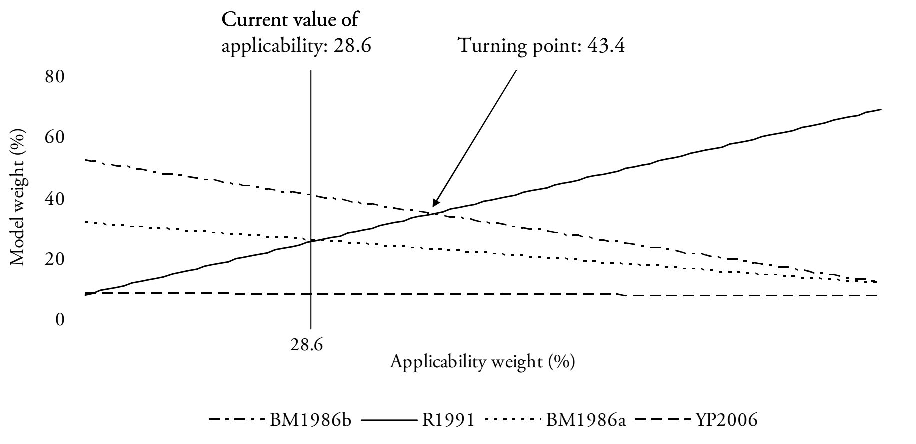
Figure 66
Impact of the changes of understandability (respondent # 3)



A similar sensitivity analysis is conducted for the criteria of applicability and acceptability.

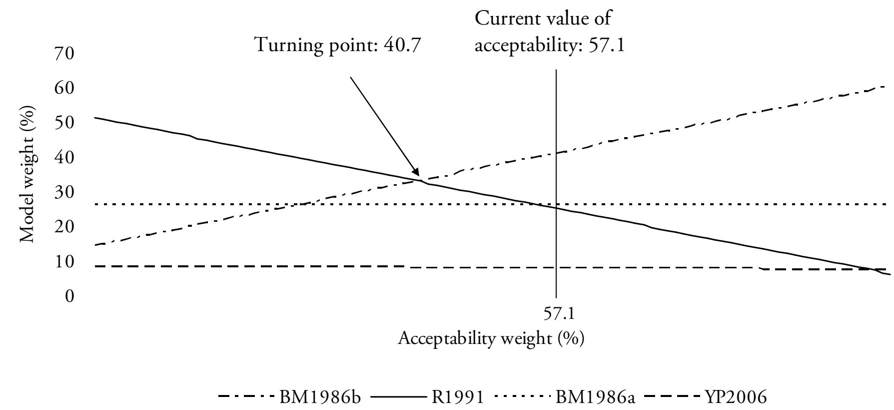
Figure 67 focuses on the criterion of applicability. The preferred model in the current situation (i.e. with a value of applicability of 28.6) is BM1986b. If the current applicability weight is reduced, then the preferred model remains BM1986b. If the current applicability weight is increased to over 43.4 (turning point on the graph), then the preferred model is no longer BM1986b but R1991.

Figure 67
Impact of the changes of applicability (respondent # 3)



Finally, Figure 68 focuses on the criterion of acceptability. The preferred model in the current situation (i.e. with a value of acceptability of 57.1) is BM1986b. If the current applicability weight is reduced to under 40.7 (turning point on the graph), then the preferred model is no longer BM1986b but R1991. If the current acceptability weight is increased, then the preferred model remains BM1986b.

Figure 68
Impact of the changes of acceptability (respondent # 3)



When applied to the aggregated results of respondents #1 to # 6, the sensitivity analysis shows that:

- BM1986a is the most suitable model if the understandability weight is higher than 12.3. In the current situation, this weight is valued at 27.1. Should the understandability weight be lower than 12.3, then the preferred model would no longer be BM1986a but R1991.
- BM1986a is the most suitable model if the applicability weight is lower than 47.2. In the current situation, this weight is valued at 36.3. Should the applicability weight be higher than 47.2, then the preferred model would no longer be BM1986a but R1991.
- BM1986a is the most suitable model if the acceptability weight is higher than 10.5. In the current situation, this weight is valued at 36.6. Should the acceptability weight be lower than 10.5, then the preferred model would no longer be BM1986a but R1991.

4.6 Further analysis

Performance measurement techniques include several methods, such as econometric or linear programming methods. They possibly deliver divergent results (Farsi & Filippini, 2005). Within the same method, such as DEA, different models also deliver divergent results (Huguenin, 2014). As far as the author is aware, multi-criteria decision analysis (MCDA) methods have never been applied with the objective of selecting the most suitable method (or model). In this sense, the current study constitutes a first step to fill this research gap. Future research could capitalize on it in order to further investigate the use of MCDA methods to select a performance measurement technique.

The identification of the selection criteria is a crucial step in AHP applications. In the current study, the establishment of a preliminary list of such criteria was made difficult by the limited information available in the literature. Further studies could focus on how to establish a preliminary list of selection criteria in such situations. Conducting workshops with experts could be a possible way.

The approach developed in the current study to identify selection criteria is a qualitative one. The web-based questionnaire collected the judgments of the respondents about the criteria. The respondents could add or remove criteria and justify their decision. As a result, the questionnaire did not ask the respondents to rank the criteria (for instance on a Likert scale). The qualitative approach is appropriate when the number of criteria is low (due in this case to the limited information available in the literature) and when the opinion of respondents is sought after, as in this study. Further research could augment the number of pre-established criteria and adopt a quantitative approach. In such a case, selection criteria would be considered as relevant if they obtain a value higher than a cutoff value fixed by the researchers (Tam & Tummala, 2001).

Once identified, the selection criteria may be considered as biased as they may be influenced by the interests of stakeholders who decided on them. More generally, the representativeness of the sample of respondents is underresearched in AHP. Even within an apparently homogeneous parent population, the representativeness of the sample should be tested if the objective of the research is to establish and validate the selection criteria as representative of the entire parent population. Testing the representativeness requires prior knowledge about the characteristics of the parent population and the collection of general information about the respondents, for instance their number of years of experience in the sector analyzed or their professional activity.

Further studies could also test if preferences expressed by different expert groups, from different backgrounds, provide convergent results about the ranking of the alternatives. In other words, do the different expert groups get the same final results or not? This issue raises the question of the generalization of AHP results, which is underresearched in AHP theory and applications.

Finally, the group aggregation process in AHP requires further research. Chwolka and Raith (2000) identify two processes of group aggregation:

- The choice harmonization process is used during workshops. This process requires compromises throughout the decision process. A consensus is reached among group members about criteria and alternative weights, meaning that the aggregation is realized by 'live discussion' among the experts.
- The choice aggregation process is used in situations when a consensus cannot be reached or when the experts are located in different locations (and cannot exchange views among each other). As a result, a mathematical aggregation of the individual judgements is needed.

As pointed out by Saaty and Vargas (2005)

> To achieve a decision with which the group is satisfied, the judgments, and ultimately the priorities, must be accepted by the group members. This requires that (a) the judgments be homogeneous, and (b) the priorities of the individual group members be compatible with the group priorities (p. 1).

Ramanathan and Ganesh (1994) show that dispersion in judgments leads to violations of the principle of Pareto optimality in AHP. In other words, only homogeneous judgments can be aggregated (Basak, 1988). However, empirical evidence suggests that the preferences expressed by even apparently homogeneous group of experts could in fact be heterogeneous (von Solms, 2009). This can lead, in the choice aggregation process, to a smoothing of the weights calculated for the criteria, local and global priorities, especially alongside an increasing number of group members. As a result, no clear alternative emerges from the process of aggregation. Further research is needed about how to handle heterogeneous judgements in the choice aggregation process.

4.7 Conclusion

This study provides an AHP-based approach to select the most suitable DEA model. In this approach, the identification of the selection criteria is a crucial step. To avoid a biased model selection and potential opportunistic behaviour from decision makers, this study argues that such criteria should not be oriented towards the results of the alternative DEA models. Still, it is a difficult task to identify valid criteria, especially if they have to be consensually accepted by the different stakeholder groups involved. In certain situations and for certain decision makers, the choice of criteria may ultimately be a political judgment.

In this study, a preliminary list of criteria, based on limited available information in the literature, has been submitted to members of the DEA community. The criteria of understandability, acceptability and applicability have been backed. However, note that it is likely that another group of decision makers, for instance a group of headteachers or a group of top public servants, would have reached a consensus on alternative criteria. This point leads to the difficulty of generalizing results produced by AHP applications. However, as long as the technique is used to produce individual decision-making (or small homogeneous group decision-making), this is not a matter of concern.

The case of individual decision-making deserves a discussion. AHP is tailored to take into account individual preferences. Nothing prevents these preferences from also being expressed towards the choice of selection criteria. In such a case, an explicit risk exists that an individual decision maker would define criteria oriented towards the expected (or wanted) results. In other words, selection criteria are dictated by expected results. The application of MCDA methods would not make sense with biased criteria. At best, the individual decision maker would seek, in this case, a scientific justification of his pre-determined decision. As a result, even in the case of the individual decision-making, the selection criteria should be backed by the literature or by an expert group.

The categorical model developed by Banker and Morey (1986a) emerges as the most suitable model when results are aggregated over the group of respondents. This model has the weakness of lessening the discriminating power of DEA when the different categories number a small number of entities. This is likely to be the case when more than one environmental variable have to be taken into account. As a result, the second most suitable model identified by this study may be the best option in all situations. This model is the two-stage DEA model developed by Ray (1991). Note that the global priority score obtain by this model (27.79) is very close to the score obtained by the categorical model (28.74). However, practitioners may find it difficult to apply, as it implies a two-stage procedure and the potential need of two different software packages. This is not the opinion of members of the DEA community, who think that the Ray (1991) model is the most applicable one among the four models considered in this study. They are probably right as both stages of the model (efficiency analysis and multiple regression analysis) can nowadays also be performed in *Excel* ® (but still require programming skills).

In terms of policy and managerial implications, the results of the current study suggest that:

- The number of selection criteria and alternatives should remain parsimonious in order to avoid the time consuming process of AHP.
- The selection criteria should be backed by the literature or by an expert group. They should not be oriented towards the results.
- A single individual decision maker should not be left alone to identify the selection criteria.

Once the most suitable DEA model is identified, the following principles should prevail:

- The principle of permanence of methods: This principle states that the retained methods (in the current case, the DEA model retained) are not modified from one period of time to the other. It allows the coherence and comparison of the efficiency results produced.
- The principle of consistency: This principle requires the decision maker to be consistent from one period of time to another in applying the same DEA model.

References

Adler, N. & Berechman, J. (2001). Measuring airport quality from the airlines' viewpoint: an application of data envelopment analysis. *Transport Policy*, *8*, 171-181.

Agasisiti, T., Bonomi, F. and Sibiano, P. (2014). Measuring the "managerial" efficiency of public schools: a case study in Italy. *International Journal of Educational Management*, *28*(2), 120-140.

Al-Harbi, K. M. (2001). Application of the AHP in project management. *International Journal of Project Management*, *19*(1), 19-27.

Al Khalil, M. I. (2002). Selecting the appropriate project delivery method using AHP. *International Journal of Project Management*, *20*(6), 469-474.

Al-Najjar, S. M. & Al-Jaybajy, M. A. (2012). Application of Data Envelopment Analysis to Measure the Technical Efficiency of Oil Refineries: A Case Study. *International Journal of Business Administration*, *3*(5), 64-77.

Al-Rawashdeh, T., Al'azzeh, F. & Al-Qatawneh, S. (2014). Evaluation of ERP Systems Quality Model Using Analytic Hierarchy Process (AHP) Technique. *Journal of Software Engineering and Applications*, *7*(4), 225-232.

Andersson, C., Manson, J. & Sund, K. (2014). Technical efficiency of Swedish employment offices. *Socio-Economic Planning Sciences*, *48*(1), 57-64.

Arbel, A. & Orgler, Y. E. (1990). An application of the AHP to bank strategic planning: The mergers and acquisitions process. *European Journal of Operational Research*, *48*(1), 27-37.

Azadeh, A., Ghaderi, S. F. & Izadbakhsh, H. (2008). Integration of DEA and AHP with computer simulation for railway system improvement and optimization. *Applied Mathematics and Computation*, *195*(2), 775-785.

Azizi, K. H., Lofti, F. H., Saati, S. & Vahidi, A. R. (2007). Ranking the Electricity Producer Companies in View of Manpower Efficiency by DEA. *Applied Mathematical Sciences*, *1*(16), 761-768.

Bana e Costa, C., De Corte, J.-M. & Vansnick, J.-C. (2005). On the mathematical foundation of MACBETH. In J. Figueira, S. Greco & M. Ehrogott (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys* (pp. 409-437). New York: Springer.

Bana e Costa, C. & De Corte, J.-M. & Vansnick, J.-C. (2003). MACBETH. *OR Working Paper No. 03.56*. London: London School of Economics and Political Science.

Bana e Costa, C. & Vansnick, J.-C. (1999). The MACBETH approach: Basic ideas, software, and an application. In N. Meskens & M. Roubens (Eds.), *Advances in Decision Analysis, Mathematical Modelling: Theory and Application* (Vol. 4, pp. 131-157). Dordrecht: Kluwer Academic Publishers.

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.

Banker, R. D. & Morey, R. C. (1986a). The Use of Categorical Variables in Data Envelopment Analysis. *Management Science*, *34*(4), 1613-1627.

Banker, R. D. & Morey, R. C. (1986b). Efficiency Analysis for Exogenously Fixed Inputs and Outputs. *Operations Research*, *32*(12), 513-521.

Barros, C. A. P. & Santos, C. A. (2006). The measurement of efficiency in Portuguese hotels using data envelopment analysis. *Journal of Hospitality & Tourism Research*, *30*(3), 378-400.

Basak, I. (1988). When to Combine Group Judgments and When Not to in the Analytic Hierarchy Process: A New Method. *Mathematical and Computer Modeling*, *10*(6), 395-404.

Bayazit, O. (2005). Use of AHP in decision-making for flexible manufacturing systems. *Journal of Manufacturing Technology Management*, *16*(7), 808-819.

Belton, V. & Gear, A. (1983). On a shortcoming of Saaty's method of analytical hierarchies. *Omega*, *11*(3), 228-230.

Bertolini, M., Braglia, M. & Carmignani, G. (2006). *International Journal of Project Management*, *24*(5), 422-430.

Borges, M. R., Nektarios, M. & Barros, C. P. (2008). Analysing The Efficiency Of The Greek Life Insurance Industry. *European Research Studies*, *11*(3), 35-52.

Butler, T. W. & Johnson, W. W. (1997). Efficiency Evaluation of Michigan Prisons Using Data Envelopment Analysis. *Criminal Justice Review*, *22*(1), 1-15.

Cai, Y. & Wu, W. (2001). Synthetic financial evaluation by a method of combining DEA with AHP. *International Transactions in Operational Research*, *8*(5), 603-609.

Caulfield, B., Bailey, D. & Mullarkey, S. (2013). Using data envelopment analysis as a public transport project appraisal tool. *Transport Policy*, *29*, 74-85.

Celik, M., Kandakoglu, A. & Er, I. D. (2009). Structuring fuzzy integrated multi-stages evaluation model on academic personnel recruitment in MET institutions. *Expert Systems with Applications*, *36*(3), 6918-6927.

Charnes, A, Cooper, W. W. & Rhodes E. L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429-444.

Chen, Y.-C. L. (2006). An Analytic Hierarchy Framework for Evaluation Balanced Scorecards Healthcare Organizations. *Canadian Journal of Administrative Sciences*, *23*(2), 85-104.

Cheng, C.-H., Yang, K.-L. & Hwang, C.-H. (1999). Evaluating attack helicopters by AHP based on linguistic variable weight. *European Journal of Operational Research*, *116*(2), 423-435.

Chwolka, A. & Raith, M. G. (2000). Supporting Group Decisions with the AHP: Harmonization vs Aggregation of Preferences. In G. Wanka (Eds.), *Decision Theory and Optimization in Theory and Practice* (pp. 17-32). Aachen: Shaker Verlag.

Coelli, T. J., Prasada Rao, D. S., O'Donnel, C. J. & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York: Springer.

Cook, C., Heath, F. & Thompson, R. L. (2000). A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement*, *60*(6), 821-836.

Cooper, W. W., Seiford, L. M. & Tone, K. (2006). *Introduction to Data Envelopment Analysis and Its Uses*. New York: Springer.

Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data Envelopment Analysis: A comprehensive Text with Models, Applications, References and DEA-Solver Software*. New York: Springer.

Cordero-Ferrara, J. M., Pedraja-Chaparro, F. & Salinas-Jiménez, J. (2008). Measuring efficiency in education: an analysis of different approaches for incorporating non-discretionary inputs. *Applied Economics*, *40*(10), 1323-1339.

Cordero, J. M., Pedraja, F. & Santín, D. (2009). Alternative approaches to include exogenous variables in DEA measures: A comparison using Monte Carlo. *Computers & Operations Research*, *36*(10), 2699-2706.

De Witte, K. & Moesen, W. (2010). Sizing the government. *Public Choice*, *145*(1), 39-55.

Dyer, J. (1990). Remarks on the analytic hierarchy process. *Management Science*, *36*(3), 249-258.

Eling, M. & Huang, W. (2010). An Efficiency Comparison of Non-life Insurance Industry in the BRIC Countries. *Working Papers on Risk Management and Insurance No. 94*. St.Gallen: University of St.Gallen.

Emrouznejad, A., Parker, B. R. & Tavares G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, *42*(3), 151-157.

Ertay, T., Ruan, D. & Tuzkaya, U. R. (2006). Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems. *Information Sciences*, *176*(3), 237-262.

Estelle, S. M., Johnson, A. L. & Ruggiero, J. (2010). Three-stage DEA models for incorporating exogenous inputs. *Computers & Operations Research*, *37*(6), 1087-1090.

Farsi, M. & Filippini, M. (2005). A benchmarking analysis of electricity distribution utilities in Switzerland. *CEPE Working Paper No. 43*. Zurich: Center for Energy Policy and Economics, Swiss Federal Institute of Technology.

Feng, Y. J., Lu, H. & Bi, K. (2004). An AHP/DEA method for measurement of the efficiency of R&D management activities in universities. *International Transactions in Operational Research*, *11*(2), 181-191.

Fethi, M. D., Jackson, P. M. & Weyman-Jones, T. G. (2000). Measuring the efficiency of European airlines: An application of data envelopment analysis and Tobit Analysis. *EPRU Discussion Papers*. Leicester: University of Leicester.

Forgionne, G. A., Kohli, R. & Jennings, D. (2002). An AHP analysis of quality in AI and DSS journals. *Omega*, *30*(3), 171-183.

Forman, E. H. & Gass, S. I. (2001). The analytical hierarchy process – an exposition. *Operations Research*, *49*(4), 469-487.

Frei, F. X. & Harker, P. T. (1999). Measuring aggregate process performance using AHP. *European Journal of Operational Research*, *116*(2), 436-442.

Friebelová, J., Friebel, L. & Marková, K. (2009). Using DEA models for evaluation of firestation efficiency. *Bohemiae Meridionalis*, *12*(2), 71-78.

Fried, H. O., Schmidt, S. S. & Yaisawarng, S. (1999). Incorporating the operating into a nonparametric measure of technical efficiency. *Journal of Productivity Analysis*, *12*(3), 249-267.

Fried, H. O., Lovell, C. A. K., Schmidt, S. S. & Yaisawarng, S. (2002). Accounting for environmental effects and statistical noise in data envelopment analysis. *Journal of Productivity Analysis*, *17*(1/2), 157-174.

Gautam, S., Hicks, L., Johnson, T. & Mishra, B. (2013). Measuring the Performance of Critical Access Hospitals in Missouri using Data Envelopment Analysis. *The Journal of Rural Health*, *29*(2), 150-158.

Giannoulis, C. & Ishizaka, A. (2010). A Web-based decision support system with ELECTRE III for a personalised ranking of British universities. *Decision Support Systems*, *48*(3), 488-497.

Gupta, S. & Kumar, U. (2012). An analytical hierarchy process (AHP)-guided decision model for underground mining method selection. *International Journal of Mining, Reclamation and Environment*, *26*(4), 324-336.

Guo, J.-Y., Liu, J. & Qiu, L. (2006, December). *Research on Supply Chain Performance Evaluation Based on DEA/AHP Model.* Paper presented at the Asia-Pacific Conference on Services Computing, Guangzhou, Guangdong, China.

Hajkowicz, S., Young, M., Wheeler, S., MacDonald, D. H. & Young, D. (2000). *Supporting decisions, understanding natural resource management assessment techniques. A report to the Land and Water Resources Research and Development Corportation* (Technical Report of June 2000). Adelaide: Land and Water Research and Development Corporation.

Handwerk, P., Carson, C. & Blackwell, K. (2000). *On-line versus paper-and-pencil surveying of students: A case study.* Paper presented at the Annual Forum of the Association for Institutional Research, Cincinnati, United States.

Harrison, J., Rouse, P. & Armstrong, J. (2012). Categorical and continuous non-discretionary variables in data envelopment analysis: a comparison of two single-stage models. *Journal of Productivity Analysis*, *37*(3), 261-276.

Harrison, J. & Rouse, P. (2014). Competition and public high school performance. *Socio-Economic Planning Science*, *48*(1), 10-19.

Ho, W. (2008). Integrated analytic hierarchy process and its applications – a literature review. *European Journal of Operational Research*, *186*(1), 211-228.

Holder, R. (1990). Some comment on the analytic hierarchy process. *Journal of the Operational Research Society*, *41*(11), 1073-1076.

Holder, R. (1991). Response to Holder's comments on the analytic hierarchy process: Response to the response. *Journal of the Operational Research Society*, *42*(10), 914-918.

Holod, D. & Lewis, H. F. (2011). Resolving the deposit dilemma: a new DEA bank efficiency model. *Journal of Banking & Finance*, *35*(11), 2801-2810.

Huang, C.-C., Chu, P.-Y. & Chiang, Y.-S. (2008). A fuzzy AHP application in government-sponsored R&D project selection. *Omega*, *36*(6), 1038-1052.

Huguenin, J.-M. (2013). Data Envelopment Analysis (DEA). In A. Ishizaka & P. Nemery (Eds.), *Multi-Criteria Decision Analysis: Methods and Software* (pp. 235-274). Chichester: John Wiley & Sons.

Huguenin, J.-M. (2014). DEA does not like positive discrimination: a comparison of alternative models based on empirical data. *IDHEAP Working Paper 7/2014*. Lausanne: Swiss Graduate School of Public Administration.

Huguenin, J.-M. (forthcoming). Determinants of school efficiency: the case of primary schools in the State of Geneva, Switzerland. *International Journal of Educational Management.*

Ichinose, D., Yamamoto, M. & Yochida, Y. (2013). Productive efficiency of public and private solid waste logistics and its implications for waste management policy. *International Association of Traffic and Safety Sciences Research*, *36*(2), 96-105.

International Public Sector Accounting Standards Board (2013). *The Conceptual Framework for General Purpose Financial Reporting by Public Sector Entities* (Technical Report of January 2013). New York: International Federation of Accountants.

Ishizaka, A., Balkenborg, D. & Kaplan, T. (2011). Does AHP help us make a choice? An experimental evaluation. *Journal of the Operational Research Society*, *62*(10), 1801-1812.

Ishizaka, A. & Labib, A. (2011). Review of the main developments in the Analytic Hierarchy Process. *Expert Systems with Applications*, *38*(11), 14336-14345.

Ishizaka, A. & Labib, A. (2009). Analytic Hierarchy Process and Expert Choice: Benefits and limitations. *OR Insight*, *22*(4), 201-220.

Ishizaka, A. & Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. Chichester: John Wiley & Sons.

Jablonsky, J. (2007). Measuring the efficiency of production units by AHP models. *Mathematical and Computer Modelling*, *46*(7-8), 1091-1098.

Johnes, J. (2004). Efficiency measurement. In G. Johnes & J. Johnes (Eds.), *International Handbook on the Economics of Education* (pp. 613-742). Cheltenham: Edward Elgar Publishing.

Kelly, E., Shalloo, L., Geary, U., Kinsella, A. & Wallace, M. (2012). Application of data envelopment analysis to measure technical efficiency on a sample of Irish dairy farms. *Irish Journal of Agricultural and Food Research*, *51*(1), 63-77.

Kong, W.-H. & Fu, T.-T. (2012). Assessing the performance of business colleges in Taiwan using data envelopment analysis and student based value-added indicators. *Omega*, *40*(5), 541-549.

Korhonen, P., Tainio, R. & Wallenius, J. (2001). Value efficiency analysis of academic research. *European Journal of Operational Research*, *130*(1), 49-65.

Korpela, J., Lehmusvaara, A. & Nisonen, J. (2007). Warehouse operator selection by combining AHP and DEA methodologies. *International Journal of Production Economics*, *108*(1-2), 135-142.

Korpela, J. & Tuominen, M. (1996). A decision aid in warehouse site selection. *International Journal of Production Economics*, *45*(1-3), 169-180.

Lai, V., Wong, B. K. & Cheung, W. (2002). Group decision making in a multiple criteria environment: A case using the AHP in the software selection. *European Journal of Operational Research*, *137*(1), 134-144.

Lam, K. & Zhao, X. (1998). An application of quality function deployment to improve the quality of teaching. *International Journal of Quality Reliability Management*, *15*(4), 389-413.

Lan, C.-H., Chuang, L.-L. & Chen, Y.-F. (2009). Performance efficiency and resource allocation for fire department with the stochastic consideration. *International Journal of Technology, Policy and Management*, *9*(3), 296-315.

Lee, S.-H. (2010). Using fuzzy AHP to develop intellectual capital evaluation model for assessing their performane contribution to a university. *Expert Systems with Applications*, *37*(7), 4941-4947.

Lee, B. L. & Worthington, A. C. (2014). Technical efficiency of mainstream airlines and low-cost carriers: New evidence using bootstrap data envelopment analysis truncated regression. *Journal of Air Transport Management*, *38*, 15-20.

Lee, T., Yeo, G.-T. & Thai, V. V. (2014). Environmental efficiency analysis of port cities: Slacks-based measure data envelopment analysis approach. *Transport Policy*, *33*, 82-88.

Lin, H.-F. (2010). An application of fuzzy AHP for evaluating course website quality. *Computers & Education*, *54*(4), 877-888.

Lin, M.-I., Lee, Y.-D. & Ho, T.-N. (2011). Applying integrated DEA/AHP to evaluate the economic performance of local governments in China. *European Journal of Operational Research*, *209*(2), 129-140.

Liu, C. H., Lin, S. J. & Lewis, C. (2010). Evaluation of thermal power plant operational performance in Taiwan by data envelopment analysis. *Energy Policy*, *38*(4), 1049-1058.

Lootsma, F. A. (1989). Conflict resolution via pairwise comparison of concessions. *European Journal of Operational Research*, *40*(1), 109-116.

Lozano, S. & Villa, G. (2009). Multiobjective target setting in data envelopment analysis using AHP. *Computers & Operations Research*, *36*(2), 549-564.

Ma, D. & Zheng, X. (1991). *9/9-9/1 scale method of AHP*. Paper presented at the International Symposium on the Analytic Hierarchy Process, Pittsburgh, United States.

Macharis, C., Springael, J. De Brucker, K. & Verbeke, A. (2004). PROMETHEE and AHP: The design of operational synergies in multicriteria analysis. Strengthening PROMETHEE with ideas of AHP. *European Journal of Operational Research*, *153*(2), 307-317.

Malhotra, R., Malhotra, D. K. & Lafond, C. A. (2010). Benchmarking large U.S. retailers using a data envelopment analysis model. In K. D. Lawrence & G. Kleinman (Eds.), *Applications in Multicriteria Decision Making, Data Envelopment Analysis, and Finance (Applications of Management Science, Volume 14)* (pp. 217-235). Bingley: Emerald Group Publishing Limited.

Mallikarjun, S., Lewis, H. F. & Sexton, T. R. (2014). Operational performance of U.S. public rail transit and implications for public policy. *Socio-Economic Planning Sciences*, *48*(1), 74-88.

Manasakis, C., Apostolakis, A. & Datseris, G. (2013). Using data envelopment analysis to measure hotel efficiency in Crete. *International Journal of Contemporary Hospitality Management*, *25*(4), 510-535.

Marques, R. & Simões, P. (2009). How far are Portuguese prisons inefficient? A non-parametric approach. *MPRA Paper*. Munich: Munich Personal RePEc Archive.

Millet, I. & Saaty, T. (2000). On the relativity of relative measures: Accomodating both rank preservation and rank reversals in the AHP. *European Journal of Operational Research*, *121*(1), 205-212.

Monroe, M. C. & Adams, D. C. (2012). Increasing Response Rates to Web-Based Surveys. *Journal of Extension*, *50*(6), 6TOT7.

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

Muñiz, M., Paradi, J., Ruggiero, J. & Yang, Z. (2006). Evaluating alternative DEA models used to control for non-discretionary inputs. *Computers & Operations Research*, *33*(5), 1173-1183.

Nachiappan, S. & Ramanathan, R. (2008, February). *Robust decision making using Data Envelopment Analytic Hierarchy Process*. Paper presented at the Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Cambridge, United Kingdom.

Nguyen, T. P. T., Roca, E. & Sharma, P. (2014). How efficient is the banking system of Asia's next economic dragon? Evidence from rolling DEA windows. *Applied Economics*, *46*(22), 2665-2684.

Oggioni, G., Riccardi, R. & Toninelli, R. (2011). Eco-efficiency of the world cement industry: A data envelopment analysis. *Energy Policy*, *39*(5), 2842-2854.

Pak, R. J. (2013). Combining Importance-Performance Analysis with Analytic Hierarchy Process for Enhancing Satisfaction. *Journal of Advanced Management Science*, *1*(4), 368-371.

Pakkar, M. S. (2012). An Integrated Approach to the DEA and AHP Methodologies in Decision Making. In V. Charles & M. Kumar (Eds.), *Data Envelopment Analysis and Its Applications to Management* (pp. 136-149). Newcastle upon Tyne: Cambridge Scholars Publishing.

Pan, N.-F. (2008). Fuzzy AHP approach for selecting the suitable bridge construction method. *Automation in Construction*, *17*(8), 958-965.

Park, S. U. & Lesourd, J. B. (2000). The efficiency of conventional fuel power plants in South Korea: a comparison of parametric and non-parametric approaches. *International Journal of Production Economics*, *63*(1), 59-67.

Park, B.-Y. & Min, H. (2011). The selection of transshipment ports using a hybrid data envelopment analysis/analytic hierarchy process. *Journal of Transportation Management*, *22*(1), 47-64.

Picazo-Tadeo, A. J., Gómez-Limón, J. A. & Reig-Martínez, E. (2011). Assessing farming eco-efficiency: A Data Envelopment Analysis Approach. *Journal of Environmental Management*, *92*(4), 1154-1164.

Rabar, D. (2013). Assessment of regional efficiency in Croatia using data envelopment analysis. *Croatian Operational Research Review*, *4*(1), 76-88.

Ramanathan, R. (2006). Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process. *Computers & Operations Research*, *33*(5), 1289-1307.

Ramanathan, R. (2001). A note on the use of the analytic hierarchy process for environmental impact assessment. *Journal of Environmental Management*, *63*(1), 27-35.

Ramanathan, R. & Ganesh, L. S. (1994). Group preference aggregation methods employed in AHP: an evaluation and an intrinsic process for deriving members' weightages. *European Journal of Operational Research*, *79*(20), 249-265.

Ray, S. C. (1988). Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. *Socio-economic Planning Sciences*, *22*(4), 167-176.

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

Rogge, N. & De Jaeger, S. (2013). Measuring and explaining the cost efficiency of municipal solid waste collection and processing services. *Omega*, *41*(4), 653-664.

Roy, B. (1981). The optimisation problem formulation: Criticism and overstepping. *Journal of the Operational Research Society*, *32*(6), 427-436.

Roy, B. & Bouyssou, D. (1993). *Aide multicritère à la décision: Méthodes et cas*. Paris: Economica.

Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, *90*(3), 553-565.

Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, *111*(3), 461-469.

Ruggiero, J. (2004). Performance Evaluation in Education. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 323-346). Dordrecht: Springer.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234-281.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New-York: McGraw-Hill.

Saaty, T. L. (1991a). Response to Holder's comments on the analytic hierarchy process. *Journal of the Operational Research Society*, *42*(10), 909-914.

Saaty, T. L. (1991b). Response to Holder's comments on the analytic hierarchy process: Response to the Response to the Response. *Journal of the Operational Research Society*, *42*(10), 918-924.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83-98.

Saaty, T. L. (2012). *Decision making for leaders: The analytic hierarchy process for decisions in a complex world*. Pittsburgh: RWS Publications.

Saaty, T. L. & Vargas, L. G. (2005). The possibility of group welfare functions. *International of Information Technology & Decision Making*, *4*(2), 167-176.

Saaty, T. L. & Vargas, L. G. (2005, July). *Dispersion of group judgments*. Paper presented at the International Symposium on Analytic Hierarchy Process, Honolulu, United States.

Saen, R. F., Memariani, A. & Lotfi, F. H. (2005). Determining relative efficiency of slightly non-homogeneous decision making units by data envelopment analysis: A case study in IROST. *Applied Mathematics and Computation*, *165*(2), 313-328.

Salo, A. A. & Hämäläinen, R. P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis*, *6*(6), 309-319.

Seifert, L. M. & Zhu, J. (1998). Identifying Excesses and Deficits in Chinese Industrial Productivity (1953-1990): a Weighted Data Envelopment Analysis Approach. *Omega*, *26*(2), 279-296.

Sevkli, M., Lenny Koh, S. C., Zaim, S., Demirbag, M. & Tatoglu, E. (2007). An application of data envelopment analytic hierarchy process for supplier selection: A case study of BEKO in Turkey. *International Journal of Production Research*, *45*(9), 1973-2003.

Shang, J. & Sueyoshi, T. (1995). A unified framework for the selection of a flexible manufacturing system. *European Journal of Operational Research*, *85*(2), 297-315.

Sharma, S. (2008). Analyzing the technical and scale efficiency performance: a case study of cement firms in India. *Journal of Advances in Management Research*, *5*(2), 56-63.

Sheldon, G. (2003). The efficiency of public employment services: A non-parametric matching function analysis for Switzerland. *Journal of Productivity Analysis*, *20*(1), 49-70.

Shih, T.-H. & Fan, X. (2008). Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods*, *20*(3), 249-271.

Sinuany-Stern, Z., Mehrez, A. & Hadad, Y. (2000). An AHP/DEA methodology for ranking decision making units. *International Transactions in Operational Research*, *7*(2), 109-124.

Stevens, S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153-181.

Sólnes, J. (2003). Environmental quality indexing of large industrial development alternatives using AHP. *Environmental Impact Assessment Review*, *23*(3), 283-303.

Soteriou, A. C., Karahanna, E., Papanastasiou, C. & Diakourakis, M. S. (1998). Using DEA to evaluate the efficiency of secondary schools: the case of Cyprus. *International Journal of Educational Management*, *12*(2), 65-73.

Stewart, T. J. (1996). Relationships between Data Envelopment Analysis and Multicriteria Decision Analysis. *The Journal of the Operational Research Society*, *47*(5), 654-665.

Sueyoshi, T. & Goto, M. (2012). Data envelopment analysis for environmental assessment: Comparison between private and public ownership in petroleum industry. *European Journal of Operational Research*, *216*(3), 668-678.

Sun, S. (2002). Measuring the relative efficiency of police precincts using data envelopment analysis. *Socio-Economic Planning Sciences*, *36*(1), 51-71.

Suzuki, S., Nijkamp, P., Pels, E. & Rietveld, P. (2014). Comparative performance analysis of European airports by means of extended data envelopment analysis. *Journal of Advanced Transportation*, *48*(3), 185-202.

Tahriri, F, Osman, M. R., Ali, A., Yusuff, R. M. & Esfandiary, A. (2008). AHP approach for supplier evaluation and selection in a steel manufacturing company. *Journal of Industrial Engineering and Management*, *1*(2), 54-76.

Takamura, Y. & Tone, K. (2003). A comparative site evaluation study for relocating Japanese government agencies out of Tokyo. *Socio-Economic Planning Sciences*, *37*(2), 85-102.

Tam, M. C. Y. & Tummala, V. M. R. (2001). An application of the AHP in vendor selection of a telecommunications system. *Omega*, *29*(2), 171-182.

Tavares, G. (2002). A bibliography of Data Envelopment Analysis (1978-2001). *Rutcor Research Report No. 01-02*. Piscataway: Rutgers University.

Thanassoulis, E. (1996). Altering the Bias in Differential School Effectiveness Using Data Envelopment Analysis. *The Journal of the Operational Research Society, 47*(7), 882-894.

Thanassoulis, E., Portela, M. C. S. & Despic, O. (2008). Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. In H. O. Fried, C. A. K. Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency and Productivity Growth* (pp. 251-420). Oxford: Oxford University Press.

Thanassoulis, E., Portela Silva, M. C. A. & Graveney, M. (2014). Using DEA to estimate potential savings at GP units at medical specialty level. *Socio-Economic Planning Sciences*, *48*(1), 38-48.

Tongzon, J. (2001). Efficiency measurement of selected Australian and other international ports using data envelopment analysis. *Transportation Research Part A: Policy and Practice*, *35*(2), 107-122.

Vaidya, O. S. & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of Operational Research*, *169*(1), 1-29.

Vaz, C. B. & Camanho, A. S. (2012). Performance comparison of retailing stores using a Malmquist-type index. *Journal of the Operational Research Society*, *63*(5), 631-645.

Verma, A., Gavirneni, S. (2006). Measuring police efficiency in India: an application of data envelopment analysis. *Policing: An International Journal of Police Strategies & Management*, *29*(1), 125-145.

Vidal, L.-A., Sahin, E., Martelli, N., Berhoun, M. & Bonan, B. (2010). Applying AHP to select drugs to be produced by anticipation in a chemotherapy compounding unit. *Expert Systems with Applications*, *37*(2), 1528-1534.

von Solms, S. (2009, July). *Homogeneity and choice aggregation in the analytic hierarchy process*. Paper presented at the International Symposium on Analytic Hierarchy Process, Pittsburgh, United States.

Wang, J. & Zhang, Y. (2014). Service Quality Evaluation of Urban Parks Based on AHP Method and SD Software. *Journal of Applied Sciences*, *14*(3), 291-295.

Wang, Y.-M. & Chin, K.-S. (2009). A new data envelopment analysis method for priority determination and group decision making in the analytic hierarchy process. *European Journal of Operational Research*, *195*(1), 239-250.

Wang, Y.-M. & Elhag, T. M. S. (2006). An approach to avoiding rank reversal in AHP. *Decision Support Systems*, *42*(3), 1474-1480.

Wang, Y.-M. & Luo, Y. (2009). On rank reversal in decision analysis. *Mathematical and Computer Modelling*, *49*(5-6), 1221-1229.

Wang, Y.-M., Parkan, C. & Luo, Y. (2007). A linear programming method for generating the most favorable weights from a pairwise comparison matrix. *Computers and Operations Research*, *35*(1), 3918-3930.

Wong, J. K. W. & Li, H. (2006). Development of a conceptual model for the selection of intelligent building systems. *Building and Environment*, *41*(8), 1106-1123.

Wong, J. K. W. & Li, H. (2008). Application of the analytic hierarchy process (AHP) in multi-criteria analysis of the selection of intelligent building systems. *Building and Environment*, *43*(1), 108-125.

Yang, T. & Kuo, C. (2003). A hierarchical AHP/DEA methodology for the facilities layout design problem. *European Journal of Operational Research*, *147*(1), 128-136.

Yang, H. & Pollitt, M. (2009). Incorporating both undesirable outputs and uncontrollable variables into DEA: The performance of Chinese coal-fired power plants. *European Journal of Operational Research*, *197*(3), 1095-1105.

Yin, Q. (2013). An Analytical Hierarchy Process Model for the Evaluation of College Experimental Teaching Quality. *Journal of Technology and Science Education*, *3*(2), 59-65.

Yin, K., Pu, Y., Liu, Z. & Zhou, B. (2014). An AHP-based Approach for Banking Data Quality Evaluation. *Information Technology Journal*, *13*(8), 1523-1531.

Yoo, H (2003). A study on the efficiency evaluation of total quality management activities in Korean companies. *Total Quality Management*, *14*(1), 119-128.

Zhu, J. (2003). *Quantitative Models for Performance Evaluation and Benchmarking*. New York: Springer.

Appendix 1

Web-based survey, step 1

How to select the most suitable DEA model?

Within DEA, several alternative models allow for an environment adjustment. The goal of the current research is to use multi-criteria decision analysis methods in order to select the most suitable DEA model.

The current survey aims to identify the criteria which will be used in order to select the most suitable model. A preliminary list of four criteria, based on the literature, is provided below. You are being asked to complete this preliminary list (if you think that it has to be completed). You may also remove one or all criteria from this preliminary list if you consider them as inappropriate. Please take into account the two following assumptions:

- The “true” efficiency is unknown. As a results, the deviation between the “true” efficiency and the estimated efficiency cannot be considered as a criterion;
- Sufficient information about discretionary and non-discretionary variables is available in order to perform all DEA models. In particular, the influence direction of the non-discretionary variables is known and the non-discretionary variables are available in categorical and continuous terms.

Preliminary list of criteria

1. Understandability
The model is simple and transparent. It is easy to understand. As a result, it is easy to communicate.

2. Applicability
The model is easy to apply. It is easy to run (or perform). Results are easy to calculate.

3. Acceptability
The model is acceptable by the various stakeholders. For instance, the model used to benchmark hospitals should be acceptable by surgeons, physicians, nurses, patients, and so on. The intrinsic features of the model make it acceptable or not. A model could be easily understandable but not acceptable as its features do not make sense (to some stakeholders).

4.Cost-benefit
Running a model imposes costs. The benefits of a model should justify its costs. Assessing whether the benefits justify the costs is a matter of judgement. The costs of a model include the costs of collecting and processing the data. The benefits of a model include the added-value in terms of results provided by the model.

1. What are the criteria you would add to the list? Please cite and provide a working definition to the added criteria.

2. What are the criteria you would remove from the list? Please mention which one and provide some explanation.

3. If you were forced to remove one criterion from the list, which one would it be? (tick one of the grey boxes).

- ○ Understandability
- ○ Applicability
- ○ Acceptability
- ○ Cost-benefit

4. General comments?

Answers to the survey are treated anonymously.

THANK YOU FOR YOUR KIND COOPERATION!

Jean-Marc Huguenin
Swiss Graduate School of Public Administration
University of Lausanne
Switzerland

Terminé

Web-based survey, step 2

How to select the most suitable DEA model?

Within DEA, several alternative models allow for an environment adjustment. The goal of this research is to use multi-criteria decision analysis methods in order to select the most suitable model. In a previous first step, criteria have been identified. In the second step of the current survey, you are being asked, on a pairwise comparison basis, to express your preferences. The preferences concern:

- first the importance of each criterion with respect to the selection of the most suitable model;
- second the importance of each model with respect to each criterion.

The three criteria retained are:

1. Understandability
The model is simple and transparent. It is easy to understand. As a result, it is easy to communicate.

2. Applicability
The model is easy to apply. It is easy to run (or perform). Results are easy to calculate.

3. Acceptability
The model is acceptable by the various stakeholders. For instance, the model used to benchmark hospitals should be acceptable by surgeons, physicians, nurses, patients, and so on. The intrinsic features of the model make it acceptable or not. A model could be easily understandable but not acceptable as its features do not make sense (to some stakeholders).

Please take into account the following assumption: sufficient information about discretionary and environmental variables is available in order to perform all DEA models; in particular, the influence direction of the environmental variables is known and the environmental variables are available in categorical and continuous terms.

Four DEA models are considered:

1. Banker & Morey 1986a: BM–CATEGORICAL
In this model, DMUs are grouped into different categories according to the condition of the environment. For instance, we consider that category 1 contains DMUs facing the most detrimental environment and category 2 contains DMUs facing the most advantageous environment. In this example, DMUs in category 1 are only evaluated against schools within this group; DMUs in category 2 are evaluated with reference to DMUs in category 1 and 2. As a result, no DMU is compared to another with a favourable environment.
Reference if needed:
Banker, R. D. & Morey, R. C. (1986a). The Use of Categorical Variables in Data Envelopment Analysis. Management Science, 34(4), 1613-1627.

2. Banker & Morey 1986b: BM–CONTINUOUS
In this model, environmental variable(s) are directly included as continuous non-discretionary variables in the linear programming formulation.
Reference if needed:
Banker, R. D. & Morey, R. C. (1986b). Efficiency Analysis for Exogenously Fixed Inputs and Outputs. Operations Research, 32(12), 513-521.

3. Ray 1991: RAY–TWO-STAGE
This model introduces a two-stage procedure. In the first stage, a basic DEA model is performed using only discretionary variables. In the second stage, the efficiency scores (obtained from the first stage) are regressed upon environmental variables.
Reference if needed:
Ray, S. C. (1991). Resource-use efficiency in public schools: a study of Connecticut data. Management Science, 37(12), 1620-1628.

4. Yang & Paradi 2006: YP–I&O ADJUSTED
This model applies ex ante a handicapping measure based on the levels of the environmental variables. DMUs with a favourable environment are penalized by the handicapping measure. In such a case, inputs are adjusted with a higher handicap (i.e. they are augmented) and/or outputs are adjusted with a lower handicap (i.e. they are reduced). As a result, adjusted inputs have a higher value than original inputs and adjusted outputs have a lower value than original outputs.
Reference if needed:
Muñiz, M., Paradi, J., Ruggiero, J. & Yang, Z. (2006). Evaluating alternative DEA models used to control for non-discretionary inputs. Computers & Operations Research, 33 (5), 1173-1183.

Suiv.

How to select the most suitable DEA model?

To select the most suitable model, which criterion is most important? (pairwise comparisons of criteria)

1. Understandability versus Applicability

Please select one option

In order to select the most suitable DEA model...

2. Understandability versus Acceptability

Please select one option

In order to select the most suitable DEA model...

3. Acceptability versus Applicability

Please select one option

In order to select the most suitable DEA model...

Préc. Suiv.

How to select the most suitable DEA model?

To select the most suitable model, which criterion is most important? (pairwise comparisons of criteria)

1. Understandability versus Applicability

Please select one option

In order to select the most suitable DEA model...

Understandability is of equal importance to Applicability
Understandability is weakly or slightly more important than Applicability
Understandability is moderately more important than Applicability
Understandability is moderately (plus) more important than Applicability
Understandability is strongly more important than Applicability
Understandability is strongly (plus) more important than Applicability
Understandability is very strongly more important than Applicability
Understandability is very very strongly more important than Applicability
Understandability is extremely more important than Applicability
Applicability is weakly or slightly more important than Understandability
Applicability is moderately more important than Understandability
Applicability is moderately (plus) more important than Understandability
Applicability is strongly more important than Understandability
Applicability is strongly (plus) more important than Understandability
Applicability is very strongly more important than Understandability
Applicability is very very strongly more important than Understandability
Applicability is extremely more important than Understandability

2. Understandability versus Acceptability

In order to select the most suitable DEA model...

3. Acceptability versus Applicability

In order to select the most suitable DEA model...

Préc. Suiv.

How to select the most suitable DEA model?

To fulfil the criterion of UNDERSTANDABILITY, what model is preferable? (pairwise comparisons of models)

4. BM-CATEGORICAL versus BM-CONTINUOUS on UNDERSTANDABILITY

Please select one option

In order to fulfill the criterion of UNDERSTANDABILITY...

5. BM-CATEGORICAL versus RAY-TWO-STAGE on UNDERSTANDABILITY

Please select one option

In order to fulfill the criterion of UNDERSTANDABILITY...

6. BM-CONTINUOUS versus RAY-TWO-STAGE on UNDERSTANDABILITY

Please select one option

In order to fulfill the criterion of UNDERSTANDABILITY...

7. BM-CATEGORICAL versus YP-I&O ADJUSTED on UNDERSTANDABILITY

Please select one option

In order to fulfill the criterion of UNDERSTANDABILITY...

8. BM-CONTINUOUS versus YP-I&O ADJUSTED on UNDERSTANDABILITY

Please select one option

In order to fulfill the criterion of UNDERSTANDABILITY...

9. RAY-TWO-STAGE versus YP-I&O ADJUSTED on UNDERSTANDABILITY

Please select one option

In order to fulfill the criterion of UNDERSTANDABILITY...

Préc. Suiv.

ow to select the most suitable DEA model?

To fulfil the criterion of UNDERSTANDABILITY, what model is preferable? (pairwise comparisons of models)

4. BM-CATEGORICAL versus BM-CONTINUOUS on UNDERSTANDABILITY

Please select one option

In order to fulfill the criterion of UNDERSTANDABILITY...

BM–CATEGORICAL is of equal preference to BM–CONTINUOUS
BM–CATEGORICAL is weakly or slightly preferred to BM–CONTINUOUS
BM–CATEGORICAL is moderately preferred to BM–CONTINUOUS
BM–CATEGORICAL is moderately (plus) preferred to BM–CONTINUOUS
BM–CATEGORICAL is strongly preferred to BM–CONTINUOUS
BM–CATEGORICAL is strongly (plus) preferred to BM–CONTINUOUS
BM–CATEGORICAL is very strongly preferred to BM–CONTINUOUS
BM–CATEGORICAL is very very strongly preferred to BM–CONTINUOUS
BM–CATEGORICAL is extremely preferred to BM–CONTINUOUS
BM–CONTINUOUS is weakly or slightly preferred to BM–CATEGORICAL
BM–CONTINUOUS is moderately preferred to BM–CATEGORICAL
BM–CONTINUOUS is moderately (plus) preferred to BM–CATEGORICAL
BM–CONTINUOUS is strongly preferred to BM–CATEGORICAL
BM–CONTINUOUS is strongly (plus) preferred to BM–CATEGORICAL
BM–CONTINUOUS is very strongly preferred to BM–CATEGORICAL
BM–CONTINUOUS is very very strongly preferred to BM–CATEGORICAL
BM–CONTINUOUS is extremely preferred to BM–CATEGORICAL

5. BM-CATEGORICAL versus RAY-TWO-STAGE on UNDERSTAND

In order to fulfill the criterion of UNDERSTANDABILITY...

6. BM-CONTINUOUS versus RAY-TWO-STAGE on UNDERSTANDA

In order to fulfill the criterion of UNDERSTANDABILITY...

7. BM-CATEGORICAL versus YP-I&O ADJUSTED on UNDERSTANDABILITY

Please select one option

In order to fulfill the criterion of UNDERSTANDABILITY...

8. BM-CONTINUOUS versus YP-I&O ADJUSTED on UNDERSTANDABILITY

How to select the most suitable DEA model?

To fulfil the criterion of APPLICABILITY, what model is preferable? (pairwise comparisons of models)

10. BM-CATEGORICAL versus BM-CONTINUOUS on APPLICABILITY

	Please select one option
In order to fulfill the criterion of APPLICABILITY...	

11. BM-CATEGORICAL versus RAY-TWO-STAGE on APPLICABILITY

	Please select one option
In order to fulfill the criterion of APPLICABILITY...	

12. BM-CONTINUOUS versus RAY-TWO-STAGE on APPLICABILITY

	Please select one option
In order to fulfill the criterion of APPLICABILITY...	

13. BM-CATEGORICAL versus YP-I&O ADJUSTED on APPLICABILITY

	Please select one option
In order to fulfill the criterion of APPLICABILITY...	

14. BM-CONTINUOUS versus YP-I&O ADJUSTED on APPLICABILITY

	Please select one option
In order to fulfill the criterion of APPLICABILITY...	

15. RAY-TWO-STAGE versus YP-I&O ADJUSTED on APPLICABILITY

	Please select one option
In order to fulfill the criterion of APPLICABILITY...	

Préc. Suiv.

How to select the most suitable DEA model?

To fulfil the criterion of ACCEPTABILITY, what model is preferable? (pairwise comparisons of models)

16. BM-CATEGORICAL versus BM-CONTINUOUS on ACCEPTABILITY

Please select one option

In order to fulfill the criterion of ACCEPTABILITY...

17. BM-CATEGORICAL versus RAY-TWO-STAGE on ACCEPTABILITY

Please select one option

In order to fulfill the criterion of ACCEPTABILITY...

18. BM-CONTINUOUS versus RAY-TWO-STAGE on ACCEPTABILITY

Please select one option

In order to fulfill the criterion of ACCEPTABILITY...

19. BM-CATEGORICAL versus YP-I&O ADJUSTED on ACCEPTABILITY

Please select one option

In order to fulfill the criterion of ACCEPTABILITY...

20. BM-CONTINUOUS versus YP-I&O ADJUSTED on ACCEPTABILITY

Please select one option

In order to fulfill the criterion of ACCEPTABILITY...

21. RAY-TWO-STAGE versus YP-I&O ADJUSTED on ACCEPTABILITY

Please select one option

In order to fulfill the criterion of ACCEPTABILITY...

22. General comments?

Answers to the survey are treated anonymously.

THANK YOU FOR YOUR KIND COOPERATION!

Jean-Marc Huguenin
Swiss Graduate School of Public Administration
University of Lausanne Switzerland

Préc. Terminé

Bibliography

Abbott, M. & Doucouliagos, C. (2000). Technical and scale efficiency of vocational education and training institutions: The case of the New Zealand polytechnics. *New Zealand Economic Paper*, *34*(1), 1-23.

Abbott, M. & Doucouliagos, C. (2003). The efficiency of Australian universities: a data envelopment analysis. *Economics of Education Review*, *22*(1), 89-97.

Adler, N. & Berechman, J. (2001). Measuring airport quality from the airlines' viewpoint: an application of data envelopment analysis. *Transport Policy*, *8*, 171-181.

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. & Salerno, C. (2007). Assessing the Cost Efficiency of Italian Universities. *Education Economics*, *15*(4), 455-471.

Agasisti, T., Bonomi, F. & Sibiano, P. (2014). Measuring the "managerial" efficiency of public schools: a case study in Italy. *International Journal of Educational Management*, *28*(2), 120-140.

Ahn, T., Arnold, V., Charnes, A. & Cooper W. W. (1989). DEA and ratio efficiency analyses for public institutions of higher learning in Texas. *Research in Governmental and Nonprofit Accounting*, *5*, 165-185.

Ahn, T. & Seiford, L. M. (1993). Sensitivity of data envelopment analysis to models and variable sets in a hypothesis test setting: the efficiency of university operations. In Y. Ijiri (Eds.), *Creative and Innovative Approaches to the Science of Management* (pp. 191-208). Westport: Quorum Books.

Al-Harbi, K. M. (2001). Application of the AHP in project management. *International Journal of Project Management*, *19*(1), 19-27.

Al Khalil, M. I. (2002). Selecting the appropriate project delivery method using AHP. *International Journal of Project Management*, *20*(6), 469-474.

Al-Najjar, S. M. & Al-Jaybajy, M. A. (2012). Application of Data Envelopment Analysis to Measure the Technical Efficiency of Oil Refineries: A Case Study. *International Journal of Business Administration*, *3*(5), 64-77.

Al-Rawashdeh, T., Al'azzeh, F. & Al-Qatawneh, S. (2014). Evaluation of ERP Systems Quality Model Using Analytic Hierarchy Process (AHP) Technique. *Journal of Software Engineering and Applications*, *7*(4), 225-232.

Alexander, W. R. J. & Jaforullah, M. (2004). Explaining efficiency differences of New Zealand secondary schools. *Economics Discussion Papers No. 0403*. Dunedin: University of Otago.

Alexander, W. R. J., Haug, A. A. & Jaforullah, M. (2010). A tow-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools. *Journal of Productivity Analysis*, *34*(2), 99-110.

Allen, R., Athanassopoulos, A., Dyson, R. G. & Thanassoulis, E. (1997). Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions. *Annals of Operations Research*, *73*(1), 13-34.

Andersen, P. & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, *39*(10), 1261-1264.

Anderson, T. R., Daim, T. U. & Lavoie, F. F. (2007). Measuring the efficiency of university technology transfer. *Technovation*, *27*(5), 306-318.

Andersson, C., Manson, J. & Sund, K. (2014). Technical efficiency of Swedish employment offices. *Socio-Economic Planning Sciences*, *48*(1), 57-64.

Andrews, R., Boyne, G. A. & Enticott, G. (2006). Performance failure in the public sector. Misfortune or mismanagement? *Public Management Review*, *8*(2), 273-296.

Arbel, A. & Orgler, Y. E. (1990). An application of the AHP to bank strategic planning: The mergers and acquisitions process. *European Journal of Operational Research*, *48*(1), 27-37.

Arcelus, F. J. & Coleman, D. F. (1997). An efficiency review of university departments. *International Journal of Systems Science*, *28*(7), 721-729.

Audit Commission (2010). The practice of performance indicators. *Management paper.*

Avkiran, N. K. (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Economic Planning Science*, *35*(1), 57-80.

Azadeh, A., Ghaderi, S. F. & Izadbakhsh, H. (2008). Integration of DEA and AHP with computer simulation for railway system improvement and optimization. *Applied Mathematics and Computation*, *195*(2), 775-785.

Azizi, K. H., Lofti, F. H., Saati, S. & Vahidi, A. R. (2007). Ranking the Electricity Producer Companies in View of Manpower Efficiency by DEA. *Applied Mathematical Sciences*, *1*(16), 761-768.

Badillo, P.-Y. & Paradi, J. C. (1999). *La méthode DEA: analyse des performances*. Paris: HERMES Science Publications.

Bana e Costa, C., De Corte, J.-M. & Vansnick, J.-C. (2005). On the mathematical foundation of MACBETH. In J. Figueira, S. Greco & M.

Ehrogott (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys* (pp. 409-437). New York: Springer.

Bana e Costa, C. & De Corte, J.-M. & Vansnick, J.-C. (2003). MACBETH. *OR Working Paper No. 03.56*. London: London School of Economics and Political Science.

Bana e Costa, C. & Vansnick, J.-C. (1999). The MACBETH approach: Basic ideas, software, and an application. In N. Meskens & M. Roubens (Eds.), *Advances in Decision Analysis, Mathematical Modelling: Theory and Application* (Vol. 4, pp. 131-157). Dordrecht: Kluwer Academic Publishers.

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.

Banker, R. D. & Morey, R. C. (1986a). The Use of Categorical Variables in Data Envelopment Analysis. *Management Science*, *34*(4), 1613-1627.

Banker, R. D. & Morey, R. C. (1986b). Efficiency Analysis for Exogenously Fixed Inputs and Outputs. *Operations Research*, *32*(12), 513-521.

Banker, R. D., Zheng, Z. & Natarajan R. (2010). DEA-based hypothesis tests for comparing two groups of decision making units. *European Journal of Operational Science*, *206*(1), 231-238.

Barnett, R. R., Glass, J. C., Snowdon, R. I. & Stringer, K. S. (2002). Size, Performance and Effectiveness: Cost-Constrained Measures of Best-Practice Performance and Secondary-School Size. *Education Economics*, *10*(3), 291-311.

Barnum, D. T. & Gleason, J. M. (2008). Bias and precision in the DEA two-stage method. *Applied Economics*, *40*(18), 2305-2311.

Barr, R. (2003). DEA Software Tools and Technology: A State-of-the-Art survey. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 539-566). Boston: Kluwer Academic Publishers.

Barros, C. P. (2004). Measuring performance in defence-sector companies in a small NATO member country. *Journal of Economic Studies*, *31*(2), 112-128.

Barros, C. A. P. & Santos, C. A. (2006). The measurement of efficiency in Portuguese hotels using data envelopment analysis. *Journal of Hospitality & Tourism Research*, *30*(3), 378-400.

Barrow, M. M. (1991). Measuring local education authority performance: A frontier approach. *Economics of Education Review*, *10*(1), 19-27.

Basak, I. (1988). When to Combine Group Judgments and When Not to in the Analytic Hierarchy Process: A New Method. *Mathematical and Computer Modeling*, *10*(6), 395-404.

Bayazit, O. (2005). Use of AHP in decision-making for flexible manufacturing systems. *Journal of Manufacturing Technology Management*, *16*(7), 808-819.

Beasley, J. E. (1990). Comparing university departments. *Omega*, *18*(2), 171-183.

Beasley, J. E. (1995). Determining the teaching and research efficiencies. *Journal of the Operational Research Society*, *46*(4), 441-452.

Behera, S. K., Faroquie, J. A. & Dash, A. P. (2011). Productivity change of coal-fired thermal power plants in India: a Malmquist index approach. *IMA Journal of Management Mathematics*, *22*(4), 387-400.

Belton, V. & Gear, A. (1983). On a shortcoming of Saaty's method of analytical hierarchies. *Omega*, *11*(3), 228-230.

Bertolini, M., Braglia, M. & Carmignani, G. (2006). *International Journal of Project Management*, *24*(5), 422-430.

Bessent, A. M. & Bessent, E. W. (1979). *Determining the Comparative Efficiency of Schools through Data Envelopment Analysis* (Research Report CCS 361). Austin: Center for Cybernetic Studies, University of Texas.

Bessent, A. M. & Bessent, E. W. (1980). Determining the Comparative Efficiency of Schools through Data Envelopment Analysis. *Educational Administration Quarterly*, *16*(2), 57-75.

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

Bifulco, R. & Bretschneider, S. (2001). Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review*, *20*(5), 417-429.

Binam, J. N., Sylla, K., Diarra, I. & Nyambi, G. (2003). Factors affecting technical efficiency among coffee farmers in Côte d'Ivoire: evidence from the centre west region. *R&D Management*, *15*(1), 66-76.

Blank, J. L. T., van Hulst, B. L., Koot, P. M. & van der Aa, R. (2012). Benchmarking overhead in education: a theoretical and empirical approach. *Benchmarking: An International Journal*, *19*(2), 239-254.

Bogethoft, P. & Otto, L. (2010). *Benchmarking with DEA, SFA, and* R. New York: Springer.

Borge, L.-E. & Naper L. R. (2006). Efficiency Potential and Efficiency Variation in Norwegian Lower Secondary Schools. *FinanzArchiv / Public Finance Analysis*, *62*(2), 221-249.

Borges, M. R., Nektarios, M. & Barros, C. P. (2008). Analysing The Efficiency Of The Greek Life Insurance Industry. *European Research Studies*, *11*(3), 35-52.

Bouckaert, G. & Halligan, J. (2008). *Managing Performance: International Comparisons*. London: Routledge.

Bowerman, B. L. & O'Connell, R. T. (1990). *Linear Statistical Models: An Applied Approach*. Boston: Duxbury Press.

Bradley, S., Johnes, J. & Little, A. (2010). Measurement and determinants of efficiency and productivity in the further education sector in England. *Bulletin of Economic Research*, *62*(1), 1-30.

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.

Brockett, P. L., Charnes, A., Cooper, W. W., Huang, Z. M. & Sun, D. B. (1991). Data Transformations in DEA Cone Ratio Envelopment Approaches for Monitoring Bank Performance. *Journal of Operational Research*, *98*(2), 250-268.

Burgat, P. & Jeanrenaud, C. (1990). Mesure de l'efficacité productive et de l'efficacité-coût : le cas des déchets ménagers en Suisse. *Working Paper No. 9002*. Neuchâtel: Institut de recherches économiques et régionales, Université de Neuchâtel.

Burgat, P. & Jeanrenaud, C. (1992). Measurement of productive efficiency: the example of household waste in Switzerland. *Working Paper No. 9209*. Neuchâtel : Institut de recherches économiques et régionales, Université de Neuchâtel.

Burgat, P. & Jeanrenaud, C. (1994). Technical Efficiency and Institutional Variables. *Swiss Journal of Economics and Statistics*, *130*(4), 709-717.

Burgess, J. F. & Wilson, P. W. (1998). Variation in inefficiency among US hospitals. *Canadian Journal of Operational Research and Information Processing (INFOR)*, *36*(3), 84-102.

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

Butler, T. W. & Johnson, W. W. (1997). Efficiency Evaluation of Michigan Prisons Using Data Envelopment Analysis. *Criminal Justice Review*, *22*(1), 1-15.

Cai, Y. & Wu, W. (2001). Synthetic financial evaluation by a method of combining DEA with AHP. *International Transactions in Operational Research*, *8*(5), 603-609.

Carrington, R., Puthucheary, N., Rose, D. & Yaisawarng, S. (1997). Performance measurement in government service provision: the case of police services in New South Wales. *Journal of Productivity Analysis*, *8*(4), 415-430.

Caulfield, B., Bailey, D. & Mullarkey, S. (2013). Using data envelopment analysis as a public transport project appraisal tool. *Transport Policy*, *29*, 74-85.

Celik, M., Kandakoglu, A. & Er, I. D. (2009). Structuring fuzzy integrated multi-stages evaluation model on academic personnel recruitment in MET institutions. *Expert Systems with Applications*, *36*(3), 6918-6927.

Chakraborty, K. (1998). *Essays on Scale Economies and Efficiency in Public Education* (Unpublished doctoral dissertation). Utah State University, USA.

Chakraborty, K., Biswas, B. & Lewis, W. C. (2001). Measurement of technical efficiency in public education: a stochastic and nonstochastic production approach. *Southern Economic Journal*, *67*(4), 889-905.

Chalos, P. (1997). An examination of budgetary inefficiency in education using data envelopment analysis. *Financial Accountability and 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*, *14*(2), 143-160

Charnes, A., Clarke, C., Cooper, W. W. & Golany, B. (1984). A development study of DEA in measuring the effect of maintenance units in the U.S. Air Force. *Annals of Operations Research*, *2*(1), 95-112.

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., Huang, Z. M. & Sun, D. B. (1990). Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. *Journal of Econometrics*, *46*(1-2), 73-91.

Charnes, A, Cooper, W. W. & Rhodes E. L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429-444.

Charnes, A., Cooper, W. W. & Rhodes, E. L. (1981). Evaluating program and managerial efficiency: An application of DEA to program follow through. *Management Science*, *27*(6), 668-697.

Charnes, A., Haag, S., Jaska, P. & Semple, J. (1992). Sensitivity of efficiency calculations in the additive model of data envelopment analysis. *International Journal of System Sciences*, *23*(5), 789-798.

Charnes, A. & Neralic, L. (1992). Sensitivity analysis in data envelopment analysis. *Glasnik Matematicki*, *27*(47), 191-201.

Chen, J.-K. & Chen, I.-S. (2011). Inno-Qual efficiency of higher education: Empirical testing using data envelopment analysis. *Expert Systems with Applications*, *38*(3), 1823-1834.

Chen, Y.-C. L. (2006). An Analytic Hierarchy Framework for Evaluation Balanced Scorecards Healthcare Organizations. *Canadian Journal of Administrative Sciences*, *23*(2), 85-104.

Cheng, C.-H., Yang, K.-L. & Hwang, C.-H. (1999). Evaluating attack helicopters by AHP based on linguistic variable weight. *European Journal of Operational Research*, *116*(2), 423-435.

Chwolka, A. & Raith, M. G. (2000). Supporting Group Decisions with the AHP: Harmonization vs Aggregation of Preferences. In G. Wanka (Eds.), *Decision Theory and Optimization in Theory and Practice* (pp. 17-32). Aachen: Shaker Verlag.

Coelli, T. J. (1996). A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program. *CEPA Working Paper 96/08*. Brisbane: Centre for Efficiency and Productivity Analysis, University of Queensland.

Coelli, T. J. (1998). A Multi-stage Methodology for the Solution of Oriented DEA Models. *Operations Research Letters*, *23*(3-5), 143-149.

Coelli, T. J. & Perelman, S. (1996). Efficiency measurement, Multiple-output Technologies and Distance Functions: With Application to European Railways. *CREPP Working Paper 96/05*. Liège: Centre de Recherche en Economie Publique et de la Population, University of Liège.

Coelli, T. J. & Perelman, S. (1999). A Comparison of Parametric and Non-Parametric Distance Functions: With Application to European Railways. *European Journal of Operational Research*, *117*(2), 326-339.

Coelli, T. J. & Prasada Rao, D. S. (2005). Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980-2000. *Agricultural Economics*, *32*(s1), 115-134.

Coelli, T. J., Prasada Rao, D. S., O'Donnel, C. J. & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York: Springer.

Colbert, A., Levary, R. R. & Shaner, M. C. (2000). Determining the relative efficiency of MBA programs using DEA. *European Journal of Operational Research*, *125*(3), 656-669.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, M. D & York, R. L. (1966). *Equality of Educational Opportunity*. Washington, D.C.: US Department of Health, Education, and Welfare, Government Printing Office.

Cook, C., Heath, F. & Thompson, R. L. (2000). A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement*, *60*(6), 821-836.

Cook, W. D. & Seiford, L. M. (2008). Data envelopment analysis (DEA) – Thirty years on. *European Journal of Operational Research, 192*(1), 1-17.

Cook, W. D. & Zhu, J. (2008). *Data Envelopment Analysis: Modeling Operational Processes and Measuring Productivity*. Seattle: CreateSpace.

Cooper, W. W., Ruiz, J. L. & Sirvent, I. (2011). Choices and Uses of DEA Weights. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 93-126). New York: Springer.

Cooper, W. W., Seiford, L. M. & Tone, K. (2006). *Introduction to Data Envelopment Analysis and Its Uses*. New York: Springer.

Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data Envelopment Analysis: A comprehensive Text with Models, Applications, References and DEA-Solver Software*. New York: Springer.

Cooper, W. W., Seiford, L. M. & Zhu, J. (2004). Data Envelopment Analysis: History, Models and Interpretations. In. W. W. Cooper, L. M. Seiford and J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 1-39). Boston: Kluwer Academic Publishers.

Cooper, W. W., Seiford, L. M. & Zhu, J. (2011). *Handbook on Data Envelopment Analysis* (2nd ed.). New York: Springer.

Cordero, J. M., Pedraja, F. & Santín, D. (2009). Alternative approaches to include exogenous variables in DEA measures: A comparison using Monte Carlo. *Computers & Operations Research*, *36*(10), 2699-2706.

Cordero-Ferrara, J. M., Pedraja-Chaparro, F. & Salinas-Jiménez, J. (2008). Measuring efficiency in education: an analysis of different approaches for incorporating non-discretionary inputs. *Applied Economics*, *40*(10), 1323-1339.

Cour des comptes (2013). *Gérer les enseignants autrement* (Rapport public thématique). Paris: Cour des comptes.

Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, *64*(8), 672-683.

Currie, J. & Goodman, J. (2010). Parental Socioeconomic Status, Child Health, and Human Capital. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of Education* (pp. 156-162). Oxford: Elsevier.

Dantzig, G. B. (1951). Maximization of a linear function of variables subject to linear inequalities. In T. C. Koopmans (Ed.), *Activity Analysis of Production and Allocation* (pp. 339-347). New York: John Wiley & Sons.

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

De Bruijn, H. (2002). *Managing Performance in the Public Sector*. London: Routledge.

De Witte, K. & Moesen, W. (2010). Sizing the government. *Public Choice, 145*(1), 39-55.

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.

De Witte, K. & Van Klaveren, C. (2014). How are teachers teaching? A nonparametric approach. *Education Economics, 22*(1), 3-23.

Demeuse, M., Frandji, D., Greger, D. & Rochex, J.-Y. (2008). *Les politiques d'éducation prioritaire en Europe : Conceptions, mises en œuvre, débats*. Lyon: Institut national de recherche pédagogique.

Demeuse, M., Frandji, D., Greger, D. & Rochex, J.-Y. (2012). *Education Policies and Inequalities in Europe*. New York: Palgrage Macmillan.

Demeuse, M. & Friant, N. (2012). Evaluer les politiques d'éducation prioritaire en Europe : un défi méthodologique. *Revue Suisse des Sciences de l'Education*, *34*(1), 39-55.

Demir, I. & Depren, Ö. (2010). Assessing Turkey's secondary schools performance by different region in 2006. *Procedia – Social and Behavioral Sciences*, *2*(1), 2305-2309.

Denaux, Z. S., Lipscomb, C. A. & Plumly, L. W. (2011). Assessing the technical efficiency of public high schools in the state of Georgia. *Review of Business Research*, *11*(5), 46-57.

Diagne, D. (2006). Mesure de l'efficience technique dans le secteur de l'éducation : une application de la méthode DEA. *Revue suisse d'économie et de statistique*, *142*(2), 231-262.

Diamond, A. & Medewitz, J. N. (1990). Use of data envelopment analysis in an evaluation of the efficiency of the DEEP program for economic education. *Journal of Economic Education*, *21*(3), 337-354.

Dorsch, J. J. & Yasin, M. M. (1998). A framework for benchmarking in the public sector. *International Journal of Public Sector Management*, *11*(2/3), 91-115.

D'Souza, S. (2013). Parent Feedback: A Critical Link in Improving Our Schools. *Public Administration Review*, *73*(3), 413-414.

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

Dusansky, R. & Wilson, P. W. (1994). Measuring efficiency in the care of developmentally disabled. *Review of Economics and Statistics, 76*(2), 340-345.

Dusansky, R. & Wilson, P. W. (1995). On the relative efficiency of alternative modesof producing public sector output: The case of developmentally disabled. *European Journal of Operational Research*, *80*(3), 608-628.

Dyer, J. (1990). Remarks on the analytic hierarchy process. *Management Science*, *36*(3), 249-258.

Dyson, R. G. & Thanassoulis, E. (1988). Reducing weight flexibility in DEA. *Journal of the Operational Research Society*, *39*(6), 563-576.

Eling, M. & Huang, W. (2010). An Efficiency Comparison of Non-life Insurance Industry in the BRIC Countries. *Working Papers on Risk Management and Insurance No. 94*. St.Gallen: University of St.Gallen.

Emrouznejad, A., Parker, B. R. & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, *42*(3), 151-157.

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

Ertay, T., Ruan, D. & Tuzkaya, U. R. (2006). Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems. *Information Sciences*, *176*(3), 237-262.

Essid, H., Ouellette, P. & Vigeant, S. (2010). Measuring efficiency of Tunisian schools in the presence of quasi-fixed inputs: A bootstrap data envelopment analysis approach. *Economics of Education Review*, *29*(4), 589-596.

Essid, H., Ouellette, P. & Vigeant, S. (2013). Small is not that beautiful after all: measuring the scale efficiency of Tunisian high schools using a DEA-bootstrap method. *Applied Economics*, *45*(9), 1109-1120.

Estelle, S. M., Johnson, A. L. & Ruggiero, J. (2010). Three-stage DEA models for incorporating exogenous inputs. *Computers & Operations Research*, *37*(6), 1087-1090.

Fandel, G. (2007). On the performance of universities in North Rhine-Westphalia, Germany: Government's redistribution of funds using DEA efficiency measures. *European Journal of Operational Research*, *176*(1), 521-533.

Färe, R., Grosskopf, S. & Margaritis, D. (2011). Malmquist Productivity Indexes and DEA. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 127-150). New York: Springer.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of Royal Statistical Society*, *120*(3), 253-281.

Farsi, M. & Filippini, M. (2005). A benchmarking analysis of electricity distribution utilities in Switzerland. *CEPE Working Paper No. 43*. Zurich: Center for Energy Policy and Economics, Swiss Federal Institute of Technology.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of Royal Statistical Society*, *120*(3), 253-281.

Favero, N. & Meier, K. J. (2013). Evaluating Urban Public Schools: Parents, Teachers, and State Assessments. *Public Administration Review*, *73*(3), 401-412.

Feng, Y. J., Lu, H. & Bi, K. (2004). An AHP/DEA method for measurement of the efficiency of R&D management activities in universities. *International Transactions in Operational Research*, *11*(2), 181-191.

Ferro-Luzzi, G., Flueckiger, Y., Ramirez, J. & Vassiliev, A. (2006). Unemployment and employment offices' efficiency: what can be done? *Socio-economic Planning Sciences*, *40*(3), 169-186.

Fethi, M. D., Jackson, P. M. & Weyman-Jones, T. G. (2000). Measuring the efficiency of European airlines: An application of data envelopment analysis and Tobit Analysis. *EPRU Discussion Papers*. Leicester: University of Leicester.

Forgionne, G. A., Kohli, R. & Jennings, D. (2002). An AHP analysis of quality in AI and DSS journals. *Omega*, *30*(3), 171-183.

Forman, E. H. & Gass, S. I. (2001). The analytical hierarchy process – an exposition. *Operations Research*, *49*(4), 469-487.

Frandji, D. (2008). Pour une comparaison des politiques d'éducation prioritaire en Europe. In M. Demeuse, D. Frandji, D. Greger & J.-Y. Rochex (Eds.), *Les politiques d'éducation prioritaire en Europe: Conceptions, mises en oeuvre, débats* (pp. 9-34). Lyon : Institut national de Recherche pédagogique.

Frei, F. X. & Harker, P. T. (1999). Measuring aggregate process performance using AHP. *European Journal of Operational Research*, *116*(2), 436-442.

Friebelová, J., Friebel, L. & Marková, K. (2009). Using DEA models for evaluation of firestation efficiency. *Bohemiae Meridionalis*, *12*(2), 71-78.

Fried, H. O., Lovell, C. A. K., Schmidt, S. S. & Yaisawarng, S. (2002). Accounting for environmental effects and statistical noise in data envelopment analysis. *Journal of Productivity Analysis*, *17*(1/2), 157-174.

Fried, H. O., Schmidt, S. S. & Yaisawarng, S. (1999). Incoporating the Operating Environment Into a Nonparametric Measure of Technical Efficiency. *Journal of Productivity Analysis*, *12*(3), 249-267.

Fuentes, A. (2011). Raising Education Outcomes in Switzerland. *OECD Economics Department Working Papers No. 838*. Paris: OECD Publishing.

Garrett, W. A. & Kwak, N. K. (2011). Performance comparisons of Missouri public schools using data envelopment analysis. In K. D. Lawrence & G. Kleinman (Eds.), *Applications in Multicriteria Decision Making, Data Envelopment Analysis, and Finance (Applications of Management Science, Volume 14)* (pp. 135-155). Bingley: Emerald Group Publishing Limited.

Gasparini, C. & Ramos, F. (2003). Efetividade e eficiencia no ensino medio brasileiro. *Economia Aplicada, 7*(2), 389-411.

Gautam, S., Hicks, L., Johnson, T. & Mishra, B. (2013). Measuring the Performance of Critical Access Hospitals in Missouri using Data Envelopment Analysis. *The Journal of Rural Health*, *29*(2), 150-158.

Giannoulis, C. & Ishizaka, A. (2010). A Web-based decision support system with ELECTRE III for a personalised ranking of British universities. *Decision Support Systems*, *48*(3), 488-497.

Grin, F. & Hanhart, S. (2003). Modalités de financement de l'éducation: balisage d'une évaluation par les résultats. *Revue Suisse des Sciences de l'Éducation*, *25*(3), 365-372.

Grosskopf, S. & Moutray, C. (2001). Evaluating performance in Chicago public high schools in the wake of decentralization. *Economics of Education Review*, *20*(1), 1-14.

Gupta, S. & Kumar, U. (2012). An analytical hierarchy process (AHP)-guided decision model for underground mining method selection. *International Journal of Mining, Reclamation and Environment*, *26*(4), 324-336.

Guo, J.-Y., Liu, J. & Qiu, L. (2006, December). *Research on Supply Chain Performance Evaluation Based on DEA/AHP Model*. Paper presented at the Asia-Pacific Conference on Services Computing, Guangzhou, Guangdong, China.

Hajkowicz, S., Young, M., Wheeler, S., MacDonald, D. H. & Young, D. (2000). *Supporting decisions, understanding natural resource management assessment techniques. A report to the Land and Water Resources Research and Development Corportation* (Technical Report of June 2000). Adelaide: Land and Water Research and Development Corporation.

Handwerk, P., Carson, C. & Blackwell, K. (2000). *On-line versus paper-and-pencil surveying of students: A case study*. Paper presented at the Annual Forum of the Association for Institutional Research, Cincinnati, United States.

Hansen, J. A. (2008). *A comparison of parametric and nonparametric techniques used to estimate school district production functions: analysis of model response to change in* *sample size and multicollinearity* (Unpublished doctoral dissertation). Indiana University, USA.

Hanushek, E. A. (2006). School Resources. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education, Volume 2* (pp. 865-903). Amsterdam: North-Holland.

Harrison, J., Rouse, P. & Armstrong, J. (2012). Categorical and continuous non-discretionary variables in data envelopment analysis: a comparison of two single-stage models. *Journal of Productivity Analysis*, *37*(3), 261-276.

Harrison, J. & Rouse, P. (2014). Competition and public high school performance. *Socio-Economic Planning Science*, *48*(1), 10-19.

Hirao, Y. (2012). Efficiency of the top 50 business schools in the United States. *Applied Economics Letters*, *19*(1), 73-78.

Ho, W. (2008). Integrated analytic hierarchy process and its applications – a literature review. *European Journal of Operational Research*, *186*(1), 211-228.

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

Holder, R. (1990). Some comment on the analytic hierarchy process. *Journal of the Operational Research Society*, *41*(11), 1073-1076.

Holder, R. (1991). Response to Holder's comments on the analytic hierarchy process: Response to the response. *Journal of the Operational Research Society*, *42*(10), 914-918.

Hollingsworth, B. & Smith, P. (2003). Use of ratios in data envelopment analysis. *Applied Economics Letters*, *10*(11), 733-735.

Holod, D. & Lewis, H. F. (2011). Resolving the deposit dilemma: a new DEA bank efficiency model. *Journal of Banking & Finance*, *35*(11), 2801-2810.

Hood, C. (1991). A public management for all seasons? *Public Administration*, *69*(1), 3-19.

Hood, C. (2012). Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses. *Public Administration Review*, *72*(S1), 85-92.

Hu, Y., Zhang, Z. & Liang, W. (2009). Efficiency of primary schools in Beijing, China: an evaluation by data envelopment analysis. *International Journal of Educational Management*, *23*(1), 34-50.

Huang, C.-C., Chu, P.-Y. & Chiang, Y.-S. (2008). A fuzzy AHP application in government-sponsored R&D project selection. *Omega*, *36*(6), 1038-1052.

Huguenin, J.-M. (2012). Data Envelopment Analysis (DEA): a pedagogical guide for decision makers in the public sector. *Cahier de l'IDHEAP No. 276*. Lausanne: Swiss Graduate School of Public Administration.

Huguenin, J.-M. (2013a). Data Envelopment Analysis (DEA): un guide pédagogique à l'intention des décideurs dans le secteur public. *Cahier de l'IDHEAP No. 278*. Lausanne : Institut de hautes études en administration publique.

Huguenin, J.-M. (2013b). Data Envelopment Analysis (DEA). In A. Ishizaka & P. Nemery (Eds.), *Multi-Criteria Decision Analysis: Methods and Software* (pp. 235-274). Chichester: John Wiley & Sons.

Huguenin, J.-M. (2014). DEA does not like positive discrimination: a comparison of alternative models based on empirical data. *IDHEAP Working Paper 7/2014*. Lausanne: Swiss Graduate School of Public Administration.

Huguenin, J.-M. (forthcoming). Determinants of school efficiency: the case of primary schools in the State of Geneva, Switzerland. *International Journal of Educational Management*.

Ichinose, D., Yamamoto, M. & Yochida, Y. (2013). Productive efficiency of public and private solid waste logistics and its implications for waste management policy. *International Association of Traffic and Safety Sciences Research*, *36*(2), 96-105.

International Public Sector Accounting Board (2012). *Handbook of Internantional Public Sector Accounting Pronouncements*. New York: International Federation of Accountants.

International Public Sector Accounting Standards Board (2013). *The Conceptual Framework for General Purpose Financial Reporting by Public Sector Entities* (Technical Report of January 2013). New York: International Federation of Accountants.

Ishizaka, A., Balkenborg, D. & Kaplan, T. (2011). Does AHP help us make a choice? An experimental evaluation. *Journal of the Operational Research Society*, *62*(10), 1801-1812.

Ishizaka, A. & Labib, A. (2011). Review of the main developments in the Analytic Hierarchy Process. *Expert Systems with Applications*, *38*(11), 14336-14345.

Ishizaka, A. & Labib, A. (2009). Analytic Hierarchy Process and Expert Choice: Benefits and limitations. *OR Insight*, *22*(4), 201-220.

Ishizaka, A. & Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. Chichester: John Wiley and Sons.

Jablonsky, J. (2007). Measuring the efficiency of production units by AHP models. *Mathematical and Computer Modelling*, *46*(7-8), 1091-1098.

Jeanrenaud, C. & Vuilleumier, M. (2006). Evaluating Technical Efficiency of Swiss Consulates. *Review of Business and Economics*, *Vol. LI*(3), 266-279.

Jennings, E. T. (2012). Organizational Culture and Effects of Performance Measurement. *Public Administration Review*, *72*(S1), 93-94.

Jeon, Y. & Shields, M. P. (2008). Integration and utilization of public education resources in remote and homogenous areas: a case study of the upper peninsula of Michigan. *Contemporary Economics Policy*, *23*(4), 601-614.

Jesson, D., Mayston, D. & Smith, P. (1987). Performance Assessment in the Education Sector: Educational and Economic Perspectives. *Oxford Review of Education*, *13*(3), 249-266.

Johnes, J. (2003). Measuring teaching efficiency in higher education: an application of data envelopment analysis to graduates from UK universities 1993. *Discussion paper EC7/03*. Lancaster: Lancaster University.

Johnes, J. (2004). Efficiency measurement. In G. Johnes & J. Johnes (Eds.), *International Handbook on the Economics of Education* (pp. 613-742). Cheltenham: Edward Elgar Publishing.

Johnes, J. (2006a). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, *25*(3), 273-288.

Johnes, J. (2006b). Measuring teaching efficiency in higher education: An application of data envelopment analysis to economics graduates from UK Universities 1993. *European Journal of Operational Research*, *174*(1), 443-456.

Johnes, J. & Taylor, J. (1990). *Performance Indicators in Higher Education: UK Universities*. Milton Keynes: Open University Press.

Johnes, J. & Yu, L. (2008). Measuring the research performance of Chinese higher education institutions using data envelopment analysis. *China Economic Review*, *19*(4), 679-696.

Joumady, O. & Ris, C. (2005). Performance in European higher education: A non-parametric production frontier approach. *Education Economics*, *13*(2), 189-205.

Kantabutra, S. & Tang, J. C. S. (2006). Urban-rural and size effects on school efficiency: The case of Northern Thailand. *Leadership and Policy in Schools*, *5*(4), 355-377.

Katharaki, M. & Katharakis, G. (2010). A comparative assessment of Greek universities' efficiency using quantitative analysis. *International Journal of Educational Research*, *49*(4), 115-128.

Kelly, E., Shalloo, L., Geary, U., Kinsella, A. & Wallace, M. (2012). Application of data envelopment analysis to measure technical efficiency on a sample of Irish dairy farms. *Irish Journal of Agricultural and Food Research*, *51*(1), 63-77.

Kempkes, G. & Pohl, C. (2010). The efficiency of German universities – some evidence from nonparametric and parametric methods. *Applied Economics*, *42*(16), 2063-2079.

Kirjavainen, T. & Loikkanen, 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.

Knoepfel, P., Larrue, C., Varone, F. & Hill, M. (2011). *Public policy analysis*. Bristol: The Policy Press.

Kong, W.-H. & Fu, T.-T. (2012). Assessing the performance of business colleges in Taiwan using data envelopment analysis and student based value-added indicators. *Omega*, *40*(5), 541-549.

Korhonen, P., Tainio, R. & Wallenius, J. (2001). Value efficiency analysis of academic research. *European Journal of Operational Research*, *130*(1), 121-132.

Korpela, J., Lehmusvaara, A. & Nisonen, J. (2007). Warehouse operator selection by combining AHP and DEA methodologies. *International Journal of Production Economics*, *108*(1-2), 135-142.

Korpela, J. & Tuominen, M. (1996). A decision aid in warehouse site selection. *International Journal of Production Economics*, *45*(1-3), 169-180.

Kroll, A. (2012, September). *Does Performance Information Use Increase Organizational Performance? Examining an Implicit Assumption.* Paper presented at the Conference of the European Group of Public Administration (EGPA), Bergen, Norway.

Lai, V., Wong, B. K. & Cheung, W. (2002). Group decision making in a multiple criteria environment: A case using the AHP in the software selection. *European Journal of Operational Research*, *137*(1), 134-144.

Lam, K. & Zhao, X. (1998). An application of quality function deployment to improve the quality of teaching. *International Journal of Quality Reliability Management*, *15*(4), 389-413.

Lan, C.-H., Chuang, L.-L. & Chen, Y.-F. (2009). Performance efficiency and resource allocation for fire department with the stochastic consideration. *International Journal of Technology, Policy and Management*, *9*(3), 296-315.

Lavado, R. F. & Cabanda, E. C. (2009). The efficiency of health and education expenditures in the Philippines. *Central European Journal of Operations Research*, *17*(3), 275-291.

Lee, J.-Y. (2008). Application of the three-stage DEA in measuring efficiency – an empirical evidence. *Applied Economics Letters*, *15*(1), 49-52.

Lee, S.-H. (2010). Using fuzzy AHP to develop intellectual capital evaluation model for assessing their performane contribution to a university. *Expert Systems with Applications*, *37*(7), 4941-4947.

Lee, B. L. & Worthington, A. C. (2014). Technical efficiency of mainstream airlines and low-cost carriers: New evidence using bootstrap data envelopment analysis truncated regression. *Journal of Air Transport Management*, *38*, 15-20.

Lee, T., Yeo, G.-T. & Thai, V. V. (2014). Environmental efficiency analysis of port cities: Slacks-based measure data envelopment analysis approach. *Transport Policy*, *33*, 82-88.

Leithwood, K. & Jantzi, D. (2009). A Review of Empirical Evidence About School Size Effects: A Policy Perspective. *Review of Educational Research*, *79*(1), 464-490.

Lin, H.-F. (2010). An application of fuzzy AHP for evaluating course website quality. *Computers & Education*, *54*(4), 877-888.

Lin, M.-I., Lee, Y.-D. & Ho, T.-N. (2011). Applying integrated DEA/AHP to evaluate the economic performance of local governments in China. *European Journal of Operational Research*, *209*(2), 129-140.

Liu, C.-C. (2009). A study of optimal weights restriction in Data Envelopment Analysis. *Applied Economics*, *41*(14), 1785-1790.

Liu, C. H., Lin, S. J. & Lewis, C. (2010). Evaluation of thermal power plant operational performance in Taiwan by data envelopment analysis. *Energy Policy*, *38*(4), 1049-1058.

Löber, G. & Staat, M. (2010). Integrating categorical variables in Data Envelopment Analysis models: A simple solution technique. *European Journal of Operational Research*, *202*(3), 810-818.

Lootsma, F. A. (1989). Conflict resolution via pairwise comparison of concessions. *European Journal of Operational Research*, *40*(1), 109-116.

Lovell, C. A. K. (1993). Production Frontiers and Productive Efficiency. In H. O. Fried, C. A. K Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications* (pp. 3-67). New York: Oxford University Press.

Lovell, C. A. K., Walters, L. C. & Wood, L. L. (1994). Stratified models of education production using modified data envelopment analysis and regression analysis. In A. Charnes, W. W. Cooper, A. Y. Lewin & L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology and applications* (pp. 329-351). Dordrecht: Kluwer Academic.

Lozano, S. & Villa, G. (2009). Multiobjective target setting in data envelopment analysis using AHP. *Computers & Operations Research*, *36*(2), 549-564.

Ma, D. & Zheng, X. (1991). *9/9-9/1 scale method of AHP*. Paper presented at the International Symposium on the Analytic Hierarchy Process, Pittsburgh, United States.

McCarty, T. A. & Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. In H. O. Fried, C. A. K Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications* (pp. 271-287). New York: Oxford University Press.

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

Macharis, C., Springael, J. De Brucker, K. & Verbeke, A. (2004). PROMETHEE and AHP: The design of operational synergies in multicriteria analysis. Strengthening PROMETHEE with ideas of AHP. *European Journal of Operational Research*, *153*(2), 307-317.

McMillan, M. L. & Datta, D. (1998). The Relative Efficiencies of Canadian Universities: A DEA Perspective. *Canadian Public Policy / Analyse de Politiques*, *24*(4), 485-511.

Malhotra, R., Malhotra, D. K. & Lafond, C. A. (2010). Benchmarking large U.S. retailers using a data envelopment analysis model. In K. D. Lawrence & G. Kleinman (Eds.), *Applications in Multicriteria Decision Making, Data Envelopment Analysis, and Finance (Applications of Management Science, Volume 14)* (pp. 217-235). Bingley: Emerald Group Publishing Limited.

Mallikarjun, S., Lewis, H. F. & Sexton, T. R. (2014). Operational performance of U.S. public rail transit and implications for public policy. *Socio-Economic Planning Sciences*, *48*(1), 74-88.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estatistica*, *4*(2), 209-242.

Manasakis, C., Apostolakis, A. & Datseris, G. (2013). Using data envelopment analysis to measure hotel efficiency in Crete. *International Journal of Contemporary Hospitality Management*, *25*(4), 510-535.

Mancebón, M. J. & Bandrés, E. (1999). Efficiency Evaluation of Secondary Schools: the key role of model specification and of ex post analysis of results. *Education Economics*, *7*(2), 131-152.

Mancebón, M. J. & Mar Molinero, C. (2000). Performance in Primary Schools. *The Journal of the Operational Research Society*, *51*(7), 843-854.

Mante, B. and O'Brien, G. (2002). Efficiency measurement of Australian public sector organisations: the case of state secondary schools in Victoria. *Journal of Educational Administration*, *40*(3), 274-296.

Marques, R. & Simões, P. (2009). How far are Portuguese prisons inefficient? A non-parametric approach. *MPRA Paper*. Munich: Munich Personal RePEc Archive.

Martin, E. (2006). Efficiency and Quality in the Current Higher Education Context in Europe: an application of the data envelopment analysis methodology to performance assessment of departments within the University of Zaragoza. *Quality in Higher Education*, *12*(1), 57-79.

Mény, Y. & Thoenig, J. C. (1989). *Politiques publiques*. Paris: Presses Universitaires de France.

Meunier, M. (2008). Are Swiss Secondary Schools Efficient? In N. C. Soguel & P. Jaccard (Eds.), *Governance and Performance of Education Systems* (pp. 187-202). Dordrecht: Springer.

Millet, I. & Saaty, T. (2000). On the relativity of relative measures: Accomodating both rank preservation and rank reversals in the AHP. *European Journal of Operational Research*, *121*(1), 205-212.

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

Monroe, M. C. & Adams, D. C. (2012). Increasing Response Rates to Web-Based Surveys. *Journal of Extension*, *50*(6), 6TOT7.

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

Muñiz, M., Paradi, J., Ruggiero, J. & Yang, Z. (2006). Evaluating alternative DEA models used to control for non-discretionary inputs. *Computers & Operations Research*, *33*(5), 1173-1183.

Myers, R. (1990). *Classical and Modern Regression with Applications*. Boston: Duxbury Press.

Nachiappan, S. & Ramanathan, R. (2008, February). *Robust decision making using Data Envelopment Analytic Hierarchy Process*. Paper presented at the

Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Cambridge, United Kingdom.

Neely, A. D., Mills, J. F., Gregory, M. J. & Platts, K. W. (1995). Performance measurement system design – literature review and research agenda. *International Journal of Operations and Production Management*, *15*(4), 80-116.

Neralic, L. (2004). Preservation of efficiency and inefficiency classification in data envelopment analysis. *Mathematical Communications*, *9*(1), 51-62.

Neralic, L. & Wendell, R. E. (2004). Sensitivity in data envelopment analysis using an approximate inverse matrix. *Journal of the Operational Research Society*, *55*(11), 1187-1193.

Nevo, D. (1995). *School-Based Evaluation: A dialogue for school improvement.* Oxford: Pergamon.

Nevo, D. (2007). Evaluation in education. In I. F. Shaw, J. C. Greene and M. M. Mark (Eds.), *The SAGE Handbook of Evaluation* (pp. 441-460). London: SAGE Publications.

Ng, Y. C. & Li, S. K. (2000). Measuring the Research Performance of Chinese Higher Education Institutions: An Application of Data Envelopment Analysis. *Education Economics*, *8*(2), 139-156.

Nguyen, T. P. T., Roca, E. & Sharma, P. (2014). How efficient is the banking system of Asia's next economic dragon? Evidence from rolling DEA windows. *Applied Economics*, *46*(22), 2665-2684.

Noreisch, K. (2007). School catchment area evasion: the case of Berlin, Germany. *Journal of Education Policy*, *22*(1), 69-90.

Observatory on Primary Education (2010). *Allocation des ressources aux établissements* (Report of December 2010). Geneva: General Direction of Primary Schools, Education Department, State of Geneva.

O'Donnell, C. J. & van der Westhuizen, G. (2002). Regional comparisons of banking performance in SouthAfrica. *South African Journal of Economics*, *70*(3), 485-518.

Oggioni, G., Riccardi, R. & Toninelli, R. (2011). Eco-efficiency of the world cement industry: A data envelopment analysis. *Energy Policy*, *39*(5), 2842-2854.

Olivares, M. & Schenker-Wicki, A. (2010). How do Swiss Universities master the reform of the last ten years? Empirical evidence from a data envelopment analysis. *Zurich Open Repository and Archive.* Zurich: University of Zurich.

Olivares, M. & Schenker-Wicki, A. (2012). The Dynamics of Productivity in the Swiss and German University Sector: A Non-Parametric Analysis That Accounts for Heterogeneous Production. *UZH Business Working Paper No. 309.* Zurich: University of Zurich.

O'Neill, J. & West-Burnham, J. (2001). Perspectives on Performance Management. In J. West-Burnham, I. Bradbury & J. O'Neill (Eds.), *Performance Management in Schools: How to Lead and Manage Staff for School Improvement* (pp. 3-14). London: Pearson Education.

Organisation for Economic Co-operation and Development (2001). *Measuring Productivity: Measurement of Aggregate and Industry-level Productivity Growth*. Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development (2007). *Education at a Glance 2007: OECD indicators*. Paris: Organisation for Economic Co-operation and development.

Orme, C. & Smith, P. (1996). The Potential for Endogeneity Bias in Data Envelopment Analysis. *Journal of Operational Research Society*, *47*(1), 73-83.

Ouellette, P. & Vierstraete, V. (2005). An evaluation of the efficiency of Québec's school boards using the Data Envelopment Analysis method. *Applied Economics*, *37*(14), 1643-1653.

Pak, R. J. (2013). Combining Importance-Performance Analysis with Analytic Hierarchy Process for Enhancing Satisfaction. *Journal of Advanced Management Science*, *1*(4), 368-371.

Pakkar, M. S. (2012). An Integrated Approach to the DEA and AHP Methodologies in Decision Making. In V. Charles & M. Kumar (Eds.), *Data Envelopment Analysis and Its Applications to Management* (pp. 136-149). Newcastle upon Tyne: Cambridge Scholars Publishing.

Pan, N.-F. (2008). Fuzzy AHP approach for selecting the suitable bridge construction method. *Automation in Construction*, *17*(8), 958-965.

Park, S. U. & Lesourd, J. B. (2000). The efficiency of conventional fuel power plants in South Korea: a comparison of parametric and non-parametric approaches. *International Journal of Production Economics*, *63*(1), 59-67.

Park, B.-Y. & Min, H. (2011). The selection of transshipment ports using a hybrid data envelopment analysis/analytic hierarchy process. *Journal of Transportation Management*, *22*(1), 47-64.

Pastor, J. M. (2002). Credit risk and efficiency in the European banking system: A three-stage analysis. *Applied Financial Economics*, *12*(12), 895-911.

Pastor, J. M., Ruiz, J. L. & Sirvent, I. (1999). A statistical test for detecting influential observations in DEA. *European Journal of Operational Research*, *115*(3), 542-554.

Paton, G. (2013, March 4). Schoolchildren losing the power to concentrate in class. *The Daily Telegraph*.

Picazo-Tadeo, A. J., Gómez-Limón, J. A. & Reig-Martínez, E. (2011). Assessing farming eco-efficiency: A Data Envelopment Analysis Approach. *Journal of Environmental Management*, *92*(4), 1154-1164.

Portela, M. C. A. S., Camanho, A. S. & Borges, D. N. (2011). BESP – benchmarking of Portuguese secondary schools. *Benchmarking: An International Journal*, *18*(2), 240-260.

Portela, M. C. A. S. & Thanassoulis, E. (2001). Decomposing school and school type efficiency. *European Journal of Operational Research*, *132*(2), 114-130.

Portela, M. C. A. S., Thanassoulis, E. & Simpson, G. P. M. (2004). Negative data in DEA: A directional distance approach applied to bank branches. *Journal of the Operational Research Society*, *55*(10), 1111-1121.

Rabar, D. (2013). Assessment of regional efficiency in Croatia using data envelopment analysis. *Croatian Operational Research Review*, *4*(1), 76-88.

Ramanathan, R. (2006). Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process. *Computers & Operations Research*, *33*(5), 1289-1307.

Ramanathan, R. (2001a). A Data Envelopment Analysis of Comparative Performance of Schools in the Netherlands. *Opsearch*, *38*(2), 160-182.

Ramanathan, R. (2001b). A note on the use of the analytic hierarchy process for environmental impact assessment. *Journal of Environmental Management*, *63*(1), 27-35.

Ramanathan, R. & Ganesh, L. S. (1994). Group preference aggregation methods employed in AHP: an evaluation and an intrinsic process for deriving members' weightages. *European Journal of Operational Research*, *79*(20), 249-265.

Rassouli-Currier, S. (2007). Assessing the Efficiency of Oklahoma Public Schools: A Data Envelopment Analysis. *Southwestern Economic Review*, *34*(1), 131-144.

Ray, S. C. (1988). Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. *Socio-economic Planning Sciences*, *22*(4), 167-176.

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

Ray, S. C. & Jeon, Y. (2008). Reputation and efficiency: A non-parametric assessment of America's top-rated MBA programs. *European Journal of Operational Research*, *189*(1), 245-268.

Reschovsky, A. (1994). Fiscal equalization and school finance. *National Tax Journal*, *47*(1), 185-197.

Rhodes, E. L. (1978). *Data envelopment analysis and approaches for measuring the efficiency of decision making units with an application to program follow through in U.S. education* (Unpublished doctoral dissertation). Carnegie-Mellon University, USA.

Rogge, N. & De Jaeger, S. (2013). Measuring and explaining the cost efficiency of municipal solid waste collection and processing services. *Omega*, *41*(4), 653-664.

Roll, Y., Cook, W. D. & Golany B. (1991). Controlling Factor Weights in Data Envelopment Analysis. *IEE Transactions*, *23*(1), 2-9.

Rothstein, R. (2010). Family Environment in the Production of Schooling. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of Education* (pp. 148-155). Oxford: Elsevier.

Roy, B. (1981). The optimisation problem formulation: Criticism and overstepping. *Journal of the Operational Research Society*, *32*(6), 427-436.

Roy, B. & Bouyssou, D. (1993). *Aide multicritère à la décision: Méthodes et cas*. Paris: Economica.

Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, *90*(3), 553-565.

Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, *111*(3), 461-469.

Ruggiero, J. (2000). Nonparametric estimation of returns to scale in the public sector with an application to the provision of educational services. *Journal of the Operational Research Society*, *51*(8), 906-912.

Ruggiero, J. (2004). Performance Evaluation in Education. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 323-346). Dordrecht: Springer.

Ruggiero, J., Miner, J. & Blanchard, L. (2002). Measuring equity of educational outcomes in the presence of inefficiency. *European Journal of Operational Research*, *142*(3), 642-652.

Ruggiero, J. & Vitaliano, D. F. (1999). Assessing the efficiency of public schools using data envelopment analysis and frontier regression. *Contemporary Economic Policy*, *17*(3), 321-331.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234-281.

Saaty, T. (1980). *The Analytic Hierarchy Process*. New-York: McGraw-Hill.

Saaty, T. L. (1991a). Response to Holder's comments on the analytic hierarchy process. *Journal of the Operational Research Society*, *42*(10), 909-914.

Saaty, T. L. (1991b). Response to Holder's comments on the analytic hierarchy process: Response to the Response to the Response. *Journal of the Operational Research Society*, *42*(10), 918-924.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83-98.

Saaty, T. L. (2012). *Decision making for leaders: The analytic hierarchy process for decisions in a complex world*. Pittsburgh: RWS Publications.

Saaty, T. L. & Vargas, L. G. (2005). The possibility of group welfare functions. *International of Information Technology & Decision Making*, *4*(2), 167-176.

Saaty, T. L. & Vargas, L. G. (2005, July). *Dispersion of group judgments*. Paper presented at the International Symposium on Analytic Hierarchy Process, Honolulu, United States.

Saen, R. F., Memariani, A. & Lotfi, F. H. (2005). Determining relative efficiency of slightly non-homogeneous decision making units by data envelopment analysis: A case study in IROST. *Applied Mathematics and Computation*, *165*(2), 313-328.

Salo, A. A. & Hämäläinen, R. P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis*, *6*(6), 309-319.

Sarica, K. & Or, I. (2007). Efficiency assessment of Turkish power plants using data envelopment analysis. *Energy*, *32*(8), 1484-1499.

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.

Sarrico, C. S., Rosa, M. J. & Coelho, I. P. (2010). The performance of Portuguese secondary schools: an exploratory study. *Quality Assurance in Education*, *18*(4), 268-303.

Sav, G. T. (2013). Four-Stage DEA Efficiency Evaluations: Financial Reforms in Public University Funding. *International Journal of Economics and Finance*, *5*(1), 24-33.

Scheel, H. (2000). *EMS: Efficiency Measurement System User's Manual.* Retrieved from http://www.wiso.tu-dortmund.de/wiso/de/fakultaet/personen/institut/or/EXT-HOSC.html

Schenker-Wicki, A. & Hürlimann, M. (2006). Universités suisses : échec ou succès du financement fondé sur les résultats ? Analyse a posteriori. *Politiques et gestion de l'enseignement supérieur*, *18*(1), 61-78.

Schoenenberger, A, Mack, A. & von Gunten, F. (2009). *Efficacité technique des exploitations forestières publiques en Suisse. Impact des subventions* (Strukturberichterstattung Nr. 42). Berne: Secrétariat d'Etat à l'économie.

Seifert, L. M. & Zhu, J. (1998). Identifying Excesses and Deficits in Chinese Industrial Productivity (1953-1990): a Weighted Data Envelopment Analysis Approach. *Omega*, *26*(2), 279-296.

Seiford, L. M. & Zhu, J. (1998a). Stability regions for maintaining efficiency in data envelopment analysis. *European Journal of Operational Research*, *108*(1), 127-139.

Seiford, L. M. & Zhu, J. (1998b). Sensitivity analysis of DEA models for simultaneous changes in all the data. *Journal of the Operational Research Society*, *49*(10), 1060-1071.

Seiford, L. M. & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, *142*(1), 16-20.

Selwood, I. & Visscher, A. J. (2008). The potential of School Information Systems for Enhancing School Improvement. In. N. C. Soguel & P. Jaccard (Eds.), *Governing and Performance of Education Systems* (pp. 269-288). Dordrecht: Springer.

Sengupta, J. K. (1990). Tests of efficiency in DEA. *Computers Operations Research*, *17*(2), 123-132.

Sevkli, M., Lenny Koh, S. C., Zaim, S., Demirbag, M. & Tatoglu, E. (2007). An application of data envelopment analytic hierarchy process for supplier selection: A case study of BEKO in Turkey. *International Journal of Production Research*, *45*(9), 1973-2003.

Shang, J. & Sueyoshi, T. (1995). A unified framework for the selection of a flexible manufacturing system. *European Journal of Operational Research*, *85*(2), 297-315.

Shang, J.-K., Hung, W.-T., Lo, C.-F. & Wang, F.-C. (2008). Ecommerce and hotel performance: three-stage DEA analysis. *The Service Industries Journal*, *28*(4), 529-540.

Sharma, S. (2008). Analyzing the technical and scale efficiency performance: a case study of cement firms in India. *Journal of Advances in Management Research*, *5*(2), 56-63.

Sheldon, G. (1995). Zur Messung der Effizienz im Bildungsbereich mit Hilfe der Data Envelopment Analysis. *Wirtschaftswissenschaftliches Zentrum der Universität Basel – Studien, Nr. 47*. Basel: University of Basel.

Sheldon, G. (2003). The efficiency of public employment services: A non-parametric matching function analysis for Switzerland. *Journal of Productivity Analysis*, *20*(1), 49-70.

Shih, T.-H. & Fan, X. (2008). Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods*, *20*(3), 249-271.

Sibiano, P. & Agasisti, T. (2013). Efficiency and heterogeneity of public spending in education among Italian regions. *Journal of Public Affairs, 13*(1), 12-22.

Sillah, B. M. S. (2012). An analysis of efficiency in Senior Secondary Schools in the Gambia 2006-2008: Educational inputs and production of credits in English and Mathematics subjects. *Africa Education Review*, *9*(1), 86-104.

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*(2), 205-218.

Singh, S., Rylander, D. H. & Mims, T. C. (2012). Efficiency of Online vs. Offline Learning: A Comparison of Inputs and Outcomes. *International Journal of Business, Humanities and Technology*, *2*(1), 93-98.

Sinuany-Stern, Z., Mehrez, A. & Hadad, Y. (2000). An AHP/DEA methodology for ranking decision making units. *International Transactions in Operational Research*, *7*(2), 109-124.

Smith, P. (2005). *An Introduction to Measuring Efficiency and Productivity in Public Sector Organisations: Course notes*. York: University of York.

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

Smith, P. C. & Street, A. (2006). *Analysis of Secondary School. Efficiency: Final report* (Research Report RR788). London: Department for Education and Skills.

Soguel, N. C. & Huguenin, J.-M. (2008). Comparer l'efficience des prestations financières de l'aide sociale : le cas des centres sociaux régionaux vaudois. In G. Bonoli & F. Bertozzi (Eds.), *Les nouveaux défis de l'Etat social* (pp. 165-184). Lausanne: Presses polytechniques et universitaires romandes.

Solaux, G., Huguenin, J.-M., Payet, J.-P. & Ramirez, J. (2011). Evaluation, concertation, décision : quelle régulation pour le système éducatif ? Le cas de l'enseignement primaire genevois. *Raisons éducatives No. 15*. Bruxelles: De Boeck.

Sólnes, J. (2003). Environmental quality indexing of large industrial development alternatives using AHP. *Environmental Impact Assessment Review*, *23*(3), 283-303.

Soteriou, A. C., Karahanna, E., Papanastasiou, C. & Diakourakis, M. S. (1998). Using DEA to evaluate the efficiency of secondary schools: the case of Cyprus. *International Journal of Educational Management*, *12*(2), 65-73.

Souci, A. & Nidegger, C. (2010). *Le réseau d'enseignement prioritaire à Genève : quels effets sur les acquis des élèves après quelques années ?* Genève: Service de la recherche en éducation.

Steinmann, L. & Zweifel, P. (2003). On the (in)efficiency of Swiss hospitals. *Applied Economics*, *35*(3), 361-370.

Stevens, P. A. (2001). The determinants of economic efficiency in English and Welsh universities. *Discussion paper No. 185*. London: National Institute of Economic and Social Research.

Stevens, S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153-181.

Stewart, T. J. (1996). Relationships between Data Envelopment Analysis and Multicriteria Decision Analysis. *The Journal of the Operational Research Society*, *47*(5), 654-665.

Stewart, J. & Walsh, K. (1994). Performance Measurement: When Performance can Never be Finally Defined. *Public Money & Management*, *14*(2), 45-49.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman H. O. & Provus, M. M. (1971). *Educational Evaluation and Decision Making*. Itasca: Peacock.

Sueyoshi, T. & Goto, M. (2012). Data envelopment analysis for environmental assessment: Comparison between private and public ownership in petroleum industry. *European Journal of Operational Research*, *216*(3), 668-678.

Sueyoshi, T., Goto, M. & Omi, Y. (2010). Corporate governance and firm performance: Evidence from Japanese manufacturing industries after the lost decade. *European Journal of Operational Research*, *203*(3), 724-736.

Summermatter, L. & Siegel, J. P. (2009, April). *Defining Performance in Public Management: Variations over time and space*. Paper presented at the Conference of the International Research Society for Public Management (IRSPM), Copenhagen, Denmark.

Sun, S. (2002). Measuring the relative efficiency of police precincts using data envelopment analysis. *Socio-Economic Planning Sciences*, *36*(1), 51-71.

Sutherland, D., Price, R., Joumard, I. & Nicq, C. (2007). Performance Indicators for Public Spending Efficiency in Primary and Secondary Education. *OECD Economics Department Working Papers No. 546*. Paris: Organisation for Economic Co-operation and Development.

Suzuki, S., Nijkamp, P., Pels, E. & Rietveld, P. (2014). Comparative performance analysis of European airports by means of extended data envelopment analysis. *Journal of Advanced Transportation*, *48*(3), 185-202.

Syrjänen, M. J. (2004). Non-discretionary and discretionary factors and scale in data envelopment analysis. *European Journal of Operational Research*, *158*(1), 20-33.

Tahriri, F, Osman, M. R., Ali, A., Yusuff, R. M. & Esfandiary, A. (2008). AHP approach for supplier evaluation and selection in a steel manufacturing company. *Journal of Industrial Engineering and Management*, *1*(2), 54-76.

Takamura, Y. & Tone, K. (2003). A comparative site evaluation study for relocating Japanese government agencies out of Tokyo. *Socio-Economic Planning Sciences*, *37*(2), 85-102.

Tam, M. C. Y. & Tummala, V. M. R. (2001). An application of the AHP in vendor selection of a telecommunications system. *Omega*, *29*(2), 171-182.

Tavares, G. (2002). A bibliography of Data Envelopment Analysis (1978-2001). *Rutcor Research Report No. 01-02*. Piscataway: Rutgers University.

Thanassoulis, E. (1996). Altering the Bias in Differential School Effectiveness Using Data Envelopment Analysis. *The Journal of the Operational Research Society*, *47*(7), 882-894.

Thanassoulis, E., Portela, M. C. S. & Allen, R. (2004). Incorporating value judgements in DEA. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 99-138). Boston: Kluwer Academic Publishers.

Thanassoulis, E., Portela, M. C. S. & Despic, O. (2008). Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. In H. O. Fried, C. A. Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency and Productivity Growth* (pp. 251-420). New York: Oxford University Press.

Thanassoulis, E., Portela Silva, M. C. A. & Graveney, M. (2014). Using DEA to estimate potential savings at GP units at medical specialty level. *Socio-Economic Planning Sciences*, *48*(1), 38-48.

Thomson, R. G., Langemeier, L. N., Lee, C.-T., Lee, E. & Thrall, R. M. (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics*, *46*(1-2), 93-108.

Thomson, R. G., Singleton Jr., F. D., Thrall, R. M. & Smith, B. A. (1986). Comparative site evaluations for locating a high-energy physics lab in Texas. *Interfaces*, *16*(6), 35-49.

Thrall, R. M. (1996a). The lack of invariance of optimal dual solutions under translation. *Annals of Operations Research*, *66*(2), 103-108.

Thrall, R. M. (1996b). Duality, classification and slacks in DEA. *Annals of Operations Research*, *66*(2), 109-138.

Tobin, J. (1958). Estimation for relationships with limited dependent variables. *Econometrica*, *26*(1), 24-36.

Tongzon, J. (2001). Efficiency measurement of selected Australian and other international ports using data envelopment analysis. *Transportation Research Part A: Policy and Practice*, *35*(2), 107-122.

Tone, K. (2004). Malmquist productivity index – Efficiency change over time. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 203-227). Boston: Kluwer Academic Publishers.

Vaidya, O. S. & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of Operational Research*, *169*(1), 1-29.

Van der Waldt, G. (2004). *Managing Performance in the Public Sector: Concepts, Considerations and Challenges*. Lansdowne: Juta and Co Ltd.

Van Helden, G. J. & Tillema, S. (2005). In search of a benchmarking theory for the public sector. *Financial Accountability & Management*, *21*(3), 337-361.

Van Zanten, A. (2003). Middle-class Parents and Social Mix in French Urban Schools: reproduction and transformation of class relations in education. *International Studies in Sociology of Education*, *13*(2), 107-123.

Varian, H. R. (2010). *Intermediate Microeconomics: a Modern Approach*. New-York: W. W. Norton & Company.

Vaz, C. B. & Camanho, A. S. (2012). Performance comparison of retailing stores using a Malmquist-type index. *Journal of the Operational Research Society*, *63*(5), 631-645.

Verma, A., Gavirneni, S. (2006). Measuring police efficiency in India: an application of data envelopment analysis. *Policing: An International Journal of Police Strategies & Management*, *29*(1), 125-145.

Vidal, L.-A., Sahin, E., Martelli, N., Berhoun, M. & Bonan, B. (2010). Applying AHP to select drugs to be produced by anticipation in a chemotherapy compounding unit. *Expert Systems with Applications*, *37*(2), 1528-1534.

Vierstraete, V. & Yergeau, E. (2011). Performance of the Different Methods of Study Financing: A Measurement through the Data Envelopment Analysis Method. *Managerial and Decision Economics*, *33*(1), 1-9.

Viger, G. (2007, June). *L'analyse comparative au service de l'amélioration et de la performance.* Note de cadrage presented at the third meeting of the Contrôle de Gestion des Programmes.

von Solms, S. (2009, July). *Homogeneity and choice aggregation in the analytic hierarchy process.* Paper presented at the International Symposium on Analytic Hierarchy Process, Pittsburgh, United States.

Waldo, S. (2007). Efficiency in Swedish Public Education: Competition and Voter Monitoring. *Education Economics*, *15*(2), 231-251.

Wang, J. & Zhang, Y. (2014). Service Quality Evaluation of Urban Parks Based on AHP Method and SD Software. *Journal of Applied Sciences*, *14*(3), 291-295.

Wang, Y.-M. & Chin, K.-S. (2009). A new data envelopment analysis method for priority determination and group decision making in the analytic hierarchy process. *European Journal of Operational Research*, *195*(1), 239-250.

Wang, Y.-M. & Elhag, T. M. S. (2006). An approach to avoiding rank reversal in AHP. *Decision Support Systems*, *42*(3), 1474-1480.

Wang, Y.-M. & Luo, Y. (2009). On rank reversal in decision analysis. *Mathematical and Computer Modelling*, *49*(5-6), 1221-1229.

Wang, Y.-M., Parkan, C. & Luo, Y. (2007). A linear programming method for generating the most favorable weights from a pairwise comparison matrix. *Computers and Operations Research*, *35*(1), 3918-3930.

Webb, R. (2003). Level of efficiency in UK retail banks: a DEA window analysis. *International Journal of the Economics of Business*, *10*(3), 305-322.

Widmer, P. & Zweifel, P. (2008). Provision of Public Goods in a Federalist Country: Tiebout Competition, Fiscal Equalization, and Incentives for Efficiency in Switzerland. *Socioeconomic Institute Working Papers No. 0804*. Zurich: University of Zurich.

Wilson, P. W. (1993). Detecting outliers in deterministic non-parametric frontier models with multiple outputs. *Journal of Business and Economic Statistics*, *11*(3), 319-323.

Wilson, P. W. (1995). Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis*, *6*(1), 27-45.

Woessmann, L. (2003). Schooling Resources, Educational Institutions and Student Performance: the International Evidece. *Oxford Bulletin of Economics and Statistics*, *65*(2), 117-170.

Wössmann, L. (2005). The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, *13*(2), 143-169.

Wolter, S. (2010). *Swiss Education Report 2010.* Aarau: Swiss Coordination Centre for Research in Education.

Wong, J. K. W. & Li, H. (2006). Development of a conceptual model for the selection of intelligent building systems. *Building and Environment*, *41*(8), 1106-1123.

Wong, J. K. W. & Li, H. (2008). Application of the analytic hierarchy process (AHP) in multi-criteria analysis of the selection of intelligent building systems. *Building and Environment*, *43*(1), 108-125.

Yang, H.-H. & Chang, C.-Y. (2009). Using DEA window analysis to measure efficiencies of Taiwan's integrated telecommunication firms. *Telecommunications Policy*, *33*(1-2), 98-108.

Yang, T. & Kuo, C. (2003). A hierarchical AHP/DEA methodology for the facilities layout design problem. *European Journal of Operational Research*, *147*(1), 128-136.

Yang, H. & Pollitt, M. (2009). Incorporating both undesirable outputs and uncontrollable variables into DEA: The performance of Chinese coal-fired power plants. *European Journal of Operational Research*, *197*(3), 1095-1105.

Yanyan, P. (2012). Commercial Bank Branch Efficiencies Based on Three-Stage DEA Model. In L. Zhang & C. Zhang (Eds.), *Engineering Education and Management* (pp. 465-470). Berlin: Springer.

Yilmaz, B. & Ali Yurdusev, M. (2011). Use of data envelopment analysis as a multi criteria decision tool – a case of irrigation management. *Mathematical and Computational Applications*, *16*(3), 669-679.

Yin, Q. (2013). An Analytical Hierarchy Process Model for the Evaluation of College Experimental Teaching Quality. *Journal of Technology and Science Education*, *3*(2), 59-65.

Yin, K., Pu, Y., Liu, Z. & Zhou, B. (2014). An AHP-based Approach for Banking Data Quality Evaluation. *Information Technology Journal*, *13*(8), 1523-1531.

Yoo, H (2003). A study on the efficiency evaluation of total quality management activities in Korean companies. *Total Quality Management*, *14*(1), 119-128.

Yue, P. (1992). Data Envelopment Analysis and Commercial Bank Performance: A Primer with Applications to Missouri Banks. Federal *Reserve Bank of St Louis Review*, *74*(1), 31-45.

Zhu, J. (1996). Robustness of the efficient DMUs in data envelopment analysis. *European Journal of Operational Research*, *90*(3), 451-460.

Zhu, J. (2001). Super-efficiency and DEA sensitivity analysis. *European Journal of Operational Research*, *129*(2), 443-455.

Zhu, J. (2003). *Quantitative models for performance evaluation and benchmarking*. New York: Springer.

Zikopoulos, P. & Eaton, C. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. New-York: McGraw-Hill Osborne Media.

Jean-Marc Huguenin
- Senior lecturer and Project manager, Swiss graduate school of public administration – IDHEAP
- Chief executive officer, Economistes indépendants

Annex

Annex 1
Data Envelopment Analysis (DEA)
A pedagogical guide for decision makers in the public sector

Preliminary remark

This chapter introduces a pedagogical guide about DEA. This guide is a modified version of the complete (long) version published by Huguenin (2012 – in English –, 2013a – in French –) and of the short version published by Huguenin (2013) in Ishizaka and Nemery (2013)[116]. The short version distinguishes itself from the complete version by the fact that it contains only the CRS model (and not the VRS model as in the complete version). These two original versions contain several practical exercises about DEA to help decision makers master the method. However, as a member of this PhD jury pointed out, such practical exercices cannot be included in a PhD thesis. As a result, they have been removed from the original versions and do not appear in the current modified version.

[116] Note that the complete version of the guide (Huguenin 2012; 2013a) has already been diffused in Australia, Canada, France, India, Morrocco and Switzerland, following the request from professors and graduate students addressed to the author.

1 Introduction

This guide introduces Data Envelopment Analysis (DEA), a performance measurement technique, in such a way as to be appropriate to decision makers with little or no background in economics and operational research. The use of mathematics is kept to a minimum. This guide therefore adopts a strong practical approach in order to allow decision makers to conduct their own efficiency analysis and to easily interpret results.

DEA helps decision makers for the following reasons:

- By calculating an efficiency score, it indicates if an entity is efficient or has capacity for improvement.
- By setting target values for input and output, it calculates how much input must be decreased or output increased in order to become efficient.
- By identifying the nature of returns to scale, it indicates if an entity has to decrease or increase its scale (or size) in order to minimize the average cost.
- By identifying a set of benchmarks, it specifies which other entities' processes need to be analysed in order to improve its own practices.

After this introduction, Section 22 presents the essentials about DEA, alongside a case study to intuitively understand its application. Section 3 introduces *Win4DEAP*, a software package that conducts efficiency analysis based on DEA methodology. Section 4 is dedicated to more demanding readers interested in the methodical background of DEA. Four advanced topics of DEA (adjustment to the environment; preferences; sensitivity analysis; time series data) are presented in Section 5. Finally, Section 6 shows how to program the Solver in *Microsoft Excel* ® in order to run a basic DEA efficiency analysis.

2 Basics of DEA

2.1 An efficiency measurement method

DEA is used to measure the performance of firms or entities (called Decision-Making Units – DMUs –) which convert multiple inputs into multiple outputs. Entity efficiency is defined as the ratio of the sum of its weighted outputs to the sum of its weighted inputs (Thanassoulis *et al.*, 2008, p. 264). DEA is suitable for the use of both private sector firms and public sector organizations (and even for entities such as regions, countries, etc.)[117]. It was formulated in Charnes *et al.* (1978, 1981) in order to evaluate a US federal government program in the education system called 'Program Follow Through'. The use of DEA then spread to other public organizations (hospitals,

---

[117] DMU stands for Decision Making Units. A DMU is an entity which converts inputs into outputs. Some authors, as Coelli, Prasada Rao, O'Donnell and Battese (2005), use the term of firm instead of DMU. However, in this pedagogical guide and following the practice of the International Public Sector Accounting Standards Board (2012), the term of entity is used instead of DMU or firm.

aged-care facilities, social service units, unemployment offices, police forces, army units, prisons, waste management services, power plants, public transportation companies, forestry companies, libraries, museums, theatres, etc.) and to the private sector (banks, insurance companies, retail stores, etc.).

Each entity's efficiency score is calculated relative to an efficiency frontier. Entities located on the efficiency frontier have an efficiency score of 1 (or 100%). Entities operating beneath the frontier have an efficiency score inferior to 1 (or 100%) and hence have the capacity to improve future performance. Note that no entity can be located above the efficiency frontier because they cannot have an efficiency score greater than 100%. Entities located on the frontier serve as benchmarks – or peers – to inefficient entities. These benchmarks (i.e. real entities with real data) are associated with best practices. DEA is therefore a powerful benchmarking technique.

2.2 Case study 1

To better understand the mechanics behind DEA, this section develops a simple practical case study. It includes only one input and one output, although DEA can handle multiple inputs and multiple outputs.

Five register offices (A to E) produce one output (total number of documents, such as marriage or birth certificates) with one input (number of full-time equivalent public servants)[118]. The data are listed in Table 1. For example, two public servants work in Register Office A. They produce one document (during a certain period of time).

Table 1
Case study 1 – Five register offices produce documents with public servants

Register Office	Input	Output
	Public servants (x)	Documents (y)
A	2	1
B	3	4
C	5	5
D	4	3
E	6	7

a) Case study 1 – Two basic DEA models

Two basic models are used in DEA, leading to the identification of two different frontiers:

- The first model assumes constant returns to scale technology (CRS model). This is appropriate when all entities are operating at an optimal scale. However, note that this is quite an ambitious assumption. To operate at an

[118] Note that DEA can handle more outputs and inputs. In order to represent this example in a two-dimensional graph, we consider a total of two outputs and inputs of two (one output, one input; no variable representing the quality of the variables).

optimal scale, entities should evolve in a perfectly competitive environment, which is seldom the case. The CRS model calculates an efficiency score called constant returns to scale technical efficiency (CRSTE).

- The second model assumes variable returns to scale technology (VRS model). This is appropriate when entities are not operating at an optimal scale. This is usually the case when entities face imperfect competition, government regulations, etc. The VRS model calculates an efficiency score called variable returns to scale technical efficiency (VRSTE).

Comparison between the two models reveals the source of inefficiency. Constant returns to scale technical efficiency corresponds to the global measure of entity performance. It is composed by a 'pure' technical efficiency measure (captured by the variable returns to scale technical efficiency score) and a scale efficiency measure (SE). Section 4.2 demonstrates how these three notions (CRSTE, VRSTE and SE) relate to each other.

b) Case study 1 – Input or output orientation

A DEA model can be input or output oriented:

- In an input orientation, DEA minimizes input for a given level of output; in other words, it indicates how much an entity can decrease its input for a given level of output.
- In an output orientation, DEA maximizes output for a given level of input; in other words, it indicates how much an entity can increase its output for a given level of input.

The efficiency frontier will be different in a CRS or a VRS model (see Section 4.2). However, within each model, the frontier will not be affected by an input or an output orientation. For example, the efficiency frontier under VRS will be exactly the same in an input or an output orientation. Entities located on the frontier in an input orientation will also be on the frontier in an output orientation.

In a CRS model, technical efficiency scores have the same values in an input or an output orientation. But these values will be different according to the model's orientation when VRS is assumed. However, Coelli and Perelman (1996, 1999) note that, in many instances, the choice of orientation has only a minor influence upon the technical efficiency scores calculated in a VRS model.

Choosing between an input or an output orientation

The model's orientation should be selected according to which variables (inputs or outputs) the decision maker has most control over. For example, a school principal will probably have more control over his teaching staff (input) than over the number of pupils (output). An input orientation will be more appropriate in this case.

In the public sector, but sometimes also in the private, a given level of input can be granted and secured to an entity. In this case, the decision maker may want to maximize the output (and therefore choose an output orientation). Alternatively, if the decision maker's task is to produce a given level of output (e.g. a quota) with the minimum input, he will opt for an input orientation.

If the decision maker is not facing any constraints and has control of both input and output, the model's orientation will depend on his objectives. Does he need to cut costs (input orientation) or does he want to maximize production (output orientation)?

c) Case study 1 – CRS efficient frontier

Figure 1 represents the efficient frontier assuming constant returns to scale technology (CRS efficient frontier). The CRS efficient frontier starts at the origin and runs through Register Office B. Register Office B happens to be the observation with the steepest slope, or the highest productivity ratio, among all register offices (4 / 3 = 1.33, meaning that one public servant produces 1.33 documents). Register Office B is on the frontier; it is 100% efficient. Register Offices A, C, D and E are beneath the frontier. Their respective efficiency scores are less than 100%. DEA assumes that the production possibility set is bounded by the frontier. This actually implies that DEA calculates relative and not absolute efficiency scores. Although entities on the efficient frontier are granted a 100% efficiency score, it is likely that they could further improve their productivity.

Figure 1
Case study 1 – Register offices beneath the efficient frontier have the capacity to improve performance.

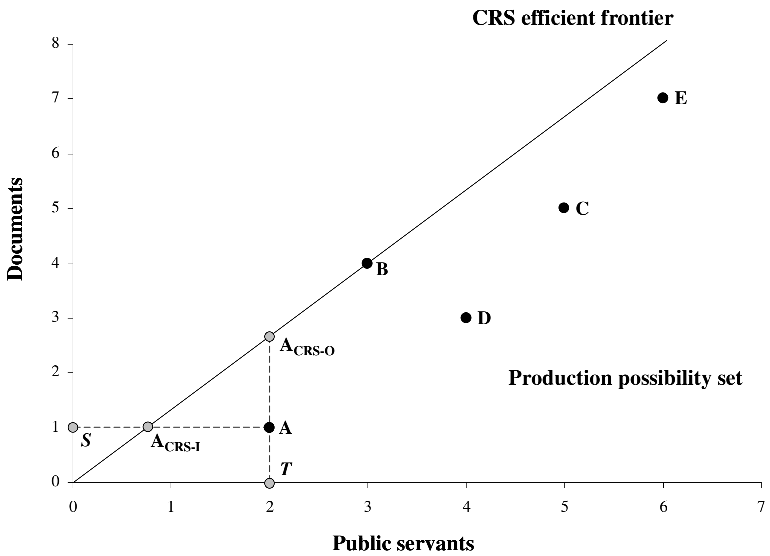


Figure 1 also illustrates how DEA measures efficiency scores. The example of Register Office A is described below:

- In an input orientation, A's efficiency score is equal to the distance $SA_{CRS\text{-}I}$ divided by the distance SA. $A_{CRS\text{-}I}$ is the projection of point A on the efficient frontier (assuming constant returns to scale – CRS – and an input orientation – I –). Note that one can easily calculate efficiency scores using a ruler and measuring the distances on the graph. A's score is 37.5%. This means that Register Office A could reduce the number of public servants employed (input) by 62.5% (100 – 37.5) and still be able to produce the same number of documents (one).
- In an output orientation, A's efficiency score is equal to the distance TA divided by the distance $TA_{CRS\text{-}O}$. $A_{CRS\text{-}O}$ is the projection of point A on the efficient frontier (assuming constant returns to scale – CRS – and an output orientation – O –). A's score is 37.5%, as in an input orientation[119]. This means that Register Office A could increase its production of documents (output) by 62.5% (100 – 37.5) whilst holding the number of public servants constant at two.

d) Case study 1 – VRS efficient frontier

Figure 2 represents the efficient frontier assuming variable returns to scale technology (VRS efficient frontier). The VRS efficient frontier is formed by

119 Note that the efficiency scores in a CRS model are always the same for an input or an output orientation.

enveloping all the observations. Register Offices A, B and E are on the frontier. They are 100% efficient. Register Offices C and D are beneath the frontier. Their respective efficiency scores are inferior to 100%. DEA assumes that the production possibility set is bounded by the frontier. Again, this implies that DEA calculates relative and not absolute efficiency scores. Although entities on the efficient frontier are granted a 100% efficiency score, it is likely that they could further improve their productivity.

Figure 2
Case study 1 – For the same level of input, Register Office D could improve its output up to the projected values of point $D_{VRS\text{-}O}$ (i.e. from 3 to 5 documents).

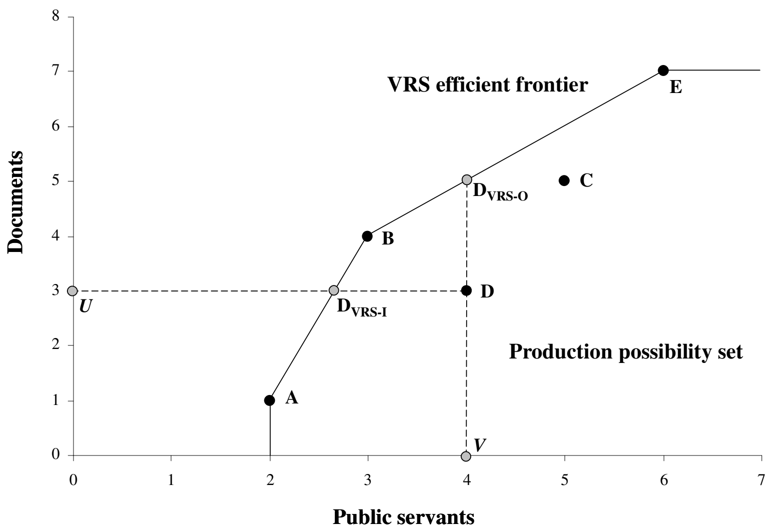


Figure 2 also illustrates how DEA measures efficiency scores. The example of Register office D is described below:

- In an input orientation, D's efficiency score is equal to the distance $UD_{VRS\text{-}I}$ divided by the distance UD. $D_{VRS\text{-}I}$ is the projection of point D on the efficient frontier (assuming variable returns to scale – VRS – and an input orientation – I –). Note that one can easily calculate efficiency scores using a ruler and measuring the distances on the graph. D's score is 66.7%. This means that Register Office D could reduce the number of public servants employed (input) by 33.3% (100 – 66.7) and still be able to produce the same number of documents (three).
- In an output orientation, D's efficiency score is equal to the distance VD divided by the distance $VD_{VRS\text{-}O}$. $D_{VRS\text{-}O}$ is the projection of point D on the efficient frontier (assuming variable returns to scale – VRS – and an output orientation – O –). D's score is 60%[120]. This means that Register Office D

120 Note that the efficiency scores in a VRS model are different for an input or an output orientation.

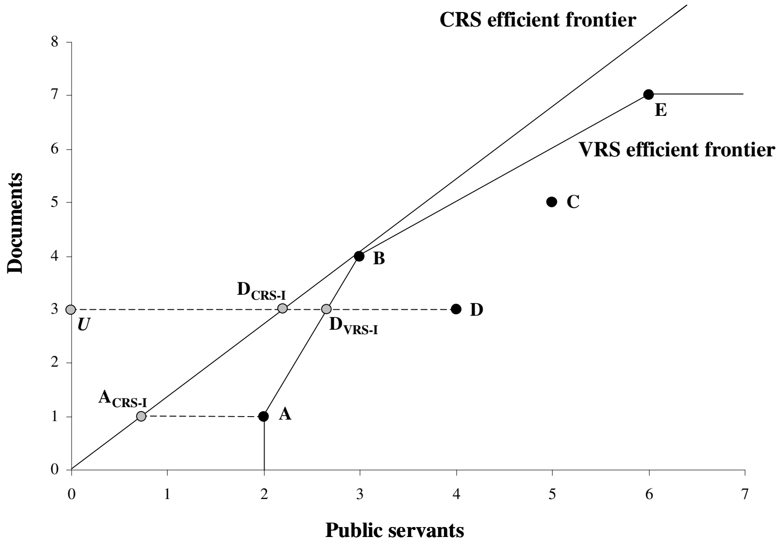
could increase its production of documents (output) by 40% (100 – 60) whilst holding the number of public servants constant at four.

> How to interpret efficiency scores according to the DEA model's output or input orientation
>
> Register Office C has an efficiency score of 75% in the CRS model. It will get the same efficiency score in an output or in an input-oriented model under the constant returns to scale assumption. However:
>
> - In the input-oriented model, the capacity to improve input (i.e. a reduction) by 25% (100 – 75) is calculated using the original input value of 5 public servants. The DEA model calculates a projected value of 3.75. The 25% improvement is then calculated according to the original value: ((5 – 3.75) / 5) x 100 = 25. From a practical point of view, the capacity to improve input by 25% means that the Register Office should reduce all of its inputs by 25% in order to become efficient.
> - In the output-oriented model, the capacity to improve output (i.e. an augmentation) by 25% (100 – 75) is calculated using the projected output value. Register Office C has an original output value of 5 documents. The DEA model calculates a projected value of 6.67 documents. The 25% improvement is calculated according to the projected value: ((6.67 – 5) / 6.667) x 100 = 25. From a practical point of view, the capacity to improve output by 25% means that the Register Office should augment all of its outputs by 25% in order to become efficient.

e) Case study 1 – CRS, VRS and scale efficiency

Figure 3 represents the CRS and the VRS efficient frontiers on the same graph. Register Office B is CRS and VRS efficient, as it is located on both frontiers. Register Offices A and E are efficient under the variable returns to scale assumption but inefficient under the constant returns to scale assumption. Finally, Register Offices D and C are both CRS and VRS inefficient; they are located neither on the CRS nor on the VRS frontiers.

Figure 3
Case study 1 – Register Offices A and E are VRS efficient but CRS inefficient.



The gap observed between the CRS and the VRS frontiers is due to a problem of scale. For example, Register Office A is VRS efficient. To become CRS efficient, Register Office A should modify its scale (or size). Only by operating at point ACRS-I would Register Office A be as productive as Register Office B, which is the only CRS efficient Register Office.

Some Register Offices (D and C) are not even located on the VRS frontier. These Register Offices not only have a scale problem but are also poorly managed. For example, Register Office D should move to point D_{VRS-I} located on the VRS frontier in order to become VRS efficient (i.e. to eliminate the inefficiency attributable to poor management). Furthermore, Register Office D should move from point D_{VRS-I} to point D_{CRS-I} located on the CRS frontier in order to become CRS efficient (i.e. to eliminate the inefficiency attributable to a problem of scale).

As a result, the CRS efficiency (also called 'total' efficiency) can be decomposed into two components: the VRS efficiency (also called 'pure' efficiency) and the scale efficiency. The following ratios represent these three types of efficiency for Register Office D (input orientation).

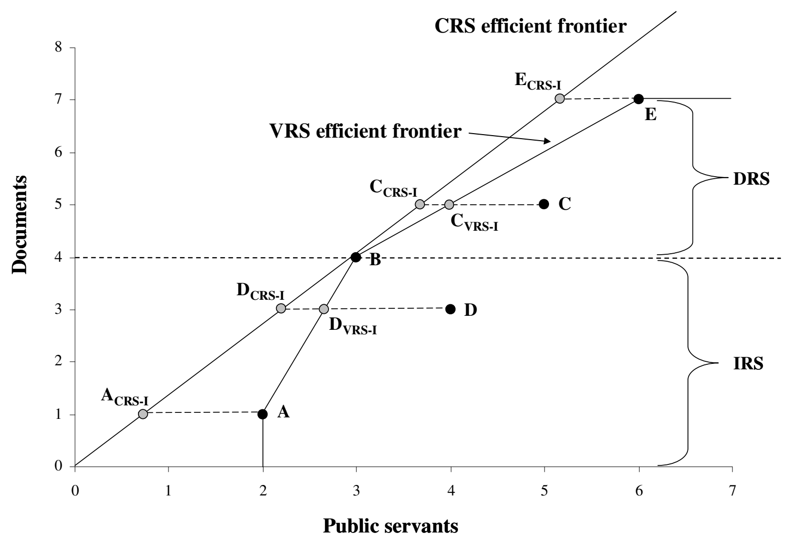
Technical efficiency of D under CRS	Technical efficiency of D under VRS	Scale efficiency of D
$TE_{CRS} = \frac{UD_{CRS-I}}{UD} = 56.3\%$	$TE_{VRS} = \frac{UD_{VRS-I}}{UD} = 66.7\%$	$SE = \frac{UD_{CRS-I}}{UD_{VRS-I}} = 84.4\%$

f) Case study 1 – Nature of returns to scale

The nature of returns to scale of register offices not located on the CRS frontier (in other words, scale inefficient) has to be identified. Figure 4 represents the

CRS efficient points A_{CRS-I} and E_{CRS-I} of Register Offices A and E (which are CRS inefficient but VRS efficient). It also represents the CRS efficient points D_{CRS-I} and C_{CRS-I} and the VRS efficient points D_{VRS-I} and C_{VRS-I} of Register Offices D and C (which are CRS and VRS inefficient).

Figure 4
Case study 1 – Register Offices A and D face increasing returns to scale – IRS – (economies of scale); C and D face decreasing returns to scale – DRS – (diseconomies of scale).



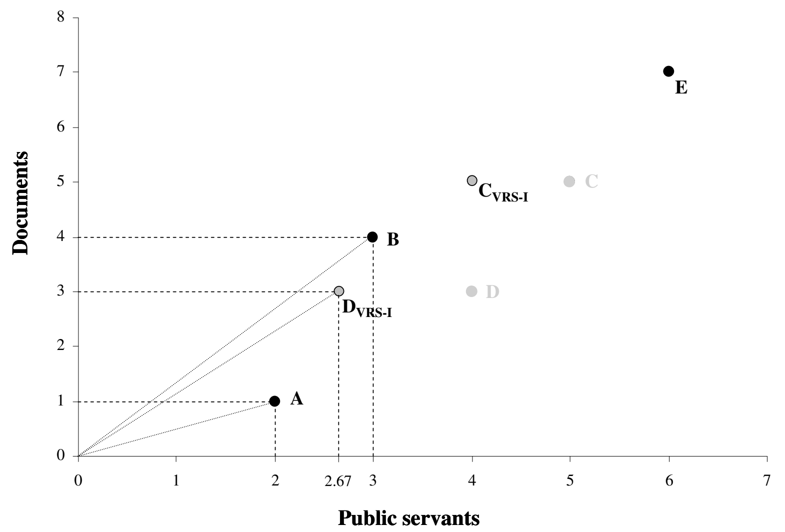
To identify the nature of returns to scale, one has to focus on the slope of the VRS efficient points A, D_{VRS-I}, B, C_{VRS-I} and E (or productivity). Three situations can occur:

- A register office is located both on the CRS and the VRS efficient frontiers (such as point B). Register Office B has the highest productivity of all VRS efficient points (4 / 3 = 1.33). It is facing constant returns to scale. Such an entity reaches its optimal size (or efficient scale)[121]. It is operating at a point where the scale (or size) has no impact on productivity. This situation occurs when the average inputs consumption is minimized and does not vary with output. In a situation of constant returns to scale, an increase in output of 1 percent requires a proportionate increase in input (i.e. 1 percent).
- A register office (or the projected point of a register office) is located at a point where the scale (or the size) has a positive impact on productivity.

[121] In the economic context, an entity operates at the optimal size (or efficient scale) when it minimizes its average cost. In the context of DEA, we can measure efficiency in physical or in monetary terms. Because cost and price information is not always available or appropriate, the use of technical efficiency is often preferred. As this latter measure is based on physical terms, we prefer to use the expression of average inputs consumption instead of average cost.

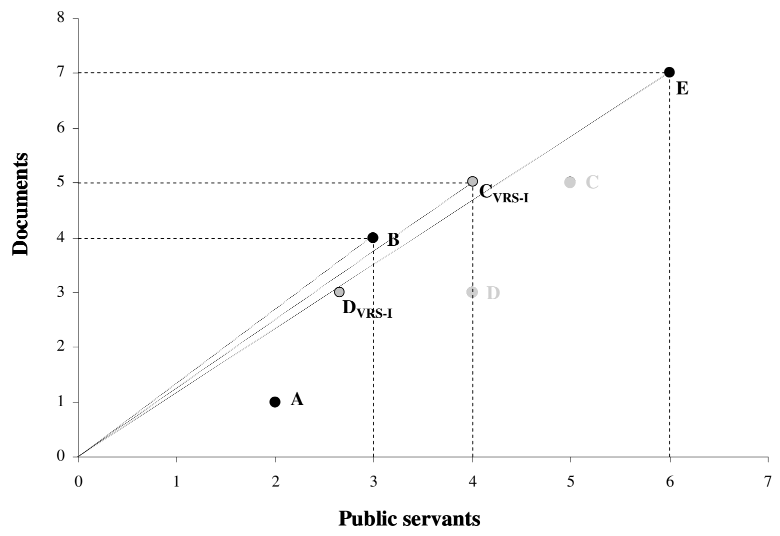
Points A and D_{VRS-I} are in such a position (see Figure 5). The productivity of A (1 / 2 = 0.5) is inferior to the productivity of D_{VRS-I} (3 / 2.67 = 1.12). The ratio of productivity is increasing with the scale. This situation occurs until point B, which has a productivity of 1.33. Register Offices A and D are therefore facing increasing returns to scale (IRS) – or economies of scale –. In this situation, the average inputs consumption declines whilst output rises. Register Offices A and D have not yet reached their optimal size (or efficient scale). To improve their scale efficiency, they have to expand their output. In a situation of economies of scale, a variation in output of 1 percent results in a variation in input of less than 1 percent. Hence, an increase in output results in a reduction of the average inputs consumption.

Figure 5
Case study 1 – The ratio of productivity is increasing with the scale.



- A register office (or the projected point of a register office) is located at a point where the scale (or the size) has a negative impact on productivity. Points C_{VRS-I} and E are in such a position (see Figure 6). The productivity of C_{VRS-I} (5 / 4 = 1.25) is superior to the productivity of E (7 / 6 = 1.17). The ratio of productivity is decreasing with the scale. This situation occurs from point B, which has a productivity of 1.33. Register Offices C and E are therefore facing decreasing returns to scale (DRS) – or diseconomies of scale –. In this situation, the average inputs consumption rises whilst output rises. Register Offices C and E have exceeded their optimal size (or efficient scale). To improve their scale efficiency, they have to reduce their output. In a situation of diseconomies of scale, a variation in output of 1 percent results in a variation in input of more than 1 percent. Hence, a decrease in output results in a reduction of the average inputs consumption.

Figure 6
Case study 1 – The ratio of productivity is decreasing with the scale.



The specific cases of the five Register offices are described below (see Figure 4):

- Register Office A is located on the VRS frontier but not on the CRS frontier. Its inefficiency is due to an inappropriate scale. A is facing increasing returns to scale. A variation in output of 1 percent results in a variation in input of less than 1 percent.
- Register Office D is neither located on the CRS nor on the VRS frontier. Its inefficiency is due to an inappropriate scale and to poor management. D is facing increasing returns to scale. A variation in output of 1 percent results in a variation in input of less than 1 percent.
- Register Office B is located both on the CRS and on the VRS frontier. It has no inefficiency at all. B is facing constant returns to scale. A variation in output of 1 percent results in a variation in input of 1 percent.
- Register Office C is neither located on the CRS nor on the VRS frontier. Its inefficiency is due to an inappropriate scale and to poor management. C is facing decreasing returns to scale. A variation in output of 1 percent results in a variation in input of more than 1 percent.
- Register office E is located on the VRS frontier (but not on the CRS frontier). Its inefficiency is due to an inappropriate scale. E is evolving in a situation of decreasing returns to scale. A variation in output of 1 percent results in a variation in input of more than 1 percent.

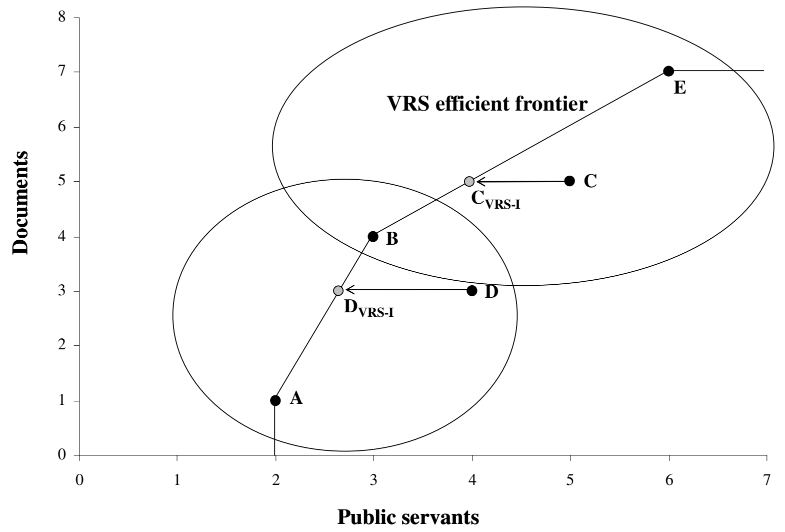
g) Case study 1 – Peers (or benchmarks)

DEA identifies, for each inefficient entity, the closest efficient entities located on the frontier. These efficient entities are called peers or benchmarks. If inefficient entities want to improve their performance, they have to look at the best practices developed by their respective peers.

Under the CRS assumption, Register Office B is the only entity located on the efficient frontier. Hence it is identified as the peer for all other inefficient register offices.

Figure 7 illustrates the peers under the VRS assumption. Three Register Offices (A, B and E) are located on the efficient frontier. Two Register Offices (C and D) are inefficient. Register Office C has two assigned peers: B and E. $C_{VRS\text{-}I}$, the projected point of C on the VRS frontier, lies between these two benchmarks. Register Office D also has two assigned peers: A and B. $D_{VRS\text{-}I}$, the projected point of D on the VRS frontier, lies between these two benchmarks.

Figure 7
Case study 1 – Register Offices A and B are peers of Register Office D; Register Offices B and E are peers of Register Office C.

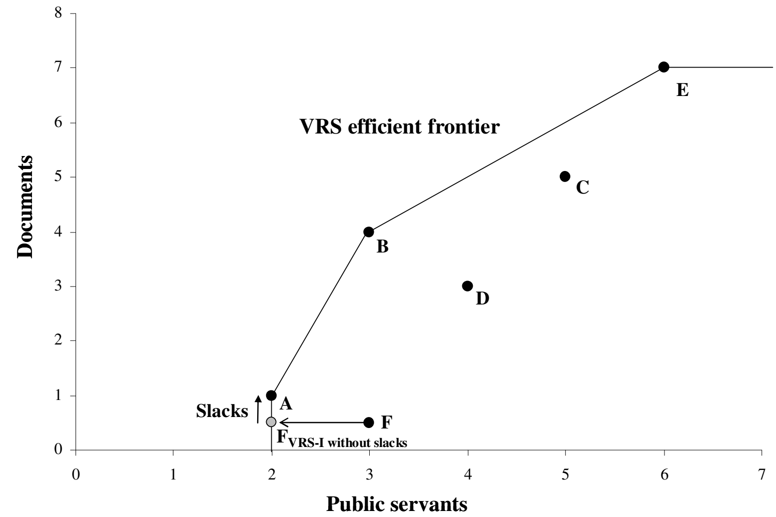


h) Case study 1 – Slacks

Particular positions located on the frontier are inefficient. Assume there is an additional register office in our sample, F. It produces 0.5 document with two public servants. Figure 8 illustrates the efficient frontier under VRS. Register Office F is not located on the frontier. In order to become efficient, it has first to move to point $F_{VRS\text{-}I\ without\ slacks}$. At this location, Register Office F should have an efficiency score of 100%, as it is located on the frontier. But Register Office A, next to him on the frontier, is also 100% efficient. The difference between F and A is striking. With the same number of inputs (two public servants), F produces 0.5 document and A produces one document (i.e. 0.5 more than F). Therefore point $F_{VRS\text{-}I\ without\ slacks}$ cannot be considered as 100% efficient, because it produces less output with the same amount of input than another register office (A). To get a 100% efficiency score, point $F_{VRS\text{-}I\ without\ slacks}$ has to move further up to point A. This additional improvement needed for an entity to become efficient is called a slack.

Indeed, every point located on the sections of the frontier which run parallel to either the x or the y axes has to be adjusted for slacks. DEA is designed to take slacks into account.

Figure 8
Case study 1 – Register Offices A and B are peers of Register Office D; Register Offices B and E are peers of Register Office C.



2.3 Multiple outputs and inputs

DEA allows multiple outputs and multiple inputs to be taken into account. For example, a shirt company uses machines, workers and tissue (three inputs) in order to produce T-shirts, pants and underwear (three outputs). DEA can account for all of these variables and even more. As a result, DEA goes far beyond the calculation of single productivity ratios such as, for example, the number of T-shirts produced per worker (one output divided by one input).

However, the total number of outputs and inputs being considered is not limitless from a practical point of view. It depends on the number of entities in the data set. If the number of entities is smaller than, roughly speaking, three times the sum of the total number of inputs and outputs, it is highly probable that several entities, if not all, will obtain a 100% score[122]. For example, a dataset containing 21 shirt companies allows a total of seven outputs and inputs to be dealt with (21 divided by 3). As Cooper *et al.* (2006) point out,

> if the number of DMUs (n) is less than the combined number of inputs and outputs (m + s), a large portion of the DMUs will be identified as efficient and efficiency discrimination among DMU is questionable

---

122 The higher the number of inputs and outputs that are taken into consideration for a given number of entities, the more probable it is that each entity will be the best producer of at least one of the outputs. Therefore, all entities could obtain a 100% efficiency score.

> due to an inadequate number of degrees of freedom. (…). Hence, it is desirable that n exceeds m + s by several times. A rough rule of numbs in the envelopment model is to choose n (= the number of DMUs) equal to or greater than max {m x s, 3 x (m + s)} (p. 106).

DEA measures entity efficiency based on multiple outputs and multiple inputs. If Shirt Company A produces a lot of T-shirts but only a few pants and underwear, DEA will automatically attribute a high weighting to the T-shirts variable in order to emphasize this strength. As a result, DEA 'automatically' optimizes the weighting of each variable in order to present each entity in the best possible light.

The particularity of DEA is that weights assigned to outputs and inputs are not decided by users. Moreover, it does not use a common set of weights for all entities. Instead, a different set of weights is calculated by a linear optimization procedure.

Unfortunately, DEA does not work with negative or zero values for inputs and outputs. However, zero values can be substituted with very low values such as 0.01.

It is also noted that each DMU must have the same number of inputs and outputs in order to be compared, otherwise DEA cannot be applied.

A distinction has to be made between variables which are under the control of management (discretionary variables) and variables which are not (non-discretionary or environmental variables). Ideally, a DEA model will exclusively include discretionary variables although some DEA models can also accommodate non-discretionary. In a second step, efficiency scores can be adjusted to account for environmental variables (i.e. such variables influence the efficiency of an entity but are not a traditional input and are not under the control of the manager).

Moreover, variables should reflect both quantitative and qualitative characteristics of entities' resources and services. Although it may not be easy to identify and to convert qualitative characteristics into numbers, it is desirable to include such variables in the model in order to appropriately benchmark entities.

2.4 Types of efficiency

The notion of efficiency refers to an optimal situation; the maximum output for a given level of input or the minimum input for a given level of output. Subject to data availability, several types of efficiency can be measured:

- Technical efficiency, in which both outputs and inputs are measured in physical terms[123].
- Cost efficiency: identical to technical efficiency, except that cost (or price) information about inputs is added to the model.
- Revenue efficiency: identical to technical efficiency, except that price information about outputs is added to the model.
- Profit efficiency: identical to technical efficiency, except that cost information about inputs and price information about outputs are added to the model.

Technical efficiency is a global measure of entity performance. However, it does not indicate the source of inefficiency. This source could be twofold:

- First, the entity could be poorly managed and operated.
- Second, it could be penalised for not operating at the right scale.

Technical efficiency can be decomposed into a 'pure' technical efficiency measure and a scale efficiency measure to reflect these two sources of inefficiency[124].

2.5 Managerial implications

DEA is a benchmarking technique. The efficiency scores provide information about an entity's capacity to improve output or input. In this sense, DEA offers strong support to decision making. To conduct an efficiency analysis and to interpret results often raises practical questions. The following list of frequently asked questions offers some advice.

- Is it advisable to involve the managers of the entities to be benchmarked in the efficiency analysis from the beginning of the process?

  Yes, it is, and for two main reasons. First, managers know the processes of their entities and the data available. Therefore they are the right persons to pertinently identify which inputs and outputs have to be integrated into the analysis. Second, managers involved from the beginning of the process are more likely to accept the results of the analysis (rather than to reject them) if they have been involved in the process.

---

123 This pedagogical guide will focus on the measurement of technical efficiency for two main reasons: first, entities in the public sector are often not responsible for the age pyramid of their employees; therefore taking into account the wages of the employees (which often grow higher alongside seniority) would unfairly alter efficiency of an entity with a greater proportion of senior employees; second, entities in the public sector do not often produce commercial goods or services with a set price.

124 The entity's management team will definitely be held responsible for the 'pure' technical efficiency score. In a situation where it does not have the discretionary power to modify the entity's size, it will likely not be accountable for the scale efficiency score. However, especially in the private sector, one has the choice of the scale at which it operates: the management team can easily downsize the entity and, with some efforts, upsize it also.

- How should one respond to managers who claim that their entities are different from others, and therefore cannot be compared to them?

  Sometimes, inefficiencies can be explained by indisputable environmental variables. But sometimes they cannot. Managers often justify the low efficiency scores of their entities by arguing that their situations are different compared to the situations of the other entities. They claim to be a 'special case' (and therefore it is acceptable to be inefficient). Actually, the majority of entities could possibly claim to be different as most possess a specificity that others do not have. However, it is likely that the difference of one entity will be compensated by the difference of another. More generally, it is up to the managers to prove that they really face a hostile environment. If they cannot prove it, management measures have to be taken to improve efficiency.

- Assume that an entity obtains an efficiency score of 86.3%. Does this number have to be strictly applied?

  Not really, it should be interpreted more as an order of magnitude. This order of magnitude informs managers that they have to increase their outputs or to decrease their inputs in order to become more efficient. But one should not focus too strictly on the capacity for 13.7% improvement. Such a number could be interpreted by practitioners as too 'accurate' and may offend their sensibilities. Therefore it is better to consider efficiency scores more as more of an objective basis to hold an open discussion about the way to improve entity efficiency rather than a number to be strictly applied.

- An entity faces increasing returns to scale. It has economies of scale. What does that concretely mean from a managerial point of view?

  Such an entity has not yet reached its optimal size. In order to reduce its average cost (or its average inputs consumption), it has to increase its size. Practically, this could be done either by internal growth (i.e. producing more output) or by merging with another entity which is also facing increasing returns to scale. If, for some reason, managers cannot influence the scale of an entity, they should not be held accountable for this source of inefficiency.

- An entity faces decreasing returns to scale. It has diseconomies of scale. What does that concretely mean from a managerial point of view?

  Such an entity is already oversized, having exceeded its optimal size. In order to reduce its average cost (or its average inputs consumption), it has to decrease its size. Practically, this could be done either by internal decay (i.e. producing less output) or by splitting the entity into two separate businesses. Note that some of the production could be transferred to an entity facing increasing returns to scale. If, for some reason, managers cannot influence the scale of an entity, they should not be held accountable for this source of inefficiency.

- Is efficiency the only criteria to assess an entity's performance?

Not necessarily. Basically, the assessment of an entity's performance will depend on the management objectives. Other criteria such as effectiveness or equity are often considered alongside efficiency. If this is the case, the overall performance should be balanced with the various criteria.

- One entity obtains a score of 100% but all the others in the dataset obtain much lower scores (for example, starting at 40% or lower). Is this realistic?

  It could be realistic, but the gap appears to be important. In such a case, data have to be carefully checked, and especially data of the efficient entity. If a data problem is not identified, such results mean that the efficient entity is likely to have completely different processes than the other entities. It should therefore be absolutely presented as a best practice model. However, even if they are realistic, such results are likely to be rejected by managers whose entities have low efficiency scores. These managers are likely to be discouraged because it is obviously unrealistic for them to improve their entity's efficiency by 60% (or more) in the short run. Therefore it is better to exclude the efficient entity from the sample and to run a new model.

- Almost all the entities obtain an efficiency score of 100%. Does that mean that all of them are really efficient?

  Yes, it could mean that all the entities are efficient. Such results would be great! But they are unlikely. Here, the total number of inputs and outputs is probably too high compared to the number of entities in the dataset. In this case, one variable has to be excluded and a new model has to be run. If the number of entities obtaining a 100% score decreases, it indicates that the number of variables was too high compared to the number of entities. If not, all the entities are just efficient and must be congratulated.

- The model does not show any results. What does that mean?

  Data has to be checked. This could happen when data with a value of zero are in the set. Zeros have to be substituted by a very small number (0.01).

3 DEA software

3.1 Existing software

The user-friendly software packages of DEA incorporate intuitive graphical user interfaces and automatic calculation of efficiency scores. Some of them are compatible with *Microsoft Excel* ®. For a survey of DEA software packages, one can refer to Barr (2004). Today, several software packages have been developed:

- Free packages include *DEAP* (Timothy Coelli, Coelli Economic Consulting Services) and *Win4DEAP* (Michel Deslierres, University of Moncton), *Benchmarking package in R* (Peter Bogetoft, Copenhagen Business School, and Lars Otto, University of Copenhagen), *Efficiency Measurement System* (Holger Scheel, University of Dortmund) or *DEA*

Solver Online (Andreas Kleine and Günter Winterholer, University of Hohenheim).

- Commercial packages include *DEAFrontier* ®[125] (Joe Zhu, Worcester Polytechnic Institute), *DEA-Solver PRO* ®[126] (Saitech, Inc.), *PIM-DEA* ® (Ali Emrouznejad, Aston Business School) or *Frontier Analyst* (Banxia Software Ltd).

This section focuses on the 'twin' DEA software packages *DEAP/Win4DEAP*[127]. These packages centre on the basics of DEA, are simple to use and are stable over time. They are freely available[128] and come with data files as examples. As *Win4DEAP* is the Windows based interface of *DEAP* (which is a DOS program), the current section refers only to *Win4DEAP*. All screenshots and icons presented in this section and coming from *DEAP* or *Win4DEAP* are reproduced by permission of Timothy Coelli and Michel Deslierres.

3.2 Case study 2

The use of *Win4DEAP* is illustrated by a case study including a sample of 15 primary schools (see Table 2 below).

The data used in this case study are fictitious (but are very similar to real ones). 15 schools produce one output (number of pupils) with three inputs (number of full-time equivalent teachers, number of full-time administrative staff and number of computers – used as a proxy for technology investment –). For example, School # 8 educates 512 pupils with 28.6 teachers, 1.3 administrative staff and 26 computers.

---

125 Zhu (2003) includes an earlier version of *DEAFrontier* ®, *DEA Excel Solver* ®, on a CD-ROM. This software works only under *Excel* ® 97, 2000 and 2003. It deals with an unlimited number of DMUs and is available at little cost.

126 Cooper *et al.* (2006) include a CD-ROM with a *DEA-Solver* ® version limited at 50 DMUs. It is available at little cost.

127 As *DEAP* is a DOS program, a user friendly Windows interface has been developed for it (*Win4DEAP*). These 'twin' software packages have to be both downloaded and extracted to the same folder. *Win4DEAP* cannot work without *DEAP*.

128 DEAP Version 2.1: http://www.uq.edu.au/economics/cepa/deap.htm
Win4DEAP Version 1.1.3: http://www8.umoncton.ca/umcm-deslierres_michel/dea/install.html

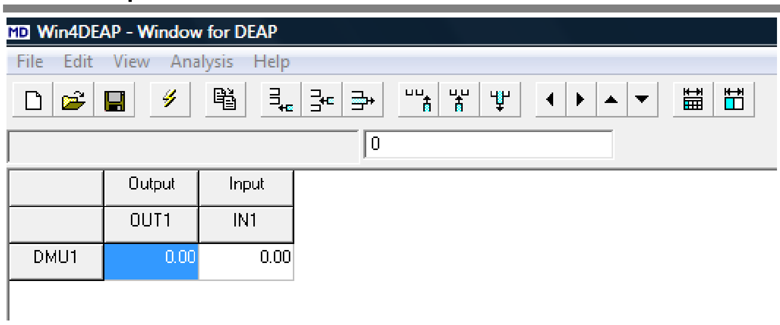
Table 2
Case study 2 – On average, each school has 393.6 pupils, 23.2 teachers, 1.3 administrative staff and 21.6 computers.

School	Input			Output
	FTE teachers	FTE adm. staff	Computers	Pupils
1	40.2	2.0	37.0	602.0
2	18.1	1.1	17.0	269.0
3	42.5	2.1	41.0	648.0
4	11.0	0.8	10.0	188.0
5	24.8	1.3	22.0	420.0
6	21.1	1.3	19.0	374.0
7	13.5	1.0	13.0	247.0
8	28.6	1.3	26.0	512.0
9	23.5	1.3	22.0	411.0
10	15.9	1.0	15.0	285.0
11	23.2	1.3	22.0	397.0
12	26.0	1.4	25.0	466.0
13	11.1	0.8	11.0	198.0
14	28.8	1.6	26.0	530.0
15	19.7	1.3	18.0	357.0

a) Case study 2 – Building a spreadsheet with *Win4DEAP*

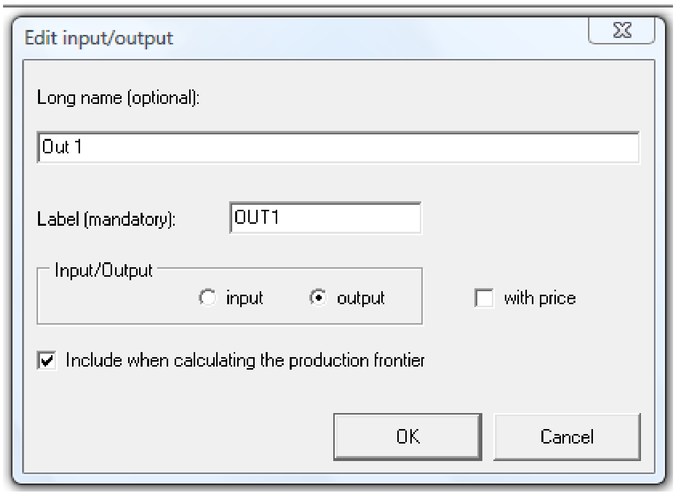
Win4DEAP is launched by clicking the MD icon (MD). Entities (called decision-making units or DMUs) are listed in the rows and variables (outputs and inputs) in the columns. The opening spreadsheet contains one decision-making unit (DMU1), one output (OUT1) and one input (IN1) by default (see Figure 9).

Figure 9
Case study 2 – The opening spreadsheet contains one DMU, one output and one input.



To edit and name entities, outputs and inputs, the user has to click the DMU1 (DMU1), OUT1 (OUT1) and IN1 (IN1) icons, respectively. The window reproduced in Figure 10 allows the user to (1) assign a long name and a label (maximum of eight characters) to any variable and (2) select the nature of the variables (either 'input' or 'output'). Finally, the user has to select the 'with price' option if he intends to measure cost, revenue or profit efficiency (i.e. a 'price' column will be added to the selected variable in the spreadsheet).

Figure 10
Case study 2 – Input and output editing.



The icons enable the user to add entities (DMUs). The icons enable the user to add variables (inputs or outputs). The icons and are used to delete any existing DMUs or variables. Finally, the following icons allow the user to reverse the order of appearance of DMUs (rows) or variables (columns).

How to import Microsoft Excel ® data into Win4DEAP

Note that data can be imported from an *Excel* ® file into *Win4DEAP* by following these steps:

- Save the *Microsoft Excel* ® data (only numbers, no names of DMUs or variables should be included) into the CSV format (Comma delimited).
- In *Win4DEAP*, first select the 'File' menu, then the 'Import' option and finally the 'New data set' application.
- Select the CSV file and open it.
- The data is now presented in the *Win4DEAP* spreadsheet, which still has to be configured (DMUs and variables have to be named and variables must be defined as inputs or outputs).

b) Case study 2 – Running a DEA model

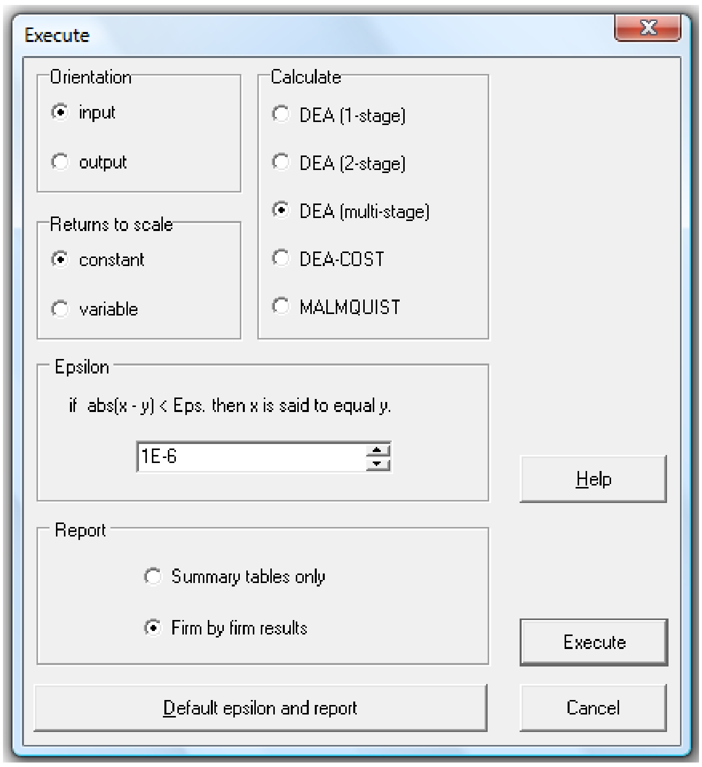
To run a DEA model, the user has to click the 'lightning' icon ( ). The window represented in Figure 11 will then appear. This window allows a calibration of the model following steps 1 to 4 described below:

1. Select an input or an output orientation (Orientation box).
2. Select the assumption about returns to scale (Returns to scale box). By ticking 'constant', one assumes constant returns to scale (CRS); by ticking 'variable', one assumes variable returns to scale (VRS). If one

cannot be certain about the fact that entities are operating at an optimal scale, running a VRS model is recommended.

3. Select a model (Calculate box). Three main options are available:
   - To calculate technical efficiency (TE) or technical (CRS), 'pure' (VRS) and scale efficiency (SE), tick 'DEA (multi-stage)'. Options 'DEA (1-stage)', 'DEA (2-stage)' and 'DEA (multi-stage)' correspond to different treatments of slacks. Following Coelli (1998), the multi-stage treatment is recommended.
   - To calculate cost, revenue or profit efficiency, tick 'DEA-COST'. For this option, cost and/or price information about variables must be available and added to the spreadsheet.
   - To calculate technical and scale efficiency when panel data are available, tick 'MALMQUIST'. See Section 5.4 to learn more about this.
4. Choose how to display the results: only summarized or reported entity by entity (Report box).
5. Click 'Execute' to run the model.

Figure 11
Case study 2 – *Win4DEAP*'s cockpit.



c) Case study 2 – Interpreting results

The model calibrated in Figure 12 is run with the data presented in Figure 13 (or in Table 2). The model presents the following characteristics:

- As the school system is heavily regulated, a variable returns to scale model is required;
- As schools are confronted with budget restrictions, an input orientation is selected;
- Finally, an obligatory school by school report is expected.

Figure 12
Case study 2 – An input oriented model calibrated for VRS.

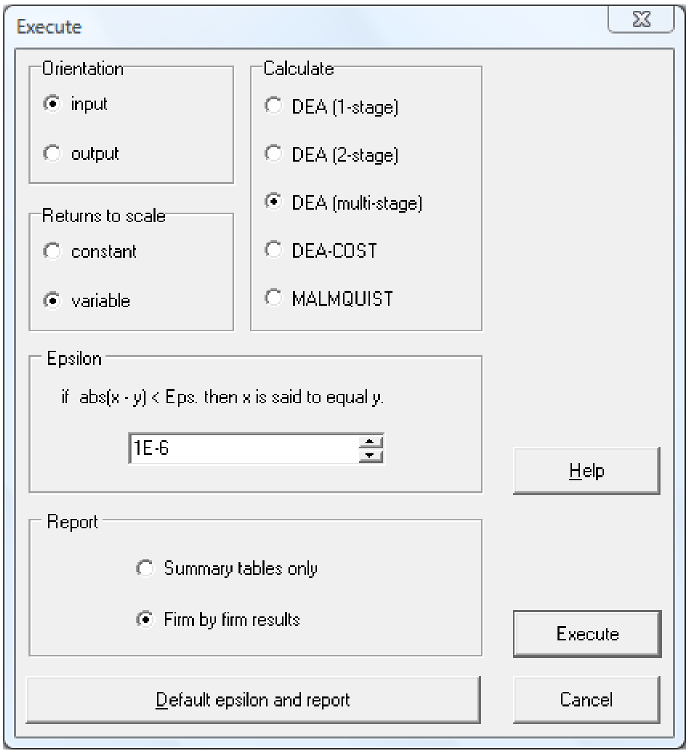
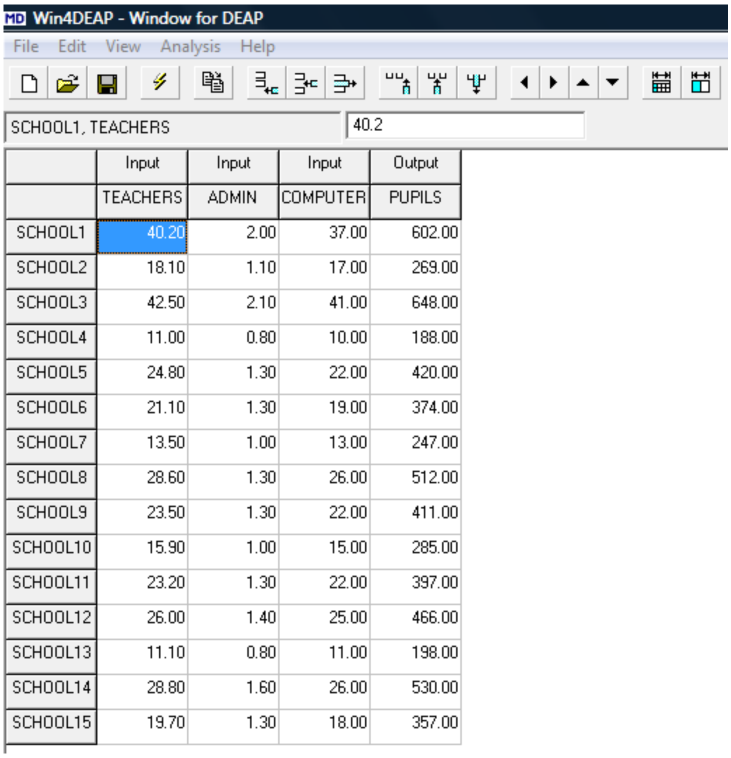


Figure 13
Case study 2 – A ready-to-use spreadsheet in *Win4DEAP*.



	Input	Input	Input	Output
	TEACHERS	ADMIN	COMPUTER	PUPILS
SCHOOL1	40.20	2.00	37.00	602.00
SCHOOL2	18.10	1.10	17.00	269.00
SCHOOL3	42.50	2.10	41.00	648.00
SCHOOL4	11.00	0.80	10.00	188.00
SCHOOL5	24.80	1.30	22.00	420.00
SCHOOL6	21.10	1.30	19.00	374.00
SCHOOL7	13.50	1.00	13.00	247.00
SCHOOL8	28.60	1.30	26.00	512.00
SCHOOL9	23.50	1.30	22.00	411.00
SCHOOL10	15.90	1.00	15.00	285.00
SCHOOL11	23.20	1.30	22.00	397.00
SCHOOL12	26.00	1.40	25.00	466.00
SCHOOL13	11.10	0.80	11.00	198.00
SCHOOL14	28.80	1.60	26.00	530.00
SCHOOL15	19.70	1.30	18.00	357.00

After executing the selected model, a short notice appears with information about Timothy Coelli, the developer of *DEAP*. Results are displayed after closing this window. It is recommendable for first time users to take some time navigating through the results file in order to become familiar with it. Some results tables are commented on in this section. Table 3 contains a list of abbreviations with the main acronyms used in the results file.

Table 3
Case study 2 – A table of abbreviations to help with reading the results file.

Acronym	Full name
DEA	Data Envelopment Analysis
CRS	Constant Returns to Scale
VRS	Variable Returns to Scale
TE	Technical Efficiency
CRSTE	Constant Returns to Scale Technical Efficiency
VRSTE	Variable Returns to Scale Technical Efficiency
SE	Scale Efficiency
IRS	Increasing Returns to Scale
DRS	Decreasing Returns to Scale

Figure 14 represents the first table to be commented on. It is an extract of the results file and features an efficiency summary. The first column contains the 15 schools (listed 1 to 15). The second one displays the constant returns to scale technical efficiency scores (CRSTE)[129]. This 'total' efficiency score is decomposed into a 'pure' technical efficiency measure (variable returns to scale technical efficiency – VRSTE – in the third column) and a scale efficiency measure (scale efficiency – SE – in the fourth column). The last column indicates the nature of returns to scale (IRS, DRS or a dash):

- Entities associated with IRS are facing increasing returns to scale (economies of scale).
- Entities associated with DRS are facing decreasing returns to scale (diseconomies of scale).
- Entities associated with a dash are facing constant returns to scale; they are operating at an optimal scale.

On average, schools efficiency scores are:

- 94% for CRSTE; overall, schools could reduce their inputs by 6% whilst educating the same number of pupils.
- 97.5% for VRSTE; a better school organization would be able to reduce input consumption by 2.5%.
- 96.4% for SE; in adjusting their scale, schools could reduce their inputs by 3.6%.

[129] Note that if you had run a constant returns to scale model instead of a variable returns to scale one, you would have obtained only one type of efficiency score in your results file (technical efficiency – TE –). Technical efficiency scores are strictly equal to constant returns to scale technical efficiency scores obtained in the CRSTE column of your variable returns to scale model.

Figure 14
Case study 2 – Technical efficiency (CRSTE) is decomposed into 'pure' technical efficiency (VRSTE) and scale efficiency (SE).

```
 EFFICIENCY SUMMARY:

  firm  crste  vrste  scale

    1   0.827  0.951  0.869 drs
    2   0.808  0.838  0.964 irs
    3   0.842  1.000  0.842 drs
    4   0.929  1.000  0.929 irs
    5   0.943  0.962  0.981 irs
    6   0.966  0.984  0.981 irs
    7   0.994  1.000  0.994 irs
    8   1.000  1.000  1.000  -
    9   0.951  0.963  0.987 irs
   10   0.974  0.995  0.978 irs
   11   0.930  0.943  0.986 irs
   12   0.978  0.984  0.994 irs
   13   0.969  1.000  0.969 irs
   14   1.000  1.000  1.000  -
   15   0.985  0.998  0.987 irs

 mean   0.940  0.975  0.964

Note: crste = technical efficiency from CRS DEA
      vrste = technical efficiency from VRS DEA
      scale = scale efficiency = crste/vrste


Note also that all subsequent tables refer to VRS results
```

All subsequent tables displayed in the results file refer to the VRSTE scores. These tables contain the following information:

- The number of the DMU under review ('Results for firm').
- The technical efficiency score ('Technical efficiency'), corresponding to the VRSTE when a VRS model has been run or to the CRSTE when a CRS model has been run.
- The scale efficiency score ('Scale efficiency'); note that the SE is mentioned only when a VRS model has been run.
- The lines of the matrix represent the outputs and the inputs of the model ('output 1', 'output 2', etc., 'input 1', 'input 2', etc.).
- The first column of the matrix recalls the original values of the variables' outputs and inputs ('original values').
- The second column of the matrix represents the movement an inefficient DMU has to take in order to be located on the frontier ('radial movement').
- The third column of the matrix is the additional movement a DMU located on a segment of the frontier running parallel to the axis has to take in order to become efficient ('slack movement').
- The fourth column of the matrix lists the values of the variables which enable the DMU to be efficient ('projected value'); these projected values take into account both the radial and the slack movements.

- Finally, the listing of peers is mentioned. Each peer is identified by a number and has an associated weight ('lambda weight') representing the relative importance of the peer.

As illustrations, three individual school tables are specifically commented on below: School # 1 (Figure 15), # 2 (Figure 16) and # 3 (Figure 17).

School # 1 (Figure 15) has a 'pure' efficiency score of 95.1% and a scale efficiency score of 86.9%. It is facing decreasing returns to scale (DRS). By improving the operation of the school, 4.9% (100 – 95.1) of inputs could be saved. By adjusting the school to its optimal size, 13.1% (100 – 86.9) of inputs could be saved.

The 'original value' column contains the original values of the school's variables: School # 1 educates 602 pupils with 40.2 teachers, 2 administrative staff and 37 computers. However, School # 1 could 'produce' the same quantity of output with fewer inputs: 37.186 teachers instead of 40.2; 1.902 administrative staff instead of 2; 35.185 computers instead of 37 (see the 'projected value' column). The decreases in inputs 2 and 3 are equal to 4.9 % of the original values: (–0.098 / 2) x 100 for input 2 and (–1.815 / 37) x 100 for input 3[130]. The case of input 1 is slightly different: to become efficient, it has to decrease not only by 4.9% (minus 1.972 from the 'radial movement' column) but also by an additional 1.042 (from the 'slack movement' column). Overall School # 1 has to decrease its first input by minus 3.014 ((–1.972) + (–1.042)) to become efficient. This represents 7.5 % ((–3.014 / 40.2) x 100).

To improve its efficiency, School # 1 has to analyse the practice of Schools # 3, # 14 and # 8, which are identified as its peers. To be a peer (or a benchmark), an entity must have a 'pure' efficiency score of 100%. The lambda weight associated with each peer corresponds to its relative importance among the peer group. Ideally, School # 1 should analyse best practice from a composite school formed by 61.2% of School # 3, 37.3% of School # 14 and 1.4% of School # 8. As such a 'virtual' school does not exist. School # 1 should concentrate its best practice analysis on the peer associated with the highest lambda value (i.e. School # 3).

130 In a VRS model, the improvement in variables (decrease in inputs or increase in outputs) is calculated according to the VRS technical efficiency score (only). In a CRS model, it is calculated according to the CRS technical efficiency score, or TE score, including not only the pure efficiency but also the scale efficiency.

Figure 15
Case study 2 – School # 1 results table.

```
Results for firm:      1
Technical efficiency = 0.951
Scale efficiency     = 0.869  (drs)
 PROJECTION SUMMARY:
  variable            original        radial         slack     projected
                         value      movement      movement         value
 output      1         602.000         0.000         0.000       602.000
 input       1          40.200        -1.972        -1.042        37.186
 input       2           2.000        -0.098         0.000         1.902
 input       3          37.000        -1.815         0.000        35.185
 LISTING OF PEERS:
  peer   lambda weight
    3       0.612
   14       0.373
    8       0.014
```

School # 2 (Figure 16) has a 'pure' efficiency score of 83.8% and a scale efficiency score of 96.4%. It is facing increasing returns to scale (IRS). By improving the operation of the school, 16.2% (100 – 83.8) of inputs could be saved. By adjusting the school to its optimal size, 3.6% (100 – 96.4) of inputs could be saved.

The 'original value' column contains the original values of the school's variables: School # 2 educates 269 pupils with 18.1 teachers, 1.1 administrative staff and 17 computers. However, School # 2 could 'produce' the same quantity of output with fewer inputs: 15.163 teachers instead of 18.1, 0.922 administrative staff instead of 1.1; 14.242 computers instead of 17 (see the 'projected value' column). The decreases in inputs 1, 2 and 3 are equal to 16.2% of the original values ('radial movement' column). No slack movement is identified.

To improve its efficiency, School # 2 has to refer to Schools # 13, # 4, # 14 and # 8, which are identified as its peers.

Figure 16
Case study 2 – School # 2 results table.

```
Results for firm:      2
Technical efficiency = 0.838
Scale efficiency     = 0.964  (irs)
 PROJECTION SUMMARY:
  variable            original        radial         slack     projected
                         value      movement      movement         value
 output      1         269.000         0.000         0.000       269.000
 input       1          18.100        -2.937         0.000        15.163
 input       2           1.100        -0.178         0.000         0.922
 input       3          17.000        -2.758         0.000        14.242
 LISTING OF PEERS:
  peer   lambda weight
   13       0.506
    4       0.261
   14       0.016
    8       0.218
```

School # 3 (Figure 17) has a 'pure' efficiency score of 100% and a scale efficiency score of 84.2%. It is facing decreasing returns to scale (DRS). This school is well managed. It cannot improve its 'pure' efficiency. The only capacity for improvement lies in a scale adjustment: 15.8% (100 – 84.2) of inputs could be saved.

The ‘original value’ column contains the original values of the school’s variables: School # 3 educates 648 pupils with 42.5 teachers, 2.1 administrative staff and 41 computers. These values are equal to the projected ones (‘pure’ efficiency = 100%).

As School # 3 is purely efficient, it acts as its own peer.

Figure 17
Case study 2 – School # 3 results table.

```
Results for firm:      3
Technical efficiency = 1.000
Scale efficiency     = 0.842  (drs)
 PROJECTION SUMMARY:
  variable           original        radial         slack     projected
                        value      movement      movement         value
 output      1        648.000         0.000         0.000       648.000
 input       1         42.500         0.000         0.000        42.500
 input       2          2.100         0.000         0.000         2.100
 input       3         41.000         0.000         0.000        41.000
 LISTING OF PEERS:
  peer   lambda weight
    3      1.000
```

4 DEA in the black box

This section describes the two principal DEA models: the constant returns to scale model (Charnes *et al.*, 1978) and the variable returns to scale model (Banker *et al.*, 1984). DEA is based on the earlier work of Dantzig (1951) and Farrell (1957), whose approach adopted an input orientation. Zhu and Cook (2008), Cooper *et al.* (2007) or Coelli *et al.* (2005) provide a comprehensive treatment of the methodology. By 2007, Emrouznejad *et al.* (2008) identified more than 4000 research articles about DEA published in scientific journals or books.

DEA is a non-parametric method. Unlike parametric methods (such as ordinary least square, maximum likelihood estimation or stochastic frontier analysis), inputs and outputs are used to compute, using linear programming methods, a hull to represent the efficiency frontier. As a result, a non-parametric method does not require specification of a functional form.

4.1 Constant returns to scale

Charnes *et al.* (1978) propose a model assuming constant returns to scale (CRS model)[131]. It is appropriate when all entities operate at the optimal scale. Efficiency is defined by Charnes *et al.* (1978, p. 430) as “the maximum of a ratio of weighted outputs to weighted inputs subject that the similar ratios for every DMU be less or equal to unity”. The following notation is adopted, as in Johnes (2004):

131 This model is also known as the Charnes, Cooper & Rhodes model (CCR model).

$$TE_k = \frac{\sum_{r=1}^{s} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \qquad (1)$$

Where:

TE_k	is the technical efficiency of entity k using m inputs to produce s outputs
y_{rk}	is the quantity of output r produced by entity k
x_{ik}	is the quantity of input i consumed by entity k
u_r	is the weight of output r
v_i	is the weight of input i
n	is the number of entities to be evaluated
s	is the number of outputs
m	is the number of inputs

The technical efficiency of entity k is maximized under two constraints. First, the weights applied to outputs and inputs of entity k cannot generate an efficiency score greater than 1 when applied to each entity in the dataset (equation # 3). Second, the weights on the outputs and on the inputs are strictly positive (equation # 4). The following linear programming problem has to be solved for each entity:

$$\text{Maximize} \quad \frac{\sum_{r=1}^{s} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \qquad (2)$$

$$\text{Subject to} \quad \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \qquad j = 1, \ldots, n \qquad (3)$$

$$u_r, v_i > 0 \qquad \forall r = 1, \ldots, s; i = 1, \ldots, m \qquad (4)$$

This linear programming problem can be dealt following two different approaches. In the first one, the weighted sums of outputs are maximized holding inputs constant (output-oriented model). In the second one, the weighted sums of inputs are minimized holding outputs constant (input-oriented model)[132]. The primal equations for each model, known as the multiplier form, are given below:

132 Note that the input and output orientations refer to the dual equations of each model (and not to the primal ones).

CRS output-oriented model Primal equation	CRS input-oriented model Primal equation
Minimize $\sum_{i=1}^{m} v_i x_{ik}$ (5)	Maximize $\sum_{r=1}^{s} u_r y_{rk}$ (9)
Subject to	Subject to
$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0 \quad j = 1,\ldots,n$ (6)	$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0 \quad j = 1,\ldots,n$ (10)
$\sum_{r=1}^{s} u_r y_{rk} = 1$ (7)	$\sum_{i=1}^{m} v_i x_{ik} = 1$ (11)
$u_r, v_i > 0 \quad \forall r = 1,\ldots,s; i = 1,\ldots,m$ (8)	$u_r, v_i > 0 \quad \forall r = 1,\ldots,s; i = 1,\ldots,m$ (12)

Using the duality in linear programming, an equivalent form, known as the envelopment form, can be derived from this problem. It is often preferable to solve the computation using the envelopment form because it contains only *s + m* constraints rather than *n + 1* constraints in the multiplier form.

CRS output-oriented model Dual equation	CRS input-oriented model Dual equation
Maximize ϕ_k (13)	Minimize θ_k (17)
Subject to	Subject to
$\phi_k y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \quad r = 1,\ldots,s$ (14)	$y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \quad r = 1,\ldots,s$ (18)
$x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} \geq 0 \quad i = 1,\ldots,m$ (15)	$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} \geq 0 \quad i = 1,\ldots,m$ (19)
$\lambda_j \geq 0 \quad \forall j = 1,\ldots,n$ (16)	$\lambda_j \geq 0 \quad \forall j = 1,\ldots,n$ (20)

Where:

Where:

$1/\phi_k$ and θ_k represent the technical efficiency of entity *k*;

λ_j represents the associated weighting of outputs and inputs of entity *j*.

Every entity located on the sections' envelope running parallel to the axes has to be adjusted for output and input slacks. However, the preceding formulation does not integrate the role of slacks in measuring efficiency. Considering output slacks, s_r, and input slacks, s_i, the above equations become:

CRS output-oriented model Dual equation with slacks	CRS input-oriented model Dual equation with slacks
Maximize $\phi_k + \varepsilon \sum_{r=1}^{s} s_r + \varepsilon \sum_{i=1}^{m} s_i$ (21)	Minimize $\theta_k - \varepsilon \sum_{r=1}^{s} s_r - \varepsilon \sum_{i=1}^{m} s_i$ (25)
Subject to	Subject to
$\phi_k y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} + s_r = 0 \quad r = 1,\ldots,s$ (22)	$y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} + s_r = 0 \quad r = 1,\ldots,s$ (26)
$x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} - s_i = 0 \quad i = 1,\ldots,m$ (23)	$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} - s_i = 0 \quad i = 1,\ldots,m$ (27)
$\lambda_j, s_r, s_i \geq 0 \ \forall j = 1,\ldots,n; r = 1,\ldots,s; i = 1,\ldots,m$ (24)	$\lambda_j, s_r, s_i \geq 0 \ \forall j = 1,\ldots,n; r = 1,\ldots,s; i = 1,\ldots,m$ (28)

Here, ε is a non-Archimedean value defined to be smaller than any positive real number. ε is greater than 0. The entity *k* is efficient only if:

- the efficiency score $TE_k = \left(\frac{1}{\phi_k} \right) = 1$ (or $TE_k = \theta_k = 1$);
- and the slacks $s_r, s_i = 0$, $\forall_r = 1, \ldots s$ and $i = 1, \ldots, m$.

For an in-depth analysis on the treatment of slacks, and especially the multi-stage methodology, see Coelli (1998).

4.2 Variable returns to scale

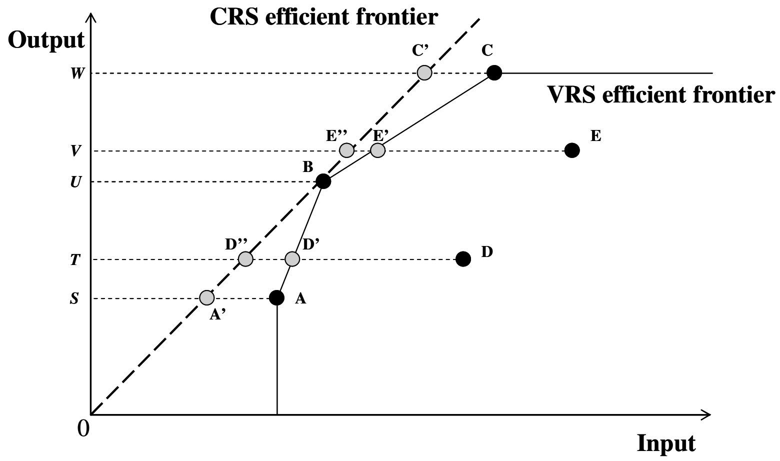
Banker *et al.* (1984) propose a model assuming variable returns to scale (VRS model)[133]. It is appropriate when all entities do not operate at optimal scale. As Coelli *et al.* (2005) point out,

> the use of the CRS specification when not all entities are operating at the optimal scale, results in measures of TE that are confounded by scale efficiencies (SE). The use of the VRS specification permits the calculation of TE devoid of these SE effects (p. 172).

The CRS model can be modified by relaxing the constant returns to scale assumption. A measure of return to scale for entity *k* is added in the primal equation (or the convexity constraint $\sum_{j=1}^{n} \lambda_j = 1$ in the dual equations).

Figure 18 represents the CRS efficiency frontier (the dashed line) and the VRS efficiency frontier (the solid line) on the same graph to illustrate a simple example with one output and one input. Only one entity, B, is located on both frontiers. A and C are 100% efficient under the VRS assumption, but inefficient under the CRS assumption. D and E are inefficient under both specifications.

Figure 18
Constant versus variable returns to scale.



---

133 This model is also known as the Banker, Charnes & Cooper model (BCC model).

The specific situations of entities D, E, A, B and C are commented on in detail below:

- Entity D is inefficient under VRS and CRS. In order to become VRS efficient, it has to move to point D'. The input-oriented VRS technical inefficiency of point D is the distance DD'. In order to become CRS efficient, entity D has to move further toward point D''. The input-oriented CRS technical inefficiency of point D is the distance DD''. The distance between D' and D'' corresponds to scale inefficiency. The ratio efficiency measures, bounded by zero and one, are as follows:

Technical efficiency of D under CRS	**Technical efficiency of D under VRS**	**Scale efficiency of D**
$TE_{CRS} = \frac{TD''}{TD}$	$TE_{VRS} = \frac{TD'}{TD}$	$SE = \frac{TD''}{TD'}$

- Entity E is inefficient under VRS and CRS. In order to become CRS efficient, it has to move toward point E''. The input-oriented CRS technical inefficiency of point E is the distance EE''. In order to become VRS efficient, it has to move to point E'. The input-oriented VRS technical inefficiency of point E is the distance EE'. The difference between these two distances, i.e. the distance E'E'', corresponds to scale inefficiency. The ratio efficiency measures, bounded by zero and one, are as follows:

Technical efficiency of E under CRS	***Technical efficiency of E under VRS***	***Scale efficiency of E***
$TE_{CRS} = \frac{VE''}{VE}$	$TE_{VRS} = \frac{VE'}{VE}$	$SE = \frac{VE''}{VE'}$

- Entity A is efficient under VRS but inefficient under CRS. In order to become CRS efficient, it has to move toward point A'. The input-oriented CRS technical inefficiency of point A is the distance AA'; this also corresponds to scale inefficiency. The ratio efficiency measures, bounded by zero and one, are as follows:

Technical efficiency of A under CRS	***Technical efficiency of A under VRS***	***Scale efficiency of A***
$TE_{CRS} = \frac{SA'}{SA}$	$TE_{VRS} = \frac{SA}{SA} = 1$	$SE = \frac{SA'}{SA}$

- Entity B is efficient both under VRS and CRS. It is operating at the optimal scale. The ratio efficiency measures, bounded by zero and one, are as follows:

Technical efficiency of B under CRS	***Technical efficiency of B under VRS***	***Scale efficiency of B***
$TE_{CRS} = \frac{UB}{UB} = 1$	$TE_{VRS} = \frac{UB}{UB} = 1$	$SE = \frac{UB}{UB} = 1$

- Entity C is efficient under VRS but inefficient under CRS. In order to become CRS efficient, it has to move toward point C'. The input-oriented CRS technical inefficiency of point C is the distance CC'; this also corresponds to scale inefficiency. The ratio efficiency measures, bounded by zero and one, are as follows:

Technical efficiency of C under CRS	Technical efficiency of C under VRS	Scale efficiency of C
$TE_{CRS} = \frac{WC'}{WC}$	$TE_{VRS} = \frac{WC}{WC} = 1$	$SE = \frac{WC'}{WC}$

Knowing TE under CRS and TE under VRS, the scale efficiency is easily calculated. As $TE_{k,CRS} = TE_{k,VRS} \times SE_k$, the scale efficiency is obtained through the division of TE under CRS by TE under VRS: $SE_k = \frac{TE_{k,CRS}}{TE_{k,VRS}}$.

The linear programming problem to be solved under VRS includes a measure of returns to scale on the variables axis, *ck*, for the entity *k*. The primal equations are as follows:

VRS output-oriented model Primal equation		VRS input-oriented model Primal equation	
Minimize $\sum_{i=1}^{m} v_i x_{ik} - c_k$	(29)	Maximize $\sum_{r=1}^{s} u_r y_{rk} + c_k$	(33)
Subject to		Subject to	
$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} - c_k \geq 0 \quad j = 1,\ldots,n$	(30)	$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} - c_k \geq 0 \quad j = 1,\ldots,n$	(34)
$\sum_{r=1}^{s} u_r y_{rk} = 1$	(31)	$\sum_{i=1}^{m} v_i x_{ik} = 1$	(35)
$u_r, v_i > 0 \quad \forall r = 1,\ldots,s; i = 1,\ldots,m$	(32)	$u_r, v_i > 0 \quad \forall r = 1,\ldots,s; i = 1,\ldots,m$	(36)

The dual linear programming models are presented hereafter.

VRS output-oriented model Dual equation		VRS input-oriented model Dual equation	
Maximize ϕ_k	(37)	Minimize θ_k	(42)
Subject to		Subject to	
$\phi_k y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \quad r = 1,\ldots,s$	(38)	$y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0 \quad r = 1,\ldots,s$	(43)
$x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} \geq 0 \quad i = 1,\ldots,m$	(39)	$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} \geq 0 \quad i = 1,\ldots,m$	(44)
$\sum_{j=1}^{n} \lambda_j = 1$	(40)	$\sum_{j=1}^{n} \lambda_j = 1$	(45)
$\lambda_j \geq 0 \quad \forall j = 1,\ldots,n$	(41)	$\lambda_j \geq 0 \quad \forall j = 1,\ldots,n$	(46)

When slacks are added into the model, the dual linear programming equations become:

VRS output-oriented model
Dual equation with slacks

Maximize $$\phi_k + \varepsilon \sum_{r=1}^{s} s_r + \varepsilon \sum_{i=1}^{m} s_i \qquad (47)$$

Subject to

$$\phi_k y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} + s_r = 0 \quad r = 1,\ldots,s \qquad (48)$$

$$x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} - s_i = 0 \quad i = 1,\ldots,m \qquad (49)$$

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad (50)$$

$$\lambda_j, s_r, s_i \geq 0 \ \forall j = 1,\ldots,n; r = 1,\ldots,s; i = 1,\ldots,m \qquad (51)$$

VRS input-oriented model
Dual equation with slacks

Minimize $$\theta_k - \varepsilon \sum_{r=1}^{s} s_r - \varepsilon \sum_{i=1}^{m} s_i \qquad (52)$$

Subject to

$$y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} + s_r = 0 \quad r = 1,\ldots,s \qquad (53)$$

$$\theta_k x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} - s_i = 0 \quad i = 1,\ldots,m \qquad (54)$$

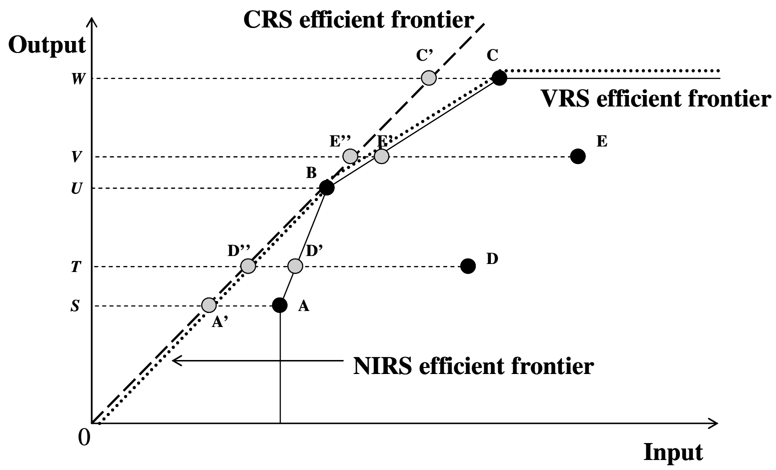
$$\sum_{j=1}^{n} \lambda_j = 1 \qquad (55)$$

$$\lambda_j, s_r, s_i \geq 0 \ \forall j = 1,\ldots,n; r = 1,\ldots,s; i = 1,\ldots,m \qquad (56)$$

A further step has to be taken in order to identify the nature of the returns to scale. This relates to another model, the non-increasing returns to scale model (NIRS), derived from the VRS model in which the $\sum_{j=1}^{n} \lambda_j = 1$ restriction is substituted by the $\sum_{j=1}^{n} \lambda_j \leq 1$ constraint (Coelli *et al.*, 2005).

In Figure 19, the NIRS efficiency frontier has been added (the dotted line). It corresponds to the CRS frontier from the origin to point B followed by the VRS frontier from point B. The nature of the scale inefficiencies for each entity can be determined by comparing technical efficiency scores under NIRS and VRS. If NIRS TE ≠ VRS TE (as for entities A and D), increasing returns to scale apply. If NIRS TE = VRS TE (but ≠ CRS TE) (as for entities E and C), decreasing returns to scale apply. Finally, if NIRS TE = VRS TE = CRS TE, as for entity B, constant returns to scale apply.

Figure 19
The nature of returns to scale is identified by comparing a NIRS and a VRS model.



5 Extensions of DEA

In this section, a selection of four extensions of DEA is shortly introduced: adjusting for the environment, preferences (weight restrictions), sensitivity analysis and time series data. For a broader overview of the major developments in DEA, see Cook and Seiford (2008). For an up-to-date review of DEA, readers will refer to Cooper *et al.* (2011).

5.1 Adjusting for the environment

Environmental variables influence the efficiency of entities but are not under the control of the management team. In DEA, several methods accommodate such variables. Those include the Charnes *et al.* (1981) approach, the categorical model (Banker & Morey, 1986a) or the non-discretionary variable model derived by Banker and Morey (1986b) (which indeed includes the environmental variable directly into the DEA model).

The most convincing of these methods, however, is the two-stage method, the advantages of which are described in Coelli *et al.* (2005, pp. 194-195) or in Pastor (2002, p. 899). The two-stage method combines a DEA model and a regression analysis. In the first stage, a traditional DEA model is conducted. This model includes only discretionary inputs and outputs. In the second stage, the efficiency scores are regressed against the environmental (i.e. non-discretionary or exogenous) variables. Tobit regression is often used in the second stage. However, recent studies have shown that ordinary least squares regression is sufficient to model the efficiency scores (Hoff, 2007) or even more appropriate than Tobit (McDonald, 2009).

The coefficients of the environmental variables, estimated by the regression, are used to model the efficiency scores to correspond to an identical condition of environment (e.g. usually the average condition). Simar and Wilson (2007, p. 32) provide a selection of studies using the two-stage method. Among those are applications in education (Chakraborty *et al.*, 2001; McMillan & Datta, 1998; McCarty & Yaisawarng, 1993), hospitals (Burgess & Wilson, 1998), defence (Barros, 2004), police (Carrington *et al.*, 1997), farming (Binam *et al.*, 2003) or banking (O'Donnell & van der Westhuizen, 2002). Sueyoshi *et al.* (2010) and Sibiano and Agasisti (2013) provide more recent applications in the sector of manufacturing sector and education.

5.2 Preferences

For different reasons (e.g. the weights assigned to variables by DEA are considered unrealistic for some entities; the management team may wish to give priority to certain variables; etc.), preferences about the relative importance of individual inputs and outputs can be set by the decision maker. This is done by placing weight restrictions onto outputs and inputs (also called multiplier restrictions). Cooper *et al.* (2011) and Thanassoulis *et al.* (2004) provide a review of models regarding the use of weights restrictions. An earlier review can be found in Allen *et al.* (1997). Generally, the imposition of weight restrictions worsens efficiency scores. Three main approaches are identified to accommodate preferences:

- Dyson and Thanassoulis (1988) propose an approach which imposes absolute upper and lower bounds on input and output weights. This technique is applied in Roll *et al.* (1991) to highway maintenance units or in Liu (2009) to garbage clearance units.
- Charnes *et al.* (1990) develop the cone-ratio method. This approach imposes a set of linear restrictions that define a convex cone, corresponding to an 'admissible' region of realistic weight restrictions. See Brockett *et al.* (1997) for an application to banks.
- Thompson *et al.* (1986, 1990) propose the assurance region method. This approach is actually a special case of the cone ratio. It imposes constraints on the relative magnitude of the weights. For example, a constraint on the ratio of weights for input 1 and input 2 can be included, such as the following: $L_{1,2} \leq \frac{v_2}{v_1} \leq U_{1,2}$, where $L_{1,2}$ and $U_{1,2}$ are lower and upper bounds for the ratio of the weight of input 2 (v_2) to the weight of input 1 (v_1). As a result, the assurance region method limits the 'region' of weights to a restricted area by prohibiting large differences in the value of those weights. An application of this model is provided by Sarica and Or (2007) in the assessment of power plants.

5.3 Sensitivity analysis

Cooper *et al.* (2006, p. 271) mention that the term 'sensitivity' corresponds to stability or robustness. For Zhu (2003, p. 217), "the calculated frontiers of DEA models are stable if the frontier DMUs that determine the DEA frontier remain on the frontier after particular data perturbations are made". Sensitivity analysis aims to identify the impact on entity efficiency when certain parameters are modified in the model.

The first way to test the sensitivity of DEA results consists in adding or extracting entities to DEA models. Dusansky and Wilson (1994, 1995) and Wilson (1993, 1995) provide different approaches to deal with this concern. The approach of Pastor *et al.* (1999) allows users to identify the observations which considerably affect the efficiency of the remaining entities. It also determines the statistical significance of efficiency variations which are due to the inclusion of a given entity in the sample.

Another way to test the sensitivity of DEA results consists in modifying the values of outputs and inputs. They focus on the maximum data variations a given entity can endure, whilst maintaining its efficiency status. Approaches include data perturbation of:

- A single variable of an efficient entity (Charnes *et al.*, 1985), data of other entities remaining fixed;

- Simultaneous proportional data perturbation of all outputs and inputs of an efficient entity (Charnes & Neralic, 1992), data of other entities remaining fixed;
- Simultaneous data perturbation of an efficient entity in a situation where outputs and inputs can be modified individually (Seiford & Zhu, 1998a, or Neralic & Wendell, 2004), data of other entities remaining fixed;
- Simultaneous proportional data perturbation of all outputs and inputs of all entities (Seiford & Zhu, 1998b).

For further review of sensitivity analysis, readers can refer to Zhu (2001).

5.4 Time series data

In DEA, panel data are considered using two methods: window analysis and the Malmquist index.

Window analysis, introduced by Charnes *et al.* (1985), examines the changes in the efficiency scores of a set of entities over time. A 'window' of time periods is chosen for each entity. The same entity is treated as if it represented a different entity in every time period. In this sense, window analysis can also be considered as a sensitivity analysis method. For instance, a model including n entities with annual data and a chosen 'window' of t years will result in $n \times t$ units to be evaluated. For each entity, t different efficiency scores will be measured. The 'window' is then shifted by one period (one year in our example) and the efficiency analysis is repeated. Yue (1992) provides a didactical application of window analysis. Other applications include Yang and Chang (2009), Avkiran (2004) or Webb (2003).

The Malmquist total factor productivity index was first introduced by Malmquist (1953) before being further developed in the frame of DEA. It is used to measure the change in productivity over time. The Malmquist index decomposes this productivity change into two components:

- The first one is called 'catch-up'. This captures the change in technical efficiency over time.
- The second one is called 'frontier-shift'. This captures the change in technology which occurs over time (i.e. the movement of efficiency frontiers over time).

Readers will refer to Färe *et al.* (2011) and Tone (2004) for actual reviews. Applications of the Malmquist index can be found in Coelli and Prasada Rao (2005) and Behera *et al.* (2011).

6 DEA with Microsoft Excel ® Solver

6.1 Microsoft Excel ® Solver

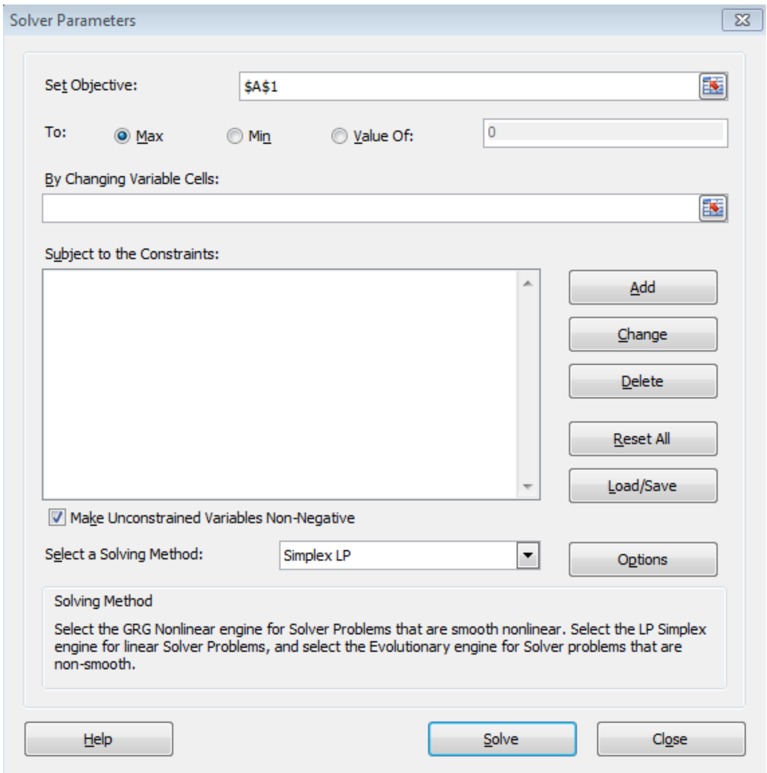
Excel ® Solver is a tool used to find the best way to do something, in other words to optimize an objective. Instructions on loading *Excel* ® Solver are easily found on the Internet[134].

Excel ® Solver allows users to solve optimization problems. An optimization model is composed of three elements: the target cell, the changing cells and the constraints. These three elements correspond to the parameters to be defined in *Excel* ® Solver (see Figure 20).

- The target cell ('Set objective') corresponds to the objective. It has to be either minimized or maximized.
- The changing variable cells are the cells which can be altered in order to optimize the target cell.
- The constraints (one or several) correspond to restrictions placed on the changing cells.

[134] In *Microsoft Excel* ® 2010, the Solver has to be loaded by clicking the File button, then the Excel Options and finally the Add-Ins button. In the Manage box, Excel Add-ins has to be selected before clicking the Go button. In the Add-Ins box, the Solver Add-in has to be selected. Finally, the OK button has to be clicked. Once the Solver is loaded, it is located in the Analysis group on the Data tab.

Figure 20
Three parameters have to be defined in *Excel* ® Solver.



6.2 Programming a CRS model

Consider five register offices (A to E) producing two outputs (birth and marriage certificates) with one input (full-time equivalent public servant). The data are listed in Table 4. For example, one full-time equivalent (FTE) public servant works in Register Office A. He produces one birth and six marriage certificates during a certain period of time.

Table 4
Five Register Offices produce birth and marriage certificates using public servants.

Register Office	**Input**	**Output**	
	Public servant (x)	Birth (y_1)	Marriage (y_2)
A	1	1	6
B	1	3	8
C	1	4	3
D	1	5	6
E	1	6	2

The use of Excel ® Solver is illustrated with the following CRS model.

CRS input-oriented model
Primal equation

$$\text{Maximize} \quad \sum_{r=1}^{s} u_r y_{rk}$$

$$\text{Subject to} \quad \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0 \quad j = 1, \ldots, n$$

$$\sum_{i=1}^{m} v_i x_{ik} = 1$$

$$u_r, v_i > 0 \quad \forall r = 1, \ldots, s; i = 1, \ldots, m$$

In this model, the objective is to maximize the weighted sum of outputs of entity *k*. Two constraints have to be considered. First, the weighted sum of inputs minus the weighted sum of outputs of entity *j* has to be greater than or equal to zero. Second, the weighted sum of inputs of entity *k* has to be equal to one.

Users have to prepare an *Excel* ® spreadsheet, such as the one appearing in Figure 21. This is divided into two parts:

- The first part comprises rows 2 and 3. This section enables users to successively calculate the efficiency of the five register offices (one at a time). To do this, data of each register office have to be entered successively in cells B2 to D2 (dark grey cells). Figure 21 already contains data on Register office C. The two outputs and one input of Register Office C are assigned weights in cells B3 to D3 (light grey cells). A value of one has been assigned to all of them in the spreadsheet. These values will be precisely modified by *Excel* ® Solver in order to maximize the register offices' efficiency scores. Cell E2 contains the weighted sum of outputs for Register Office C. The formula associated with cell E2 is (B2*$B$3) + (C2*$C$3). Cell F2 contains the weighted sum of the input for Register Office C. The formula associated with cell F2 is (D2*$D$3). Finally, cell G2 contains the efficiency score of Register office C as a percentage (light grey cell). The formula associated with cell G2 is (E2/F2)*100. Note that the score of 700% appearing in the spreadsheet is calculated using weighted values of 1 and without any constraints. In other words, this score has not yet been optimized under varying constraints.
- The second part comprises rows 6 to 10. It contains the data for register offices A to E (output 1 = column B, output 2 = column C, input = column D, weighted sum of outputs = column E, weighted sum of the input = column F). The same formulae as above apply to the weighted sums of outputs and the input. An additional column, G, is added in the spreadsheet. It is a working column which will be used by *Excel* ® Solver.

Column G contains the weighted sum of the input minus the weighted sum of outputs to adequately reflect the $\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0$ constraint. The formula associated with cell G6 is F6 – D6, the formula associated with cell G7 is F7 – D7, etc.

Figure 21
An *Excel* ® spreadsheet ready to use with *Excel* ® Solver.

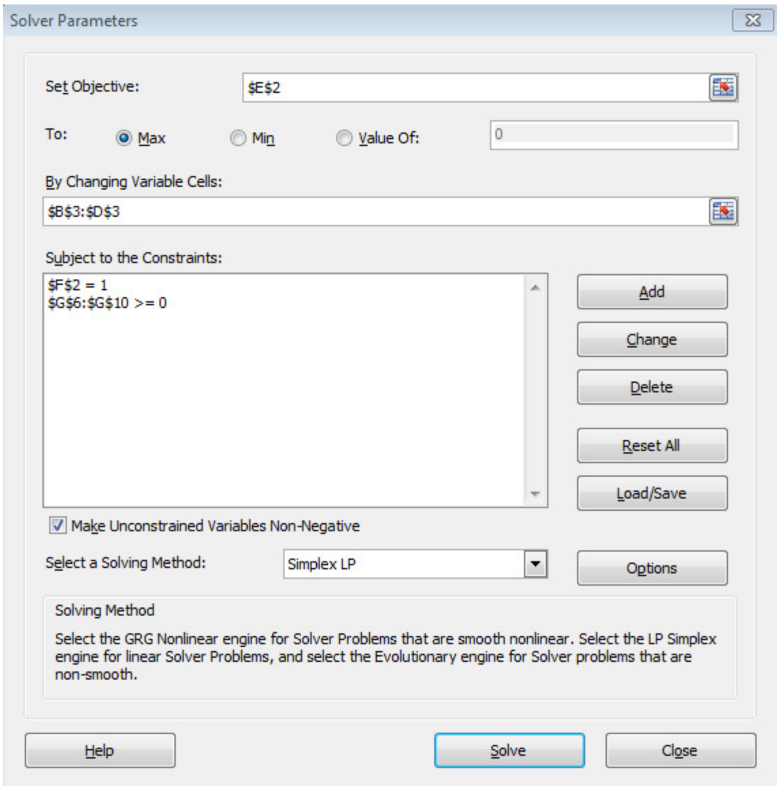
	A	B	C	D	E	F	G
1		Output 1 Birth	Output 2 Marriage	Input 1 Public servant	Weighted output	Weighted input	Efficiency (%)
2	Register office	4	3	1	7.00	1.00	700.00
3	Weight	1.00	1.00	1.00			
4							
5		Output 1 Birth	Output 2 Marriage	Input 1 Public servant	Weighted output	Weighted input	Weighted input minus Weighted output
6	Register office A	1	6	1	7.00	1.00	-6.00
7	Register office B	3	8	1	11.00	1.00	-10.00
8	Register office C	4	3	1	7.00	1.00	-6.00
9	Register office D	5	6	1	11.00	1.00	-10.00
10	Register office E	6	2	1	8.00	1.00	-7.00

Once the spreadsheet is ready, the parameters of *Excel* ® Solver have to be specified in the following way:

- The objective is to maximize the weighted sum of outputs of Register Office *k* ($\sum_{r=1}^{s} u_r y_{rk}$). In the objective parameter, cell $E$2 has to be specified. The Max option has to be ticked.
- To optimize the objective, the changing variable cells have to be specified. They correspond to the weights associated with outputs and inputs. In the changing variable cells parameter, cells $B$3 to $D$3 ($B3:$D$3) have to be specified.
- Finally, the restrictions placed on the changing cells have to be introduced as constraints. A constraint is added by clicking the Add button. In the Add Constraint box, three parameters have to be specified: the cell reference, the sign of the constraint (<=, = or >=) and the value of the constraint. The first constraint of the CRS model ($\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0$) is therefore specified as follows: $G$6:$G$10>=0 (where the cell reference is $G$6:$G$10, the sign is >= and the constraint is 0). The second constraint ($\sum_{i=1}^{m} v_i x_{ij} = 1$) is specified as follows: $F$2=1 (where the cell reference is $F$2, the sign is = and the constraint is 1). Note that this constraint means that the given level of input is kept constant.

Figure 22 represents the Solver Parameters defined above.

Figure 22
The Solver parameters are specified.



Finally, a Simplex LP solving method has to be selected and the 'Make Unconstrained Variables Non-Negative box' has to be ticked. This indicates that a linear model with non-negative variables is appropriate (and therefore the third and last 'constraint' $u_r, v_i > 0$ is taken into account).

The Solve button should be clicked in order to execute *Excel* ® Solver. *Excel* ® Solver will search every feasible solution to determine the solution which has the best target cell value. Register Office C obtains an efficiency score of 73.08% (cell G2). This score is obtained using weights of 0.15, 0.04 and 1 assigned to output 1, output 2 and input 1, respectively (cells B3, C3 and D3). A Solver Results box appears after solving the model. Before solving the model again for the other register offices, 'Restore Originals Values' has to be ticked before clicking the OK button.

To measure the efficiency of Register Office A (for example), it is necessary to replace the values of cells B3 to D3 (which currently refer to Register Office C) with the values of cells B6 to D6 (which refer to Register Office A). Solving the model will calculate an efficiency score of 75% for Register Office A.

References

Allen, R., Athanassopoulos, A., Dyson, R. G. & Thanassoulis, E. (1997). Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions. *Annals of Operations Research*, *73*(1), 13-34.

Avkiran, N. K. (2004). Decomposing technical efficiency and window analysis. *Studies in Economics and Finance*, *22*(1), 61-91.

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.

Banker, R. D. & Morey, R. C. (1986a). Efficiency Analysis for Exogenously Fixed Inputs and Outputs. *Operations Research*, 34(4), 513-521.

Banker, R. D. & Morey, R. C. (1986b). The Use of Categorical Variables in Data Envelopment Analysis. *Management Science*, *32*(12), 1613-1627.

Barr, R. (2003). DEA Software Tools and Technology: A State-of-the-Art survey. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 539-566). Boston: Kluwer Academic Publishers.

Barros, C. P. (2004). Measuring performance in defence-sector companies in a small NATO member country. *Journal of Economic Studies*, *31*(2), 112-128.

Behera, S. K., Faroquie, J. A. & Dash, A. P. (2011). Productivity change of coal-fired thermal power plants in India: a Malmquist index approach. *IMA Journal of Management Mathematics*, *22*(4), 387-400.

Binam, J. N., Sylla, K., Diarra, I. & Nyambi, G. (2003). Factors affecting technical efficiency among coffee farmers in Côte d'Ivoire: evidence from the centre west region. *R&D Management*, *15*(1), 66-76.

Bogethoft, P. & Otto, L. (2010). *Benchmarking with DEA, SFA, and R*. New York: Springer.

Brockett, P. L., Charnes, A., Cooper, W. W., Huang, Z. M. & Sun, D. B. (1991). Data Transformations in DEA Cone Ratio Envelopment Approaches for Monitoring Bank Performance. *Journal of Operational Research*, *98*(2), 250-268.

Burgess, J. F. & Wilson, P. W. (1998). Variation in inefficiency among US hospitals. *Canadian Journal of Operational Research and Information Processing (INFOR)*, *36*(3), 84-102.

Carrington, R., Puthucheary, N., Rose, D. & Yaisawarng, S. (1997). Performance measurement in government service provision: the case of police services in New South Wales. *Journal of Productivity Analysis*, *8*(4), 415-430.

Chakraborty, K., Biswas, B. & Lewis, W. C. (2001). Measurement of technical efficiency in public education: a stochastic and nonstochastic production approach. *Southern Economic Journal*, *67*(4), 889-905.

Charnes, A., Clarke, C., Cooper, W. W. & Golany, B. (1984). A development study of DEA in measuring the effect of maintenance units in the U.S. Air Force. *Annals of Operations Research*, *2*(1), 95-112.

Charnes, A., Cooper, W. W., Huang, Z. M. & Sun, D. B. (1990). Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. *Journal of Econometrics*, *46*(1-2), 73-91.

Charnes, A, Cooper, W. W. & Rhodes E. L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429-444.

Charnes, A, Cooper, W. W. & Rhodes E. L. (1981). Evaluating program and managerial efficiency: An application of DEA to program follow through. *Management Science*, *27*(6), 668-697.

Charnes, A., Haag, S., Jaska, P. & Semple, J. (1992). Sensitivity of efficiency calculations in the additive model of data envelopment analysis. *International Journal of System Sciences*, *23*(5), 789-798.

Charnes, A. & Neralic, L. (1992). Sensitivity analysis in data envelopment analysis. *Glasnik Matematicki*, *27*(47), 191-201.

Coelli, T. J. (1996). A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program. *CEPA Working Paper 96/08*. Brisbane: Centre for Efficiency and Productivity Analysis, University of Queensland.

Coelli, T. J. (1998). A Multi-stage Methodology for the Solution of Oriented DEA Models. *Operations Research Letters*, *23*(3-5), 143-149.

Coelli, T. J. & Perelman, S. (1996). Efficiency measurement, Multiple-output Technologies and Distance Functions: With Application to European Railways. *CREPP Working Paper 96/05*. Liège: Centre de Recherche en Economie Publique et de la Population, University of Liège.

Coelli, T. J. & Perelman, S. (1999). A Comparison of Parametric and Non-Parametric Distance Functions: With Application to European Railways. *European Journal of Operational Research*, *117*(2), 326-339.

Coelli, T. J. & Prasada Rao, D. S. (2005). Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980-2000. *Agricultural Economics*, *32*(s1), 115-134.

Coelli, T. J., Prasada Rao, D. S., O'Donnel, C. J. & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York: Springer.

Cook, W. D. & Seiford, L. M. (2008). Data envelopment analysis (DEA) – Thirty years on. *European Journal of Operational Research, 192*(1), 1-17.

Cook, W. D. & Zhu, J. (2008). *Data Envelopment Analysis: Modeling Operational Processes and Measuring Productivity*. Seattle: CreateSpace.

Cooper, W. W., Ruiz J. L. & Sirvent, I. (2011). Choices and Uses of DEA Weights. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 93-126). New York: Springer.

Cooper, W. W., Seiford, L. M. & Tone, K. (2006). *Introduction to Data Envelopment Analysis and Its Uses*. New York: Springer.

Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data Envelopment Analysis: A comprehensive Text with Models, Applications, References and DEA-Solver Software*. New York: Springer.

Cooper, W. W., Seiford, L. M. & Zhu, J. (2004). *Handbook on Data Envelopment Analysis* (1st ed.). Boston: Kluwer Academic Publishers.

Cooper, W. W., Seiford, L. M. & Zhu, J. (2011). *Handbook on Data Envelopment Analysis* (2nd ed.). New York: Springer.

Dantzig, G. B. (1951). Maximization of a linear function of variables subject to linear inequalities. In T. C. Koopmans (Ed.), *Activity Analysis of Production and Allocation* (pp. 339-347). New York: John Wiley & Sons.

Dusansky, R. & Wilson, P. W. (1994). Measuring efficiency in the care of developmentally disabled. *Review of Economics and Statistics, 76*(2), 340-345.

Dusansky, R. & Wilson, P. W. (1995). On the relative efficiency of alternative modesof producing public sector output: The case of developmentally disabled. *European Journal of Operational Research*, *80*(3), 608-628.

Dyson, R. G. & Thanassoulis, E. (1988). Reducing weight flexibility in DEA. *Journal of the Operational Research Society*, *39*(6), 563-576.

Emrouznejad, A., Parker, B. R. & Tavares G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, *42*(3), 151-157.

Färe, R., Grosskopf, S. & Margaritis, D. (2011). Malmquist Productivity Indexes and DEA. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 127-150). New York: Springer.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of Royal Statistical Society*, *120*(3), 253-281.

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

Huguenin, J.-M. (2012). Data Envelopment Analysis (DEA): a pedagogical guide for decision makers in the public sector. *Cahier de l'IDHEAP No. 276*. Lausanne: Swiss Graduate School of Public Administration.

Huguenin, J.-M. (2013a). Data Envelopment Analysis (DEA): un guide pédagogique à l'intention des décideurs dans le secteur public. *Cahier de l'IDHEAP No. 278*. Lausanne: Institut de hautes études en administration publique.

Huguenin, J.-M. (2013b). Data Envelopment Analysis (DEA). In A. Ishizaka & P. Nemery (Eds.), *Multi-Criteria Decision Analysis: Methods and Software* (pp. 235-274). Chichester: John Wiley & Sons.

International Public Sector Accounting Board (2012). *Handbook of Internantional Public Sector Accounting Pronouncements*. New York: International Federation of Accountants.

Ishizaka, A. & Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. Chichester: John Wiley and Sons.

Johnes, J. (2004). Efficiency measurement. In G. Johnes & J. Johnes (Eds.), *International Handbook on the Economics of Education* (pp. 613-742). Cheltenham: Edward Elgar Publishing.

Liu, C.-C. (2009). A study of optimal weights restriction in Data Envelopment Analysis. *Applied Economics*, *41*(14), 1785-1790.

McCarty, T. A. & Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. In H. O. Fried, C. A. K Lovell & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications* (pp. 271-287). Oxford University Press: New York.

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

McMillan, M. L. & Datta, D. (1998). The relative efficiencies of Canadian universitites: a DEA perspective. *Canadian Public Policy-Analyse de Politiques*, *24*(4), 485-511.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estatistica*, *4*(2), 209-242.

Neralic, L. (2004). Preservation of efficiency and inefficiency classification in data envelopment analysis. *Mathematical Communications*, *9*(1), 51-62.

Neralic, L. & Wendell, R. E. (2004). Sensitivity in data envelopment analysis using an approximate inverse matrix. *Journal of the Operational Research Society*, *55*(11), 1187-1193.

O'Donnell, C. J. & van der Westhuizen, G. (2002). Regional comparisons of banking performance in SouthAfrica. *South African Journal of Economics*, *70*(3), 485-518.

Pastor, J. M. (2002). Credit risk and efficiency in the European banking system: A three-stage analysis. *Applied Financial Economics*, *12*(12), 895-911.

Pastor, J. M., Ruiz, J. L. & Sirvent, I. (1999). A statistical test for detecting influential observations in DEA. *European Journal of Operational Research*, *115*(3), 542-554.

Rhodes, E. L. (1978). *Data envelopment analysis and approaches for measuring the efficiency of decision making units with an application to program follow through in U.S. education* (Unpublished doctoral dissertation). Carnegie-Mellon University, USA.

Roll, Y., Cook, W. D. & Golany B. (1991). Controlling Factor Weights in Data Envelopment Analysis. *IEE Transactions*, *23*(1), 2-9.

Sarica, K. & Or, I. (2007). Efficiency assessment of Turkish power plants using data envelopment analysis. *Energy*, *32*(8), 1484-1499.

Scheel, H. (2000). *EMS: Efficiency Measurement System User's Manual.* Retrieved from http://www.wiso.tu-dortmund.de/wiso/de/fakultaet/personen/institut/or/EXT-HOSC.html

Seiford, L. M. & Zhu, J. (1998a). Stability regions for maintaining efficiency in data envelopment analysis. *European Journal of Operational Research*, *108*(1), 127-139.

Seiford, L. M. & Zhu, J. (1998b). Sensitivity analysis of DEA models for simultaneous changes in all the data. *Journal of the Operational Research Society*, *49*(10), 1060-1071.

Sibiano, P. & Agasisti, T. (2013). Efficiency and heterogeneity of public spending in education among Italian regions. *Journal of Public Affairs, 13*(1), 12-22.

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.

Sueyoshi, T., Goto, M. & Omi, Y. (2010). Corporate governance and firm performance: Evidence from Japanese manufacturing industries after the lost decade. *European Journal of Operational Research*, *203*(3), 724-736.

Thanassoulis, E, Portela, M. C. S. & Allen, R. (2004). Incorporating value judgements in DEA. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 99-138). Boston: Kluwer Academic Publishers.

Thanassoulis, E., Portela, M. C. S., & Despic, O. (2008). Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. In H. O. Fried, C. A. Lovell, & S. S. Schmidt, *The Measurement of Productive Efficiency and Productivity Growth* (pp. 251-420). New York: Oxford University Press.

Thomson, R. G., Langemeier, L. N., Lee, C.-T., Lee, E. & Thrall, R. M. (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics*, *46*(1-2), 93-108.

Thomson, R. G., Singleton Jr., F. D., Thrall, R. M. & Smith, B. A. (1986). Comparative site evaluations for locating a high-energy physics lab in Texas. *Interfaces*, *16*(6), 35-49.

Tone, K. (2004). Malmquist productivity index – Efficiency change over time. In W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 203-227). Boston: Kluwer Academic Publishers.

Webb, R. (2003). Level of efficiency in UK retail banks: a DEA window analysis. *International Journal of the Economics of Business*, *10*(3), 305-322.

Wilson, P. W. (1993). Detecting outliers in deterministic non-parametric frontier models with multiple outputs. *Journal of Business and Economic Statistics*, *11*(3), 319-323.

Wilson, P. W. (1995). Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis*, *6*(1), 27-45.

Yang, H.-H. & Chang, C.-Y. (2009). Using DEA window analysis to measure efficiencies of Taiwan's integrated telecommunication firms. *Telecommunications Policy*, *33*(1-2), 98-108.

Yue, P. (1992). Data Envelopment Analysis and Commercial Bank Performance: A Primer with Applications to Missouri Banks. Federal *Reserve Bank of St Louis Review*, *74*(1), 31-45.

Zhu, J. (2001). Super-efficiency and DEA sensitivity analysis. *European Journal of Operational Research*, *129*(2), 443-455.

Zhu, J. (2003). *Quantitative Models for Performance Evaluation and Benchmarking.* New York: Springer.