Lancaster University Management School

Submitted in fulfilment of the requirements of the degree of

Doctor of Philosophy

By

Dr Maria Papadimitriou

June 2018

Drivers of Efficiency in Higher Education in England

Dr Maria Papadimitriou

Bachelor of Science in Economics, University of Macedonia, Thessaloniki, Greece

Master of Science in Economics, University of Macedonia, Thessaloniki, Greece

Department of Economics
Lancaster University
Management School
(LUMS)

June 2018

Statement of Originality

I hereby declare that this thesis is my own work and has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Maria Papadimitriou

____(Signed) _______

Dedication

To the memory of my father

Father Thomas Papadimitriou

Abstract

The context of this thesis explores efficiency measures in higher education (HE) in England. Measures of efficiency serve as a crucial link between the economic sustainability of the HE sector and the policymaking establishment. Given that the idea of efficient allocation of resources, in a period of tighter budget constraints, curtailed government funding and increasing competition for a greater share for research funding and number of students has such a powerful influence, the concept of efficiency becomes meaningful serving as a basis for decisions to improve resource allocation. Understanding the nature of efficiency aims to put in place a simpler and more efficient HE and research system in England that encourages competition and choice, enhances quality, and ensures greater accountability and value for money.

The thesis unfolds two main disciplines of technical and cost efficiency in HE in England. Therefore the research objectives discussed are driven upon that conceptualization of efficiency and provide further insights into first, the effects of merger activity on efficiency, and second on whether permanent or transient (in) efficiency dominates the English HE sector. Those topics are key aspects and critically important for both policy change and ongoing institutional and structural reform and as thus are explored in the lines of this thesis.

Regarding the first research objective on the potential effect of mergers on Higher Education Institutions' (HEIs) efficiency, in a first stage analysis efficiency scores of English universities are derived for a 17-year period using the frontier estimation method data envelopment analysis (DEA). A second stage analysis explores the effect of merger and other factors on efficiency. We find that mean efficiency for the sector has varied around 60 percent to 70 percent, but that the efficiency levels of the vast majority of individual HEIs are not significantly different from each other. Merged HEIs have efficiency which is 5 percentage points higher post-merger than non-merging HEIs holding all else constant; moreover the efficiency impact of merger comes within 2 years of the merger taking place. Of the other factors included in the second stage analysis, pre-1992 universities have lower efficiency than other types of institution. In addition, having a higher proportion of income from government sources is an incentive to greater efficiency. Finally, a sensitivity analysis was conducted which exposed the post-merger efficiency results to a different method assessment as a validation test of the proposed policy implications. The sensitivity analysis resulted in confirming the main findings of efficiency improvements in the units received the treatment of merger.

Turning to the second research objective, a common weakness in most of the models dealing with efficiency is their deficiency to account for unobserved heterogeneity that finally lead to biased efficiency estimates. Most of the cost efficiency frontier models, focused either on the transient or on the persistent part of cost inefficiency, confounding firm effects (that are not part of inefficiency) with persistent inefficiency or blending

persistent inefficiency with latent heterogeneity. However a decomposition of the two parts, persistent (long –term) and transient/residual (short-term) inefficiencies, provides an in-depth analysis of whether short term practises or more long term structural changes within colleges and universities affect the degree of cost efficiency in the English HE sector.

This distinction seems to be further appealing to the policy makers as a regulatory asset that aims at improving the efficiency of the sector through incentive reforms. Hence, more recent developments in panel data[1] allow a further appealing distinction in the cost efficiency of HEIs in which unobserved firm effects (firm heterogeneity) can be disentangled from time invariant and time varying inefficiency. Hence the purpose of this thesis is partly to assess the level of persistent and transient inefficiency in the English HE sector from 2008/09 to 2013/14 by using a four-way error component model (persistent and transient inefficiency, random firm effects and noise) and so as to retain the apparatus of statistical inference stemming from a generalised true random effects (GTRE) model based on maximum simulated likelihood (MSL) techniques.

In order to provide evidence that the aforementioned method ameliorates the predicted power of the model we offer a comparative study through the fundamental models applied in the literature so far. Consequently, statistical inference will be attempted by countering the efficiency estimates of a GRTE model with a random effects (RE) model proposed by Pitt and Lee (1981), informative on the persistent part, and a true random effects model (TRE) proposed by Greene (2005a, 2005b), enlightening the transient part. Finally, omitted variables bias will be controlled by implementing a MGTRE model, minimizing the resulting heterogeneity bias.

The comparison reinforces the validity of the GTRE model since it captures every single component of inefficacy while heterogeneity is controlled. For the English HE inefficiency is considered as persistent since short-run efficiency estimates are proven to be higher than the long-run. This gives further rise for more comprehensive and structural changes rather than simple mechanisms for short-term cost savings.

[1] SFA models by Colombi et al. (2014) and Fillipini and Greene, (2016).

Acknowledgements

The composition and completion of this thesis would not have been possible without the support and encouragement of several special people. Hence, I would like to take this opportunity to express my gratitude to those who have assist me in a myriad of ways.

The thesis is dedicated to the memory of my father who is no longer with me, but I owe a debt of gratitude to him for contributing to the decision of starting a PhD. He could not support me through this endeavour because he left once I began, however I know he would be truly ecstatic to see my progress today. He always encouraged me to be what I truly want in life and he always wanted to become a scientist. I know that he saw himself in me in many different ways and thanks to him I am what I am today. I hope he would be sincerely proud to see me accomplishing this goal.

First and for most, I am indebted to my principal supervisor, Professor Jill Johnes, who was the first person who put her faith on my academic skills and for selecting me for completing this promising project. I would like to express my sincere gratitude for her continuous support of my Ph.D. study and related research, for her patience, motivation, encouragement and immense knowledge. A more supportive and considerate supervisor I could not have asked for. There was many times where I had reached the 'crossroads' and each time Jill was there to steer me towards the right path. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study. Her willingness to offer me so much of her time and intellect is the major reason this thesis was completed. Thank you for helping me to grow my ability of academic writing.

Besides my principal supervisor, I would like to thank my second advisor Professor Efthimios Tsionas and the rest of my thesis committee, Dr Hilary Ingham for their insightful comments and encouragement, but also for their hard questions which incentivized me to widen my research from various perspectives.

Special thanks to…

My family to whom I owe a great deal. I would like to thank my mother Iliani for being a constant source of support and encouragement for me every time I lost faith on myself and for making numerous of sacrifices for the entire family. I am blessed to have her since she was both mother and father after his death, a role that deserved two. Great appreciation and eternal thanks to my sister Christina who is by my side through good times and bad. Your extensive support have been ever-present in this important time of my life. I am grateful for your stubbornness since you have inspired me to hold tight to my position no matter how complicated it gets. To my brother Lazaros for putting me up when I needed a break from the thesis work and for being so patient with me even on times of insane nervousness. Finally, I would like to extend my sincere gratitude to our wider family under the name ‘’ Tsiolas family’’ for supporting me in many different ways and for making so many things possible.

Many thanks to…
All my colleagues at LUMS for creating a cordial working environment, and especially to Pari, Jekaterina, Ambra, Pantelis and Alberto for their friendship, editing advice and support throughout this challenging journey.

My friends and roommates, thank you for listening, offering me advice, for the wonderful times we shared, especially the Saturday night dinners, for supporting me through this entire process and for giving me the necessary distractions from my research and made my stay in England memorable.

My friends scattered around Greece, for their thoughts, their everlasting support, their endless skype and phone calls, e-mails, texts, visits and being there whenever I needed a friend.

Acknowledgement of Financial Support

The work of this thesis was financially supported by the Economic and Social Research Council, UK [grant number ES/J500094/1] and partly supported by Lancaster University Management School, UK. I also gratefully acknowledge their generous financial support for all the conferences, workshops and training courses attended throughout my studies. Finally, I am also sincerely thankful to the Dowager Countess Eleanor Peel Trust Fund, Lancaster University for offering me a two year financial award (2014/15-2015/16).

Acknowledgement of Data Provision

Department
for Business
Innovation & Skills

This thesis supported by an ESRC Collaborative (CASE) Studentship entitled “Drivers of efficiency in higher education”. As part of this collaboration I was engaged in close cooperation with the Department of Business, Innovation and Skills[2] [DBIS], London, UK. Data admission by HESA [HE Business and Community Interaction Survey] has been enabled through DBIS. A special report on ‘’Efficiency in English universities: A production function approach’’ has been also produced as part of my PhD being completed in partnership with DBIS and funded by the ESRC.

[2] Note that: The Department for Business, Innovation and Skills (BIS) and the Department of Energy and Climate Change (DECC) have merged to form the Department for Business, Energy and Industrial Strategy (BEIS) in July 2016.

Acknowledgement of Published Papers and Conferences Produced by this Thesis

Published Paper(s):

Papadimitriou, M., and Johnes, J. (2018). Does merging improve efficiency? A study of English universities, *Studies in Higher Education*, [A* ranked in ERA; A ranked in ABDC] DOI: 10.1080/03075079.2018.1450851

Conference Presentations:

July, 2017. **5th International PhD Meeting in Economics,** University of Macedonia Thessaloniki, Greece. **Presentation Title:** Persistent and transient cost inefficiency in the English higher education sector: A Generalised True Random Effects Model.

May, 2017. **PhD Seminars Series, LUMS,** Department of Economics, Lancaster, UK.

Presentation Title: Persistent and transient cost inefficiency in the English higher education sector: A Generalised True Random Effects Model.

Oct, 2016. **4th Workshop on Efficiency in Education**, Milan Italy. **Organized by:** Politecnico di Milano – Department of Management Economics and Industrial Engineering. **Presentation Title:** Persistent and transient cost inefficiency in the English higher education sector: A Generalised True Random Effects Model.

Sept, 2016. **OR58 Conference**, Annual conference, Portsmouth, UK. **Organized by:** Operational Research Society. **Presentation Title:** Persistent and transient cost inefficiency in the English higher education sector: A Generalised True Random Effects Model.

Aug, 2015. **DEA 2015**, 13th International Conference on Data Envelopment Analysis. **Organized by:** Technische Universitat Braunschweig International Data Envelopment Analysis Society (iDEAs). **Presentation Title:** Assessing the efficiency of non-, pre- and post-merging HEIs in England using an output distance function.

June, 2015. **Department of Business Innovation and Skills (BIS),** London. Partner organization of the case studentship. **Presentation Title:** Efficiency in English universities: A production function approach.

May, 2015. **NWDTC Economics PhD Conference,** Liverpool, UK**. Organized by:** North West Doctoral Training Centre. **Presentation Title:** Assessing the efficiency of non-, pre- and post-merging HEIs in England using an output distance function.

Jan, 2014. **PhD Seminars Series, LUMS,** Department of Economics, Lancaster, UK.

Presentation Title: Assessing the efficiency of non-, pre- and post-merging HEIs in England using an output distance function.

Table of Contents

List of Tables

List of Figures

List of Graphs

List of Picture(s)

List of Acronyms

Allocative Efficiency [AE]

Allocative Inefficiency [AI]

Alternative Providers [APs]

Augmented Inverse Probability Weighted [AIPW]

Average treatment effects [ATE]

Average Treatment on the Treated [ATT]

Central Limit Theorem [CLT]

Cobb-Douglas [CD]

Colleges and Universities [CUs]

Committee of University Chairs [CUC]

Conditional Average Treatment Effect [CATE]

Conditional Independence Assumption [CIA]

Constant Elasticity of Substitution [CES]

Constant Returns to Scale [CRS]

Convex Nonparametric Least Squares [CNLS]

Correlated Ordinary Least Squares [COLS]

Correlated Random Effects [CRE]

Data Envelopment Analysis [DEA]

Data-Generating Process [DGP]

Decision Making Unit [DMU]

Decreasing Returns to Scale [DRS]

Department of Business, Innovation and Skills [DBIS]

Economic Efficiency [EE]

European Union [EU]

Feasible Generalised Least Squares [FGLS]

Free Disposal Hull [FDH]

Full-time-equivalent [FTE]

Further Education Colleges [FECs]

Further Education Institutions [FEIs]

Generalised Least Squares [GLS]

Generalised True Random Effects model [GTRE]

Gross domestic Product [GDP]

Higher Education [HE]

Higher Education Funding Council for England [HEFCE]

Higher Education Funding Council for Wales [HEFCW]

Higher Education Institutions [HEIs]

Higher Education Statistics Agency [HESA]

Increasing Returns to Scale [IRS]

Information and Communication Technologies [ICT]

Inverse probability of Treatment Weighting [IPTW]

Inverse Probability Weighted [IPW]

Inverse Probability Weighted with Regression Adjustment [IPWRA]

Least Significant Difference [LSD]

Least Square Dummy Variable [LSDV]

Marginal Cost [MC]

Maximum Likelihood [ML]

Maximum Likelihood Estimation [MLE]

Maximum Simulated Likelihood [MSL]

Mean Technical Efficiency [MTE]

Modified Ordinary Least Squares [MOLS]

Mundlak Adjustment GTRE [MGTRE]

Nearest Neighbour [NN]

Nearest Neighbour Matching [NNM]

Observatory of European University [OEU]

Office for Fair Access [OFFA]

Ordinary Least Squares [OLS]

Organisation for Economic Co-operation and Development [OECD]

Probability Density Function [PDF]

Production Possibility Set [PPS]

Product-Specific Returns to Scale [PSRS]

Product-Specific Scope Economies [PSOE]

Propensity Score Matching [PSM]

Pure Technical Efficiency Change [PTEC]

Quadratic Cost Function [QCF]

Ray Economies of Scale [SRAY]

Ray Returns to Scale [RRS]

Research and Development [R&D]

Returns to Scale [RTS]

Scale Efficiency [SE]

Scale Efficiency Change [SEC]

Stochastic Frontier Analysis [SFA]

Stochastic Nonparametric Envelopment of Data [StoNED]

Technical Efficiency [TE]

Technological Change [TC]

Total Economic Efficiency [TEE]

Total Factor Productivity [TFP]

Transcendental Logarithmic [Translog]

True Fixed Effects [TFE]

True Random Effects [TRE]

UK Performance Indicators [UKPIs]

United Kingdom [UK]

United States [US]

United States of America [USA]

University Grants Committee [UGC]

University-Enterprise Partnerships [UEPs]

Variable Returns to Scale [VRS]

1. Chapter 1: Introduction

1.1 Thesis Background

In a shifting socio-economic environment, the side effects that flow into the HE sector are accountable. Hence, the HE sector is always looking for ways to become more efficient. In recent decades, concepts such as efficiency and productivity have become central topics in discussions on the sustainability, costs, and quality of the HE sector (NRC, 2012). Universities and colleges struggle to find ways to make their operations more efficient to secure their long-term futures and to ensure that the public investment in HE provides value for money.

The United Kingdom (UK) in general has a world-class reputation for HE and is a popular study destination among international students. The direct contribution[3] of the UK's universities to the economy is considerable, reaching 2.8 percent of UK gross domestic product (GDP) in 2011, 2.3 percent up from 2007 (UUK, 2014). What is sometimes disregarded is that UK universities implicitly have wider macroeconomic effects in the UK economy as they operate as large enterprises by generating substantial economic activity, employment opportunities and overseas investment.

Compared with other leading HE systems, UK universities and colleges deliver teaching and research by spending significantly less money in education as a percentage of the country's GDP. According to reports for 2011–2012, less than half of revenues received from UK universities were from public sources (UUK, 2014). The role of public financial support is still apparent since universities continue to underpin economic growth and form a core part of the economic infrastructure.

Institutions in the UK received just over one quarter of their total income[4] from direct government sources for the period 2014–15. The distribution of the total income was mainly focused on financing the 'full economic cost' of teaching,[5] world-class research,[6] and innovation activities. While less than one fifth of income for teaching comes in the form of direct government grants, 66 percent of income for research comes from government. However, the UK invests gradually less in research and development (R&D) than many other countries. According to a UUK report, the total UK R&D

[3] The HE sector makes substantial contributions to economic activity, and generated over £73 billion of output (both direct and indirect effects) for the year 2011–2012.

[4] The total income for UK universities amounts to £33.2bn. There were 164 higher education institutions (HEIs) in the UK that received an element of public funding, with 130 located in England.

[5] Cost of teaching: This includes costs for staff, equipment, and services. It also includes the costs of replacing infrastructure and investing in innovation to meet the future needs of students, employers and society (UUK, 2016).

[6] Cost of conducting research: This includes academic staff, training of postgraduate research students, fieldwork, and laboratory and studio work. It also includes maintaining and replacing infrastructure, and investing in innovation (UUK, 2016).

expenditure was 1.7 percent of GDP in 2014, well below the Organisation for Economic Co-operation and Development (OECD) average of 2.4 percent (UUK, 2016).

In the aftermath of the economic crisis, 2008 and after, most of the OECD countries adopted austerity measures, as an immediate response. Throughout the OECD from 2000 to 2012, the average share of government funding for HEIs dropped from 68.8 percent to 64.5 percent (OECD, 2015). Government funding for HE and research has been significantly curtailed, as outlined in the 2015 Comprehensive Spending Review. An indication of the overstretched climate is that teaching funding from direct government grants will fall by £120 million in cash terms by 2019–20, which will lead to sharp budget cuts (UUK, 2016).

Cuts in public funding to UK HE make it imperative that universities utilise their funds efficiently. Increasing efficiency became a government priority more than ever, whereby efficiency and productivity indicators have been highlighted as central tools in this policy. Consequently, the British government has recommended the calculation and publication of performance indicators. These indicators aim to measure how HE providers perform objectively and consistently. The UK Performance Indicators (UKPIs) for HE provide information on the nature and performance of the HE sector in the UK. According to the Higher Education Statistics Agency (HESA), data has been collected to define UKPIs, such as institutional indicators, sector indicators, and associated benchmark values (HEFCE, 2013).

The discussion should begin with a clear understanding of how efficiency measures are important concepts for universities that aim to bring down costs and make their resources go further. By maintaining control over pay costs, making better use of estates, and sharing assets and services, institutions gain a constructive allay to anticipate key policy questions around public funding. In terms of efficiency and cost savings, English universities have achieved £2.4 billion over the past decade.[7] Therefore, public cuts can result in better allocation of government funds, gains in efficiency, and economic dynamism.

What still needs to be clarified is what drives efficiency in universities. Multiple factors can drive universities to ensure that efficiency and value for money are core strategic and operational priorities. Beyond a simple response to austerity measures and adaptation in a new public funding environment, universities intensify their market position by being more competitive in a constrained fiscal environment. One of the key factors in this attempt is the increasing amount of investment in people and infrastructure (UUK, 2015). The existing reports produced by institutions in England provide a valuable resource for understanding efficiency and cost savings in the HE sector. However, a more robust, accessible, and comprehensive extrapolation of the key findings is needed.

[7] Between 2005–06 and 2013–14, it is estimated that UK universities delivered £2.38 billion of efficiency savings (UUK, 2015).

1.2 Thesis Motivation

The value of efficiency analysis for planning purposes maintains focus on university governing bodies in order to understand how universities perform in terms of efficiency, effectiveness, and value for money in a way that potentially creates a more thorough and balanced accounting framework (NRC, 2012). Monitoring institutional sustainability in all its forms[8] is a key role for governing bodies (UUK, 2015). Specifically, the UK HE sector is committed to developing robust mechanisms to evidence its success in delivering efficiency and cost savings. To date, many studies have been conducted to provide such information about the HE sector.[9] Even so, there is space for further research since recent issues await further study.

By opening the dialogue on literature oversights in efficiency topics, one of the initial research gaps are the effects of merger activity on efficiency. The so-called 'merger fever' as a means of restructuring HE is not a new phenomenon; universities have been forming alliances of one type or another, or complete mergers, since the 1980s (Curaj, 2015). The main rationale of such partnerships is to increase didactic and scientific performance and the efficiency of the educational system as a whole. The 'black box' of the merger process, drivers leading to mergers, and short- and long-term effects of merger outcomes have been extensively discussed in the merger literature (Pinheiro et al., 2016). However, there is still a lack of analysis in the literature on how mergers as a policy instrument can act as an enhancing factor for efficiency, and whether this is the case in the English HE system.

Another topic that still needs to be explored further, though it is an emergent phenomenon gaining ground, is universities' engagement with the external environment. Along with the universities' core missions of teaching and research, an economic development mission has come to the forefront, the so-called 'third mission'. Universities have to establish a secure linkage with the external environment to demonstrate their relevance and secure additional funding (Koryakina et al., 2015). Therefore, any interconnections and interdependencies between HE, society, and the economy are supplied through the close links between education, research, and innovation – the three sides of the 'knowledge triangle' (European Commission, 2008).

The social impact of research in an environment in which universities have been considered as engines of economic growth is conclusive. While the contributing role of universities in the transition to a knowledge economy has been well discussed in the literature (Laredo, 2007; Montesinos et al., 2008; Jongbloed et al., 2008; European Commission, 2008; DBIS, 2009), the potential links and impact on efficiency are not well explored. Addressing this deficiency of previous studies in HE and adding to the existing studies by Johnes et al. (2008) and Thanassoulis et al. (2011), interactions between third-stream activities and efficiency will be explored through a second-stage analysis.

[8] Including both academic and financial.

[9] Kosor (2013) provides a well-informed review of the available conceptual basis for addressing efficiency in HE.

A growing number of studies that have applied a two-stage DEA approach can be traced, wherein efficiency is estimated in the first stage, and then the efficiency estimates are regressed on covariates[10] that are viewed as representing environmental variables. In the HE zone, Wolszczak-Derlacz and Parteka (2011) have extensively explored efficiency and its determinants in a set of HEIs in several European countries, utilising nonparametric frontier techniques combined with bootstrapped truncated regression analysis in a second step. However, researchers have underlined that problems arise, since true DEA scores are unobserved and replaced by the efficiency estimates generated in the first stage. Those estimates are likely to be serially correlated in many unforeseen ways. Simar and Wilson (2007) have exposed the various pathogeneses that occur in such approaches since the error term tends to be correlated with the environmental variables. Also the efficiency estimates produced lack a coherent data-generating process (DGP) which is problematic.

A comprehensive bibliography is offered by Simar and Wilson (2007a) on two-stage DEA studies, which points out how the standard approaches to inference are incapable of producing consistent inference. Hence, as a robustness check on the inference derived from the conventional two-stage DEA approach, as well as to avoid potential endogeneity problems, propensity score matching (PSM) techniques will be applied. These are quasi-experimental methods that aim to contribute to and improve current knowledge on the estimating effects of treatments,[11] interventions, and exposure on outcomes attributable to a particular programme (Austin, 2011).

The production relationship of transforming essential inputs, the factors of production, into desired output(s) and the measurement of efficiency in production originated with the work of Koopmans (1951) and Debreu (1951), and was empirically explored further by Farrell (1957). In the HE sector, it is usual to access and label efficiency outcomes through production frontiers[12] (Mc Millan and Chan, 2006; Johnes, 2006; Agasisti, 2011; Sav, 2012a, 2012b; Das and Das, 2014) since production theory is the most widely used conceptual framework to examine the efficiency of for-profit and not-for-profit institutions (Titus and Eagan, 2016). A growing number of universities are encountering financial difficulties and, in the UK, competitive context identification of strategies to reduce production costs is of immediate priority. A mechanism to achieve such a goal of lower costs is the analysis of the cost efficiency of the unit, i.e. a university's cost (in)efficiency part can be further decomposed into two parts: persistent and transient, as mentioned earlier by Colombi et al. (2014) and Filippini and Greene (2016).

From an economic policy point of view, the identification of both types of cost inefficiency is crucial since public financial resources should be allocated progressively

[10] Typically, potentially exogenous, different from those used in the first stage.

[11] A treatment here can be considered the decision to merge, or not.

[12] In his preliminary work on the HE production function and its limitations, Hopkins (1990) offers a clear discussion of the relevant work in the field and its strengths and limitations. In terms of the shortcomings, he points out that a production function for HE will never be fully specified because 'there are simply too many intangibles relating to the abilities of various key actors to contribute to the process of education and research in ways that never will be very well understood' (p. 32).

to universities already operating with high degrees of efficiency. A considerable knowledge gap still exists around cost efficiency aspects since, thus far, cost efficiency studies in HE have tended to blend the short-term (transient) and the long-term (persistent) components of inefficiency, or assess only one attribute. Therefore, the aim is to improve on the received literature within these dimensions and launch the first study within the literature in the English HE sector.

1.3 Thesis Objectives

This study has three major objectives: to present an analytically well-defined concept of efficiency in HE in England, to assess whether university mergers as a policy tool can have a conducive effect on university performance in terms of efficiency, and to discuss the topic of cost efficiency in the English HE sector since it warrants research attention in a climate of curtailed public funding.

The extent to which the HE sector is already aligning itself to be more effective and efficient is not always concrete (UUK, 2011). An overview of the existing climate and a more sector-wide approach to identifying any significant developments or efficiency progress at the institutional level might be necessary. Therefore, the present thesis attempts to crystallise the blurred factors that influence efficiency in HE in England; but first, it attempts to present a well-structured landmark for what has been done already in terms of efficiency measurement in HE in the archival research.

A deeper understanding of institutional mergers' effect on efficiency may be both necessary and possible, since, in countries with many universities and colleges developed on very small and frequently narrow scales for historical reasons, it is commonplace to use mergers to realise potentially significant savings (Johnstone and Marcucci, 2007). Thus, two alternative approaches are recommended so as to produce empirically valid inferences from the results. In the context of this thesis, a two stage-DEA approach is applied, as well as a propensity score matching (PSM) method, since this uses observational studies to estimate the effects of treatments (mergers) on outcomes (efficiency) (Austin, 2011). In addition to its obvious policy and research value, the dynamics and financial implications of any merger activity may generate insights that lead potentially to enhanced departmental, institutional, or systemic educational processes.

Similar to other sectors of the economy, HE has experienced the challenge of installing a functional funding model for the sector. This is a common trait among English institutions since they need to comply with financial pressures. Little attention has been devoted to the impact of the various financial risks that universities are required to manage. Previous research has suggested that universities respond to such challenges in a number of ways, by following prudent financial management, generating new sources of cash for investment, eliminating costs, and intensifying efficiency and cost savings. In particular, in England, these efficiency savings have increased to £2.4 billion over the past decade (UUK, 2017). In pursuit of these dimensions, the objective of this thesis is to explore further the nature of the cost inefficiencies straining the sector by extracting accurate inferences of whether those inefficiencies have a permanent or a temporary effect and, by extension, providing valid policy guidance.

1.4 Thesis Outline

The remainder of this introductory section provides a short overview of the structure of the thesis. This thesis explores efficiency issues in the English HE sector and is structured as follows. As a starting point and an auxiliary tool to enable the reader to become acquainted with the English HE sector and the various transformations and reforms sustained, a short summary of the leading dimensions is presented in Chapter 2.

After clarifying the peculiarities of the English HE sector, Chapter 3 aims to present a critical review of the available approaches to measuring efficiency, the possible estimation methods, the model specifications, and the findings from studies on efficiency in HE. Therefore, Chapter 3 serves as an operationally practical guideline for measuring efficiency and aims initially to analyse the production relationship between the involved inputs in the HE process. A better understanding of the production process and the efficiency levels of HEIs is imperative so that public funds can be used more effectively. The main concern in this developing literature, is focused mainly on which technique is appropriate for estimating efficiency, since there are no accepted criteria for choosing a broadly acceptable approach and there are many alternative and conflicting approaches. Therefore, this chapter, apart from displaying the core theory used in efficiency studies, will enhance readers' knowledge of the research gaps in this area.

Chapter 4 accentuates the realisation that little is known of the effects on efficiency of merger activity in the HE sector. Publicly funded sectors are under pressure to deliver more for less, and none more so than the English HE sector. There has been speculation that funding cuts can be absorbed by efficiency savings (Mandelson, 2009), which might be achieved to some extent by the closures or mergers of some universities (Griffiths, 2010). The chapter explores the emerging literature to find instances of proposed mergers following cuts in public funding in UK HE (Baker, 2011; Matthews, 2011) and analyses the effects of mergers by using the production surface of the English HE context based on up-to-date data, including all types of HEI.

Clearly, there is a need for even more detailed and robust methods when treatment effects are assessed, since mergers are identified as a potential policy tool for adapting to failure (Browne, 2010). Given this orientation, PSM techniques are utilised in Chapter 5 as a robustness check of the outcomes of the previous chapter. This technique improves on the received literature on mergers in HE in England since inverse probability of treatment weighting (IPTW) using the propensity score is applied for first time, allowing us to obtain unbiased estimates of causal treatment effects using observational data.

In the next chapter, the cost structures of universities will be explored, with a variety of frontier methods based mainly on stochastic frontier analysis (SFA). One of the common deficiencies in most studies analysing cost efficiency in HE is that they tend to confound or to separately and independently identify only one of the two parts of the productive efficiency. Therefore, the aim of Chapter 6 is to cover the existing gap in the literature regarding the persistent and transient parts of cost efficiency since the distinction seems to be appealing also for regulators. In particular, during the last decade, HE has witnessed a wave of regulatory reforms aimed at improving efficiency through incentive regulation. Most of these regulation schemes use benchmarking, namely measuring institutional efficiency and rewarding universities accordingly. The main goal of this chapter is to assess the level of persistent and transient (in) efficiency in the English HE sector and to examine further implications for each university.

Throughout this thesis, the focus is on analysing the production and cost relationship in the English HE sector. Productive efficiency in UK HE has been observed by many researchers (Athanassopoulos and Shale, 1997; Flegg et al., 2004; Flegg and Allen, 2007a, 2007b; Johnes 2008), in terms of DEA, the disadvantages of which are well known. However, this is less well explored compared to the considerable body of literature on the cost efficiency of universities (Johnes et al., 2005; Stevens, 2005; Johnes et al., 2008; Johnes and Johnes, 2009; Thanassoulis et al., 2011), wherein parametric techniques such as SFA are used frequently. In the last part of this thesis, Chapter 7, a summary of the main findings and of the principal issues and suggestions that have arisen in this thesis are provided. Also, most importantly, this chapter gives implications and policy recommendations for the English HE sector since it has changed in size and shape over the past decade. The combination of the findings provides some support for the conceptual premise that the English HE sector is indeed a sector with various demographic, economic, technological, and political changes that are likely to have implications for future patterns and trends in the composition and finances of HEIs (UUK, 2017). This chapter also highlights the limitations of the thesis and indicates future directions for research in this area.

2. Chapter 2: The English Higher Education Sector

2.1 Introduction

HE comprises education provided by institutions such as universities, colleges, and other various academies that award academic degrees. HE consists of the obtained basic knowledge beyond secondary or elementary education. In England, it is evident that significant development of the prominent universities has been realised, with the further establishment of academies outside university walls (Young Yoon, 2008). There are several types of HEI in the UK, which can be classified as universities and university colleges or other HE providers, i.e. HEIs, further education institutions (FEIs), and alternative providers (APs).

The term ‘university’ is a constituted name for only some institutions that have the right to use it. If an institution hopes to gain university status, there are two possible routes: either through the Privy Council,[13] since this formal body of advisers (senior politicians) is responsible for approving the use of the word ‘university’ (including ‘university college’), or through the provision of the Companies Act 2006.[14] Higher education institutions (HEIs) is a term from the Further and Higher Education Act 1992 (HEFCE Glossary)[15]. The Act, defines as such any provider which is one or more of the following: a UK university; a higher education corporation; a designated institution. In the same sense, HEIs can be split into: a) universities; b) institutions run by a higher education corporation; or c) institutions designated as eligible to receive financial support administered by the Higher Education Funding Council for England (HEFCE) (EED, 2017). At present, all English HEIs draw upon funding administered by the HEFCE, which directly funds 128 HEIs. HEFCE may choose to fund HEIs for teaching and research if certain conditions of grant are met.

FEIs offer a range of HE programmes provided in over 250 further education colleges (FECs). In the current year 2017, 241 FECs deliver HE, one part of which is directly funded by the HEFCE, while others deliver HE through a sub-contractual arrangement (HEFCE, 2017). Further education courses fall into the two categories of academic and vocational, and are prominently respected by employers and academics worldwide. The last category is APs, which do not receive direct annual funding from the HEFCE, do

[13] This is possible through the Further and Higher Education Act 1992. The Privy Council is a formal body of advisers to the Sovereign in the UK. Its members mostly are senior figures who are (or have been) members of the House of Commons or the House of Lords (http://www.hefce.ac.uk/Glossary/#letterP).

[14] The Companies Act 2006 is an Act of Parliament of the UK that forms the primary source of UK company law. For a formal overview of the act, see http://www.legislation.gov.uk/ukpga/2006/46/contents.

[15] See http://www.hefce.ac.uk/Glossary/ for definition.

not receive direct recurrent public funding from local authorities, or from the Secretary of State for Education or its equivalent bodies in the devolved administrations, and are not embodied into FECs. These institutions might be independent private institutions, including both for-profit and not-for-profit organisations, and are not FECs. These other providers of HE programmes may use other institutional titles such as 'college'. The use of such titles is not regulated by law. As of 13 March 2017, there were 115 APs with specific course designation for 2016–17. Seven APs have degree-awarding powers of some description and HESA reports that there were 52,675 students on designated courses at APs in 2015–16 (HEFCE, 2017). Furthermore, more than 700 colleges and other institutions in the UK are not able to award degrees but provide complete courses leading to recognised UK degrees. These courses, once completed, are validated by institutions that have degree awarding powers.

The remainder of this chapter is structured as follows. It focuses primarily on the history of English HE and the stages of development the sector has undergone. During the expansion of the sector, significant structural financial changes took place so as to ensure the financial sustainability of the system. Finally, HE participation has expanded dramatically in England over the last half century. Therefore, addressing inequality of access to university for socio-economically disadvantaged students is still a major policy challenge and, as such, the final part of this chapter explores this issue.

2.2 History

The roots of English education lie in the middle ages, initiated by monasteries in the sixth century, where students were trained to be members of the clergy or were members of the ruling class. However, this form of education was elementary and only began to be shaped into a more advanced and secular framework two centuries later. The first action in HE in England is attributed to Alfred the Great around the 9th century. From that time and onwards since the establishment of the University of Oxford in the 12th century, HE has been variously developed. Almost a century later, the University of Cambridge was established when a number of students left Oxford after a dispute with the townspeople. In the same period, the first colleges were established also. Various colleges owned lands and were faithful to the church, such as Merton College, Exeter College, Oriel College, and Queens College; however, some colleges expressed their independence, such as Winchester and Eton, which were established in the late 14th and early 15th centuries (Young Yoon, 2008). However, HE was not free from religious control at this time, so the subject areas were limited. Normally, the structure of the curriculum and the emphasis were focused mainly on Latin grammar and literature, with arithmetic, geometry, music, and astronomy having a secondary role after students had been awarded the degree of Bachelor of Arts.

Generally, HE in the UK has a long history.[16] Apart from the Oxford University, which documents teaching from 1096, making it the oldest university in the English-speaking world, the University of Cambridge celebrated its 800th anniversary in 2009; three Scottish universities – St. Andrew's, Glasgow, and Aberdeen – were founded by papal bull in the 15th century; and the University of Edinburgh was established by royal charter in 1583. In the following years, great expansion and a shift towards HE in the UK occurred in the 19th century due to industrialisation and, in the latter part of the century, medical, science, and engineering colleges in England's major industrial cities were installed. Part of those colleges were eventually amalgamated to become the so-called 'redbrick' universities of Birmingham, Bristol, Leeds, Liverpool, Manchester, and Sheffield. The redbricks constitute the main body of what comes under the name of Russell Group universities (research intensive universities foremost), with some exceptions of few post-1992 entrants.

By the end of the 19th century, despite the extraordinary expansion on the participation rates, HE had a limited role only to provide a well-skilled and educated workforce for British industry or to serve national issues (Hayton and Paczuska, 2002). Even in the early 20th century, HE was still overwhelmed by the upper-middle classes, maintaining an elitist system. However, to a certain extent, significant steps for wider participation have been made, overriding financial, cultural, and social barriers to accessing HE (Archer et al., 2002). British HE was fundamentally a private endeavour in the 19th century; but a century later, it has been transformed into state-dependent institutions that are generally considered to be public.

The second expansion wave in the HE sector occurred during the 1950s and 1960s. This was a consequence of the high demands of an overgrowing population combined with the new requirement for a society adjusted to new technological advances. Institutions such as Aston, Bath, Bradford, Brunel, City, Loughborough, Salford, and Surrey were all awarded university status in 1966,[17] since they had previously functioned as new colleges of advanced technology, established in 1956. Following the same pattern, 13 more UK institutions, including Hull and Leicester, both former university colleges, obtained university status during the 1950s and 1960s. At the same time, this expansion replenished the sector with seven new institutions, i.e. the universities of East Anglia, Essex, Kent, Lancaster, Sussex, Warwick and York[18].

The Robbins Report on HE, published in October 1963, became the basis for many of the changes in British HE. Since then, the dramatic expansion of the HE system has brought various transformations, some more successful than others. The report implied the immediate expansion of universities, and greater reform within the sector by giving

[16] For a more thorough analysis and a review of English HE, see Gillard (2011), *Education in England: A brief history* (www.educationengland.org.uk/history).

[17] In this way, the University of Wales Institute of Science and Technology went on to become a constituent part of what is now Cardiff University in 1988.

[18] For further details on the history follow the link: https://www.google.gr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiqgZn_qMfWAhUDOhoKHRvED2EQFggmMAA&url=https%3A%2F%2Fwww.britishcouncil.in%2Fsites%2Fdefault%2Ffiles%2Fhigher_education_system_of_uk.pdf&usg=AFQjCNGDiKj1ht_QLe65XIUTKkQIW2MOpg

university status to colleges of advanced technology. The great HE expansion was reflected also in the establishment of a new body of institutions, the well-known polytechnics. These entities did not acquire university status and they were locally controlled and financed. Their main objective was to provide more practical and technical subjects, and, due to their inability to award degrees on their own right, they coordinated with national awarding bodies[19] under a formal recognition arrangement. Through the 1980s, polytechnics gained autonomy from local governmental control and approached the British university model (UUK, 2005). In light of the Robins report proposals, the number of full-time university students increased from 197,000 in the 1967–68 academic year to 217,000 in 1973–74, with further big expansion thereafter. The report also stressed the necessity for wider participation in university since, 'Courses of higher education should be available for all those who are qualified by ability and attainment to pursue them and who wish to do so' (Robbins, 1963).

According to the report, institutions should have four main objectives essential to any properly balanced system, the so-called 'Robbins principles', which constituted a core guideline for universities' research and teaching activities (Robbins, 1963). Robbins gave prominence to the value of HE as an investment action and the central role of tuition fees and government support. He enhanced and outlined the structure of governing bodies, the balance between teaching and research, the place of business and management studies, the value of study in modern languages, the need for flexibility in curricula, the importance of postgraduate study for UK students, and the danger of university selection processes that used excessively narrow criteria and pushed secondary schools to narrow their own curricula (Barr and Glennerster, 2014).

The Robbins Report can be deemed as an articulate research effort to support evidence-based policy. At that time, new institutions joined the HE sector and the economic insolvencies of low-income students began to be supported through a national system of state support. The massive growth in the overall participation rate in the following 40 years was remarkable since, by 1966, there were HEIs with more than 44,500 enrolled students (Whitty et al., 2015). Although there have been other significant inquiries, HE policy has often lacked such an evidentiary basis. However, not every recommendation of the Robbins Report was adopted. Therefore, remarkably, many of the issues it raised remain central to HE debates today.

2.3 Development

The publication of the Robbins Report in 1963 signified a pivotal shift from an 'elite' to a 'mass' system of HE in England (Trow, 1974; Scott, 1995). This expansion was even more evident after the Further and Higher Education Act 1992,[20] when UK HE was endowed with 35 more institutions previously carrying the statuses of polytechnics or colleges of higher and further education. This was the year in which John Major's government granted university charters to a number of former polytechnics and colleges

[19] National Councils for Academic Awards.
[20] See http://www.legislation.gov.uk/ukpga/1992/13/contents.

of HE. While the ‘new’ or post-1992 universities do not have the prestige or ranking of their older counterparts, they are quickly gaining ground to compete with the older institutions mainly in terms of the facilities’ infrastructures. While the nexus of Oxbridge still maintains the elite power in academia, the great expansion in the HE sector signalled in 1992 and afterwards has contributed to many more opportunities for HE. The main element of the 1992 act focused on the idea of increasing the quality and scope of research in subjects seen as applied or vocational, such as art, education, and information technology. This act has continued to be applied since 1992; however, in the last decade, the sector has been unable to flourish at the same pace as previously due to difficulties and mergers.

Between 2001 and 2013, 31 more universities were created. Those universities were an amalgam of those resulting from the break-up[21] of the federal University of Wales and a further ten university colleges that had had their applications for university status put forward to the Privy Council for formal approval. Those universities joining the sector but previously serving as vocational institutions have collectively become known by the terms ‘post-92’ or ‘modern’ universities. HE in the UK is now provided by a diverse range of organisations. 166 institutions currently have their own degree-awarding powers[22]. Seldom most of the institutions that raised to ‘university’ title satisfy certain prerequisite criteria, but, at the same time, there are enlisted institutions that do not have the power to award their own degrees.[23] Where this is the case, such institutions have the right to render complete courses leading to recognised UK degrees. However, such courses are validated by institutions that have degree-awarding powers.[24] According to official state estimates in 2011, in addition to degree-awarding institutions, there were more than 1,600 bodies, including 250 FECs, offering some form of HE provision.

HEIs that had university status before the provisions of the Further and Higher Education Act 1992 came into force are included with the pre-1992 universities. These universities are widely known as ‘old’ institutions that were part of the university sector prior to 1992, including universities created by ancient usage, Act of Parliament, or Royal Charter, and full colleges of the federal universities of London and Wales in 1992. Their mission is slightly different since they are traditional universities, with an articulate focus on rewarding research (Parker, 2008). From a deeper perspective, institutions that formally have been universities since medieval times belong to the category of ‘ancient’ universities. Some of the most typical cases are the universities of Oxford, Cambridge, St. Andrew’s, Glasgow, Aberdeen, and Edinburgh, the dates of foundation of which range from 1096 to 1582.

[21] Merged institutions already possessing the title of ‘university’ are not considered.

[22] The names of institutions with their own degree-awarding powers (‘Recognised Bodies’) are available for download at: http://www.dcsf.gov.uk/recognisedukdegrees/index.cfm?fuseaction=institutes.list&InstituteCategoryID=1.

[23] Since 2005, institutions that have had degree-awarding powers and at least 4,000 full-time equivalent students, of whom at least 3,000 are registered on degree-level (including foundation degree) courses have also been permitted to apply to use the title ‘university’. According to a white paper (BIS, 2011), the student requirement was reduced to 1,000 full-time students. Also, institutions that award taught degrees but which do not meet the numerical criteria for the university title may apply to use the title ‘university college’, although not all choose to do so.

[24] Since 2008 in England, and 2010 in Wales, FEIs have been able to apply to the Privy Council for powers to award their own ‘foundation degrees’ (typically, vocationally focused and equivalent to two thirds of a full honours degree).

The Robbins Report on HE gave further space to the UK government for a further endowment of new institutions since it declared that there was a lack of universities in the UK. The report gave rise to the creation of a whole new group of universities known as the plate glass or 1960s universities. It is worth noting that the original plate glass[25] universities were established following decisions by the University Grants Committee (UGC) in the late 1950s and early 1960s. This is a group with mixed composition that embodies some of the most prestigious institutions, such as the University of Warwick, while others coast in the middle of the league tables.

However, universities in the UK vary in many aspects, from age to size to prestige, and this diversity provides choice. The UK HE system has so much to offer and one of the choices is the type of institution at which to study postgraduate courses without attending an official university. As the name suggests, specialist higher education colleges offer a small number of postgraduate programmes. This means that, apart from the exceptional teaching level, they offer even the most unusual subjects in different fields and the entire college will be orientated towards that certain subject. Some examples of this kind of college are the specialist agricultural colleges and those that specialise in the creative and performing arts. The powerful capability of these bodies is that they often offer courses that sometimes are not available at fully independent universities, allowing students to obtain professional expertise in subjects about which they are passionate.

University groupings are informative in many different ways, but mainly in terms of subject choice. Institutions formed around the same time are often similar in style and prestige. Universities tend to show solidarity by forming official organisations so as to lobby government over research funding. One such groups that is regarded as very prominent is the Russell Group, which is considered an elite group of the most prestigious universities with high-quality research strength. Indeed, on average, universities in the Russell Group demonstrate considerable excellence in research and account for two thirds of university grants awarded in the UK. Members consist of mainly the ancient and 19th-century universities, with a few of the larger civic universities. However, it should be stated here that this does not mean that a university is excellent in all areas of research when it is classified as a member of the Russell Group, and it certainly does not mean that these universities are superior to the rest. A university should show excellence in many different aspects, i.e. employment rates, student satisfaction, and lecture quality. Indeed, there are many other universities in the UK that are not part of the Russell Group but still generate valuable research.

If we go further on this topic, there is one more group, the 1994 group. This set of universities incorporates smaller, research-intensive, and often campus-based universities. While the 1994 group also contains some of the top universities, it is considered superior to the Russell Group. The MillionPlus group[26] is a coalition of modern universities, oriented to the influences of the 21st century, and it provides

[25] For more information, see: http://www.wow.com/wiki/Plate_glass_university?s_chn=94.
[26] For more information about the members and their mission, see http://www.millionplus.ac.uk/who-we-are/our-role.

valuable advice and analysis of HE policies. This organisation represents most of the former polytechnic universities.

It should also be noted that HE policy, for a number of reasons, follows separate strategies in each of the countries making up the UK, with the Scottish government, the Welsh Assembly, and the Northern Ireland Executive each having specific and differing responsibilities for certain parts of HE and student policies. However, this chapter focuses mainly on HE in England, unless otherwise stated.

2.4 Structural Financial Reform

Certain aspects of the UK's HE system are overseen by central bodies, whereas other powers are devolved to the individual nations that make up the UK. The HEFCE funds and regulates universities and colleges in England.[27] While funding is distributed nationally, quality assurance standards are UK wide. Standards of HE are monitored and advised by the Quality Assurance Agency (QAA), which covers the whole of the UK. Generally, universities in the UK are independent and self-governing, but they maintain close links with the central government since they absorb a quantifiable proportion of their income from public funds. The financial aid from the state is not only unilateral for universities but also for students. Government endowments to cover tuition fees are also received by middle- or low-income students who receive state-funded grants inversely related to their parents' income. In addition, some students are awarded tuition reductions as a bonus based on their academic performance. Also, most students[28] receive state-funded loans to cover living expenses and these are redeemed after the student graduates and obtains a job. Public funds spent in HE combine recurrent grants, tuition fees from domestic or foreign (international) students, and income from various private sector sources.

In the aftermath of the great expansion in the 1960s, the system came under extreme pressure in the 1980s. Public funding was no longer sufficient and the grant system was too weak to support future student projects. More specifically, during the rapid expansion period between 1989 and 1997, public funding per student was downscaled by around 36 percent, putting considerable pressure on universities and colleges. From another perspective, this significant expansion in students' induction rates was not in accordance with increased funding (Eurydice, 2009). By 2005–06, total government spending on HE reached £6.5 billion (UUK, 2006). The increasing pressure on public funding in the 1980s and 1990s triggered a climate of tuition fees enforcement for UK and EU undergraduate students in the UK in 1997. In England, the fee cap was increased from £1,000 to £3,000 per annum in 2006 in recognition of the fact that despite the

[27] The Higher Education Funding Council for Wales (HEFCW) carries out similar duties in Wales. It is responsible for regulating degree levels, ensuring a quality framework, and scrutinising the relative performance of universities. The Scottish Funding Council (SFC) is the equivalent body for Scotland. It funds 25 colleges and 19 universities and its main aim is to generate high-quality learning and teaching, world-leading research, greater innovation, and widening participation in HE. Universities in Northern Ireland receive their funding from the Department for Employment and Learning (DELNI).

[28] It should be stated that foreign students and British students taking a degree at an overseas university are not generally eligible for public funding.

austerity imposition on resources, universities should preserve the quality assurance of the system as a whole.

In an attempt to release that pressure, policy officials introduced the present system of student loans.[29] Also, a regulatory system of HE funding councils in each UK territory was established to manage the system efficiently. The HE sector in England is currently in a stable financial position, reaping the benefits of positive cash flows. This does not ensure that financial performance is equivalent within the sector since there are institutions with continuously dwindling figures in student recruitment that further imply compressed income (HEFCE, 2013). The tight government budget constraints have given rise to a novel policy, with tuition fees[30] enforcement for HEIs in England reaching a maximum of £9,000 per year in 2012–13 from a previous ceiling of £3,290 per annum (Callender, 2015).

Under this context, an increase in fee income for home and EU students was observed in 2012–2013. However, the significant fall in public funding through the HEFCE grants was counterbalanced (HEFCE, 2013). In 2016, the upper limit of tuition rose further, to £9,250, with plans to allow further increases to keep up with inflation. Undergraduate fees can be up to £9,000 per year at English, Northern Irish, and Welsh institutions. At the same time, in Scotland, fees can be much lower for Scottish and EU students, at around £1,800 for a first degree, but can remain at up to £9,000 for students from elsewhere in the UK. The situation is slightly different in Northern Ireland, where student fees are concomitant to the rate of inflation.

The reform of English HE was circumscribed by a report[31] by the Independent Committee on Student Fees and Funding (ICSFF), chaired by Lord Browne, who set out a blueprint for ‘a new paradigm in English HE’ (ICSFF, 2010). Only half a year after the reform, the white paper ‘Higher Education: Putting Students at the Heart of the System’ (BIS, 2011) schematised the policies the government planned to adopt to implement the committee’s other recommendations and produce a workable system for a new funding regime (Scott, 2014). Consequently, the white paper offered more than an interim stage in the reform process, generating major inferences since its publication; however, substantial matters remain unresolved (Scott, 2013).

According to Scott (2013), there have been 11 major policy interventions in HE since 1960. The directional lines have been set thorough major reports, white papers, green papers, ‘letters of guidance’ to funding agencies, select committee reports, enquiry reports, planning papers on student numbers, and significant pieces of legislation. Other

[29] These are usually from devolved government agencies. English students receive financial aid from the Student Loans Company, a government-owned agency. Students usually take out a tuition fee loan that is paid directly to the relevant HEI and a maintenance loan to cover living costs, which are paid directly to the student. A generous repayment regime has been established, under which graduates are not required to make any contribution to the repayment of their loans until their incomes reach £20,000; a cap is set on the percentage of their income that can be used to make repayments, and any remaining amount is written off after 20 (or 30) years (Scott, 2014).

[30] In the 1998–1999 academic year, the tuition fee of £1,000 was introduced for full-time UK and EU undergraduate students. The fees were disbursed according to families’ residual income and students from low-income families were exempted from these charges.

[31] This report was compiled in October 2010.

intermediate agencies[32] have also developed complete reviews regulating policy interventions, in the form of circular letters, consultation documents, and annual reports. Scott (2013) offers an informative chronological review of the main HE interventions[33] in the UK. Generally, there is a tight fiscal dependency between HEIs in the UK and government income; however, they still maintain their legal autonomy. Each institution defines its own admissions policies and coordinates with its funding council regarding student number targets and the funding scheme. The underlined rationale beyond governmental funding in the UK is distributed on the basis of competitive research funding exercises or according to competitive, peer-reviewed proposals, as in the United States (US) (Johnstone and Marcucci, 2007).

In a climate of intense competition for students between UK HEIs, alongside the 2010 fee reforms, the recruitment quotas in England have changed hands from government regulation to the sole control of each university. The flexibility in student recruitment rights stemmed from the fact that HE costs were no longer only a state affair but also an individual concern. This trajectory policy, combined with the relaxation of the student cap in 2014–15, permitted further considerable expansion for some institutions (plans for expansion), or reductions in others. Note that, implicitly, the intention behind the Browne committee's recommendation in line with the white paper's proposals for a new tuition fee regime with higher fees was to fund continued growth in student numbers (Scott, 2013). Therefore, instead of restricting student's participation or reducing the per capita funding, the committee came up with the induction of a new flexible regime for tuition fees.

2.5 Fair Access and Widening Participation

It is not an overstatement to say that fair access and widening participation are mainly defined by government practices. In particular, in cases in which the HE system is heavily supported by public funds, there is minor control over student numbers. The period of great expansion, when participation rates doubled between the late 1980s and the early 1990s, is an exception, since the policy officials were prepared to increase public expenditure with increasing demand, though not in proportion to the increasing numbers of students (Thompson and Bekhradnia, 2011). In a formal expression, the Office for Fair Access[34] (OFFA) defined the widening participation as 'Removing the barriers to higher education, including financial barriers that students from lower income and other under-represented backgrounds face'. This means an improved representation pattern at a national level that supports not only young people from low-income backgrounds but every kind of underrepresented group in HE, i.e. individuals facing disabilities, ethnic minorities, part-time and mature students, people estranged

[32] Those agencies include UGC and the National Advisory Body (for public sector HE), the Universities Funding Council and Polytechnics and Colleges Funding Council, the Higher Education Funding Council for England (HEFCE), the Scottish Funding Council and Higher Education Funding Council for Wales.
[33] An analytical guide of the interventions proposed by Scott is available on appendix 1 Chapter 2 of this thesis.
[34] OFFA is an independent public body, established by the 2004 Act. The main scope of this organisation is to ascertain that the new tuition fee regime will not be detrimental to widening and increasing student participation. See https://www.offa.org.uk/ for further details.

from their families, people from gypsy and traveller communities, refugees, students with mental health problems, specific learning difficulties, and/or who are on the autism spectrum, and students from military families.

There are two responsible bodies that plan the cap for the number of students in the HE sector in England, and these are the government officials for HE policy issues and the HEFCE. The HEFCE has the power after a cycle of procedural talks with the stakeholders and sufficient government guidance to set a limit on the control of student numbers. This procedure ensures an adequate distribution of the level of publicly funded student loans and grants for fees and maintenance. However, universities and colleges are permitted to recruit as many students with high grades (currently ABB or above at A-Level, and certain equivalent qualifications) as they wish since those students are not counted as part of their permitted cap, which expands and improves students' choices (EuroEducation, 2017). However, reducing the 'high-achieving' threshold could be a risk for the sustainability of the growth rate of participation since those with lower qualifications have lower HE participation rates, so the uncertainty of the irregular distribution of students is correspondingly greater (Thompson and Bekhradnia, 2011).

If we take stock of the debate on the interactions among income, social and cultural factors, or price elasticities that can affect HE participation, the results are promising (Callender, 2011). Research on income effects reveals that participation of students from lower-income families is prone to downward shifts when tuition prices increase (Mundel, 2008). Therefore, a strong emphasis has been placed in English HE on the fact that increased grants and maintenance loans can potentially reduce the drop-out phenomenon of low-income students and can directly increase their chances of qualifying with a good university degree. In this way, participation in HE becomes more attractive, especially for those at the bottom end of the income spectrum (Thompson and Bekhradnia, 2011). The criteria under which awards or financial support are provided can be a decisive factor in promoting fair access and retention in HE.

Fair access is defined by the Office for Fair Access (OFFA) as 'equality of opportunity for all those who have the potential to benefit from HE, irrespective of their background, schooling or income'. In a more rigorous expression, fair access is considered equal access for those who are equally well qualified (Boliver, 2013). The notion of fair access in most cases is related to admission to the most selective institutions that tend to apply strict overall entry requirements. Therefore, the probability of under-represented students or applicants from disadvantaged backgrounds being accepted at such institutions is relatively small. More precisely, entry rates to the Russell Group, other old, and new universities by social class, school background, and ethnic group implies a concrete discrepancy, with an apparent social class gradient in terms of rates of entry to other old and Russel Group universities, with the former being much less steep than the latter (Boliver, 2013).

2.6 Concluding Remarks

Analysing the sector with scrutiny, there have been successive waves of HE expansion and regulatory reforms in the UK over the last century. Typically, regulatory reforms were held up as an 'off-the-shelf' solution for efficiency, contributing to systematically reducing costs associated with compliance and facilitating increased efficiency (UUK, 2011). This expansion has been considered imperative, and the climate of growth has prospered independently of any political changes. However, any kind of restructure in the sector has been followed by widespread fears regarding the future effects. Thus, it should be pointed out that expansion tensions have faced distrust, with the fear of diminishing the value of a degree. Apart from the quality effects, the new tuition fee regime in England has been faced with hesitation, since it may deter applicants, particularly those from low-income backgrounds. However, as an answer to any views of reluctance on the novel regime, the number of full-time enrolments rose to historically high levels, with recruitment on full-time courses between 2010–11 and 2013–14 growing by 3.2 percent (UUK, 2014). This enhanced participation has been smooth in the disadvantaged and underrepresented groups as well. The general HE contribution on industrial, financial, educational and societal aspects cannot be undermined and is significant, with a more pronounced effect in the last decades. In particular, between 1982 and 2005, the increase in graduate skills in the UK economy contributed around 20 percent of GDP growth (Holland, 2013).

3. Chapter 3: Efficiency Measurement, Methods, Estimation, Model Specification and Previous Literature

3.1 Chapter Background

Measurement of efficiency in HE has been well explored in the literature for more than 20 years. The growing public concern regarding performance and efficiency measures in the HE sector can be explained by the massive expansion of the HE systems worldwide. Additionally, the financial constraints stemming from the current economic challenges associated with tight government budgets and increasing pressure for greater autonomy of HEIs have contribute to this end (Cunha and Rocha, 2012).

HE financial policy commonly goes hand in hand with numerous subsidies and grants supported by political authorities. The main aim of HEIs is to obtain at least some of their income from public funds; therefore, it is essential, in the interests of accountability, to measure inter- and intra-institutional efficiency (Johnes, 2005). The HE sector, however, has characteristics that make it difficult to assess efficiency: it is non-profit making; there is an absence of output and input prices; and HEIs produce multiple outputs from multiple inputs (Johnes, 2006).

Economic efficiency (EE) is concerned with the optimal production and distribution of scarce resources and measures, and whether those resources (i.e. agricultural research and extension, tertiary healthcare, and HE, etc.) are being used to get the best value for money. Both productivity and efficiency measures have been defined as the ratio between output and input (Sengupta, 1995; Cooper et al., 2000). However, instead of defining efficiency as the ratio between outputs and inputs, it can be seen as a distance between the quantity of input and output. In particular, the quantity of input and output defines a frontier, the optimal frontier for a DMU relative to other units of its cluster (Daraio and Simar, 2007).

More than 60 years ago, Key (1940) laid down a challenge for economists to resolve the ‘basic budgeting problem’, namely, the scarcity of public funds and public expenditure management (Fozzard, 2001). Such considerations feature efficiency as a methodological tool to explore the relation between resource inputs[35] (i.e. costs, in the form of labour, non-labour expenditures in capital stock, buildings, equipment, and student services) and either intermediate[36] outputs (average attainment scores at the end

[35] Intermediate or not.

[36] Economic evaluations should focus on final educational outcomes rather than intermediate outputs as a measure of efficiency, which may lead to suboptimal results.

of each key stage,[37] university instruction hours, etc.) or final educational outcomes (degrees gained, number of credit hours, etc.). Solutions for the basic budgeting problem adopting the criterion of EE implies that resource allocation decisions are the result of technical analysis and political processes (Fozzard, 2001). Policymakers make choices that maximise the educational outcomes gained from the resources allocated to HE, stressing the importance of transparency in the process itself. Inefficiency exists when resources could be reallocated in a way that would further increase the educational outcomes attained.

Economic efficiency can be further discerned into different types, which are not exactly equivalent depending on the assumptions made for the optimal production, consumption, and distribution of scarce resources. The mutual assumption made in all these definitions of efficiency is the idea that a system is efficient if nothing more can be achieved given the available resources. The principal definition of productive efficiency encompasses the production of goods and services with the optimal combination of inputs[38] to produce the maximum output for the minimum cost. To be productively efficient means that a firm must be producing on its production possibility frontier (i.e. it is impossible to produce more of one good without producing less of another). Thus, productive efficiency is concerned with producing at the lowest point on the short-run average cost curve.

Productive efficiency is closely related to the concept of technical efficiency, since productive efficiency requires technical efficiency. According to Farell (1957), technical efficiency (TE) reflects the ability of a firm to produce the maximum output from the minimum quantity of inputs, such as labour, capital, and technology. TE requires the input-output combinations be on the isoquant. Production cost efficiency requires TE, and the level of inputs used depends on the prices paid for the inputs (Wagner, 2012). However, a firm is said to be totally economically efficient only if it is technically efficient and, at the same time, allocatively efficient. This means that, apart from using the optimal proportions of inputs, a firm should also distribute the goods and services according to consumer preferences. A firm could be productively efficient but produce goods people do not need; this would be allocative inefficient. Therefore, the firm should reallocate production with strict boundaries on the respective prices and the production technology so that the price of the good should be equal to the marginal cost (MC) of production.[39] In a formal representation, allocative efficiency (AE) is achieved when a firm employs factors of production up to the point at which the marginal rate of technical substitution between any two of its inputs equals the ratio of corresponding input prices (Huang and Wang, 2002).

A considerable body of literature exists on the measurement of TE in the HE sector (Worthington, 2001; Johnes, 2004; De Witte and López-Torres, 2015; Johnes, 2015;

[37] For further details follow Garniss (2006): *www.oecd.org/std/na/37562338.ppt*

[38] Given amount of inputs.

[39] A more precise definition of AE is at an output level, where the price equals the MC of production. This is because the price that consumers are willing to pay is equivalent to the marginal utility that they get. Therefore, the optimal distribution is achieved when the marginal utility of the good equals the MC. Firms in perfect competition are said to produce at an allocatively efficient level, while monopolies can increase the price above the MC of production and are allocatively inefficient.

Thanassoulis et al., 2016). Efficiency and productivity measures, as well as the Malmquist index, are concepts that have been analysed radically in the past decades in an attempt to assess the performance of universities. Early studies of TE in UK HE focused on individual departments such as accounting (Tomkins and Green, 1988), chemistry and physics (Beasley, 1990, 1995), economics (Johnes and Johnes 1993), and business schools (Doyle et al., 1996), or departments within a university (Sinuany-Stern et al., 1994).

Broadly speaking, two main camps have emerged for assessing efficiency among the proposed approaches. These are classified into parametric or econometric approaches and non-parametric techniques or programming approaches. Those that estimate maximal output and attribute all departures from this as inefficient are known as DEA,[40] and those that allow for both unobserved variation in output due to shocks and measurement error as well as inefficiency are known as SFA (Parmeter and Kumbhakar, 2014).

Both methods seek to characterise and quantify notions of efficiency; however, they are fundamentally different in their construction and underlying assumptions. Given that each possesses its own strengths and limitations, neither is generally regarded to be superior to the other (Salerno, 2003). Lovell (1993) demonstrates a taxonomy of parametric and non-parametric methods, making the assumption that, when using statistical approaches, the functional form of the production possibility set is the link between inputs and outputs, while, in non-parametric techniques, the input and output data themselves are used to compute the production possibility frontier, by using linear programming methods.

More recently, DEA has been applied at the HEI level to produce measures of efficiency for all HEIs in the sector (Athanassopulos and Shale, 1997; Flegg et al., 2004; Glass et al., 2006; Johnes, 2006; Flegg and Allen, 2007a, 2007b; Johnes, 2008; Flegg and Allen, 2009; Johnes, 2014). These studies differ in terms of the time period covered, model specification (i.e. inputs and outputs), returns to scale (RTS) assumed, and the HEIs included in the analysis. Early studies concentrate on a particular sub-sector (such as pre-1992 universities or post-1992 HEIs) and find that, on average, efficiency is remarkably high with average TE levels between around 80 percent and 95 percent (Athanassopulos and Shale, 1997; Flegg et al., 2004; Glass et al., 2006; Johnes, 2006; Flegg and Allen 2007a, 2007b, 2009; Johnes, 2014). Later studies that extend the dataset to include the complete HE sector that we observe in the UK today find a much wider range in mean TE at around 0.75 to 0.95.

The objective of this chapter is not to develop a formal theory or definition of production methodology. Rather, the aim is to detail the important econometric area of efficiency estimation in both HE past approaches as well as new methodologies. Beginning with the seminal work of Farrell (1957), various approaches to discerning output shortfall have been developed.

[40] For an authoritative review of DEA methods and their statistical underpinnings, see Simar and Wilson (2013).

3.2 The Economic Model

Economic modelling is at the heart of economic theory. Through the main axioms on which the economic model underlining the measurement of efficiency is based, the economist can experiment, at least logically, producing different scenarios, attempting to evaluate the effect of alternative policy options, or weighing the logical integrity of arguments presented in prose. Much empirical evidence suggests that, although universities may indeed attempt to optimise the distribution of the available resources, they do not always succeed. Therefore, utilising the minimum inputs required to produce the educational outputs, given a level of technology, is paramount. In light of the evident optimisation failure, certain types of model are extremely useful for presenting visually the essence of inexpediency to attain the optimal efficiency.

The economic theory of production places emphasis on efficient production and its consequences, so it is desirable to shift the analysis of production away from the traditional regression-based production function approaches toward frontier-based approaches. The traditional least squares statistical methodology offers estimated functions that intersect the data; however, the enveloping properties concerned with the estimation of frontiers envelop the data, and they are far more appealing in this case (Fried et al., 1993).

Measuring efficiency for any dataset requires a definition of the boundary of the production set; then, the distance between any observed point and the boundary of the production set should be measured (Daraio and Simar, 2007). Given a list of N inputs and M outputs, in economic analysis, the operations of any productive organisation can be defined by means of a set of points, S, the production set in a multi-input, multi-output technology,[41] defined as follows in the Euclidean space $\mathcal{R}_+^{N+M}$:

$$S = \{(x, q) | x \in \mathcal{R}_+^N, q \in \mathcal{R}_+^M, (x, q)\ is\ feasible\} = \{(x, q): x\ can\ produce\ q\}$$

x is defined as the input vector, q is the output vector and ‘feasibility’ of the vector (x, q), which means that, within the organisation under consideration, it is physically possible to obtain the output quantities $q_1, \dots \dots q_M$ when the input quantities $x_1, \dots \dots x_N$ are being used.

The production possibility set, S, can be defined in terms of its sub-sections, defined as the images of a relation between the input and the output vectors that are the elements of S. The input requirement set $C(q)$ consists of all input vectors that can produce the output vector $q \in \mathcal{R}_+^M$ for $\forall\ q\ \in$ S. It is defined as (Daraio and Simar, 2007):

$$C(q) = \{x \in \mathcal{R}_+^N | (x, q) \in S\} = \{x: x\ can\ produce\ q\}$$

[41] In the description of the production technology above, the explicit assumption underlined in all properties is that time does not count. If time is considered, then the superscript ^{t} is necessary as a label to define the time period each time, i.e. $S^t, P^t(x), C^t(q)$.

The output correspondence set $P(x)$ consists of all output vectors that can be produced by a given input vector $x \in \mathcal{R}_+^N$ for $\forall\, x \ \in$ S. It is defined as:

$$P(x) = \{q \in \mathcal{R}_+^M | (x, q) \in \mathrm{S}\} = \{q{:}\, x\ can\ produce\ q\}$$

The output and input sets are equivalent representations of the technology, as is Ψ, so it holds that:

$$(x, q) \in \mathrm{S} \Leftrightarrow x \in C(q), q \in P(x)$$

The axiomatic assumptions of the production technology and the scale of operation of the organisation are given in appendix 2 chapter 3, aiming to provide enough structure to create meaningful and useful technologies. Daraio and Simar (2007) offer an informative work on how we define the efficient subsets of S and how we define the efficiency measure of a DMU from the frontier using radial distances from it. In the next section, a graphical representation of the production technology is presented, offering a visual approximation of efficiency measurement.

3.3 The Theory of Measuring Efficiency–Farrell's Approach

In production theory, a firm's input and output combinations are depicted using a production function. This function can show the maximum output that can be achieved with any possible combination of inputs, so the construction of a production technology frontier is feasible. Farell (1957)[42] extended the work initiated by Koopmans (1951) and Debreu (1951) by noting that production efficiency has a second component reflecting the ability of producers to select the 'right' technically efficient input-output vector in light of prevailing input and output prices (Daraio and Simar, 2007). This led Farrell to define overall productive efficiency as the product of technical and allocative efficiency. Some decades later, in a more formal formulation, Lovell (1993) defined the efficiency of a production unit in terms of a comparison between observed and optimal values of its output and input. The comparison can take the form of the ratio of observed to maximum potential output obtainable from the given input, or the ratio of minimum potential to observed input required to produce the given output. In these two comparisons, the optimum is defined in terms of production possibilities, and efficiency is technical. Since the first empirical application of Farrell (1957), several different

[42] Forsund and Sarafoglou (2002) offer an interesting historical reconstruction of the literature developments subsequent to Farrell's seminal paper that led to the introduction of the DEA methodology.

methods for efficient frontier estimation and efficiency score calculation have been developed in the literature (Coelli et al., 1998; Thanassoulis, 2001).

Building on the ideas of Dantzig (1951) and Farrell (1957), the seminal work 'Measuring the efficiency of decision making units' by Charnes, Cooper and Rhodes (1978) applies linear programming to estimate an empirical production technology surface; in other words, a piece-wise frontier over the data. This was the first application of DEA and was attributed to the authors due to their important contribution to stating the principals of DEA. Formally, DEA is a methodology directed to frontiers rather than central tendencies. Therefore, rather than fitting a regression plane through the centre of the data, as in traditional regression, with DEA, a piecewise linear surface to rest on top of the observations is drifted (Cooper et al., 2011).

The first attempt to introduce a frontier production model was made by Farell (1957), as mentioned earlier. The original framework has been built upon measuring (EE) or production efficiency that can be further decomposed into (TE) and (AE). The combination of these two components determines the level of total economic efficiency (TEE). Farrell (1957) extended the Pareto-Koopmans property by using the performance of other DMUs to evaluate the behaviour of each DMU relative to the outputs and the inputs they all used. This made it possible to proceed empirically to determine their relative efficiencies. The resulting measure that is referred to as the 'Farrell measure of efficiency' was regarded by Farrell as restricted to meaning 'technical efficiency' or the amount of 'waste' that can be eliminated without worsening any input or output (Cooper et al., 2011). The above notation has been discerned by Farrell from 'allocative' and 'scale' efficiencies, as exist nowadays in economics literature.

This idea emerged initially from the use of a firm's framework with two inputs (x_1, x_2) to produce a single output (q) as the outcome of the production process given the production technology.[43] CRS is the underlying assumption, so it was feasible for the technology to be represented by the unit isoquant. Beyond the different representations of efficiency, there is the same underlying objective to quantify the relative performance of a unit[44] or to quantify the unit's progress towards meeting policy objectives.

The efficiency framework allows specialists to identify low-, middle-, or high-performing units whose processes might be potentially adapted by others and enable regulators to develop targets and incentives effectively (Mugisha et al., 2007). Performance evaluation by means of efficiency assessment is a metric approach that allows quantitative measurement of relative performance (*CE*, *TE*, *SE*, *AE*, and efficiency change). From the preceding discussion, asymmetries in performance emerge in the units of interest and rankings can be based on the analysis of production patterns and cost structures (Berg and Padowski, 2010). As a result, a review of the several

[43] In the appendix 3 chapter 3, a graphical presentation of the input-output-orientated measures is available, following Coelli (1996a).

[44] Controlling for external conditions.

methodologies and types of metric used by analysts to estimate comparative performance through efficiency evaluation is attempted in the next section.

3.4 Methods for Estimating Efficiency-A Taxonomy of Frontier Models

The following section cites some of the principal techniques in the literature for estimating efficiency level. Therefore, the aim of this section is to propose a general taxonomy of efficient frontier models that gives an overview of the different approaches presented in the literature for estimating the efficient frontier of a PPS and to explore the methodological advances applied in the English HE system. Analysis of the existent literature is a necessary step for the advancement of a discipline. This is particularly evident in the field of efficiency and productivity research that, in the last decades, has experienced an exponential increase in the number of methodological and applied works (Tavares, 2003; Daraio and Simar, 2007).

The first distinction in a methodological grounding for efficiency estimation approaches is between the statistical (or econometric) approach and the non-statistical (or programming) approach. The distinction between the two approaches derives from the underlying assumptions. The non-statistical approach makes no assumptions regarding the distribution of inefficiencies. In addition, it is often (but not always) non-parametric, which means that the input and output data are used to compute a convex hull to represent the efficiency frontier (Sengupta, 1999) using linear programming methods. The non-parametric frontier approach, based on envelopment techniques (DEA, FDH[45]) has been used extensively for estimating the efficiency of DMUs as it relies on very few assumptions for the production possibility set (PPS). The non-parametric approach relies on linear programming or some other form of mathematical programming to characterise the set of efficient producers and then derive estimates of efficiency for inefficient observations based on how far they deviate from the most efficient ones, rather than estimating values for selected parameters. Another competitive superiority of the non-statistical, non-parametric approach is the lack of misspecification problems,[46] since neither distributions are specified, nor is there a particular functional form[47] for the frontier function (Johnes, 2004).

Furthermore, programming methods can easily be used in a production situation in which multiple inputs and multiple outputs are handled and ensure robustness in model choice. For a comprehensive DEA bibliography covering 1978–1992, see Seiford (1994, 1996), and for an extension until 2001, see Gattoufi et al. (2004). More than 1,500 DEA references are reported by Cooper et al. (2000), despite the highlighted disadvantages of the non-statistical, non-parametric approaches since they are barren of

[45] The Free Disposal Technique introduced by Deprins et al. (1984) relies only on the free disposability assumption of the PPS and does not restrict itself to convex technologies. The FDH estimator, proposed by Deprins, et al. (1984), is a more general version of the DEA estimator.
[46] Both in the production function and the distribution of efficiencies.
[47] See appendix 4 chapter 3 for a useful insight into different functional forms.

estimates or significance tests of parameters (Geva-May, 2001). Another limitation is that the convex hull is defined using information on only a small number of observations in the sample. Further shortcomings shared by many non-parametric methods concern the curse of dimensionality. This is to avoid large variances and wide confidence interval estimates; therefore, the analysis becomes 'hungry for data', which means that a large amount of data is needed (Daraio and Simar, 2007).

The statistical approach is often (but not always) parametric, which means that a specific functional form[48] for the production frontier function $g(x,\beta)$ is assumed (Sengupta, 1999). Therefore, statistical, parametric methods use a simple mathematical form depending on some k unknown parameters, since $\beta \in \mathcal{R}^k$ represents the production technology set S. Hence, it provides estimates on the parameters of the frontier, the significance of which can be tested using standard errors (Schmidt, 1985–6). The main methodological advances of this approach are the economic interpretation of parameters and the statistical properties of estimators; more critical are the choice of the function $g(x,\beta)$ and the handling of multiple input and multiple output cases (Daraio and Simar, 2007).

A further classification between alternative methods is based on the criterion of noise presence. Hence, the distinction between deterministic and stochastic models is attributed to whether deviations from the production function are a consequence not only of inefficiency. In terms of the deterministic approach, deviation in observed output from the production frontier is solely a consequence of inefficiency (Lovell, 1993; Ondrich and Ruggiero, 2001) since it assumes that all observations (x_i, q_i) belong to the production set, so:

$$Prob\{(x_i, q_i) \in \mathrm{S}\} = 1 \; \forall \; i = 1 \dots \dots \dots n$$

The main limitation of this approach is that any errors in measurement or stochastic errors are incorporated into the measurement of efficiency and, therefore, the event of sensitivity to 'super-efficient' outliers[49] is standard. With regard to stochastic models, there might be noise in the data, i.e. some observations might lie outside S. The main weakness on this ground is the identification of noise from inefficiency. According to the stochastic approach, deviations from the production function have a bilateral explanation since they are not attributed solely to inefficiency. The objective of stochastic models is, therefore, to decompose the residual into two components: one stemming from inefficiency and one random. In practice, this implies an assumption of a specific distribution for each error component, which constitutes an important limitation. Stochastic methods are preferable on events of random shocks or measurement errors, giving them the comparative advantage of curtailing any distortions on the efficiency estimates; however, they may be affected by misspecification errors.

[48] Various specifications can be used: e.g. the CBD, the flexible fixed quadratic function, the hybrid translog function, the CES function, etc.
[49] Robust estimators are able to overcome this drawback.

In particular, SFA allows the presence of noise, but it demands parametric restrictions on the shape of the frontier and on the data-generating process (DGP) in order to permit the identification of noise from inefficiency and the estimation of the frontier. The statistical approach assumes that inefficiencies (the difference between the firm's observed output and the output that could be achieved if it was producing on the production frontier) follow a specific distribution (Førsund et al., 1980). However, any misspecification errors (either of the production function or of the inefficiency distribution) are incorporated in the efficiency measure (Lovell, 1993). Furthermore, the statistical, parametric approach is not easily applied in a situation in which there are multiple inputs and multiple outputs (Johnes, 2004).

Both techniques have strengths and limitations, so an extensive review and updated presentation of both approaches is considered appropriate (Fried et al., 2006). In the next section, an overview of the DEA and SFA methodologies is discussed. A statistical approach that unifies the parametric and non-parametric approaches can be also found in Simar and Wilson (2006b).

From a theoretical perspective, the available methodologies for measuring efficiency vary from statistical to non-statistical, parametric to non-parametric, and deterministic to stochastic.[50] Among the various versions in the literature, two are the most frequently used approaches: statistical parametric methods (deterministic or stochastic) and deterministic non-statistical non-parametric methods (Johnes, 2004). In the next section, the most common methods, including more recent developments in the context of efficiency measurement in HE, are discussed as an introduction to the whole scope of this thesis.

3.5 Mathematical Presentation of DEA

The mathematical programming approach to the construction of frontiers and the measurement of efficiency relative to the constructed frontiers goes by the descriptive title of DEA (Fried et al., 2006). This is a deterministic non-statistical non-parametric method developed by Charnes et al. (1978) following the work of Dantzig (1951) and Farrell (1957), and aims to estimate a production possibility frontier and, hence, to assess the TE of the decision-making unit (DMU), relative to the frontier. Charnes and Cooper (1961) made considerable theoretical and applied contributions in the development of linear programming, and disseminated its application in DEA in the late 1970s.[51] The DEA estimator relies on the convexity assumption, so the data points are enveloped with linear segments. The programming approach reveals the structure of frontier technology without imposing a specific functional form on either technology or deviations from it. Subject to certain assumptions about the structure of production technology, it envelops the data as tightly as possible; however, it makes no

[50] For non-parametric stochastic models:

1. For cross-sectional data see: Hall and Simar (2002), Simar (2003), and Kumbhakar et al. (2007).
2. For panel data see: Kneip and Simar (1996) and Henderson and Simar (2005).

[51] See Charnes et al. (1978) for an analytical review.

accommodation for noise, and so does not ‘nearly’ envelop a data set in the same way as the deterministic kernel of a stochastic frontier.

In Germany, the procedure was used mainly to estimate the marginal productivity of R&D and other factors of production (Brockhoff, 1970). The main feature of DEA is to compare efficiency levels across DMUs within an organisation; DEA has also been used to compare efficiency across firms. DEA comprises two basic models, with the most basic being the CCR model introduced by Charnes, Cooper and Rhodes (1978) with a constant return to scale assumption, and the DEA Banker-Charnes-Cooper model, with a variable return to scale assumption (Banker et al., 1984). The main developments of DEA in the 1970s and 1980s are documented by Fare et al. (1994), Seiford and Thrall (1990), Lovell (1994), Charnes et al. (1995), Seifford (1996), Cooper et al. (2000), and Thanassoulis (2001). Throughout the years a large number of books and journal articles have been written on DEA or applying DEA to various sets of problems.

DEA is a non-statistical and non-parametric approach that makes no assumptions regarding the distribution of inefficiencies or the functional form of the production (or distance) function (Fare et al., 1994). The only imposed restrictions that can be traced in DEA are technical restrictions, such as monotonicity, homogeneity, and convexity. The lack of assumptions in DEA regarding statistical distributions, however, means that there are no estimates or significance tests of the parameters of the function, which might be assumed to be a potentially serious limitation if results are sensitive to the specification of inputs and outputs.

In the DEA context, any deviations from the production function are deterministic, stemming solely from inefficiency. This can be a serious deficiency in a context in which stochastic errors, measurement errors, and random shocks are common. In addition, it cannot provide parameter estimates from which information on, for example, elasticities can be derived. Finally, an additional point of concern is that estimates of efficiency can be distorted by outliers or by the choice of inputs and outputs (Johnes, 2012). Considering the wide diversity in English HE, the existence of outliers in the DEA model may distort the efficiency estimates.

Using the statistical approach, the efficient frontier models using programming are classified according to three main criteria: specification (or not) of the form of the frontier; presence of noise in the estimation procedure; and the type of data analysed (cross-section or panel data). Another parameter to deal with are the types of variable available (quantities only, or quantities and prices). Since in cases in which only quantities are available, only TE can be estimated, while with quantities and prices, EE can be estimated and decomposed into its technical and allocative components. DEA was developed in a public sector, not-for-profit environment, in which prices are suspect at best and missing at worst; ergo, the vast majority of DEA studies use quantity data only (Fried et al., 2006).

The efficiency of each unit is measured as the ratio of weighted output to weighted input, where the weights used are calculated by the technique itself, and not defined *a priori*. These weights are such to reflect the unit at its most efficient level relative to all other units in the dataset. In a multi-output, multi-input space,[52] DEA provides estimates of the distance function (Shephard, 1970),[53] which is a generalisation of the single output production function. The distance function approach has two chiefly advantageous elements, in the sense that there is no need to set any behavioural assumptions about the firms, such as cost minimisation or profit maximisation, which might be especially regarded as irrelevant in the HE context; secondly, *a priori* information on the input and output prices is not available in the HE context (Johnes, 2006).

In its simplest form, DEA assumes CRS. Charnes et al. (1978) proposed a model that had an input orientation and assumed CRS. Subsequent papers have considered alternative sets of assumptions in which variable returns of scale (VRS) models have been presented (Fare et al., 1983; Banker et al., 1984). Proceeding to an overview of the CCR model, Charnes et al. (1978) assumed (N) inputs and (M) outputs for each of (I) firms. Under CRS scale conditions, the DEA-derived input-and-output-oriented measures of efficiency of a DMU are identical. An intuitive way to introduce efficiency of DMU k in a DEA framework is via the ratio form. Consequently, for each firm, a measure of the ratio of the weighted sum of all outputs over the weighted sum of all inputs is derived as:

$$E_k \frac{\sum_{r=1}^{M} u_r q_{rk}}{\sum_{i=1}^{N} v_i x_{ik}}$$

Where (u_r) is a $M \times 1$ vector of output weights applied to output r and (v_i) is a $N \times 1$ vector of input weights applied to input i; q_{rk} is the amount of output r used by DMU k; x_{ik} is the amount of input i used by DMU k. This measure of efficiency is relative to all other DMUs in the dataset.

The optimal weights when DMU k maximises its efficiency score can be derived from solving the mathematical programming problem subject to certain constrains. Therefore, for each of the I DMUs in the dataset, the following linear programming problem must be solved:

$$max \frac{\sum_{r=1}^{M} u_r q_{rk}}{\sum_{i=1}^{N} v_i x_{ik}}$$

Subject to:

i. $\frac{\sum_{r=1}^{M} u_r q_{rj}}{\sum_{i=1}^{N} v_i x_{ij}} \leq 1$ where $j = 1, \dots I$ the weights are universal (universality constraint) i.e. the weights used by DMU k when applied to each DMU in the dataset cannot produce an efficiency score exceeding unity.

ii. u_r and $v_i > 0 \; \forall \; r = 1, \dots M; i = 1, \dots N$ weights on the outputs and weights on the input are strictly positive.

[52] This is the case of the HE production possibility set.

[53] An analytical review of the distance function framework is offered in appendix 5 chapter 3.

iii. $v_i x_{ij} = 1$. Due to the infinite number of solutions, the multiplier form is broadly used.

DEA models have been developed that measure efficiency in different ways. These fall largely into the categories of being either input-or-output-oriented models. In practice, when we make computations, the dual equations are used. The dual version of the linear programming problem is always more tractable and, since economic data are most frequently in price and monetary terms, cost functions are generally more accessible than corresponding empirical investigations of production functions (Shephard, 1953). The advantage of the dual over the number of the imposed constraints is straightforward since only $M + N$ constraints are used instead of $I + 1$. For the mathematical representation of the models (both input-output-oriented specifications under CRS and VRS are available) in appendix 6 chapter 3.

When all decision-making units are operating at an optimal scale, CRS is the appropriate assumption. The efficiency models performing under CRS, in other words, calculate efficiency where feasible for all institutions so as to double their output if all inputs are doubled. This is a rather rigorous assumption; particularly where institutions can vary notably in size, it may be more instructive to relax this condition. Banker et al. (1984) introduced one flexible alternative concerning estimating efficiency under the assumption of VRS due to various restrictions in the market, such as imperfect competition government regulations, constraints on finance, and decision-making units operating on a non-optimal scale. Therefore, from an early stage, research turned to adjusting the CRS DEA model to account for VRS situations. Pioneering work in this specification is attributed to Afriat (1972), Fare et al. (1983), and Banker et al. (1984). The concept underlying VRS is that any existing scale effects cannot be confounded with TE when DMUs are not operating optimally. Therefore, the VRS specification allows for TE devoid of scale effects (Coelli et al., 2005).

The imposition of a convexity restriction $\sum_{j=1}^{I} \lambda_j = 1$ ensures that an inefficient DMU is 'benchmarked' against DMU with similar size, so the DEA frontier is a convex combination of DMUs. This is contrary to the CRS case, in which a DMU may be 'benchmarked' against firms that are substantially larger or smaller than it, with the λ-weight's sum being even less or greater than one.

Economic theory suggests that, in the long run, competitive firms will continue to adjust their scale size to the point that they operate at CRS. This means that an institution is operating at CRS if a proportion increase of all inputs results in the same proportion expansion to the output. Therefore, deviations from CRS cause scale inefficiency in institutions that do not operate at CRS; subsequently, different forms of scale inefficiency arise. Therefore, if, on the one hand, output increases by less than the proportional change in inputs, there is DRS. On the other hand, if output increases by more than the proportional change in inputs, there is IRS. The envelopment model under VRS can be derived from that under CRS simply by adding a constraint to ensure that the sum of lambdas in the technical input (output) efficiency model adds up to 1; this is known as the convexity constraint. For an analytical presentation of the VRS DEA model see appendix 7 chapter 3.

While choice of orientation does not affect efficiencies under CRS (the two measures provide the same value under CRS), it does under the assumption of VRS (unequal when VRS is assumed) (Coelli, et al., 1998). Given that linear programming does not suffer from statistical problems, such as simultaneous-equations bias, misspecification of the model, etc., the choice of an appropriate orientation is not as crucial as it is in the case of econometric estimation Coelli et al. (1999). It has been shown only to have a minor influence upon the scores obtained Coelli and Perelman (1999). Although there are studies (Glass et al., 2006; Flegg and Allen, 2007) that experimented with both methodological orientations in previous analysis, Mc Millan and Datta (1998) claim that selection of orientation did not affect the results significantly. In practice, whether the input- or output-oriented measure is more appropriate correlates with whether input conservation is more important than output augmentation.

In some cases, DMUs may be given a fixed quantity of resources and asked to produce as much output as possible. In this case, an output orientation would be more appropriate; hence, many studies[54] have conformed to an output-orientation framework (Beasley, 1990; Glass et al., 1997, 2007; Tomkins and Green, 1998; Dyson, 2000; Sarrico and Dyson, 2004; Flegg et al., 2004, 2007; Johnes, 2006, 2008, 2010, 2012; Flegg and Allen, 2009; Bradley et al., 2010). Essentially, one should select the orientation according to which quantities (inputs or outputs) the managers have most control over (Coelli et al., 2005).

In a number of studies in the literature (Doyle and Green, 1994a; Beasley, 1995; Johnes, 1995, 1999; Turner, 2005; Emrouznejad and Thanassoulis, 2005), input-oriented models tend to be selected since many DMUs have particular orders to fill, and, hence, the input quantities appear to be the primary decision variables. However, in some other sectors, the bundle of inputs cannot be adjusted or controlled, so the objective here is the maximum expansion of the output given a set of inputs.

Although this argument of input-oriented approaches may not be as strong in all industries, in HE, the output orientation is deemed more appropriate, since the quantity of the inputs, such as number of undergraduate and postgraduate entrants, expenditures in academic and central services, as well as the number of employees, can be considered fixed, at least in the short run, and can be readjusted in the long run. The institutional outcomes are the joint product of entering student characteristics, resource inputs, and institutional processes. Consequently, institutional performance is the product of the joint inputs, stemming from individuals, institutions, and local governments and, more importantly, many of these inputs are beyond the control of the institution (Bailey and Xu, 2011).

DEA has been developed with the hope of addressing and remedying some of its shortcomings. Thus, tests of significance have been suggested, and bootstrapping

[54] The literature here covers studies for UK HE.

methods have been used to assess the validity of the efficiency estimates. A summary of the possible extensions of DEA in the context of the HE sector is attempted by Johnes (2006). In this study, by using data from English universities, the aim is to present the assessment derived from three alternative methods for:

- Assessing the relevance of input(s) and/or output(s) included in a DE A (Pastor et al., 2002). This test provides further insight into whether changes in the values of the efficiency scores are significant. Spearman's rank correlation coefficient can be used in addition to Pastor et al.'s (2002) test to provide complementary information.
- Testing for significant differences in the efficiency distributions of different subgroups of DMUs (Charnes et al., 1981).
- Deriving confidence intervals for the efficiency scores of individual DMUs, by applying bootstrapping procedures (Simar and Wilson, 1998, 1999).

In line with previous findings, the efficiency levels of HE in England maintain a prominent position. No significant differences between HEI types in terms of efficiency, with which inputs are transformed into outputs, have been derived from the study. However, bootstrapping applications enhance the perspective of important differences in efficiency between the worst- and best-performing English HEIs.

3.6 Limitations on Traditional Non-parametric Envelopment Estimators-Advanced Methods

At the forefront of the non-parametric deterministic frontier models are the DEA and FDH approaches. However as discussed in short earlier there are a number of challenges and limitations when those models are in practical use despite their promising theoretical properties. FDH and DEA estimators are identified as sensitive to "super-efficient" outliers or extreme observations. These extreme values may disproportionately (depending on the location of the outlier(s)), and perhaps misleadingly, influence the evaluation of the performance of other DMUs since the efficiency estimates are distorted for one or more DMUs to an arbitrarily large degree. Thus, this sensitivity is determinable due to the fact that the efficient frontier is determined by sample observations which are extreme points (Simar and Wilson, 2011).

When deterministic technologies are applied, robustness at the presence of outliers can be improved by not enveloping the most extreme observations. Simar, (1996) points out the need for identifying and eliminating outliers when deterministic models are under use. So the first, and most conceivable, approach to tackle outliers is to identify them in the data, and then perhaps delete them if they result from corrupted data. To this end, a number of techniques exist in the literature to detect outliers in frontier settings (Wilson, 1993, 1995; Simar, 2003; Porembski et al., 2005). Although, in cases when outliers are not identifiable, the use of stochastic frontier models is recommended. However, Wheelock and Wilson (2008), Wilson (2011), Simar and Vanhems (2012), and Simar et al. (2012) have developed robust alternatives to the traditional FDH and DEA estimators. Those robust new estimators are able to overcome this drawback.

Hence, this limitation can be addressed by using two classes of partial frontiers, order$-m$ and order$-\alpha$ partial (quantile) frontier estimators[55] and the corresponding efficiency estimators[56] (Aragon et al., 2005; Cazals et al., 2002; Daouia and Simar 2007). Order$-m$ frontiers, where m can be viewed as a trimming parameter, and order$-\alpha$ quantile frontiers, analogous to traditional quantile functions but adapted to the frontier problem are considered to use a "partial" and "less extreme" frontier, while the traditional idea of a "full" frontier envelops all the data (Simar and Wilson, 2011). So the advantage of not enveloping all the data in finite samples makes the new estimators much more robust with respect to outliers and extreme points rather than the classical FDH or DEA estimators.

In an order$-m$ frontier according to Cazals et al. (2002) a unit (x, q) is benchmarked against the average maximal output reached by m peers randomly drawn from the population of units using less input than x. So asymptotically as $m \to \infty$ the order$-m$ frontier converges to the full frontier. In the same sense, in an order$-\alpha$ quantile a unit (x, q) is benchmarked against the output level not exceeded by $100(1-a)\%$ of firms in the population of units using less input than x. So as $a \to 1$, order$-\alpha$ frontier converges to the full frontier (Aragon, et al., 2005; Daouia and Simar, 2007).

Another angle which is mitigated with conditional order$-m$ and order$-\alpha$ frontiers is the curse of dimensionality, which causes convergence rates to be slower than the root$-n$ rate typically obtained with parametric estimators. So efficiency estimators while fully nonparametric, achieve root$-n$ rates of convergence and have Gaussian limiting distributions (Simar and Wilson, 2011). Also, many statistical limitations inherent in estimating a full frontier are overcome with partial frontiers since they construct a useful benchmark against comparable only DMUs and concurrently are consistent estimators of the full frontier[57]. Cazals et al. (2002), Aragon et al. (2002) and Daouia and Simar (2005) proposed all those robust alternatives of the traditional DEA and FDH models so as to handle extreme values and/or outliers since they do not envelop all the data. However, they still lack the ability to capture efficiently noise presence since they still rely heavily on the deterministic assumption.

A second inexpediency in DEA and FDH models is the absence of statistical noise due to their deterministic nature. As pointed out in Simar and Wilson (2007), introducing

---

[55] The analytical presentation of the estimators is beyond the scope of this thesis. However the analytical framework and the theoretical considerations of the estimators are well described in the literature in:
Order$-m$ frontiers: Cazals et al. (2002), Simar (2003), Daouia, et al. (2009).
Order$-\alpha$ frontiers: Aragon, et al. (2005), Daouia and Simar (2005, 2007), Daouia et al. (2009, 2010).
Both estimators: Simar and Wilson (2011).

[56] Nonparametric estimators of these partial frontiers tend to be very easy and fast to compute (Simar and Wilson, 2011).
For the analytical properties of these estimators see Daouia and Gijbels (2011) and Daouia and Ruiz-Gazen (2006).

[57] Those ''partial-order'' frontiers and their estimator allow the order of the frontier (m or α) to grow at an appropriate rate along with the sample size see: Daouia et al. (2010) and Daouia et al. (2012).

noise in DEA/FDH framework is still a challenge and an open issue of research. In the literature, chance-constrained programming to the DEA problem (Land et al., 1993; Olesen and Petersen, 1995) and fuzzy programming approaches to DEA and FDH efficiency measurement (Sengupta, 1992; Seaver and Triantis, 1992; Girod and Triantis, 1999; Triantis and Girod, 1998; Kao and Liu, 1999; Triantis and Vanden Eeckaut, 2000) have been proposed so as to make DEA stochastic.

However, all these alternatives suffer from various side incapacities such as requirements for vastly large amount of data, strong distributional assumptions when constrains are violated, information on expected values of all variables for all DMUs, variance-covariance matrices for each variable across all DMUs and inaccuracy in the data due to mismatch between the measurement system and the nature of the data needed in production studies (Daraio and Simar, 2007). So without some restrictions on the classical DEA and FDH models, a stochastic nonparametric model cannot identified. Hall and Simar, (2002) and Simar (2003, 2007)[58] explored and developed stochastic versions of DEA/FDH estimators when the noise is of moderate size. They describe a DGP that in a full multivariate setup build a new nonparametric stochastic frontier estimator.

With a more general setup, with no restrictions on the size of the noise Kumbhakar, et al. (2007) proposed a different setup using local MLE for estimating the production frontier without a parametric assumption, but still using semi-parametric assumptions about the stochastic and the inefficiency terms. So the method is based on the local ML principles which are nonparametric in the sense that the parameters of a given local polynomial model (linear or quadratic) are localized with respect to the covariates of the model[59.] Their estimator (order$-m$ local polynomial estimator) is obtained through a one-step maximization procedure. Both u and v are independently conditionally on X[60]. An extension of the local ML estimation theory to the truncated case is available by Park et al. (2008) who provide asymptotic results for the derivatives of the regression function, and treat the curse of dimensionality problem by using an unknown constant as a shape parameter of the error distribution which achieves root$-n$ consistency[61].

Simar and Zelenyuk, (2011) improved the Kumbhakar et al. (2007) model by making the resulting frontier smoother, monotonic and concave when needed. The authors so far have utilized cross-sectional data, so in the case of a panel of data, much more information is available and the identification problem can be handled more easily. Kneip and Simar, (1996) and Henderson and Simar, (2005) have widely used panel data in this particular non-parametric setup. However, in practice, a large number of time periods is needed for getting sensible results. New directions in this area have been proposed by Simar et al. (2017) who attempted a much easier, faster and more robust

[58] He expanded the stochastic non-parametric framework in a multivariate setup that is also resistant to outliers and extreme values.
[59] Gozalo and Linton (2000) point out that localizing can be viewed as a way of non-parametrically encompassing a parametric ‘‘anchorage’’ model.
[60] Follow Kumbhakar et al. (2007) for the analytical presentation of the estimator on simulated data sets and the asymptotic properties.
[61] For further developments and most flexible semi-parametric models see: Kneip et al. (2012).

estimation with less assumptions than the local MLE approach proposed earlier. The novelty of this approach which can be viewed as a non- or semi-parametric version of the ‘‘modified OLS’’ (MOLS) method is that the assumptions made here will provide estimators of the frontier that are more robust than the one obtained with the local MLE approach, since in this simpler version only local moment restrictions on u and v are used and not their full local distributions.

A retrospective limitation of non-parametric models is the lack of economic interpretation in terms of the shape of the production process, returns to scale, elasticities etc. Some alternatives to cover that negligence have been proposed in the literature, either with the use of slacks proposed by Fare et al. (1994) or by using full theory for parametric approximations of non-parametric frontiers the so called semi-parametric approaches (Fan, et al., 1996; Huang and Fu, 1999; Kuosmanen and Kortelainen, 2012) which tend to lie in between as a useful compromise. Those approaches tend to retain the essential framework of the stochastic frontier, but relax the assumption of a specific distribution for v or u or both.

Nonparametric estimation of the stochastic production frontier was introduced by Banker and Maindiratta (1992)[62] and Fan et al. (1996). While Fan et al. (1996) is commonly thought of as pioneer to lessen parametric assumptions in the SF model Banker and Maindiratta (1992) have a similar approach to DEA enriched with both noise and inefficiency in a ML framework while Fan et al. (1996) used standard kernel methods coupled with MLE. So in its simpler version in a semi-parametric framework the production function may be left unspecified, but a parametric density for the inefficiency term and an independent Gaussian process for the noise are still specified.

The work of Banker and Maindiratta (1992) and Fan et al. (1996) has been extended in various dimensions so a number of estimators have been proposed in the literature that build upon their work. So later studies that explored a variety of distributional assumptions in panel data routines are those of Park and Simar (1994), Park et al. (1998), and Park, Sickles and Simar, (2003; 2007), Adams et al. (1999), Sickles et al. (2002), Sickles (2005). Recently, as an alternative, Martins-Filho and Yao (2015) proposed an estimator which jointly estimates the distributional parameters and the unknown frontier. This is an approach that relies on local likelihood estimation. Parmeter and Racine (2013) propose imposing monotonicity and convexity constraints within the confines of the Fan et al. (1996) estimator and, substitute any distributional assumptions with a variant which simply deploys either COLS or MOLS.

Despite the promising virtues of the semi-parametric approach, it still remains restrictive in the sense that both the homoscedasticity assumption (for both the inefficiency and noise processes) and the parametric density for the inefficiency term seems problematic and open to criticism since the analysis is prone to parametric misspecification, and statistical inconsistency Greene (2008). Albeit the great strides

[62] The critiques on Banker and Maindiratta’s (1992) model was, the inability to reliably implement their estimator by that time as well as the non-smoothness of the resultant estimator (e.g. returns to scale calculation) (Parmeter and Kumbhakar, 2014).

that have been made in nonparametric and semi-parametric approaches to efficiency estimation in recent years, a number of core issues remain open to debate and stimulate the focus of current, ongoing research.

All these years, bridging the gap between axiomatic DEA and stochastic SFA was one of the most challenging processes in the efficiency analysis field. The original framework of the DEA and SFA applications has been expanded and conflated over the past decades in a unified framework. The full integration of DEA and SFA into a unified framework of productivity analysis comes under the name Stochastic Nonparametric Envelopment of Data (StoNED) methods[63]. Those are state of the art presentations of the frontier analysis that combine the existing tools of efficiency analysis in a unified framework across the DEA-SFA spectrum, facilitating a new era for further methodological development.

Recent works on convex nonparametric least squares (CNLS) embed DEA in a standard regression setting by Kuosmanen (2008), Kuosmanen and Johnson (2010), and Kuosmanen and Kortelainen (2012). This is a non-smooth approach which has ties back to Banker and Maindiratta (1992), and has been developed to the most promising new tools for axiomatic non-parametric frontier estimation and efficiency analysis, when stochastic noise is present[64]. Kuosmanen and Kortelainen (2012) who built upon the former work of Banker and Maindiratta (1992), apply a similar piecewise linear framework based on monotonicity and concavity assumptions but relied on minimizing a sum of squared errors criterion instead of maximizing a likelihood function. The distributional parameters are either recovered using MOLS or with a similar approach as in Fan et al. (1996).

The development of StoNED models enables researchers to model noise presence while axioms of production theory are imposed. So beyond the technical innovation in the efficiency analysis field StoNED offers deeper insights into the economic intuition and foundations of DEA and SFA. Therefore according to Kuosmanen, et al. (2015) StoNED renders a more general and flexible platform for efficiency analysis and related themes such as frontier estimation and production analysis. Of course, those models are open to better improvements to incorporate several dimensions to the more general case of input and output multiplicity and heteroscedasticity inclusion.

Another issue commonly discussed in efficiency analysis is the heterogeneity issue which ties quite well how environmental/external factors can explain (in) efficiency. Therefore efficiency differences may be related to differences in ownership type or structure, regulatory constraints, business environment, competition and so on among the DMUs under analysis (Simar and Wilson, 2013). For many years, researchers developed several approaches to incorporate environmental variables affecting the production process. Conditions described by Z may have a twofold effect on the

[63] For a more detailed discussion about the theoretical properties and extensions of CNLS and StoNED models, we refer the reader to Kuosmanen et al. (2015). Also see Johnson and Kuosmanen (2015) for detailed examples of computational codes and further developments on the mathematical modeling environment.

[64] No need for *a priori* distributional assumptions on the error term to estimate the production frontier.

production process or might be completely independent of (x, q), since the effect of Z is unknown and must be estimated appropriately. However, in most of the cases may affect the shape of the boundary of the attainable set since Z include factors sufficient to alter the range of attainable values for the inputs and outputs or may affect the distribution of inefficiencies inside the attainable set. Of course, the case of affecting both cannot be *a priori* excluded (Simar and Wilson, 2015). Those factors are considered as not being under the control of firms' managers but nonetheless, they may influence the production process, hence, from a public policy perspective a cost efficient regulatory framework is of critical importance if DMUs' performance need improvements.

In practice it is often not possible to include several environmental factors and for those included a concrete decision is needed on whether only either continuous or categorical exogenous variables will be used. When discrete categorical variables are considered, DMUs can be divided into different groups and then test differences in average efficiency, across or within groups. Among the testable parameters here is whether the various groups adopt the same technology, or whether the distributions of efficiencies are the same across groups if they do share the same technology (Simar and Wilson, 2013). When continuous environmental variables are considered, either two-stage approaches have been widely used in the literature or conditional efficiency measures as a novel method based on estimating partial frontiers.

Two are the main methodological tools for tracking the effects of continuous environmental variables on efficiency[65]. According to the first, a two-stage approach in which estimated efficiencies are regressed on environmental variables has been widely used in the literature. Those two-stage methods (Simar and Wilson, 2007; 2011), use traditional nonparametric or robust nonparametric estimation techniques[66] (DEA or FDH) for defining the efficient frontier (frontier of best-practice against which efficiency of a DMU can be measured). Then in a later step the analysis introduces potential exogenous variables (Z) that are beyond the producer's control but still affect the production process. They are handled neither as inputs, nor as outputs, by fitting a standard regression model, mainly an appropriate parametric model[67] (censored tobit regression, truncated regression, logistic, etc.) on the obtained nonparametric frontier.

[65] There is an additional approach: One-stage approach introduced by Banker and Morey (1986), Fare et al. (1989) and later discussed by Coelli et al., (1998) where the basic statistical model is augmented with environmental factors (Z variables) as free disposal inputs or outputs (if Z is a vector of variables, some elements of Z might be treated as inputs, while others might be treated as outputs) that contribute to defining the attainable set of production. The FDH and DEA estimators of the PPS are defined as long as the variables Z are embodied. However there are hindered considerations when a-priori decision is made about the role, direction and effect of Z as well as whether assumptions such as free disposability, convexity and RTS imposed in DEA and FDH are still valid since for many environmental variables, it is difficult to find arguments for why these assumptions might be appropriate, see: Simar and Wilson (2013).

[66] Other techniques used in this first stage include a leave-one-out estimator (LOO estimator) of efficiency originally suggested by Andersen and Petersen (1993).

[67] Various transformations on the bounded estimates of the first stage efficiency using log, logistic, or log-normal transformations are possible, and in some cases adding or subtracting an arbitrary constant to avoid division by zero or taking the log of zero (Simar and Wilson, 2013).

These two-stage techniques have been generalized also in two-stage semi-parametric approaches in which a part of the model is parametric and the rest is nonparametric. The first stage is somewhat preparatory for the second and it is used to filter the data by eliminating inefficient DMUs, and then a desired parametric model is estimated (Simar, 1992; Florens and Simar, 2005; Daouia et al., 2008). Specifically, Florens and Simar (2005) utilized the desired properties of FDH and order-m estimators in the first stage while Daouia et al. (2008) used order-α estimators to filter the data. Then in a second stage analysis, a parametric model is fitted by least squares using the projected, "efficient" DMUs. In that sense, they implicitly extract more information for the shape of the frontier by utilizing an artificial sample of k efficient DMUs than the original sample of K DMUs where inefficient DMUSs are present as well (Simar and Wilson, 2013).

However, inference is dubious or invalid in most of the two-stage generalizations since they lack of a proper definition of the DGP that would make a second-stage regression sensible (Simar and Wilson, 2007). The main limitations of the two-stage approaches have been well-summarized by Fried et al. (2008). The most restricted among the prerequisite assumptions is the separability condition between the production space for inputs and outputs, PPS, and the space of the environmental variables. In other words, the operational environment should not influence the attainable input–output set hence, the variables Z lie in a space apart from PPS without affecting the shape or boundary of PPS. However in most of the applied work the Z variables do not only influence on efficiency, but also on the PPS. Therefore, there is often no uncompounded separability between inputs, outputs and environmental variables. Also, another common drawback shared both in one-or-two-stage approaches is the *a-priori* decision of the researcher on whether to model the environmental variable(s) as an input or as an output. Further limitations constrain the effect of the environmental variable(s) to be monotone in the production process.

Last but not least, it is notable that in all two-stage applications the estimated efficiency of the first stage $\widehat{\lambda_\kappa}$ is a biased estimator[68] of the λ_κ, hence the ML estimations of the parameters in the second stage may also be biased and perhaps inefficient in a statistical sense. According to Simar and Wilson (2007)[69] by developing a bootstrap approach, by constructing bias-corrected estimates, it yields valid inference in the second stage regression when such regressions are appropriate and the problem might be mitigated.

An approach discharged of any *a priori* assumption on the effect of Z on efficiency as in existing one-stage approaches and absence of a separability condition as imposed in most of the two-stage approaches (Fried et al., 2008) is of principal importance.

[68] Other known pathogenesis of two-stage analysis involves serial correlation among $\widehat{\lambda_\kappa}$ estimates of efficiency in a complicated and unknown way. Another problem is the correlation of the Z with the error term e_κ, since x_k and q_k are correlated with z_k.

[69] Simar and Wilson (2007) point out that a second, parametric bootstrap is appropriate to obtain valid confidence interval estimates for the parameters in the second-stage regression. They held Monte-Carlo experiments and suggest that the double bootstrap performs very well, both in terms of coverage for estimated confidence intervals as well as root mean square error. Follow Simar and Wilson (2007) for the entire procedure.

Therefore, Daraio and Simar (2005) overlaid those limitations by extending the ideas of Cazals et al. (2002) of the probabilistic formulation of the production process suggesting a nonparametric approach, the so-called conditional efficiency[70] model. This is an intuitive way to introduce environmental factors into the production process and account for heterogeneity in performance assessments by defining a conditional efficiency measure. The consistency and the asymptotic properties of different conditional efficiency estimators have been also explored (Cazals et al., 2002; Daraio and Simar 2005, 2006, 2007; 2007a; Jeong et al., 2010; Badin, et al., 2010; 2012), allowing non-parametric estimation of efficiency conditional on some explanatory, contextual, environmental variables that are neither inputs nor outputs in the production process.

Daraio and Simar (2005) developed the method with robust order-m efficiency scores in simulated data, confirming the validity of the method on detecting the true effect of Z on the production process. In a later stage, Daouia and Simar (2007) adapted robust order-α conditional estimators by extending the framework to convex technologies. Also they introduced conditional DEA efficiency scores along the same lines as the conditional FDH scores developed in Simar and Wilson (2013). Note here that any potential shift of the frontier due to the Z effect or any shifts in the inefficiency distributions are likely to be traced through the comparison between the conditional and unconditional measures Badin et al. (2012, 2014).

The asymptotic properties of these non-parametric conditional efficiency estimators are derived by Jeong et al. (2010) while a full bootstrap test detecting the significance of environmental factors on the conditional efficiency scores is available by Daraio and Simar (2014). Also, conditional measures have been extended to unconditional hyperbolic order$-\alpha$ distances towards the frontier as discussed by Wheelock and Wilson (2008) and to order$-m$ radial partial frontiers along with their estimators as explored by Wilson (2011), Simar and Vanhems (2012), Simar, et al. (2012). Those are conditional directional distance functions, conditional to environmental factors.

Badin et al. (2010) eliminate most of the influence of Z on the estimated efficiency $\left(\hat{\lambda}(x,q\,\middle|\,z)\right)$ by using a flexible location-scale nonparametric model and optimal bandwidth selection by data-driven methods. Concurrently, the process allows to rank DMUs facing different operating conditions. Two very flexible location-scale nonparametric models have also been used by Florens et al. (2014) so as to eliminate dependence of Z on inputs X and outputs Q and obtain pure inputs and outputs. This is a novel method to obtain conditional efficiency scores but without explicitly estimate a non-standard conditional distribution. The merits of the method have been well developed using US banking data which shared great diversity in terms of size and services offered in the production process. The conditional efficiency approach is fully nonparametric and flexible enough to detect various possibilities, so it is increasingly used for several different research questions. In the education sector the method has

[70] A full presentation of both conditional and unconditional estimators is available by Simar and Wilson (2013, 2015).

been utilised by Cherchye et al. (2010), De Witte and Rogge (2011) and De Witte and Kortelainen (2013).

Latest developments permit inference about mean efficiency using asymptotic normal approximations Kneip et al. (2015). Since conventional central limit theorem (CLT) results do not hold for sample means of unconditional DEA and FDH estimates they developed new CLTs for means of nonparametric efficiency estimators. By extending those results Kneip et al. (2016) enable for methods testing differences in mean efficiency across groups of producers, as well as further model features like returns to scale (in the absence of environmental variables) or convexity of the production set.

What still remains ambiguous in most of the applied efficiency estimation settings is the separability condition strength. As discussed earlier, it is important to disentangle precisely, how the environmental variables might affect the production process. Conceivably, practitioners have to hypothesize whether environmental variables might affect only the distribution of efficiency among DMUs, the entire PPS of DMUs, or might affect both the distribution of efficiency as well as production possibilities.

In practice we need to test whether the separability condition described by Simar and Wilson (2007) holds. The condition implies that environmental variables only affect the distribution of efficiency and do not affect production possibilities, so as the unconditional DEA and FDH estimators remain meaningful in terms of interpretation. Daraio et al. (2018) develop a test of the restrictive separability condition as described by Simar and Wilson (2007) that enables to test separability empirically. The consistency of the test and its local power has been tested through Monte Carlo experiments, which confirm the validity of the performance of the test, for a variety of sample sizes and dimensionalities.

The presentation of the analytical methodological sketch of those conditional and unconditional efficiency measures is beyond the scope of this thesis, therefore for an insightful and well explored review of the technical parts the interested reader is referred to Simar and Wilson (2013; 2015).

3.7 Statistical Parametric Representation of the Production Possibility Set

3.7.1 Deterministic Frontier

Moving from transformation functions to frontiers, data points could be enveloped using an arbitrarily chosen function (Coelli et al., 2005). Early economists assumed that all producers were efficient (i.e. production happened on the frontier), and perfect competition[71] implies a market free of inefficiency. If this is the case, then estimation processes would be facilitated by using simple regression analysis since the residual would only capture random error (noise). However, the transformation process can and

[71] If the producer was inefficient then it would be impossible to recover its costs and it would be forced out of the market.

does diverge from the ideal hypothesis of perfect competition[72] and so not all producers are able to achieve potential (ideal) output. Potentially, there are many cases in which perfect competition is not a manageable hypothesis, so any divergences should be measured. Therefore, if we want to portray accurately the real world, there is a need to account for inefficiencies. By using a deterministic frontier, all deviations from the frontier are attributed to technical inefficiency since there is no account of measurement errors and other sources of statistical noise.

Under the statistical approach, the production function can be represented by:

$$q_k = f(x_{k1} + \dots + x_{kN})e^{-u_k} \text{ (1)}$$

Where q_k is the output of producer k and x_{ki} is the amount of the i-th input$(i = 1, \dots N)$ used by producer k. The exponential $u_k \geq 0$ and u_k represents the inefficiency factor of producer k (Lovell, 1993), and a specific distribution is assumed for u_k (Førsund et al., 1980). If we take a log-linear version (CD technology) of the equation, this can be written as:

$$\ln(q_k) = ln[f(x_{k1} + \dots + x_{kN}) - u_k] = \beta_0 + \sum_{i=1}^{N} \beta_i ln\, x_{ki} - u_k$$

$u_k \geq 0$ and represents the efficiency of producer k. TE of firm k TE_k is the ratio of the actual (observed) output of producer k to the maximum possible output (ideal) that it could achieve, as represented by the production frontier. Thus, technical efficiency[73] is then measured by the equation:

$$TE_k = \frac{q_k}{f(x_{k1} + \dots + x_{kN})}$$

$$TE_k = \frac{q_k}{f(x_{k1} + \dots + x_{kN})} = e^{-u_k}$$

$$\ln q_k = lnf(x_{k1} + \dots + x_{kN}) + lnTE_k = ln(x_{k1} + \dots + x_{kN}) - u_k$$

[72] Many situations may potentially prevent the competition from being perfect:

i. Imperfect markets (monopolies, oligopolies, market power, or markets with excessive entry barriers)
ii. Information asymmetries (price information is not always available prior to production)
iii. Agency issues and misaligned incentive between owners and executives

[73] Here we use a production frontier framework. If cost is the centre of attention, then, instead of TE, cost efficiency is calculated. This type of efficiency applied to cost functions is a similar notion to technical efficiency. Cost efficiency is the ratio of potential costs to observed costs, i.e.

$$CE_k = \frac{f(q_k; w_k)}{C_k}$$

Where $0 < CE_k \leq 1$. So $lnC_k = lnf(q_k; w_k) - lnCE_k = lnf(q_k; w_k) + u_k$.
Note that u_k reflects cost inefficiency so $u_k = -lnCE_k \approx 1 - CE_k$. $u_k \geq 0$. Larger values denote lower cost efficiency.
Cost inefficiency is the percentage by which the observed costs need to decrease in order for the DMU to attain 100 percent cost efficiency (produce observed output at minimum cost).

Note that u_k represents technical inefficiency, so $u_k = -lnTE_k \approx 1 - TE_k$ and cannot be negative with larger values to denote lower technical efficiency. Technical inefficiency is the percentage by which the observed output needs to grow (increase) in order for the DMU to become 100 percent technically efficient. $TE_k = e^{-u_k} = \exp(-u_k)$ where $u_k \geq 0$. Note as well that $0 < TE_k \leq 1$.

A particular functional form is assumed for the production function in Equation (1). A variety of econometric techniques can be used in the estimation process of inefficiency ($\mathrm{u_k}$), including corrected ordinary least squares (COLS), modified ordinary least squares (MOLS), and maximum likelihood estimation (MLE) (Lovell, 1993). However, some caveats should be considered when applying parametric techniques, regarding possible misspecification of the models used, despite the testable estimates of the parameters of the frontier. Also, these types of method cannot handle a situation with multiple inputs and multiple outputs, which is the case in the HE context. In addition, OLS estimates introduce a deficiency regarding the displacement of the constant term (intercept); therefore, if we want to continue with regression models, we have no alternative but to 'fix' the regression model (Greene, 2008). Two approaches have been suggested in the literature to bridge this gap in the OLS. Both COLS and MOLS are based on the result that the OLS estimator of the slope parameters is consistent and unbiased, so the OLS residuals are pointwise consistent estimators of linear translations of the original $u_k s$.

The first attempt to estimate a Cobb-Douglas (CD) production frontier utilising cross-section data on firms was made by Aigner and Chu (1968).[74] Later, Afriat (1972) assumed that $u_k s$ were gamma distributed random variables and applied MLE for estimation purposes. Hence, the main issue here is how to estimate u_k. By using a simple OLS setting to estimate the parameters, the regression line is shifted up[75] (production) until all residuals are non-positive (ensuring that u_k are non-negative) and at least one is zero, on which we hang the function, so that it envelops all observations and is possible since the slope parameters of OLS are consistent when the residual is non-normal. This approach is referred to here as COLS.

$$\beta_{COLS} = \beta^* + max_k e_k$$

The COLS residuals are $e_{k,COLS} = e_k - max_k e_k$, and technical inefficiency[76] u_j for the j DMU is given by $\max(e_k) - e_j$. The logic of the estimator was first suggested by Winsten (1957), and much later the consistency of the COLS estimator was proved by Gabrielsen (1975) and Greene (1980a). A lengthy application with an extension to panel data[77] appears in Simar (1992). A couple of methodological problems have been identified; however, the method used to be a popular approach in the analysis of panel data (see Cornwell et al. (1990) and Evans et al. (2000a, 2000b)). It should be stressed

[74] They used linear and quadratic programming and the actual task was to minimise the sum of $u_k = lnq_k - x'_k\beta \ s.t \ u_k \geq 0$.
[75] In a cost framework is shifted down.
[76] Cost inefficiency for DMU j (u_j) is given by: $e_j - mine_k$.
[77] For cross-section and panel data, see Schmidt (1976) and Greene (1980).

that no distribution is specified for the residual term, and the entire deviation from the frontier for a particular DMU is attributed to inefficiency (Johnes, 2004).

An alternative to COLS was introduced by Richmond (1974), namely, modified OLS (MOLS) (Lovell, 1993), and instead of shifting the regression line by the maximum or minimum (cost frontier) residual, it shifts based on the model's residual sum of squares. The OLS residuals e^*, of the transformation function, save for the constant displacement, are pointwise consistent estimates of inefficiency, u_k (Greene, 2008). The variance of e^*[78] of the residuals, since the displacement is constant,[79] is a consistent estimator of the variance of inefficiency (u_k). The variance of e^* is known and is given by the model's residual sum of squares. In this way, we can use this information to derive an estimate of $E(u_k)$, if we assume that u_k follows one parameter distribution.[80] This is commonly a half-normal distribution, if we make the assumption that higher inefficiency is less likely than lower inefficiency, although an exponential distribution might alternatively be used (Lovell, 1993). The technical inefficiency[81] of DMU j (u_j) is given by $\hat{E}(u_k)$[82] $- e_j$ where the regression line is shifted by $E(u_k)$. MOLS is less severe than COLS but it requires more restrictive assumptions for the distribution of the residuals that cannot be testable.

Thus, the parameters of the regression are identified by using OLS and an additional parameter, namely, the mean of the inefficiency $E(u_k)$ is also estimated and identified through the variance of the residuals. The estimated frontier function can now be displaced upward by this estimate of $E(u_k)$. Apart from the known limitations of deterministic residuals, MOLS has the disadvantage that the production function is not necessarily shifted far enough to ensure that all observations lie on or below the frontier, and so some residuals may have the wrong sign (Førsund et al., 1980; Lovell, 1993). The MOLS method is a little less orthodox than the COLS since it is unable to result in a full set of negative residuals.

MLE production (or cost) frontiers differ significantly from those produced by the classical OLS regression approach since the relationship underlining inputs and outputs is non-linear. Therefore, this allows efficient observations (that is, those lying on the frontier) to differ in terms of technology compared to observations lying inside the frontier (Lovell, 1993). The computation logic underlying ML was established on the idea that a sample of observations is more likely to have been generated from some distributions than from others. So, the ML estimate of an unknown parameter is attributed to that value of the parameter that maximises the probability/likelihood of randomly drawing a particular sample of observations (Coelli et al., 2005). Therefore,

---

[78] The mean of e^* is by construction zero, so useless.

[79] We assume that the shift from the average production (or cost) to the frontier is constant.

[80] One parameter distribution in this setting means that the expected value (mean) of the distribution depends only on the variance of the distribution.

[81] Cost inefficiency of DMU j (u_j) is given by $\hat{E}(u_k) + e_j$.

[82] Exponentially distributed inefficiency: $\hat{E}(u_k) = \widehat{\sigma_u}$, $\widehat{\sigma_u} = \sqrt{\frac{RSS}{n-(K+1)}}$.

Half-normally distributed inefficiency: $\hat{E}(u_k) = \left(\frac{\sqrt{2}}{\sqrt{\pi}}\right)\widehat{\sigma_u}$.

utilising panel data in this context gives rise to what has been stressed by Cornwell and Schmidt (1996): ‘repeated observation of the same firm makes it possible to estimate its level of efficiency more precisely’.

The joint probability density function (PDF), known as the likelihood function, for a vector of observations $q_k = (q_1, q_2, \dots \dots q_K)'$:

$$L(q|\beta, \sigma) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{K}(q_k - x_k{}'\beta)^2\right\}$$

This is the likelihood of observing the sample observations as a function of the unknown parameters β and σ^2.

In log forms, we maximise the log-likelihood function with respect to β:

$$lnL = -\frac{K}{2}\ln(2\pi) - \frac{K}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{\kappa=1}^{K}(q_k - x_k{}'\beta)^2$$

When a linear regression model is used with errors distributed normally $v_k \sim iidN(0, \sigma^2)$, the ML estimate is equivalent to the OLS estimate.

All these methods suffer from the disadvantages of statistical parametric deterministic models since they do not take into account the stochastic element of the transformation process. Although COLS and MOLS have their place since they are trivial in estimations with solid theoretical foundations even in small datasets, it should be stressed that they produce an identical ranking of producers to OLS. In addition, the methods might be unsuitable for applications in which there are multiple outputs as well as multiple inputs, despite their robustness in providing efficiency estimates under modest measurement error (Johnes, 2004). An informative comparison of these three deterministic methods can be found in Lovell (1993).[83]

The frontier functions specified above, and labelled as deterministic frontier functions, assume that the econometric model is perfectly specified and the data are free of error. Thus, any deviation of an observation from the theoretical maximum is attributed solely to the inefficiency of the DMU. Due to the absence of any stochastic element[84] in the discussed methods, there is need for a specification of the frontier in which the maximum output that a producer can obtain is assumed to be determined both by the production function and by random events. This gives further fringes to recast the models to what is labelled extensively in the literature as stochastic frontier production models.

[83] See appendix 8 chapter 3 for the graphical illustration.

[84] Random external factors such as luck or unexpected disturbances in a related market.

3.7.2 Stochastic Frontier Analysis

In practice models are always imperfect and normally the data are noisy due to random (exogenous) or endogenous events. Ideally the analysis should include a stochastic element that can capture the effects of these random factors. Stochastic frontier estimators[85] provide parametric estimates of efficiency and have been independently proposed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977). SFA is an econometric technique for efficiency analysis based on regression analysis, that requires strong parametric assumptions for the functional form in terms of linking output and inputs and also distributional assumptions for noise and inefficiency, The main advantage of SFA is that it allows for noise in the data and makes possible stochastic inferences, while DEA basically assumes that data are noise-free, so without parameter estimates the method is deprived of providing inferences about elasticities or economies of scope (Thanassoulis et al., 2011).

In this case, the parameters of a model are first specified and then estimated using sample or simulated data (Salerno, 2003). SFA assumes, that the residual is separated into two components, one which illustrates the result of inefficiency and a second, that is considered as random. In practice, this involves assuming a specific distribution for each error component. Thus, the SFA production function can be written as, (Aigner et al., 1977):

$$q_k = f(x_{k1} + \cdots + x_{kN})e^{E_k}$$

In log-forms:

$$\ln(q_k) = ln[f(x_{k1} + \cdots + x_{kN}) + E_k] = \beta_0 + \sum_{i=1}^{N} \beta_i ln\, x_{ki} + E_k$$

Where $E_k = v_k - u_k$ and $u_k \sim N(0, \sigma_u^2); u_k \geq 0$, v_k and u_k are statistically independent[86]. The first component of the residuals v_k is normal and is attributed to measurement error and random fluctuations, while the second component u_k is one-

[85] These models fall into the parametric stochastic model caste and most of these techniques are based on the ML principle

[86] The noise component v_k has identical properties to the noise component of a linear regression model. The same properties are valid for the inefficiency component except it has a non-zero mean $u_k \geq 0$. Both errors are uncorrelated to the explanatory variables x_{ki}. The main properties can be summarised into:

i. $E(v_k) = 0$
ii. $E({v_k}^2) = \sigma_v^2$ (homoscedastic)
iii. $E(v_k v_j) = 0 \; \forall \; k \neq j$ (uncorrelated)
iv. $E(u_k^2) = constant$
v. $E(u_k u_j) = 0 \; \forall \; k \neq j$ (uncorrelated)

sided typically exponential or half-normal[87] and is attributed to technical inefficiency[88]. The parameters of the function can be estimated using MOLS (Førsund et al., 1980); Lovell, (1993) or MLE methods since it is a log linear operation and as such cannot be achieved using OLS. MLE estimators are asymptotically consistent and efficient estimators but the TE estimator may be inconsistent in some cases. In a stochastic frontier framework in the form introduced by Aigner et al. (1977), the maximum log-likelihood function considering a half normal distribution[89] $u_k \sim N^+(0, \sigma_u)$ takes the form:

$$\ln \mathrm{L}(\mathrm{q}|\alpha, \beta, \lambda, \sigma^2) = Kln\frac{\sqrt{2}}{\sqrt{\pi}} + Kln\frac{1}{\sigma} + \sum_{k=1}^{K} \ln\left[1 - \Phi(E_k\lambda\frac{1}{\sigma})\right] - \frac{1}{2\sigma^2}\sum_{k=1}^{K} E_k^2$$

With $\lambda = \frac{\sigma_u}{\sigma_v}$ reflecting the asymmetry of the distribution of the error term. The larger the value of λ[90], the more pronounced the asymmetry will be. If $\lambda = 0$ then the symmetric error component v_{it} dominates the one-side error component u_k in the determination of E_k. Thus, the complete error term is determined solely by the random disturbance that is distributed normally (Mastromarco, 2008).

According to Greene (1980a, 1980b) the distribution of the composed error term is asymmetric since it incorporates the inefficiency term. Hence, Greene's argument adopts an ML estimator that takes into consideration this information so more efficient estimates are produced, at least asymptotically. The Gamma distribution has been adopted to model the inefficiency error term due to its high flexibility but almost always the shapes of statistical noise and inefficiency are barely distinguishable.

Therefore, a stochastic approach produces efficiency measures that are separated from random shocks or measurement errors; however they are still potentially affected by misspecification errors. The imposition of a particular distributional form (e.g. half-normal or exponential) on that component of the residual that is attributed to technical inefficiency is an assumption that has no theoretical basis. Due to the allowance of stochastic errors and parameter estimation, parametric approaches, give a further insight into useful information such as, returns to scale and scope, and elasticities.

---

[87] For a half normal distribution $u_k \sim iid\ N^+(\mu, \sigma^2)$
For an exponential distribution the variance of u_k equals to σ_u^2 introduced by Meeusen and Van den Broeck (1977) and Aigner et al., (1977).
Other distributions used in the literature are more flexible but more difficult to estimate:

i. Truncated normal where $u_k \sim iid\ N^+(\mu, \sigma_u^2)$ (Stevenson, 1980)
ii. Gamma $u_k \sim iid\ G(\lambda, m)$ (gamma with mean λ and degrees of freedom m) Greene (1990)
iii. Exponential with mean λ , $u_k \sim iid\ G(\lambda, 0)$

[88] The stochastic element and the inefficiency element are independent from each other. Inefficiency is randomly distributed across DMUs similar to the deterministic frontier.
[89] For an exponential or a truncated normal distribution see Mastromarco (2008).
[90] For a formal test of whether $\lambda = 0$ a Wald statistic test or a likelihood ratio test both based on the ML estimator can be used.

The important features of the stochastic frontier model can be split into the deterministic part $[\exp(f(x_{k1}+\cdots+x_{kN})]$, the stochastic element $[\exp(v_k)]$ which is the symmetric part of the error and the TE component $[\exp(-u_k)]$ that represents the skewed part of the error term. The most common output-oriented measure of technical efficiency[91] is the ratio of the observed output to the ideal output corresponding to the stochastic frontier output.

$$TE_k = \frac{q_k}{\exp[f(x_{k1}+\cdots+x_{kN})+v_k]} = \frac{\exp[f(x_{k1}+\cdots+x_{kN})+v_k-u_k]}{\exp[f(x_{k1}+\cdots+x_{kN})+v_k]} = \exp(-u_k)$$

The first step in predicting TE_k is to estimate the parameters of the stochastic production frontier model. The measure of TE_k varies between $0 < TE_k < 1$; this is an indicative measure of the output of the k-th DMU relative to the output produced by a fully efficient DMU being located to the frontier utilising the same mix of inputs. For a recent review see Kumbhakar and Lovell (2000) and in the context of panel data, stochastic models follow Schmidt and Sickles, (1984) and Cornwell et al., (1990). From the analysis a milestone step is the selection of the inefficiency distribution since it forms a fundamental assumption and not a decision based on the model's characteristics. Generally, it is an *a-priori* decision and not testable. Nevertheless in most empirical work the various inefficiency estimates from different distributional assumptions are broadly more or less similar to each other.

By using MLE a direct estimate for the u_k is not feasible since the inefficiency parameter is unobservable. Therefore, the distribution of the inefficiency component provides sufficient information and can be deployed to get an estimate of the conditional mean of inefficiency $E(u_k|E_k)$. The main issue here is that there is no single way to generate the conditional mean, but there are two major estimators[92] in the efficiency estimation literature (Kim and Schmidt 2000). The first is based on the early work of the Jondrow, Lovell, Materov, and Schmidt (1982), (JLMS) estimator. Calculating technical inefficiency can be based either on the mode of the distribution $M(u_k|E_k)$ (the value of u with the largest probability) or based on the mean of the distribution

$$E(u_k|E_k) = \left[\frac{\sigma\lambda}{1+\lambda^2}\right]\left[\tilde{\mu}_k + \frac{\phi(\tilde{\mu}_k)}{\Phi\,(\tilde{\mu}_k)}\right]$$

Where $\tilde{\mu}_k = \frac{-\lambda E_k}{\sigma}$ and $\phi(.)$ and $\Phi(.)$ are the density and CDF of the standard normal distribution (Greene, 2008).

[91] Exactly the same logic follows a cost frontier with the only difference being the sign of the inefficiency term:

$$CE_k = \frac{f(q_k;w_k)\exp(v_k)}{C_k}$$

Where $0 < CE_k \leq 1$. So $lnC_k = lnf(q_k;w_k) + v_k - lnCE_k = lnf(q_k;w_k) + v_k + u_k$

[92] The characteristics (strengths and weaknesses) of the JLMS and Battese and Coelli estimators can be seen in Greene (2008).

However, Battese and Coelli (1988) suggest an alternative estimator that calculates TE by the mean of the distribution of $E(TE_k)|E_k) = E(exp(-u_k)|E_k)$. For the truncated normal model (which includes the half-normal case), this is:

$$\mathrm{E}(\exp(-\mathrm{u_k})|\mathrm{E_k}) = \frac{\Phi\left[\frac{\mu_k^*}{\sigma_*} - \sigma_*\right]}{\Phi\left[\frac{\mu_k^*}{\sigma_*}\right]} \exp\left[-\mu_k^* + \frac{1}{2}\sigma_*^2\right]$$

Where $\sigma_*^2 = \frac{\sigma_v^2\sigma_u^2}{\sigma^2}$ and $\mu_k^* = \tilde{\mu}_k + \frac{\mu\,\sigma_u^2}{\sigma^2}$. The academic community has not yet settled on which method to recommend since all methods produce estimates that are statistically inconsistent i.e. the estimate of $\mathrm{u_k}$ does not necessarily convert to the true value since the estimator is conditioned on a specific set of data.

SFA is an ingrained approach in economic theory and due to its statistical nature and various empirical applications, is quite a popular technique. However, there are a couple of constraints and limitations that should be stressed since, despite the computational facility of the simulation processes, the distributional assumption issue in every application is yet a major concern since the imposition of a particular distributional form remains an assumption that has no grounding in theory. Also, it requires large samples[93] to ensure accurate results and any misspecification errors are incorporated into the measure of efficiency. Estimates are statistically inconsistent and this is reflected in the confidence intervals that attach to the inefficiency estimates that might be too wide for the method to gain credibility in practice (Johnes, 2004). However, in large samples by the central limit theorem we might expect the distribution of DMU efficiencies to be normal. Finally there are competing estimators to predict TE, which do not always manage to converge in reality.

3.8 Panel Data Models of Efficiency Measurement

Econometric approaches to efficiency and productivity measurement vary relative to the type of data analysed, i.e. cross-section or panel data. In cross-sectional models, the data sample comprises observations on k DMUs: $S = \{(x_k q_k|k = 1, \ldots \ldots K))\}$ since they are limited to a single time period, while, in panel data models observations on k, DMUs are available over T periods of time: $S = \{(x_{kt} q_{kt}|k = 1, \ldots \ldots K; t = 1, \ldots \ldots T)\}$. The data are available across a number of periods and enable the measurement of productivity change. Also due to the enlarged dataset since the available data points increase[94] estimation of technical progress or regress is feasible (Daraio and Simar, 2007). The literature on panel data estimation of frontier models accommodates producer-specific effects (time-invariant heterogeneity)

[93] Preferably samples with $n > 100$

[94] The number of data points increases due to the time dimension that inserts the analysis; however at the same time, the number of estimated parameters increases as well.

and addresses fundamental questions such as how and whether inefficiency varies over time (Greene, 2008).

Utilising panel data, new options emerge for measuring efficiency. Panel data achieves relaxation of the assumption of independence between technical inefficiency and inputs and the assumptions on the distributional forms of statistical noise and technical inefficiency imposed (Mastromarco, 2008). However they are not compatible with simple methods of estimation, i.e. OLS since the main assumption here is that the residuals should be uncorrelated with each other[95]. Pooled OLS shares the cross-sectional data over time into one dataset and models the data in the standard cross-sectional OLS way. In this way the time and university dimension of the dataset is ignored since each observation is treated as a different DMU (OFGEM, 2013). The traditional framework of COLS and MOLS is endowed with a pooled-OLS framework that is the same as the cross sectional model but includes time effects:

$$\ln(q_{kt}) = \beta + \sum_{i=1}^{N} \beta_i x_{kti} + \gamma t - u_{kt}$$

Where $i = (1, \ldots \ldots, N)$ denotes inputs, $k = 1, \ldots \ldots K$ denotes numbers of cross sections, and $t = 1, \ldots \ldots., T$ are the time periods. However pooled OLS can be misleading when there are company-specific effects or when inefficiency varies over time, and if inefficiency depends on some exogenous variables (observed or unobserved). Therefore, the residuals might be inconsistent if firm effects are ignored and a pooled model is used (OFGEM, 2013). In the same lave wedge inefficiency levels of the DMUs are misleading in the sense that we cannot separate inefficiency from unit specific effects and noise. Although pooled OLS estimates will be consistent in the presence of random firm effects, the estimated standard errors will be incorrect for hypothesis testing. As it turns out, the proposed use of pooled OLS can be recast or completely abandoned for more robust estimation techniques. These techniques have been developed in the efficiency literature, many of which have been used by regulators, policymakers and researchers.

3.8.1 Panel Data Stochastic Frontier Models

In its simplest version a pooled SFA model retains the same logic with those using cross section data but due to the panel nature of the data time can be included as well.

$$\ln(q_{kt}) = \beta + \sum_{i=1}^{N} \beta_i x_{kti} + \gamma t + v_{kt} - u_{kt}$$

Where $i = (1, \ldots \ldots, N)$ denotes inputs, $k = 1, \ldots \ldots . K$ denotes numbers of cross sections, and $t = 1, \ldots \ldots., T$ are the time periods, and v_{kt}, u_{kt} are distributed as it

[95] Conventional regression-based strategies to address correlated errors (McManus, 2011)

i. Cluster-consistent covariance matrix estimator to adjust standard errors.
ii. Generalised least squares instead of OLS to exploit correlation structure.

follows $v_{kt} \sim \mathrm{N}(0, \sigma_v^2), \mathrm{u_{kt}} \sim |U|\ and\ U \sim N(0, \sigma_u^2)$. This is the normal-half-normal model that forms the basic form of an SF model. SFA and panel data models allow for direct interpretation of the residuals. In all SFA models, the statistical significance of inefficiency is a testable condition[96]. This model is estimated by ML, as discussed earlier. The JLMS estimator is used to estimate $\mathrm{u_{kt}}$ (Greene, 2008). The SFA literature has evolved steadily and further extensions in an SFA panel setting allow for explicit interpretation of the results in terms of latent heterogeneity,[97] noise in the data or specification errors, and persistent and/or time-varying inefficiency.

Furthermore, with panel data, each unit is observed at several different points in time, so we expect the constructed estimates of the efficiency levels of each unit to be consistent as the number of observations per unit increases. This means that inefficiency can be estimated more precisely and some rigidities faced with cross-sectional models can be removed. Those rigidities concern the distributional assumptions[98] used to estimate parameters and the inefficiency estimates using the JLMS formula, the independence assumption between the technical inefficiency component and the regressor(s) (Mundlak, 1961), and the inconsistency of the JLMS estimator. Before proceeding with the presentation of the available stochastic frontier methods for panel data, a distinction concerning the time dimension[99] of the inefficiency term has to be made. First, the most restrictive of the models in terms of assumed behaviour of inefficiency where inefficiency, will be kept constant over time for each unit, are presented; finally, recent developments composed of both time-invariant and time-varying inefficiency components will be presented.

3.8.2 Time-Invariant Technical Inefficiency Models

The stochastic panel model with time-invariant inefficiency can be estimated under either the fixed effects or random effects framework (Wooldridge, 2010). The selection of the framework is dependent on the level of relationship permitted between inefficiency and the explanatory variables of the model (Parmeter and Kumbhakar, 2014). The fixed effects approach permits correlation between x_{kt} and u_k , whereas the random effects approach does not.

The received literature on the fixed effects model in the frontier modelling framework is based on Schmidt and Sickles's (1984) modification on the linear regression model which incorporates a unit-specific intercept in the basic linear model framework. They propose a model that estimates the persistent part of the inefficiency without specifying an explicit distribution of the inefficiency, labelled the distribution free approach.

[96] This is not the case with OLS (COLS, MOLS).
[97] A key question related to latent heterogeneity (the time-invariant individual effects) is whether the individual effects represent (persistent) inefficiency, or whether the effects are independent of the inefficiency and reflect (persistent) unobserved heterogeneity.
[98] Some of the strong distributional assumptions used to disentangle the separate effects of inefficiency and noise can be relaxed (Coelli et al., 2005).
[99] Assumptions made on the temporal behaviour of inefficiency.

Classic fixed effects models take advantage of the panel structure to increase the explanatory power of the model by incorporating a unit-specific, time-invariant effect. So, the linear fixed effects model can be reinterpreted as:

$$\ln(\mathrm{q_{kt}}) = \beta_0 + \sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{\beta_i x_{kti}} + v_{kt} \ - u_k$$

$$\ln(\mathrm{q_{kt}}) = \beta_0 - u_k + \sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{\beta_i x_{kti}} + v_{kt}$$

$$\ln(\mathrm{q_{kt}}) = \alpha_k + \sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{\beta_i x_{kti}} + v_{kt}$$

Which can be estimated consistently and efficiently by OLS after including individual dummies as regressors[100] for $\alpha_k \equiv \beta_0 - u_k$. The model is reinterpreted by treating α_k as the firm-specific inefficiency effect. The purpose of the effect is to capture the impact of all the factors that are specific to the unit and constant over time, and which have not been included already in the model. This means that the time-invariant units' heterogeneity, e.g. location characteristics (urban or rural), a person's ability (when modelling income), prevailing environmental conditions, etc. is reflected in α_k. To retain the flavor of the frontier model once $\widehat{\alpha_k}$ is available, the DMUs are compared on the basis of the following transformation to obtain an estimated value of u_k (Schmidt and Sickles, 1984):

For production functions: $\widehat{u_k} = \max_k\{\widehat{\alpha_k}\} - \widehat{\alpha_k}$, $k = 1, \dots \dots \dots, K$, where $\widehat{\alpha_k}$ is the k −th fixed effects estimate in the within-groups fixed effects linear regression model. This formulation implicitly assumes that the most efficient unit in the sample is 100 percent efficient. Therefore, the estimated inefficiency in the fixed-effects model is relative to the best unit in the sample. The unit-specific TE estimate equals $TE_k = \exp(-\widehat{u_k})$. For cost functions: $\widehat{u_k} = \widehat{\alpha_k} - \min_k\{\widehat{\alpha_k}\}$, $k = 1, \dots \dots \dots, K$ and the cost efficiency estimate equals $CE_k = \exp(-\widehat{u_k})$.

The fixed effects approach empoweres the model with an important implication to allow for correlation[101] between x_{kt} and u_k.This may be a desirable property for empirical applications in which inefficiency is believed to be correlated with the inputs used (Mundlak, 1961). However, the model does not allow for separate identification of

[100] This technique is often referred to as the least square dummy variable (LSDV) method. The coefficients of the dummies are the estimates of α_k.

[101] An important limitation of the FE is that no other time-invariant variables, such as gender, race, region, etc., can be included in x_{kt} because doing so entails perfect multicollinearity between the a_k and the time-invariant regressors.

inefficiency and individual heterogeneity, and this constitutes an important limitation that the recent literature aimed to capture, i.e. the true fixed effects (TFE) and true random effects (TRE) models (Greene, 2002).

The model proposed by Schmidt and Sickles (1984) is extended by Cornwell et al. (1990) in order to include a time-varying effect, without specifying an explicit distribution of the inefficiency. Early work on the model suggested direct manipulation of the fixed effects term; in other words, the time-varying part of the inefficiency term is defined as:

$$\alpha_{kt} = \theta_{k0} + \theta_{k1}t + \theta_{k2}t^2$$

Despite the desirable decomposition between the time-invariant component θ_{k0} and the time-varying component $\theta_{k1}t + \theta_{k2}t^2$ some further caveats need to be mentioned. First, it does not leave space for time-invariant heterogeneity that is not inefficiency; second, it assumes a unit-specific quadratic function of time[102] to explain the time-varying part, which might be quite restrictive. Recent panel data literature has tried to relax the assumption of a time-invariant inefficiency in two components (Cornwell and Schmidt, 1996).

Turning to the random effects model developed by Pitt and Lee (1981) the inefficiency term u_k is assumed to be constant through time and randomly distributed since it must be uncorrelated with independent variables. If so, time-invariant regressors such as gender, race, etc., can be included in the model without leading to collinearity with α_k. Also, the RE framework facilitates cases in which independent variables show very little variation between time periods; in such cases, the FE may fail to identify the statistical significance of those variables since all variation is captured by the effects. However, it should be stressed that the RE model is restrictive in the sense that is does not allow for correlation between the RE unobserved time-invariant inefficiency and the independent variables (i.e. the regressors) and noise. When the assumption of no correlation between the covariates and university efficiency is indeed correct, then estimation of the stochastic frontier panel data model by using the RE framework provides more efficient estimates than estimation under the FE framework.

According to Pitt and Lee (1981), distributional assumptions on the random components of the model can be imposed and then estimation of the parameters of the model is feasible through ML.[103] The relevant log-likelihood for a random effects model with a half-normal distribution is derived by Lee and Tyler (1978) and discussed further by Battese and Coelli (1988). Inefficiency is not directly estimated via MLE, so once the parameters are estimated, JLMS-type conditional mean estimators can be used to receive an estimate for the unit-specific inefficiency (Kumbhakar, 1987).

[102] Han et al. (2005) propose factor analytic forms for modelling a_{kt}

[103] An alternative to MLE is the use of generalised least squares (GLS) estimator; see Baltagi (2013) for an analytical review of the method.

The likelihood function for the $k-$ th observation is Pitt and Lee (1981):

$$lnL_k = contant + ln\Phi\left(\frac{\mu_{k*}}{\sigma_*}\right) + \frac{1}{2}\ln(\sigma_*^2) - \frac{1}{2}\left\{\frac{\Sigma_t E_{kt}^2}{\sigma_v^2} + \left(\frac{\mu}{\sigma_u}\right)^2 - \left(\frac{\mu_{k*}}{\sigma_*}\right)^2\right\} - Tln(\sigma_v) - \ln(\sigma_u) - ln\Phi\left(\frac{\mu}{\sigma_u}\right)$$

With $\mu_{k*} = \frac{\mu\sigma_v^2 - \sigma_u^2\Sigma_t E_{kt}}{\sigma_v^2 + T\sigma_u^2}$ and $\sigma_*^2 = \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + T\sigma_u^2}$, T sample size. For RE no assumptions on the PDF of the inefficiency are made other than that inefficiency is an independently distributed random variable with non-negative values, i.e. $u_k \sim iid \geq 0$ i.e. a half normal distribution. The stochastic error component is distributed as $v_{kt} \sim iid\ N(0, \sigma_v^2)$. The estimated parameters from the MLE estimation process are utilised in the next step, where the extended JLMS estimator of inefficiency is used to obtain an estimate of inefficiency.

$$E(u_k|E_k) = \mu_{k*} + \sigma_*\left[\frac{\varphi\left(-\frac{\mu_{k*}}{\sigma_*}\right)}{1 - \Phi\left(-\frac{\mu_{k*}}{\sigma_*}\right)}\right]$$

If $u_k \sim |U_k|, U_k \sim N[0, \sigma_u^2]$ is distributed half-normally,[104] then $\mu = 0$. Also, in some cases, μ may be a function of covariates of exogenous determinants of inefficiency, i.e. $\mu = (z'_k\delta)$.

As previously mentioned, the inefficiency term here has a time-invariant interpretation since inefficiency levels may vary for different individuals, but they do not change over time therefore, time variation is an issue to be accommodated in the literature. The array of models introduced in the next section aim to give a time-varying dimension to inefficiency since the implications of the time-invariant hypothesis are too restrictive. This implies that an inefficient unit (e.g., a university) does learn over time; therefore, if the latent goal is productivity and efficiency improvements, a time-varying inefficiency framework should be structured.

3.8.3 Time-Varying Technical Inefficiency Models

Most of the primal approaches used to handle time-varying inefficiency have specified it as a product of deterministic functions of time and the random effects, u_k, now reflecting the time-varying part of inefficiency. Note that Kumbhakar (1990) gives in the inefficiency term the following specification: $u_{kt} = [1 + exp(bt + ct^2)]^{-1}|U_k|$, while the Battese and Coelli (1992) formulation of inefficiency is $u_{kt} = \exp[-\eta(t - T)]\,|U_k|$. Moving towards a time-decaying inefficiency framework, inefficiency will be a function of time and as such Battesse and Coelli (1992) suggest $u_{kt} =$

[104] Stevenson (1980) adopted a truncated normal distribution for u_k, $u_k \sim |N(\mu_\kappa, \sigma_u^2)|$.

$\exp[-\eta(t-T)+(-\eta)(t-T)^2]\,|U_k|$. The decay parameter determines whether inefficiency increases or decreases over time and remains constant across all units. These approaches require distributional assumptions for the inefficiency term and the most common candidate is the truncated normal $u_k \sim iidN^+(\mu, \sigma_u^2)$.

A more general case of the two is the version given by Lee and Schmidt (1993), in quite a flexible version without assuming any parametric function for the inefficiency term. Here, the behaviour of inefficiency is given by $u_{kt} = u_k\,\lambda_t\ with\ t = 1, \dots \dots \dots, T,$ as time-specific effects to be estimated. Both components u_k and λ_t are deterministic, but in the estimation process u_k is considered to be random. It should be noted, that the temporal pattern of inefficiency is exactly the same for all units, which might be quite restrictive compared to other specifications. Placing all these models in a unified framework, a generic formula[92] can be used where $u_{kt} = G(t)u_k,\ G(t) > 0$ is a function of time (t) representing a non-stochastic component common across units, while u_k is a unit-specific stochastic one. This type of time-dependent inefficiency varies over time and across individuals (Parmeter and Kumbhakar, 2014).

What follows in the analysis is the identification of cavities that need to be treated and a discussion of the models developed in literature to bridge such gaps. First, making reference to the distinction between any unobserved time-invariant individual specific heterogeneity and inefficiency is essential since the two are undistinguished so far. Hence, latent heterogeneity is confounded with inefficiency, so $\widehat{u_k}$ indicates heterogeneity in addition to, or even instead of, inefficiency (Greene 2005b). Second, the time-invariant nature of inefficiency is somehow misleading in long panel data since units in the long term should identify and treat any signs of inefficiency, otherwise a viable position in the market cannot be assured. Third, an undisputed question pertains to whether the time-invariant component should be considered persistent inefficiency or individual heterogeneity that captures the effects of unobserved time-invariant covariates having nothing to do with inefficiency.

Greene (2004b) argues that u_k would be absorbing a large amount of cross country heterogeneity that would inappropriately be measured as inefficiency. Hence, the ‘true’ fixed effects model in which inefficiency is time-varying irrespective of whether the time-invariant component is treated as inefficiency (persistent) or as an individual-specific effect (heterogeneity) has been developed, providing information on the transient part. The generic formula by Greene (2005a) is:

$$
\begin{gathered}
q_{kt} = a_k + x'_{kt}\beta + v_{kt} - u_{kt} \\
v_{kt} \sim N[0, \sigma_v^2] \\
u_{kt} = |U_{kt}| \\
U_{kt} \sim N[0, \sigma_u^2]
\end{gathered}
$$

In the TFE model, a_k is a random variable that might be correlated with x_{kt} and represents time-invariant heterogeneity. This is a simply pooled SFA model with unit-specific dummies capturing firm effects. TFE models have been reviewed extensively by Chen et al. (2014), who fitted the model by MLE estimator. Note that the inefficiency component here is only time-varying lacking of any measure of persistent inefficiency. The JLMS estimator is used directly for u_{kt}.

The technical difficulty with TFE models is what is known in the literature as the incidental parameters problem (Neyman and Scott, 1948); Lancaster, (2000). This technical burden derives when the number of parameters to be estimated increases with the number of cross-sectional units in the data, since when $K \rightarrow \infty$ (number of cross sections increases), the number of α_k increases with K. In the ML framework the number of regressors is fixed, but in a fixed effects case, it increases with K so that the desirable asymptotic properties of the MLE are violated with biased and poorly estimated parameters when T (time periods) is small. This leads to a persistent bias in the MLE of the parameters in many fixed effects models estimated by ML (Greene, 2007). So, actually the problem with fixed effects is that the number of parameters grows with the number of observations and, therefore, the parameter estimates can never converge to their true value as the sample size increases (Hahn and Newey, 1994). Thus, the parameter estimates are severely unreliable. However, there have been recent advanced econometric developments[105] using transformed or first-difference versions of the fixed effects framework to avoid entirely the incidental parameter problem described by Greene, (2005b).

In the TRE[106] case a_k is treated as uncorrelated with x_{kt}. Contrary to the simple RE case by Pitt and Lee, the inefficiency term does not contain any other time-invariant unmeasured sources of heterogeneity since these effects in the TRE models appear in a separate term labelled w_k and u_{kt} picks up the inefficiency. The TRE model is:

$$q_{kt} = a_k + x'_{kt}\beta + v_{kt} - u_{kt}$$

$$v_{kt} \sim N[0, \sigma_v^2]$$

$$U_{kt} \sim N[0, \sigma_u^2], u_{kt} = |U_{kt}|$$

$$a_k = \alpha + w_k, w_k \sim N[0, \sigma_w^2]$$

We can handle the model as a form of the random parameters (RP) model in which the only random parameter in the model is the constant term. The estimation technique of

[105] See Chen et al. (2014) and Wang and Ho (2010) for the likelihood function of the within transformed and the first-difference model and their closed form expressions.

[106] Here due to the fact that the time-invariant and unobserved heterogeneity appears in w_k the estimated inefficiencies would be indeed lower that the traditional RE model by Pitt and Lee.

the parameters for TRE models[107] is the maximum simulated likelihood (MSL) method. To obtain an efficiency estimate the JLMS estimator is utilized indirectly by integrating w_k out of $E(u_{kt}|E_{kt}(w_k))$; in other words, E_{kt} is a function of w_k and then w_k is integrated out of u_{kt}. For a Bayesian framework, the applied methods estimating an SF model have been analytically developed by Koop et al. (1997), Kim and Schmidt (2000) and Tsionas (2002).

A significant inexpediency commonly shared among the RE and TRE frameworks is the omitted variables bias, since unobserved variables may be correlated with the regressors. Mundlak (1978) suggests an auxiliary equation to treat this econometric issue of unobserved heterogeneity bias stemming from the questionable orthogonality assumptions of the random effects model. The idea of using an auxiliary equation dependent on a vector of the units' means of all the time varying explanatory variables can be found in the literature under the term 'correlated random effects (CRE) approach'. The formulation is as follows:

$$q_{kt} = \beta' x_{kt} + \alpha_k + E_{kt}$$

$$\alpha_k = \alpha + \gamma' \overline{x_k} + w_k$$

With the assumption that $E[w_k x_{kt}] = 0$. The auxiliary equation can be interpreted as a conditional mean function or as a projection (Greene, 2007). The method has been used extensively for various premises, i.e. robust tests[108] controlling for correlation between heterogeneity and covariates on nonlinear models, aim to treat the incidental parameters problem, and average partial effects can be identified through CRE, etc. (Wooldridge, 2005).

It is vital to make a meaningful distinction among models with time-varying inefficiency components. Therefore, apart from the division between unobserved heterogeneity and time-varying inefficiency, it is crucial to discern the persistent part of inefficiency that might inaccurately distort our estimates. There are effects from unobserved inputs or inputs such as management that vary across units but not over time (Mundlak, 1961). Hence, estimating the magnitude of persistent inefficiency is vital, especially in short panels. Persistent inefficiency can change only occasionally since it entails structural or/and operational decisions to be made, while, time-varying efficiency can change over time due to a better reallocation of the resources in the short run. Let us consider the model by (Parmeter and Kumbhakar, 2014).

$$q_{kt} = \beta_0 + x'_{kt}\beta + E_{kt} = \beta_0 + x'_{kt}\beta + v_{kt} - u_{kt} = \beta_0 + x'_{kt}\beta + v_{kt} - (u_k + \tau_{kt})$$

[107] In an application of the Swiss nursing homes Farsi et al. (2005) expressed a preference for the TRE specification.
[108] Hausman test comparing random effects (RE) and fixed effects in a linear model.

Here the error component E_{kt} is decomposed to $v_{kt} - u_{kt}$ and further the technical inefficiency part into $u_k + \tau_{kt}$ where u_k denotes persistent inefficiency (time-invariant part) and τ_{kt} denotes the time-varying part of inefficiency (residual or transient). Both components are non-negative however the former is only unit specific, while the latter is both unit- and time-specific. Both components are quite informative in terms of the policy implications, since high values of u_k are of more concern from a long-term point of view, because of its persistent nature, than high values of τ_{kt}. Regarding the estimation process, the model can be written as:

$$q_{kt} = \alpha_k + x'_{kt}\beta + \omega_{kt} = \beta_0 - u_k - E(\tau_{kt}) + x'_{kt}\beta + v_{kt} - [\tau_{kt} - E(\tau_{kt})]$$

Therefore, it can perfectly fit a standard panel data model with unit-specific effects. The estimation technique here is twofold, either by the LSDV approach under the FE framework or by GLS under the RE framework. This model treats all time constant effects as persistent inefficiency even if some have time-invariant, unit-specific heterogeneity. If this is the case, then the model is likely to produce an upward bias in inefficiency since unobserved heterogeneity is treated as persistent inefficiency. These models were developed by Kumbhakar (1991); Kumbhakar and Heshmati, (1995); Kumbhakar and Hjalmarsson (1993,1995) in the SF literature.

3.8.4 Persistent and Transient Inefficiency Plus Unit Effects Models

Based on the aforementioned, there is a gap in literature for a four-error component model which could incorporate long- and short-term inefficiency but also take into account any unit-specific individual effects. Therefore, several of the limitations of the models discussed are overcome due to the pioneering work of Colombi (2010); Colombi et al.(2011, 2014), Kumbhakar and Tsionas(2012), Kumbhakar et al. (2014); Tsionas and Kumbhakar (2014); and their contributions to account for different effects in one sole model. The model can be specified as:

$$q_{kt} = \beta_0 + x'_{kt}\beta + w_k - h_k + v_{kt} - u_{kt}$$

The two components $h_k > 0$ and $u_{kt} > 0$, are strictly positive since they reflect time constant and time-varying inefficiency, respectively, $h_k \sim N^+(0, \sigma_h^2)$ and $u_{kt} \sim N^+(0, \sigma_u^2)$, while w_k captures unobserved, time-constant unit heterogeneity, and v_{kt} is the classical random shock. Estimation is feasible in a single-stage ML method based on distributional assumptions on the four components (Colombi et al., 2011) or in a multi-step procedure (Kumbhakar et al., 2014).[109] An analysis of the MSL version of the model in a cost framework is presented in Chapter 6. In the next section an overview of the English HE literature will be presented in reference to efficiency studies.[110]

3.9 Previous Efficiency Studies in the HE Sector

A comprehensive review of the overall literature orientated in the HE sector in England will be presented in this section. Having discussed the alternative methods of efficiency measurement and the various implications underlying efficiency estimation processes, we ratiocinate the sensitivity of extrapolating the results when different hypothesis applied. A crucial distinction exists between frontier and non-frontier methods, but further distinctions can be made depending on the specification and measurement of the inputs and outputs,[111] the level of data used, and on other assumptions of the model applied, i.e. the functional form of the transformation function, the returns of scale, and the scope hypothesis (Johnes and Johnes, 2007).

Talking from a methodological point of view, efficiency in HE has been assessed by studies that used ordinary least square (OLS) regression methods (Johnes and Taylor, 1987, 1989a, 1989b, 1990, 1992; Johnes, 1996; Kokkelenberg et al., 2008), or frontier methods such as DEA[112] (Tomkins and Green, 1988; Ahn et al., 1989; Beasley, 1990, 1995; Johnes and Johnes, 1992, 1993; Athanassopoulos and Shale, 1997; Madden et al., 1997; Sarrico et al., 1997; Mc Millan and Datta, 1998; Coelli et al., 1998; Sarrico and Dyson, 2000; Avkiran, 2001; Korhonen et al., 2001; Raty, 2002; Abbott and Doucouliagos, 2003; Siegel et al., 2003; Flegg et al., 2004; Johnes, 2006; Afonso et al., 2008; Kounetas et al., 2011; Thanasoulis et al., 2011; Halkos and Tzeremes, 2012) and SFA (Johnes, 1998; Robst, 2000; Izadi et al., 2002; Stevens, 2005; Johnes and Salas-Valesco, 2007; Johnes et al., 2008; Lenton, 2008; Johnes and Johnes, 2009; Abbot and Doucoulianos, 2009; Agasisti and Johnes 2010; Johnes and Schwarzenberger, 2011; Johnes, 2014 or both methods (Johnes, 1999; Chapple et al., 2005; Kempkes and Pohl, 2010; Castano and Cabanda, 2007; Johnes, 2012).

No more than a hundred studies have employed parametric methods to estimate a multi-output multi-input distance function because of the data demands. The multiple-input multiple-output nature of production in HE and the concomitant absence of prices for both inputs and outputs has made DEA an attractive choice of methodology in the HE context, despite its shortcomings. Within the wide array of methods, DEA is the most

[109] The model can be extended to account for persistent and time-varying inefficiency that has non-zero mean as well as allowing for heteroscedasticity in both types of inefficiency.
[110] Studies based on aggregate level data, individual or/ and institution level data and subject level data.
[111] There are inputs that are not under the control of the institutions that need to be treated as partly different.
[112] An informative table regarding the available studies on efficiency measurement in HE is presented in appendix 9 Chapter 3.

frequently applied method; despite its limitations and drawbacks, the advancements in the technique contribute to eliminating many of them. Additionally, more recently, both statistical tests and bootstrapping methods for confidence intervals on DEA efficiencies and sampling variability have been developed (Banker and Natarajan, 2004; Simar and Wilson, 2008).

The level of aggregation is a variant dimension since there are distinguishable levels of analysis[113] according to the available data or the desired outcome. Most of the UK HE studies tend to focus on the institutional level of analysis (Johnes, 1998, 1999; Izadi et al., 2002; Sarrico and Dyson, 2004; Flegg et al., 2004; Thanassoulis et al., 2005; Turner, 2005; Emrouznejad and Thanassoulis, 2005; Glass et al. 2006; Johnes, 2006a, 2006b, 2006c, 2008, 2012, 2014; Flegg and Allen, 2007; Johnes et al., 2008; Johnes and Johnes 2009; Flegg and Allen, 2009). However there are studies that aim to analyse departments. More specifically, Johnes and Johnes (1993, 1995) deal with economics departments, while Beasley (1995) studies physics and chemistry departments. In the same line, Tomkins and Green, (1988) Beasley, (1990), Doyle and Green, (1994), Johnes, (1995), Sarrico and Dyson (2000), Casu and Thanassoulis (2006) concentrate on HEI departments, while Casu et al., (2005) focus on the assessment of institutions' administration services.

Research and innovation are cornerstones of UK HEIs' development policy, as university technology transfer offices can be the level of analysis in some cases (Chapple et al., 2005). Also, in the UK HE sector, due to the significant changes in its structure and funding, there are three broad groups of institution based on historical background; therefore the level of analysis in some cases is limited only to former colleges of HE (Lenton, 2008; Bradley et al., 2010). Expanding the level of analysis further, researchers have used the national or country level, such as Toth (2009) utilising data on private institutions; Agasisti, (2011) exploring 18 EU countries; Aristovnik and Obadić (2011) focusing on 22 countries; Obadić and Aristovnik (2011) studying OECD countries; Parteka and Wolszczak-Derlacz (2013) analysing the public institutions of seven European countries.

The scales of operation are definitely informative in terms of universities' developmental policies. Economies of scale can be a guidance tool on whether larger institutions incur lower costs than smaller institutions and thus determine the optimal size at which an HEI performs at an optimal level of efficiency (Patterson, 2000). Most of the empirical studies in HE dealing with efficiency include in their analysis diverse estimation techniques in determining the economies of scale and the optimal size at which average costs are the lowest (Tirivayi et al. 2014). The size-cost tie can be reflected in many ways, including through calculations of the overall and product-specific economies of scale and scope, ray returns to scale (RRS) and global returns to scope (GRS). [114]

[113] Other levels used in the literature are: subject, vocational studies, and library.
[114] See appendix 10 chapter 3 for the definitions of each category.

According to multiple previous studies, public institutions in the UK[115] continued to enjoy cost savings in terms of scale and scope economies from 1981 to 2003. For the UK, Glass et al. (1995a, 1995b) developed pioneering work on 61 public institutions for the period 1989-2003 in two individual studies exploring not only ray economies of scale and global economies of scope for the 100 percent of the output mean but also product-specific economies of scale and scope. Later, Johnes (1996) endorsed parameter estimates for scale economies in HE with SFA and compared the results with the regression results. The coefficient estimates of scale economies were more or less similar due to the normality of residuals in the regression model, but economies of scope did not exist in the results on SFA. Utilising the same kind of data for 99 public institutions for the period 1994-95, Izadi et al. (2002) and Johnes (1997) applied a nonlinear in the coefficients' CES cost function for calculating the scale and scope economies of UK universities using SFA and regression analysis techniques, respectively.

Economies of scale and of scope can be a predictive factor for universities since they explain much of the preference of certain institutions to specialise in certain subject areas. This is a fact for the UK case, since, according to Johnes (1998), product-specific scale economies do exist for two of the six outputs of the analysis; namely, for the provision of postgraduate tuition and research in the sciences. The remaining outputs exhibit constant product-specific returns to scale (PSRS) while ray economies of scale remain unexhausted and economies of scope are ubiquitous.

Through the estimates of scale and scope economies, further insight in the cost structures in different types of institution is likely. Lenton (2008) examined 96 UK HEIs and compared them with a sample of 956 US further education intuitions, for the period 2000-2002. He shed light on output expansion beyond 100 percent of the output mean level calculating ray and product-specific economies of scale. However, expansion beyond that level would only be appropriate for the institutions in the UK (up to 200 percent of the output mean). Therefore, further education intuitions in the UK could save even more costs if they adopt and adjust their operation in a joint production framework among the different types of instruction.

Significant differences in terms of scale and scope economies exist based on the ownership status of the universities. According to Johnes et al. (2008), there are ray economies of scale up to the mean output level in the public sector, and up to six times the mean output level in the private sector. This is in contrast to Cohn et al. (1989), who estimated scale and scope economies using a sample of 1,195 public and 692 private HEIs in the US; their results indicate that private institutions are more capable of enjoying the economies of scale and scope than public institutions at higher percentages of the output mean. For the English universities, product-specific economies of scale are observed only in the public sector, and only for postgraduate teaching and research

[115] This is the case for the US as well.

output, while economies of scope are found in both the public and the private sectors of HE (Johnes et al. 2008). However, Johnes and Johnes (2009) identified the solemn and sensitive effect on the findings on RRS and on returns to scope due to the choice of estimation methodology. In general, most studies found that the ray economies of scale exist at 100 percent of the mean output, implying that HEIs reap the cost savings benefits at the present output mean level.

The literature suggests that the development of efficiency analysis is particularly high on the public agenda since researchers reclaim issues or bring up issues to shed empirical light on the theoretical issues outlined above. In the next section, the literature review will explore the composition of input and output bundles used in efficiency modelling in HE.

3.9.1 Specification of Inputs and Outputs Used in Higher Education

A crucial decision for researchers dealing with efficiency measurement in HE has been the specification of the most appropriate measures of inputs and outputs. A substantial amount of research has been undertaken with regard to the effect of input and output specification on efficiency scores, much of it in the context of DEA. DEA,[116] despite its comparative advantage over alternative methods (statistical techniques), cannot provide in its basic form the significance of a set of inputs or outputs, significance tests for comparing different models, or for drawing a parallel between efficiency scores of individual groups or DMUs.

According to Johnes and Johnes, (2004), in the context of HE, the conclusions in the results range from rankings being reasonably stable regardless of input and/or output specification (Tomkins and Green, 1988; Abbot and Doucouliagos, 2003; Johnes, 2003) to results being prone to specification errors (Johnes and Johnes, 1992; Ahn and Seiford, 1993). Based on existing studies, there are considerable problems with defining and measuring the inputs and outputs of the HE production process, since, apart from the specification problem, there is a second issue regarding the importance of each of the inputs and outputs in the DEA model.

Some further concerns arise, in the process of separating inputs fully self-controlled by each university and environmental factors that may differentiate or affect the efficiency outcome. Analysts try to cover this angle by either including all inputs, whether controllable or not, in the efficiency analysis (Grosskopf, 1996); Cubbin and Tzanidakis, 1998) or by adopting a two-stage procedure, in which DEA results are derived using a sub-set of controllable inputs, and then the efficiencies from this stage are analysed at a second stage in relation to the non-controllable inputs. In practice, the

[116] DEA can be easily applied in a multiple-input multiple-output production context.

first approach occurs more frequently in the HE sector, despite the overestimation of the results. However, both approaches have been identified having several shortcomings, since misspecification errors in the second stage, or serial correlation in DEA estimates, are some of the potential limitations, making standard methods of inference invalid, (Simar and Wilson, 2004). However, the capacity to assess the performance of HEIs and systems is an even more complicated process, due to the fact that inputs and outputs in the production process are difficult to define and quantify (NRC, 2012).

Traditionally, similar to previous studies, (Worthington, 2001; Abbott and Doucouliagos, 2003; Worthington and Lee, 2008; Abbot and Doucouliagos, 2009; Glass et al., 2009); Worthington and Higgs, 2011), the input-output framework, for organising and measuring the multiple inputs and outputs in HE[117] follows a production approach to modelling university behaviour; that is, universities combine raw materials (such as students), energy (utilities), materials (e.g. paper, pens, computers if not capitalised), labour (academic staff, academic-related staff, and/or other staff), and non-labour factors (physical and financial capital) of production and produce outputs in the form of two main outputs of teaching and research (research output, research income, and research students) (Glass et al., 2002).

From a more rigorous perspective, the practical burdens of measuring labour inputs differs in the HE sector since, even if HE is largely a non-market activity, its workforce emerges from a competitive market in which faculty and other employees have a range of different options. In most cases the quantity of labour can be approached by the number of hours or full-time-equivalent workers. However, the main limitation here is the assumption that all workers have the same skills and so inherently are paid equivalent wages. Indeed, this is an unstable hypothesis and remains true only in situations in which changes and variations in the skill level of the workforce are known to be negligible (NRC, 2012). As a labour proxy in the literature, it is common to use academic and non-academic staff (Avkiran, 2001; Abbot and Doucoulianos 2003; Agasisti and Salerno, 2007) enrolments of undergraduate/postgraduate students, (Agasisti and Perez-Esparrells, 2010; Abbou-Warda, 2011), FTE of undergraduate/-postgraduate students (Arcelus and Coleman, 1997), and FTE of total number of teaching and non-teaching staff and student's own time and effort[118] such as credit hours operating (actual hours offered by each department) (Agha et al. 2011). Note here that, in research-led institutions, the time and cost of faculty and administrative personnel must be divided between research and instruction.

Turning to capital inputs, an intriguing feature is their durable nature and, as such, they generate a stream or flow of services over an extended period. Therefore, the

[117] See appendix 11 chapter 3 for an input-output list

[118] Significant concerns have been expressed regarding this type of measure since student effort should be treated as both an input and an output, this fact is the so-called co-production phenomenon (NRC, 2012).

contribution of capital can be better approached as a measure of service or rental flow, i.e. the cost of using once for one period of time and not by their price for the acquisition. These rental rates are equivalent to a wage rate and can be used in the same way to aggregate across different types of capital service and as a measure of capital income in aggregating the various inputs to production (NRC, 2012). Common proxies are expenditures on library, computing and other learning resources, subsidies, facilities required for teaching (Abbot and Doucoulianos, 2002), governance, administration and staff development, funding for research, total number of places available in teaching rooms, libraries and laboratories space, equipment, and IT, highly-qualified human resources, and library budgets.

A further classification of inputs that is deemed to be definite is between instructional and non-instructional inputs. The first class of input involves regular faculty, adjunct faculty, and graduate student instructors, while non-instructional and indirect costs encompass any administration, athletics, entertainment, student amenities, services, hospital operation, R&D, student housing and transportation, etc. (NRC, 2012).

Turning to the output specifications in HE, these tend to be organised into four different categories: instructional outputs, institutional environment outputs, research outputs, and public service outputs (Breneman, 2001). The most frequently occurring outputs are number of graduates (teaching output), and research output (i.e. income received for research purposes, funding council grants plus income from research grants and contracts). Alternative choices for research output may be research books, book chapters, and journal articles (Abbot and Doucoulianos, 2009), medical-and non-medical research funding (Abbot and Doucoulianos, 2003), student contact hours, number of publications, contribution to publications, and citations (as research output). Tertiary education qualifies graduates for jobs or additional training, intensifying their competence and analytical capacities. In this sense, they acquire advance qualifications that boost their professional education with concurrent direct income effects, increased social mobility, and health as well as other indirect effects. Additional metrics to be mentioned as suitable measures are the success rates of undergraduate students, number of doctoral dissertations, number of students enrolled on PhD courses, foreign students enrolled as a percentage of all students, and revenues from financed activities, etc.

Finally, the amount of external resources attracted to research activities (grants, consultancies, etc.), promotions (number of promotions attained by the academic staff of each department, public service activities (number of workshops, conferences, training courses and other activities by the teaching staff of each department), could be vital proxies for output measures. As mentioned previously, apart from the two traditional outputs of teaching and research, universities have developed a third output that reflects their involvement with wider society; thus, in the next section an introduction to third-stream activities in HE is presented.

3.9.2 Universities’ Third Mission

Universities have accepted, traditionally, two main missions: teaching and research. However, recently, another role has gained increasing recognition worldwide, reflecting universities’ involvement in society and industry: the well-known third mission. This mission is intensively associated with the role universities can play towards economic growth and social progress in the modern world, by constructing a growing ‘knowledge society’, which is one of the main principals of the Lisbon Agenda development plan, for the economy of the European Union. As a result, the traditional roles of teaching and research are being expanded to involve activities that facilitate universities’ engagement with society and industry (Etzkowitz, 2000; Vorley and Nelles, 2008). Policymakers have been keen to encourage universities to make contributions to society, to develop the strongest connections between knowledge and social welfare.

The third mission should no longer be formulated in terms of ‘best practices’ but in terms of a more broadly accessible and productive communication path between universities and third parties (Vendetti et al., 2011). Universities are an example of a multitask cluster, since they should preserve teaching and research at excellent levels to not only be entrepreneurial and competitive but also to show concern for their students and communities. Only through cooperation with other knowledge providers are universities able to achieve broader horizons and the comparative advantage of fulfilling accountability requirements (Watson, 2003).

Despite the recent growing recognition of the third mission, this topic is not new on the agenda. At a very early stage, in 1998, the UK government took the initiative to introduce the concept of wealth creation as a third ambition of universities, as an additional parameter to the two previously large-scale developed activities of teaching and research. This pioneering concept was incentivised with £50 million annually funding spent, as a prompting motion towards universities (Klein, 2002; Martin and Tang, 2007; Mollas-Gallart et al., 2002; Venditti et al., 2011). However, the origins behind the concept of organising a third mission can be traced back to the land-grant universities in the US in the 19th century (Clark, 1998; Etzkowith, 2002; Venditti et al., 2011). While several proposals exist for measuring research and teaching activities, there are few consistent approaches to evaluate and measure third mission activity (Montesinos et al., 2008).

In the same vein, Görason et al. (2009) highlight that the types of function that should be included in the definition of the third mission vary significantly in different countries and contexts (e.g. Germany focuses on technology transfer from universities to enterprises, whereas Latin American adopts a broader concept in which universities serve community needs). The need for a conceptual framework and a set of indicators is addressed. A discussion defining third mission indicators as well as an identification process for selecting the most relevant metrics is ongoing. Indicators should ideally reflect third mission activities within the institutions. Therefore, many projects have traced the identification process of which activities are currently part of the third mission[119]. This chapter has offered an overview of the available methods used in the literature for efficiency measurement in HE. In addition, it has offered an articulate summary of the existing literature in HE, focusing on the English HE sector.

[119] A thorough analysis on the promising projects intending to identify the crucial third mission dimensions is available in appendix 12 chapter 3. Moreover a critique on the methods used to identify third mission indicators is offered in appendix 13 chapter 3.

4. Chapter 4: Mergers in Higher Education in England

4.1 Chapter Background

Merger activity can be observed in both the private and public sectors. There is a considerable body of literature on the causes and consequences of mergers in the private sector (Field and Peck, 2003). Our interest, however, is in the much-less-researched area of mergers in the public sector and, in particular, in the HE sector. However, what is a merger and how can it be interpreted in the HE context? There is more than one terminology, but, according to the definition framed by the HEFCE:

> Merger exists, when two or more partners have combined to create a single institution, which may retain the name and legal status of one of them or be an entirely new legal entity. In the 'holding company' model, one institution can have subsidiaries that retain separate names, brands and operations, to varying degrees. Federations can be seen as a more flexible version of full merger' (HEFCE, 2012).
>
> Mergers of institutions involve the dissolution of one or more partners and integration into another partner. Another possible pattern is the dissolution of all partners and the creation of a new institution, but this is less common, (HEFCE, 2010).

Mergers in HE can be traced back to the 1960s and early 1970s in different countries (Skodvin, 1999). Australia, Great Britain, Germany, and Sweden pioneered mergers during the 1970s. A binary, or a two-fold, HE system was formed in Australia and the UK, establishing colleges of advanced education and polytechnics as the main alternatives to universities. Germany, in this context, introduced a combination of the two systems of German engineering education, the so-called Gesamthochschule (GH), which means comprehensive university, and combined these systems into the so-called Y-model. In the late 1970s and early 1980s, mergers developed more commonly as the key measure to make teachers' training more efficient. In other regions such as the US, a significant 'wave' of mergers was achieved and the arising challenges and benefits are well described in the literature (Millet, 1976; Martin and Samels, 1994; McBain, 2009; Thelin, 2011).

In the US, mergers have been relatively common among private and public HEIs since the 1960s (Skodvin, 1999). The main intention of mergers has been to strengthen weak institutions by diversifying their programmes rather than to get rid of duplicate programmes. In the public sector, this is particularly the case among community colleges, and merging public institutions and state colleges into state-wide, multi-campus institutions (Harman and Harman, 2003) could save a college from permanent closure. The majority of mergers between private sector institutions are strategic, aiming to strengthening the position of the institution. As such, they require close scrutiny of institutions and system operations to understand what might be consolidated. They have not only involved 'strong' institutions but also, in some cases, a mix of 'strong' and 'weak' (Harman and Harman, 2008). However, strategic mergers or 'mergers for mutual growth' are taking place not only in the US but elsewhere as well.

The most radical changes in HE systems due to merger activity started in the mid-1980s and continued until the 1990s. According to Skodvin (1999), this is the case for the Dutch restructuring of the college sector,[120] the consolidation of the Australian HE system in the 1990s, the reorganisation of the Norwegian college sector in 1994, and the amalgamation process in the Flemish college sector. Harman and Harman (2008) found that, from the 1960s to the 1990s, governments in various countries, such as Australia, Canada, the Netherlands, and Norway, attempted to address the issue of the fragmentation of small institutions through mergers; this was also the case in the UK. Ω Since then, mergers have stemmed from a mix of government and institutional activities in all of these countries.

In the UK HE system,[121] two key periods of intense merger activity have occurred, while a considerable number of mergers have taken place in the past 10 years. The reasons for mergers have been focused mainly on the economic problems faced by some institutions. The first phase occurred in the late 1980s to the early 1990s, and it was triggered mainly by financial difficulties experienced by former polytechnics and colleges of HE. These institutions were vulnerable due to their small-scale operations, so they were no longer financially viable, and the very best alternative was to be acquired by larger partners (HEFCE, 2010). From the mid-1990s to the early 2000s, a second phase occurred, and the driving force was the protection of HE provision. The role of the HEFCE was not only to provide adequate funding resources for the mergers but also to put in touch small and vulnerable institutions with potential merger partners.

During the mid-1990s, principal changes in merger activity can be seen compared to the previous phase, since much larger multi-faculty institutions were engaged in mergers. This was the case in the London medical schools' reorganisation as part of a rationalisation of the provision. The fundamental element of these types of merger was that they were politically driven and, consequently, they received considerable public funding. Concurrently, the HEFCE continued to pursue merger partners for and supported vulnerable institutions, where required to do so, by providing expertise for both parties.

Forty mergers occurred among UK HEIs during the period 1994–2008, with many more between HEIs and FECs (HEFCE, 2010). An 11 percent reduction in the number of

[120] This reform happened between 1983 and 1987.
[121] An extensive review on mergers in the English HE sector is provided in appendix 14 chapter 4.

HEIs can been seen, since their number has reduced gradually from 186 to 166 since 1994–95 (UUK, 2009). The main merger ‘pattern’ followed was the absorption of specialist colleges into pre-1992 universities (HEFCE, 2010). Twenty-nine mergers took place in England from 1996–97 to 2012–13 in the HE sector, reducing the number of institutions to 123 in the academic year 2012–13 from 138 in 1996–97.

Mergers vary in a number of ways, which can affect the experience of both the merger process and its outcomes (Harman and Harman, 2003, as cited in HEFCE, 2010). Mergers can be voluntary or involuntary, a consolidation or a takeover, single sector or cross-sector, two-partner or multi-partner, and between partners with similar or different academic profiles (HEFCE, 2010). The incentives behind a merger cannot be limited to the obvious natural explanation that the participants generally think that it is more advantageous than disadvantageous, since there are far more expected gains.

Beyond the organisational changes, mergers can bring about administrative, economic, and academic benefits, by merging several smaller institutions into a larger unit. Skodvin (1999) mentions that the reasons for merging in different countries can vary, from resolving financial exigency to more strategic reasons, including expectations to alleviate an institution’s position in the HE hierarchy at the national or regional level. Many institutions are obliged to merge to avoid closure or bankruptcy; therefore, survival and/or growth reasons exist for at least one of the parties (Pritchard, 1993; Rowley, 1997; Harman and Meek, 2002; Harman and Harman, 2003). This was a common strategy among the private HEIs in the US that tried to take precautions against breakdown (McBain, 2009).

The aim of mergers is the better management and use of the available administrative resources, since the intention is to achieve economies of scale with regard to the number of administrators (Harman and Harman, 2003). By achieving more professional and efficient administration, a better allocation of the financial resources is possible since saving money is one of the desirable goals of mergers. In considering international experiences of mergers in HE, Harman and Harman (2003) found that, at the national level, the drivers are major restructuring, financial and academic viability, and quality and inefficiency issues. Individual institutions have also used mergers to tackle financial problems, increased competition, often in research, and falling demand (Harman and Harman, 2003).

4.2 Literature Review-Mergers[122]

4.2.1 Mergers in the Higher Education Sector

Efficiency studies represent interesting snapshots of the efficiency of UK or English HEIs in particular time periods. None, however, examines which factors might affect the underlying efficiencies, and this a serious weakness. If we know the factors that are likely to improve the TE of HEIs, then policy can attempt to create the favourable

[122] Mergers are a common policy in many other public sectors. For the interested reader follow appendix 15 chapter 4.

conditions required to achieve greater efficiency. One obvious issue on which there is a dearth of evidence is the effect that merging universities is likely to have on subsequent efficiency.

UK HEIs have been under increasing pressure since the HE sector has faced increased competition for resources, both within the sector and from other sectors benefiting from public money. As a result, UK HE has been under pressure to provide its services as efficiently as possible, despite the huge changes in size and structure. This is also the case for the HE sectors of many other developed economies.

Mergers of HEIs have been seen as a potential response to the current drivers for change in the HE sector (Browne, 2010). There is an expectation that mergers should result in increased efficiency. Despite the apparent importance of mergers as a policy tool for adapting to failure, there is very little research into the precise effects on efficiency of mergers in HE. The results in the private sector are not very promising since they suggest that approximately 50–70 percent of mergers fail or do not deliver on the anticipated benefits. It is clear that there is a need for more detailed study on the efficiency effects of mergers in HE, so extensive research on the existing literature will be of great interest as a first step.

Mergers can be described as a combination of two or more organisations or institutions, either to create one new organisation, or to retain the identity of one of the original organisations. In the English literature on university mergers, two mainstream concepts are synonymous – mergers and/or amalgamations –both of which reflect the merging of two or more previously separate institutions into a new single institution (Skodvin, 1999). In UK HE, there have been two key periods of intense merger activity, the first between the late 1980s and the early 1990s and the second from the late 1990s to the early 2000s. It is anticipated that a new phase of mergers is imminent due to the recession or in response to other political pressures. Consequently, financial and/or political imperatives may make mergers unavoidable (Harman and Meek, 2002).

In terms of the benefits of mergers in HE, there is no clear answer on the success of the merger process. There are certainly factors, such as leadership, strategic planning, a well-developed network, and a good balance between the networks' units that give an advantage to some mergers to be successful (Skodvin, 1999). However, mergers are costly in terms of money and resources; therefore, undertaking a merger for financial reasons alone is a significant risk. The short-term benefits seem to be insignificant, and the long-term benefits take many more years than expected to develop.

Skodvin (1999) tried to elaborate on the experiences of mergers in HE in Australia, the US, and some western European countries, and discriminated between forced and voluntary mergers. The author found that there are two reasons why mergers take place, either as a reaction to educational policy, or as a result of competition between HEIs. The driving force behind mergers is usually the fear of weakened general access to resources, which tends to motivate voluntary mergers. Some voluntary mergers in HE can be seen in the Netherlands, the US, Sweden, and Canada. Between the different regions, there are similarities and dissimilarities in terms of the reasons for merging.

In addition, the reasons why the decision to merge is taken are justified. A natural explanation is that it is generally considered more advantageous than disadvantageous. The paper by Skodvin (1999) reported mergers that had been resolved under financial exigency, or even following more strategic reasons such as ambitions to improve an institution's position in the HE hierarchy. These institutional changes, as well as the organisational modifications that stem from mergers, attempt to increase efficiency or improve performance indexes, as well as improve the ranking of a university, either nationally or regionally.

The administrative, economic, and academic gains are the dominant goals of forming more efficient institutions in which money can be saved by constraining superficial expenditures. The merger process may contain small or large problems or conflicts; hence, Skodvin offers a transparent dissection between the integration and diversification strategy of merging. The former focuses more on academic integration and cooperation (such as creating new inter-disciplinary programmes) and is somehow more controversial than the latter, which tends to diversify the academic profile of the newly formed institution when two or more complimentary participants are joined.

Finally, Skodvin (1999) analyses the merger process from a theoretical point of view, and identifies the key points of success or failure. A clear answer to the question of whether mergers lead to success or failure is attempted. According to the paper, the answer depends on to whom one is talking, and the stance and perspective that one takes. In this context of clarifying the successful merger, there are three main types of similarity that may be recognised between the counterparts (i.e. the merging partners).

Firstly, there is a clear quest for structural and cultural similarities in cases in which the institutions involved in the merger tend to be unequal partners. In addition, geographical proximity seems to have a vital role in successful mergers, since the closer the institutions, the more successful the merger. Secondly, voluntary mergers tend to be more successful than forced mergers, which lack positive profitability. Finally, mergers can occur as a reaction to public policies or competitive changes within HEIs, or a quest for increased economic and administrative efficiency, which are the main objectives in the education sector. There is a distinction between academic and economic objectives, since mergers that arise for academic reasons may struggle to achieve administrative and efficiency gains. However those efficiency improvements may not be realised at all and in the opposite case they may take time to achieve their goal. Therefore, there are problems in assessing merger success, if mergers have multiple objectives.

Experiences from the US, Australia, and the Netherlands reveal that, in the short term, many resources are required, as well as constructive planning before, during, and after the merger. Only in the long term are the results clear. Consequently, mergers between HEIs can be considered a dynamic process that induces changes in inter-organisational relations.

Botha (2001) constructed a useful and analytical theoretical model, presenting the principles and a navigation scheme beyond the merger process. He suggests that, when considering a merger, both structural and process models need to be used. In this model, the basic assumption is that the envisaged merger is not a forced one, but a merger of choice. This model is a step-by-step process model, wherein each step has its own criteria to identify potential merger partners and to arrange them in priority order. However, according to the author, merging is a complex process, with much opportunity for failure, and this should be taken into account.

In a more recent study by Harman and Harman (2003), mergers seem to have been successful, especially in the context of fragmentation and non-viable HEIs; the authors identified potential and substantial long-term benefits. Mergers work better if they are voluntary rather than imposed and if there are fewer institutions to merge. The same was found in voluntary mergers by Skodvin (1999). Within national systems of HE, a need for increased efficiency and effectiveness, particularly in the context of rapid enrolment expansion, may be the primary driver for mergers. Mergers may take a variety of different forms, in particular following patterns of structures that are more likely to be successful. Harman and Harman (2003), classified five different merger forms. However, there is no clear justification of how this clarification arose.

Cultural differences are pointed out by Harman and Harman (2003) along five dimensions and, more specifically, academic roles, professional loyalties, teaching versus research, reward structures, and styles of governance that are typically evident in universities and HE colleges. Finally, the authors underline the important and definite role of governments and government agencies, which play a highly constructive support role in merger planning and implementation, in various dimensions and, more significantly, through imposing and enacting the relevant legislation. The negotiation process in mergers should be far-reaching in order to be acceptable to all parties, but this does not mean that negotiations should be achieved without due attention to principles.

The theoretical background of mergers has been extensively researched over the last decade. Different types and forms of merger, as well as reasons and strategic plans beyond mergers, exist. However, of primary importance are the more empirical studies, which may focus more on actual results and performance indexes stemming from the institutions pre- and post-merger. The number of purely empirical papers that present results pre- and post-merger in the HE sector are limited; however, they are very common in other fields, especially in the banking and health sectors. The methods applied in these sectors can be transferred to the university sector to assess the effects on efficiency stemming from merger activity.

Hu and Liang (2008) investigated dynamic changes in the research productivity of Chinese HEIs before and after merging. They opted to estimate a Malmquist index as well as the total factor productivity change indexes both pre- and post-merger. In this study, the research sample covered 25 universities. The data was longitudinally for four years between 1999 and 2002, i.e. one year before the merger, the merging year (2000), and the two years following the merger. The authors underline the significant upward changes in the TFP index, as well as the catching-up effect and the scale effect that many universities achieved, to a certain extent, in the years after the merger. The authors tried to decompose the TFP change index into technological change (TC), pure technical efficiency change (PTEC) and scale efficiency change (SEC). According to the findings, technological advancement contributes to the enhancement of overall scientific research productivity.

Therefore, universities should take advantage of the opportunity of institutional mergers to establish reasonable developmental strategies, and to boost their overall scientific research productivity promotion. The total scientific research productivity (TFP) fluctuation is large, which means that merger results are uncertain and ambiguous in terms of reform. However, the authors suggest that final evaluation of a merger in efficiency terms one or two years after it has taken place is not always feasible. This is owe to the newly formed university becomes a rather complex system. The main problem with this paper is the small sample relative to inputs and outputs, and the lack of information on how pre-merger data was selected (i.e. the sample size was 25 both pre- and post-merger).

Mao et al. (2009) examined the efficiency of university mergers from the perspective of knowledge production. They calculated the Z score, which represents the change of comprehensive research (Z score's mean) of the merged and non-merged institutions and found that the Z score of merged CUs (colleges and universities) has increased year on year since 2002, which means that the overall 'research capabilities' of these universities has improved. By comparing the Z score of merged CUs to that of non-merged CUs, they found that merger reform had a positive influence on improvements in research capabilities.

They also conducted a comparative analysis of the annual factor scores of the universities, for which the main methodological approach used was time series analysis. The comparison of different years was conducted using the single factor variance analysis method. The actual goal here was a comparison of the research capability different universities in different years, especially pre- and post-merger. At a later stage, the authors also tried to compare merged and non-merged universities. The analysis here was again based on the annual factor scores of the merged and non-merged CUs, using multi-factor variance analysis, named LSD (least significant difference) and S-N-K (q test), which are useful tests to compare the research capability of merged and non-merged universities in the pre- and post-merger periods, i.e. comparison of groups. The findings indicated that, by the end of 2005, mergers had a positive impact on universities' knowledge production, since the research input factors between merged and non-merged institutions had increased over the years. Additionally, the Z score of merged universities showed a greater increase than that of the non-merged universities.

Johnes (2014) used the DEA and SFA methods of estimation to utilise an output distance function environment that incorporated measures of both quantity and quality of teaching and research inputs and outputs over a 13-year period. The advantage of this study is that it compared the efficiency estimates derived from various estimation methods, and used the results to provide guidance for researchers, managers, and policy-makers on undertaking efficiency studies. The conclusions derived from this study suggest that the estimation methods matter only in terms of the discrimination between the very highest- and lowest-performing universities, since those in the middle are not significantly different in terms of the method used. In terms of the merger activity, institutions that had engaged in a merger presented higher efficiency scores than the average of the non-merged. Consequently, these findings enforce the perspective that increased efficiency is the result of merger activity rather than a crisis in efficiency causing the merger. However, the results involve a caveat, since the efficiency differences may stem from the wide discrepancy between the three different types (pre-1992, post-1992, and former colleges) of HEIs and not due to mergers.

A relatively new contribution to the literature on mergers is by Johnes (2010). In this working paper, the merger activity is examined in terms of efficiency. There are two main questions in this study, which mainly explores which types of institution (in terms of efficiency) are involved in mergers, and examines the effect (in terms of efficiency) of mergers. Also, the lowest- and highest-efficiency quartiles are compared in each group of institutions. The main methodology used in the paper is a parametric technique that allows for stochastic errors, since the main purpose here is the estimation of the parameters as well as to derive measures of substitutability and computing efficiency scores. The findings here suggest that the typical university involved in a merger has efficiency scores similar to the average non-merging university. The typical post-merger HEI is significantly more efficient compared to the pre-merger period or non-merged HEIs. Therefore, mergers between adequately performing institutions seem to have a beneficial effect on efficiency.

In a recent survey of university vice chancellors in the UK, 56 percent of respondents were reasonably or very confident that the UK would see significant rationalisation through HEI mergers and takeovers in the next five to 10 years (Boxall and Woodgates, 2014). However, while policy has pointed increasingly to HEIs specialising in their comparative strengths, such as research or teaching (Glass et al., 2006), very little work of a statistical nature has been undertaken to evaluate the impact on efficiency of mergers in HE. Assessing the efficiency levels of HEIs is not a current trend in the literature. What is actually considered as novel are the implications of mergers on efficiency levels in the post-merger period.

4.3 Data Outline

In this study, the data covers a 17 year period from 1996-97 to 2012-13. The data contains information about all English publicly funded HEIs. The sample is an unbalanced panel of data, for various reasons. First, 28 mergers took place among HEIs during the study period. Due to the adoption of the merger regime, the new institution has been treated as a completely new entity without being connected to its counterpart institutions, which had been merged earlier to create the new entity. This approach has previously followed by Cuesta and Orea, (2002). Second, in the period under examining, some new HEIs penetrated the sector. These units were newly-formed and had not existed previously. Finally, there are some deductions in the number of institutions utilised in the study since four HEIs were removed entirely from the sample: the Open University was removed because of its large size and the unique nature of its teaching provision; the University of London (institutes and activities) was also excluded on the grounds that the composition of the component HEIs recorded under this umbrella changed over time; the University of Buckingham was deleted because it is not publicly funded; and Heythrop College was not used in the data because it only became publicly funded during the time period under consideration. The number of HEIs included in each year therefore varies from 138 in 1996-97 to 125 in 2012-13, after a number of mergers, new establishments, or complete closures, so the panel totals 2,197 observations.

4.3.1 Data

The input and output variables used in this study were constructed from annual statistics for all HEIs in England published by HESA. HESA collects a range of data, from UK universities, HE colleges and other alternatively funded providers of HE every year. The main body of the dataset from 1996-97 to 2008-09 was used previously by Johnes (2014). In this study, a four-year extension of the whole dataset was made, adding years 2009-10 to 2012-13.

In general, HEIs are seen to use labour, capital, and 'raw materials' to produce teaching, research, and third mission activity, and we specify the inputs and outputs to align with this general model. In this study, five measures of inputs have been specified; the number of full-time equivalent (FTE) undergraduates (UGINPUT) and the number of FTE postgraduates (PGINPUT) represent the institution's 'raw materials' or primary inputs.

These inputs develop some variation by institution in terms of quality on entry; however, quality issues are not within the scope of this study due to the limitations of finding representative weights for quality assessment (Massy, 2011). Quality evaluation in both inputs and outputs is a significant issue in the literature. Common proxies of quality in literature are the input prices. However, this aspect cannot serve as a measure of value since the transactions under which prices are generated are questionable. Due to the uncompetitive nature of markets as well as the absence of sufficient information diffusion, prices do not convert to appropriate kinds of value needed for weighting productivity and efficiency metrics (Massy, 2011).

An alternative approach to price involvement is to measure the quality directly and use the results to adjust the inputs and outputs. Inputs and outputs, in this case, are adjusted to reflect their relative values in production and their relative values to users, respectively. Due to differentiations in the quality measures used by each institution or group of institutions, the comparison of productivity and efficiency statistics may be incompatible. Therefore, national quality metrics will need to be developed, to be used for benchmarking and accountability before direct quality adjustments. In the field of HE, there are some attempts to take into account the quality component in the production function, (Athanassopoulos and Shale, 1997; Flegg et al., 2004; Carrington et al., 2005; Bonaccorsi et al., 2006; Fu and Lu, 2006; Flegg and Allen, 2007; Agasisti and Salerno, 2007; Abramo et al. 2008; Agasisti and Perez-Esparrells, 2010; Johnes, 2012). A potential drawback of the model considered in this application is the lack of a quality measure, particularly for teaching inputs and outputs. However, since quality is likely to vary between HEIs but remain relatively stable over time, the random effects estimation model applied in the second stage takes into account unobserved heterogeneity from all sources, including quality differences

The number of academic staff (STAFF) is calculated by the number of full-time academic staff plus 0.5 times the number of part-time staff, employed in UK HEIs. The expenditure on administration and central services (ADMIN) involves expenditure in respect of central administrative staff, faculty officers, and other administrative work and, as such, represents the administrative staff input.[123] This category also includes expenditure associated with the running costs of an administrative computer system and the cost of other facilities or general educational expenditure. Finally, capital inputs are measured by expenditure on library, computing, and other learning resources (ACSERV).

With respect to the output measures, there are three outputs included in this model. There are two measures of the teaching output. The first is the number of undergraduate first degree qualifications (UGOUTPUT) and the second is the number of postgraduate degree qualifications (PGOUTPUT). The research output is covered by income received for research purposes, i.e. research grants and contracts (RESEARCH). The use of research income to measure research output has been questioned on the premise that it is an input rather than an output. The justification for using such a measure to reflect research output is threefold. First, it is a current measure of a university's research reputation and quality since research grants are competitively won. Second, it is easily available and is generally accepted as a reflection of research output. Third, it is typically highly correlated with publications and citation measures which might be considered more appropriate reflections of research output (Johnes and Johnes, 2013). It also has the advantage that it is more current, whereas publications and citations are inevitably backward-looking in nature. Therefore, the choice of research grants as a measure of research output has been made since it is not vulnerable to time lags between the input used in the production process and the output stemming from the process. It, therefore, provides and captures the most current trends in the research output of universities in each academic year (Flegg et al., 2004; Flegg and Allen, 2007a; 2007b; Worthington and Lee, 2008; Worthington and Higgs, 2011). Other measures of

[123] It represents academic and non-academic labour inputs, respectively.

research activity, such as number of publications, citation counts, or patents, are prone to lag differentiation between the research time and the publication time, and, also are not easily available.

The variables, that contain values in pounds (ADMIN), (RESEARCH), and (ACSERV) are deflated to December 2012 values using a consumer price index for the UK economy published by OECD.StatExtracts (an organisation for economic cooperation and development), since the HE pay and prices index for UK universities no longer provides data. This particular specification of inputs and outputs is open to criticism and we address these in the following. Here, both students as inputs and graduates as outputs have been included in the model, in an attempt to cover incidences of non-completion of studies at both postgraduate or undergraduate levels, and also to capture the effect of correlation between postgraduate and undergraduate teaching inputs and outputs, respectively (HEFCE, 2010; HESA, 2013).

Subject differences between universities are not reflected in either teaching or research outputs. The reason why aggregation of outputs by broad subject areas is applied is the avoidance of missing a lot of degrees of freedom in the DEA model. However, the inclusion of HEI-type dummies in the second stage of analysis, allows for exploration of the inter-university variations in the subject mix, which can be also traced via the DEA weights in the first stage, which are unique to each DMU in the dataset.

Another dimension of English HE are the three different types of HEI. Although, in current HE, the legal status of each type of institution is diverse and can take different forms in terms of origin, size and organisation, they share some common characteristics of being (Committee of University Chairs [CUC], CUC, 2009):

- Legally independent corporate institutions
- Bodies with charitable status
- Accountable through a governing body that carries ultimate responsibility for all aspects of the institution.

First, pre-1992 HEIs are traditional universities that had university status prior to the Further and Higher Education Act of 1992. The structure of governance for each university is laid down in the instruments of its incorporation. Unlike other types of institution, they undertake teaching (undergraduate and postgraduate) and research in a whole range of subjects, including medical and veterinary sciences (Johnes, 2014). The second group of post-1992 HEIs are former polytechnics. The Further and Higher Education Act 1992 enabled these institutions to award degrees in their own right, and to acquire the title of university. The third group of HEIs are institutions that are, or have recently been, university colleges and colleges of HE. The use of the title of ‘University College’ indicates that the college has been granted the power to award its own degrees. Since 2003, these colleges have been allowed to apply for university (and degree-awarding) status. Some of these HEIs are specialist institutions concentrating on a particular discipline. The colleges can be divided into two groups: general colleges and specialist colleges. General colleges normally offer a range of courses that may be narrower than in the universities. Their curriculum emphasis is on business, management, humanities, and education (CUC, 2009).

Finally, we do not include a measure of third mission activity in our first stage model. While the *HE Business and Community Interaction Survey* produced by HESA provides a vast array of data from which measures of universities' third mission activity might be constructed, the data is only available from 2008-09 onwards. We are, therefore, unable to include a measure of this output in the DEA model, and leave the exploration of the possible effects of this to the second stage.

4.4 Descriptive Statistics Analysis

Full definitions of the first stage inputs and outputs are provided in appendix 16 chapter 4, while descriptive statistics for the inputs and outputs used in the DEA are presented in Table 1. On average, over the study period, HEIs produced more than 1,000 graduates from postgraduate degrees, over 2,500 graduates from undergraduate degrees, and just over £72 million in research. These were produced from nearly 2,000 postgraduate students, 7,000 undergraduates, 850 FTE staff, and £19 million spent on administration and £10.5 million on academic services.

Descriptive statistics for the inputs and outputs used in the DEA are presented in Table 1. On average HEIs produce more than 1,000 graduates from postgraduate degrees and over 2,500 undergraduates graduated these 17 years. Turning to the research output, institutions received on average just over £ 72 million in research grants. Inputs formed by raw materials and labour (staff) involved in the HE production process accumulated approximately in 2,000 postgraduates and 7,000 undergraduates, plus 850 members of full- and part-time academic staff. Expenditures on academic and administration services by universities reached more than £10.5 and £19 million, respectively.

Table 1: Descriptive statistics of the data used in the DEA

Descriptive Statistics from 1996/1997 to 2012/2013 Number of Observations=2197		
Variable	Mean	Std. Deviation
OUTPUTS		
PGOUTPUT	1,142.40	1,054.48
UGOUTPUT	2,572.94	2,062.10
RESEARCH	22,942.53	50,54.13

INPUTS		
PGINPUT	1,881.77	1,647.22
UGIPNPUT	6,853.22	5303.32
STAFF	845.23	872.76
ACSERV	10,499.77	10,182.90
ADMIN	18,872.28	16,433.89

Note: All input and output variables measured in monetary values are deflated to December 2012 values using a consumer price index for the UK economy. Full definitions of the variables are provided in appendix 16 chapter 4.

According to Figure 1 and Figure 2, there was an upward overall trend with respect to both inputs and outputs over time. Only the research output seemed to demonstrate a downward trend after the academic year 2009-10, reaching a peak of £ 92.5 million on average. From 2010 onwards, there was a significant downward trend, falling to nearly £74 million in 2013. The plateauing of research output since 2009-10 is likely to be a consequence of the effect of the financial crisis on public expenditure. Expenditure on total administration and central services saw a gradual rise until the academic year 2008-09. In this particular academic period the expenditure rose to nearly £27 million on average, and maintain this pace until 2013.

Figure 1: Mean values of outputs used in DEA for the period 1996-97 to 2012-13

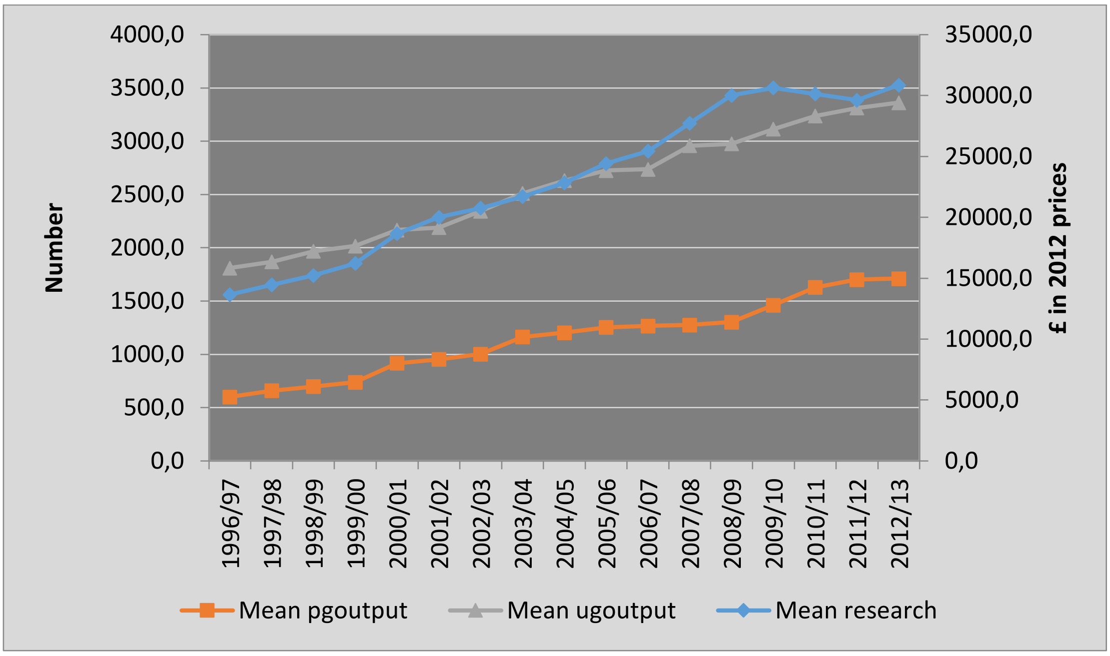
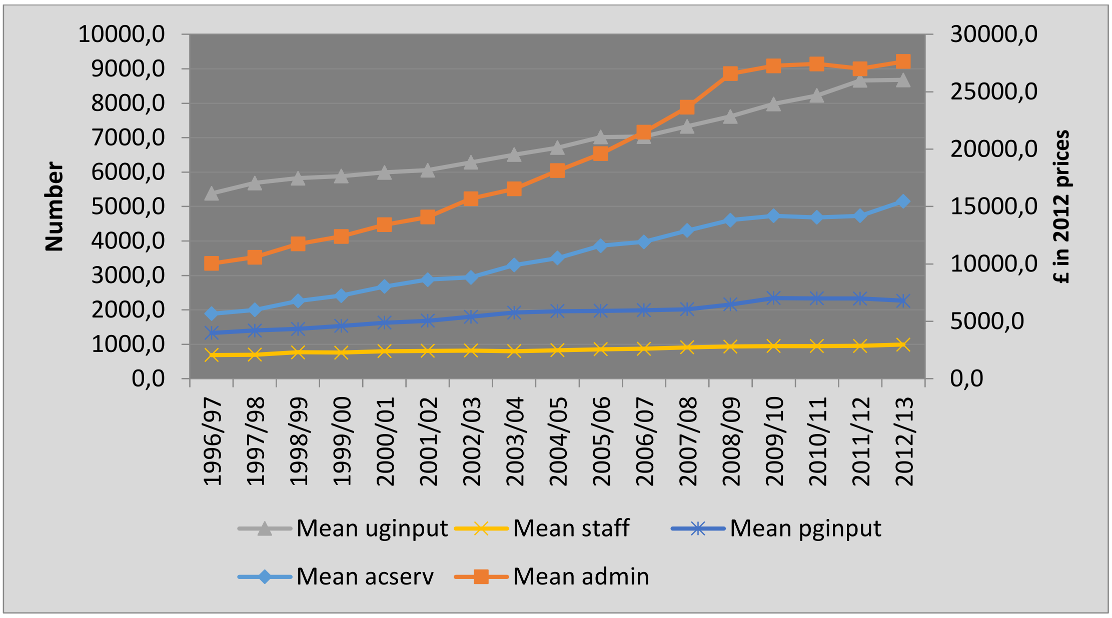


Figure 2: Mean values of inputs used in DEA for the period 1996-97 to 2012-13



Generally, the total volume of inputs and outputs over the explored period, displayed upward trends; however, these have been less significant in the past four academic years. These years after the financial economic crisis of 2008-09 tend to be more stable compared to the rapidly growing modes of the previous years. In addition the discrepancy among the standard deviation numbers established the widely existing diversity between the HEIs in England not only in size but also in the acquired research grants. In Table 2 further insight into the inputs and outputs of HEIs by type is provided, taking a further step in the analysis.

Table 2: Descriptive statistics by type of HEI from 1996-97 to 2012-13

Descriptive Statistics by Type from 1996/1997 to 2012/2013 Number of Observations=2,197						
Type of University	**Post 1992 HEIs (n= 532)**		**Pre 1992 HEIs (n=894)**		**Colleges of Higher Education (n=771)**	
Variable(s)	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
OUTPUTS						
PGOUTPUT	1,439.71	740.74	1,644.69	1,186.80	354.82	429.19

UGOUTPUT	4,848.01	1,555.24	2,440.82	1,744.73	1,155.80	1,148.63
RESEARCH	4,991.84	4,628.96	52,035.44	69,606.37	1,594.55	7,150.98
INPUTS						
PGINPUT	2,288.93	887.04	2,760.25	1,874.81	582.19	646.24
UGIPNPUT	11,633.05	3,678.68	7,526.35	5,207.19	2,774.55	2,542.60
STAFF	925.93	291.74	1,319.61	1,112.72	239.50	211.16
ACSERV	13,491.74	6,224.2	14,744.21	12,400.74	3,513.69	3,833.28
ADMIN	24,007.24	10,083.17	25,419.71	20,112.27	7,737.14	6,233.71

The number of universities under the umbrella of the pre-1992 universities varies from 55 institutions in 1996-97 to 51 institutions in 2012-13, revealing that this category is larger in size compared to almost 45 former colleges of HE and only 32 on average post-1992 institutions spread throughout the seventeen years of examination. A deeper examination reveals that the amount of inputs and outputs tended to be higher in the pre-1992 universities. There are cases in which this discrepancy between the pre-1992 universities and former colleges more than doubled, such as the number of undergraduates produced, and, in some cases the discrepancy even tripled, such as the acquired volume of research grants. The research income obtained in pre-1992 HEIs also outperformed the post-1992 HEIs, which size was on average halved.

Two additional points of review are first, the number of undergraduate students obtaining a degree in post-HEIs, was two times the number of undergraduate students graduating from pre-1992 HEIs, and, second, the almost equal level of administrative expenditures in both pre-and post-1992 HEIs, despite their fundamental differences. The inputs and outputs produced as a total of the production activities held in each type reveal a huge gap between pre-1992 and the former colleges of HE in terms of both inputs utilisation and output generation. Most of the uneven descriptive characteristics observed on the different types can be attributed mainly in the special characteristics and activities each HEI develops.

However what remains still vague are the descriptive dynamics of the merged HEIs compared to those non-merged. Therefore, in table 3 some descriptive information is offered so as to inform the reader on how inputs and outputs have deployed through the years for those taking the decision to merge (before and after the merger) and those kept the same status.

Table 3: Descriptive statistics by status of HEI from 1996-97 to 2012-13

Descriptive Statistics by Status from 1996/1997 to 2012/2013						
Number of Observations=2,197						
Status of University	**Merged HEIs (n= 268)**		**Non-merged HEIs (n=1651)**		**Pre-merged HEIs (n=278)**	
Variable(s)	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
OUTPUTS						
PGOUTPUT	2,170.84	1,423.44	1,068.75	915.41	588.33	711.88
UGOUTPUT	4,095.53	2,068.75	2,450.47	1,944.93	1,832.43	2,042.18
RESEARCH	74,025.72	95,931.38	16,451.78	36,116.48	12,244.45	29,844.63
INPUTS						
PGINPUT	3,468.78	2,258.28	1,752.43	1,398.23	1,119.95	1,352.95
UGIPNPUT	10,896.48	5,257.72	6,529.14	5,030.05	4,880.04	5,026.16
STAFF	1,733.78	1,313.77	738.65	664.35	621.64	929.63
ACSERV	20,332.77	13,596.05	9,643.86	8899.30	6,103.52	7,402.85
ADMIN	36,637.36	23,696.54	17,335.05	13,556.43	10,875.62	11,023.73

The first point of interest in table 3 is the average supremacy on the values of both inputs and outputs used and produced on a typical merged university (column 1) compared to those being subject to merger (column 5). This is an obvious fact in the produced outputs, where the difference in the number of graduates (both PGs and UGs) on merged HEIs is quadruple the number of those graduated from institutions imminent to merger. The research output is the one with the widest difference since it is six fold (74,025.72) the amount of money received in a typical university prone to merger (12,244.45). This is an indication of a more active ‘turnover’ system following a merger.

However the high standard deviation means-literally-that the average variation around the mean is large (95,931.38). Since there is no objective standard for small and large standard deviations, we can only judge whether the average deviation is small or large depending on context so according to the min and max values observed this difference might be a matter of the fact of extreme values rather than a broadly difference. Therefore, we still do not know how and if those higher rates affect subsequent organizational performance, nor do we fully are certain what is responsible for the variance.

Though what has to be mentioned here is the same supremacy in terms of inputs utilization and output production on the merged HEIs relative to the non-merged units. Specifically, the inputs usage is showed to be in a triple largest scale of operation with the merged units to employ on overage more than 1,700 salaried employees. This is an interesting fact against public perceptions which traditionally associate mergers and acquisitions with employment losses. Also some interesting features of discussion are raised regarding the number of different aspects can affect the scale of job losses besides mergers.

The inputs snapshots regarding academic and administrative expenditures on merged units remain higher indicating an even more prevalent role of administration and academic services in the post-merger period. This is a point of great interest since the underlay mechanism of cost savings can arise from economies of scale and scope. When universities merge, administrative costs can be spread over increased outputs, buildings or sites can be shed leading to lower maintenance and capital costs and small duplicate programmes across separate institutions can be eliminated, allowing staff costs to be spread over more students (Johnes, 2016). So an increase in administrative costs might lead to greater academic efficiency if it frees up staff time (Hogan, 2011). However, despite the interesting points raised by the descriptive analysis additional research should aim to understand both the origins and consequences following HEIs mergers.

4.5 Methodological Approach

The analysis was performed into two stages. In the first stage, efficiency scores for the pooled model were derived, by applying the non-parametric technique, of DEA. Consequently, an inter-temporal efficiency frontier covering all units in all years was estimated; this means a single DEA frontier (pooled-DEA) for estimating technical efficiency. In the second stage, variation in HE efficiency is explained through a multivariate regression analysis model with 'merger' being one of the independent variables.

4.5.1 First Stage

The first aspect when discussing DEA is the fundamental assumption of the absence of a causal relation between coincident input–output levels in DEA models developed to date. The single period assessment (static DEA model) fails in the presence of inter-temporal input–output dependencies; therefore, Emrouznejad and Thanassoulis (2005) developed a dynamic DEA model for the HE sector. In this transformative DEA model, the problem of the lack of correspondence between coincident input–output levels is overcome by establishing an inter-temporal relationship between inputs and outputs in

HE. Consequently, the authors tried to assess the dependencies within the path of input–output levels generated by a DMU. Despite the signals of volatilities in the static model in the presence of inter-temporal effects, here, due to limitations in the appropriate length of the assessment window and on its subdivision into periods in order to gain an accurate measure of performance, the chosen model is the static one.

A second aspect pertaining to the specification of the model is the overtime efficiency and how productivity and efficiency changes can be captured inter-temporally. Within the 17 years of examination, it might be possible for HEIs to undergo some productivity changes, either due to efficiency changes that lead to movements in relation to the frontier or due to technology improvements, which lead to a shift in the shaped frontier. The DEA model is estimated initially using a common frontier (pooled-DEA) and, in a later stage by taking a within-year approach, so technological change can be feasible in this case. Attention should be paid to the within-year approach since the dramatic reduction in the sample size may provide inaccurate and overestimated efficiency measures (Alirezaee et al., 1998). According to the findings of previous studies by (Flegg et al., 2004 and Johnes, 2008) the productivity improvements in English HE can be attributed to technological advances, rather than technical efficiency, but even these advances in technology cannot account for dramatic or fundamental changes through the years.

To sum up, in this first stage, a DEA model utilised to assess the TE of English universities. An underlying assumption of production analysis is that technology is constant for the period over which the production relationship is being estimated. Over a long period of time, this assumption is questionable. We, therefore, address the issue of time as follows. We estimate the DEA model assuming a common production frontier over time (i.e. common technology throughout the study period).[124] In the second stage, we include time dummies (see next section) to allow for efficiency differences caused by technological change over time.

We use a CRS DEA model as the resulting efficiency score incorporates inefficiencies due to both size of operations and managerial competence, and efficiency benefits from mergers are likely to arise from either of these. In this stage, DEA provides a useful insight in terms of the TE of the university k at time t, so, according to the initial notion expressed by Charnes, (1978).

$$TE_{kt} = \frac{\sum_{m=1}^{M} a_{mt}\, q_{mkt}}{\sum_{n=1}^{N} b_{nt} x_{nkt}}$$

Where q_{mkt} denotes output m $(m = 1, \dots, M)$ produced by HEI k at time t $(t = 1, \dots, T)$; and x_{nkt} denotes input n $(n = 1, \dots, N)$ used by HEI k in time t. The

[124] We could assume from the outset that the production frontier changes over time and conduct the DEA within-year, thereby allowing for technology change over time. The *caveat* of this approach is that the smaller sample size can bias the efficiency scores upwards compared to the pooled estimation approach. Johnes (2014) finds that the pooled-DEA is preferable for within-year estimation because the resulting efficiencies are closely correlated with those derived from alternative parametric methods of estimation.

parameters a_{mt} and b_{nt} are the weights applied to output m in time t and the weight applied to input n in time t, respectively. The weights can be calculated for each HEI (DMU) by maximising efficiency subject into two restrictions: a) the weights must be non-zero; b) the weights must be universal (for a more detailed presentation of the DEA methodology, see Coelli et al., (2005)). To make the efficiency ratios sensitive to the input and output mix, we would have to weigh the inputs and outputs by their relative values. Only after both inputs and outputs are weighted by relative values and costs, does the ratio reflect one DMU as more efficient. The relative weights[125] needed to value inputs (and outputs) are often not available. This is a fact in most of the service organizations, included HE. In the absence of those weights, ratio analysis may be only marginally helpful and possibly misleading in multiple-output, multiple-input applications (Cooper et al., 2011). This inability to identify reliable relative weights for different inputs and outputs limits the ability to use operating ratios to gain insights into ways to manage and improve performance. Therefore, DEA overcome this limitation since it has the ability to access relative performance by incorporating multiple inputs and outputs in their natural units when such weights are not accessible (Sherman and Zue, 2006). This attribute makes DEA uniquely suited for evaluating many service organizations and providers and particularly effective in different service environments.

DEA is performed in the context of both CRS and VRS assumption of scale. The university efficiency is assessed in terms of a unit efficiency measure, so the universities located at the frontier are considered efficient in time t and ranked with a unit efficiency score. Consequently, for efficient DMUs, TE will be equalised to unit:

$$TE_{kt} = D_{kt}(x, q) = 1$$

Although in the literature several conceptual frameworks are proposed in an effort to explain the dynamics of efficiency evaluation, focus on DEA. The vast majority of studies apply non-parametric techniques due to the fact that DEA is the most well-established non-parametric method in the field and does not have any prerequisite assumptions regarding the functional form. In this first step, DEA efficiency estimates are produced with both constant and variable returns of scale, deploying an output distance function. Both pooled and within-year DEA estimates are generated, plus bootstrapped estimates. Bootstrapping reduces the degree of uncertainty since DEA efficiencies get rid of the bias and at the same time, confidence intervals are estimated for them, recognising that our data is subject to random noise. The underlying logic is based on repeated sampling from the observed DEA efficiencies, so an empirical sampling distribution[126] for the DEA efficiencies of units is constructed, which is used to estimate confidence intervals on the DEA efficiencies.

[125] Further discussion on weights restrictions and value judgements in DEA is available by Allen et al. (1997).

[126] The concept behind this method is that the bootstrap samples are to the original sample what the original sample is to the population (Bissofi et al., 2017).

4.5.2 Second Stage

Let us denote the efficiency score of HEI k in time t estimated in the first stage by TE_{kt}. In the second stage, we are interested in the following relationship:

$$TE_{kt} = f(z_{1t}, z_{2t}, \dots, z_{jt})$$

Where $z_{1kt}, z_{2kt}, \dots, z_{jkt}$ represent a set of explanatory variables that might affect the efficiency with which an HEI can convert its inputs into outputs. Here a two-stage approach is followed where the non-parametric estimates of productive efficiency are regressed on environmental variables in a second-stage procedure to account for exogenous factors that might affect HEIs' performance. The advantages and disadvantages of the two-stage model have been well discussed earlier in chapter 3 section 3.6. Also a single bootstrap procedure is followed as proposed earlier by Simar and Wilson (2007) so as to improve valid inference, and improve statistical efficiency in the second-stage regression.

The data forms a panel comprising English HEIs in each year of the study period and this has the advantage that we can use in the second stage dedicated panel data[127] estimation methods that correct for unobserved heterogeneity across institutions (such as quality). This analysis employs a random effects estimation method for two reasons: the number of parameters to be estimated is far less than for a fixed effects model, thus preserving both degrees of freedom and information; and it yields estimates of all coefficients, including those of time-invariant explanatory variables. The latter is a particularly important point given that we are interested in the effects of characteristics such as merger and university types, which would not be estimated using fixed effects. The model[128] estimated is specified as:

$$TE_{kt} = \alpha + z_{kt}\beta + v_k + \varepsilon_{kt}$$

[127] Panel data is repeated measurements at different points in time on the same individual unit, such as person, firm, state, country or university as in our case. Panel data captures both variation over units, similar to regression on cross-section data, and variation over time, since they have both cross-sectional and time–series dimensions (Cameron and Trivedi, 2010). Normally, panel data includes K units observed at T regular time periods. So there are two instances of balanced and unbalanced panel data, where in the former case all units are observed in all time periods $T_k = T$, for all k and in the latter units are not observed in all time periods $T_k \neq T$. In this study the dataset is an unbalanced panel due to a number of mergers, new establishments or complete closures occurring in these 17 years.

[128] There are three types of panel data model: the pooled model, the fixed effects, model, and the random effect model. The pooled model specifies fixed coefficients (usual assumption for a cross section model).This model uses ordinary least square (OLS) regression analysis ignoring the fact that the data are a panel; hence it will be avoided, since it is the most restrictive panel data model in the literature. The fixed and random effects models make different assumptions for the unobserved heterogeneity across individuals v_k (Cameron and Trivedi, 2010). The main difference between the two models is whether the individual-specific effect v_k is correlated with the regressors or not. If this is the case we have a fixed effects model, since it allows for a limited form of endogeneity, if they are not we have a random effects model. In the fixed effects case, the error is $u_{kt} = v_k + \varepsilon_{kt}$ and permits x_{kt} to be correlated with the time-invariant component of the error v_k while continuing to be uncorrelated with the idiosyncratic error ε_{kt}. Each individual has a different intercept term and the same slope parameters. The individual-specific-effects v_k account for the leftover variation in the dependent variable that cannot be explained by the regressors (Cameron and Trivedi, 2005). The random effects model assumes that the individual-specific-effects are distributed independently with the regressors so we include v_k in the error term. Each individual here has the same slope parameters and a composite error term $u_{kt} = v_k + \varepsilon_{kt}$. So here v_k is purely random, where a stronger assumption usually imposed is that v_k is uncorrelated also with the regressors.

$v_k + \varepsilon_{kt}$ is the residual

$v_k \sim IID(0, \sigma_v^2)$ is the time-invarying unit-specific residual (random)

$\varepsilon_{kt} \sim IID(0, \sigma_\varepsilon^2)$ is uncorrelated over time (idiosyncratic error)

z_{kt} is the matrix of k explanatory variables (not including a constant)

α is the intercept term denoting the mean of the unobserved heterogeneity

This analysis concentrates on a random effects model, since it yields estimates of all coefficients and, hence, marginal effects, even those of time invariant regre ssors. The model can be estimated by using a feasible generalised least squares estimator (FGLS).

Correlation over time for a given unit, i.e. HEI, is assumed due to the fact that each unit concentrates on the same kind of characteristics over time; however, there is independence over units (HEI). Efficiency for the same university is correlated over time, but it is independent across universities. Here, we have a narrow panel with a reasonably long time dimension since it contains many time periods (17 academic years) and many units (HEI), with more than 120 English universities. In this second stage of analysis, the variation in efficiency estimates of the first stage will be explained by using a random effects model with merger serving as an independent variable.

The effect of merger is assessed in several distinct ways. First, a simple pre-merger, post-merger and non-merging distinction is made by including two dummy variables (PREMERGER and POSTMERGER). PREMERGER takes the value of 1 for HEIs that will merge (in all time periods prior to merger), and zero otherwise; POSTMERGER takes the value of 1 for HEIs that have merged (in all time periods following merger), and zero otherwise. The comparison group is, therefore, non-merging HEIs. For reasons previously outlined, we are unsure, *a priori*, of the direction of the relationship between POSTMERGER and efficiency.

In a separate model, we investigate the possibility that efficiency effects from merger vary over time by including the pre-merger dummy (PREMERGER) combined with dummy variables to reflect the year of merger (MERGERt), and each of the four years following the merger (MERGERt+1, MERGERt+2, MERGERt+3, and MERGERt+4). Finally, we investigate the possibility that efficiency prior to merger might also differ over time and include (instead of the pre-merger dummy, PREMERGER) separate dummy variables for three, two and one year prior to merger (MERGERt-3, MERGERt-2, and MERGERt-1).

Subject differences between universities are not reflected in either teaching or research outputs. It is possible that they are accounted for (at least to some extent) by the DEA estimation method which allows each unit to be assessed relative to others with a similar input–output mix. Thus universities should not be disadvantaged by being different. In addition, the random effects panel data estimation method allows for time-invariant unobserved heterogeneity between HEIs including subject mix differences. A further exploration on the possible effect of the subject mix of universities on their estimated efficiency in several possible ways is possible.

First, we include the ratio of the number of students undertaking medicine and veterinary studies to the total number of FTE students (MEDICINE). One hypothesis is that these courses are longer and more resource-intensive; therefore an HEI with a relatively large number of this type of students might appear less efficient than others. A competing hypothesis is that these students are often the most academically able, with high entry scores, and this in turn has a positive effect on degree completion and performance. Thus, a relatively high number of students in these subjects might have a positive effect on efficiency.

Second, the composition of the student body might be expected to affect efficiency. In particular, a high proportion of overseas students might permit greater opportunities for subsidisation of research, and this in turn might lead to greater measured efficiency in our model than would otherwise be the case. On the other hand, overseas students can require greater resources to mentor through the English HE system, and could therefore be negatively related to efficiency.

Third, we include HEI type dummies[129] to represent pre-1992 HEIs (PRE1992) and post-1992 HEIs (POST1992), which are measured relative to the base group of former colleges of HE. These are intended to reflect differences in mission both in terms of outputs produced (research, teaching or third mission) and/or in terms of subject mix. In some models we split the pre-1992 group of universities into Russell Group HEIs[130] (RUSSELL) and other pre-1992 HEIs (OTHERPRE1992), as the former have a strong research mission.

The precise effect of HEI type on efficiency is difficult to predict *a priori*. Pre-1992 (and especially Russell Group) HEIs might be involved in more resource-intensive activities and, hence, their efficiencies might appear low. On the other hand, these are the universities that are likely to have the highest quality inputs (not taken into account in the DEA model) and, hence, the greatest success at transforming inputs into outputs, thus, Russell group universities appear as more efficient.

Funding sources have been found to be important in determining university efficiency and productivity (see, for example, Bolli and Somogyi (2011)). We include here the proportion of income from the government in the form of funding body grants (GOVT) in order to check whether or not source of funding affects efficiency. Previous research suggests that this might have a positive effect on efficiency[131] (Sav 2012a,2012b; 2013).

[129] These are time-invariant regressors since they are unit (HEI) specific regressors that do not change over time

[130] The current Russell Group universities can be found here: http://russellgroup.ac.uk/about/our-universities/.

[131] Evidence from the USA suggests that reliance on tuition revenue has a negative effect on operating efficiency (Sav 2013) while government funding has a positive effect on both operating efficiency and cost efficiency (Sav 2012; 2013) in public universities, but a negative effect on cost efficiency in private universities (Sav 2012). However the funding systems for higher education in the USA and England are different; mean proportion of income derived from government sources in the USA is 0.3 (Sav 2012; 2013). It is therefore difficult to predict the direction of the effect will be similar in English higher education. The increased competitive pressures caused by receiving a lower share of income from government sources (and hence a greater share from student fees, for example) might have a positive effect on efficiency.

Finally, we include in the main analysis time dummies to allow for shifts in the frontier over time. Previous studies have found increases in productivity in English HE (Flegg et al., 2004; Johnes, 2008; 2014); positive effect on efficiency in the USA (Sav 2012b, 2013), but negative in a European context (Wolszczak-Derlacz and Parteka 2011)

Finally, we include the total number of FTE students (and its square) to examine whether efficiency is related to the size of the university. The square is included to assess whether the relationship is non-linear. So, the size variable is divided by 1,000 to find the coefficient that looks more sensible. This means that the size squared is divided by 1,000,000

Variables that might also affect the efficiency of HEIs include the involvement of the institution in overseas campuses and the production of third mission output. Unfortunately, data for the construction of possible measures of these is only available from 2008-09 onwards. Therefore, we end the empirical work with a preliminary analysis of the possible effects of these variables on a restricted data set from 2008-09 to 2012-13, which precludes the inclusion of merger dummies (since few happened during this period). We construct for each HEI the ratio of students studying in overseas campuses to the total number of students in UK institution (OVERSEAS).[132] We hypothesise that universities with considerable involvement overseas might have lower efficiency, all else being equal. We test two possible measures of third mission activity:[133] the first is the proportion of total income from continuing professional development (CPD) courses and continuing education (CE) (INCOMECPDCE), and the second is the proportion of total income from technology transfer and innovation (INCOMETT).[134] We might expect that universities with a strong third mission component will appear less efficient (*ceteris paribus*) in the first stage because the DEA model does not take into account their third mission contribution.

Full definitions of all the variables in the second stage analysis are provided in appendix 17 chapter 4. Descriptive statistics for the second stage variables can be found in table 3. These confirms the diversity observed in the English HE sector (see Daraio *et al.* (2011) for more on the diverse nature of the UK HE sector). For example, while the typical HEI has nearly 9,000 students, this varies from under 300 to nearly 36,000.

[132] Due to data limitations the number of students studying overseas is not available for the whole examination period. Therefore, their effect will be assessed in a nested period of time including years after 2007-08.

[133] These activities are the main vehicle for the generation, use, application and exploitation of knowledge, generated in the university framework. It also includes other university capabilities outside academic environments, and distinct from core activities of teaching and research (HESA, 2014).

[134] See appendix 17 chapter 4 for more detail.

Table 4: Descriptive statistics of the explanatory variables of the second stage model

Descriptive Statistics from 1996/1997 to 2012/2013						
Variable	Mean	Std. Deviation	Minimum	Maximum	Cases	Missing values
INCOMETT	0.049	0.048	0.0	0.295	629	1568
INCOMECPDCE	0.024	0.034	0.0	0.332	629	1568
GOVT	0.413	0.130	0.0	0.842	2197	0
SIZE	8.732	6.661	0.026	35.86	2197	0
MEDICINE	0.053	0.192	0.0	1.984	2197	0
OVERSEAS	0.193	1.36	0.0	21.10	756	1441

Note: See appendix 17 chapter 4 for definitions of variables

4.6 Results

4.6.1 First Stage Results

As mentioned earlier, in the first stage of this study, efficiency estimates for more than 130 English public universities were calculated by applying a non-parametric estimation technique (DEA) assuming an output distance function in a multi-input multi-output space. Increased variation was discovered around the parametric estimates produced in this stage, not only over the years but also between the different assumptions of the DEA model (e.g. VRS or CRS, pooled-DEA versus within-year DEA).

Table 5: Pooled- and within-year average efficiency of English HEIs performed by DEA

Mean Efficiencies of English Universities						
YEAR	Pooled-DEA CRS	Within-year DEA CRS	Pooled-DEA VRS	Within-year DEA VRS	Bootstrapped DEA CRS	Bootstrapped DEA VRS
1996-1997	0.610	0.907	0.764	0.945	0.624	0.736
1997-1998	0.609	0.885	0.762	0.924	0.614	0.737
1998-1999	0.600	0.909	0.752	0.944	0.607	0.721
1999-2000	0.599	0.884	0.744	0.929	0.614	0.720

2000-2001	0.630	0.836	0.772	0.903	0.645	0.742
2001-2002	0.640	0.786	0.776	0.926	0.660	0.757
2002-2003	0.660	0.804	0.789	0.913	0.683	0.770
2003-2004	0.678	0.796	0.801	0.892	0.703	0.779
2004-2005	0.682	0.780	0.806	0.897	0.705	0.777
2005-2006	0.685	0.831	0.818	0.912	0.706	0.790
2006-2007	0.682	0.807	0.813	0.910	0.702	0.790
2007-2008	0.679	0.806	0.809	0.880	0.697	0.776
2008-2009	0.663	0.832	0.793	0.903	0.686	0.758
2009-2010	0.677	0.839	0.805	0.913	0.693	0.762
2010-2011	0.702	0.866	0.833	0.911	0.726	0.790
2011-2012	0.716	0.863	0.838	0.924	0.756	0.805
2012-2013	0.693	0.876	0.812	0.927	0.709	0.777

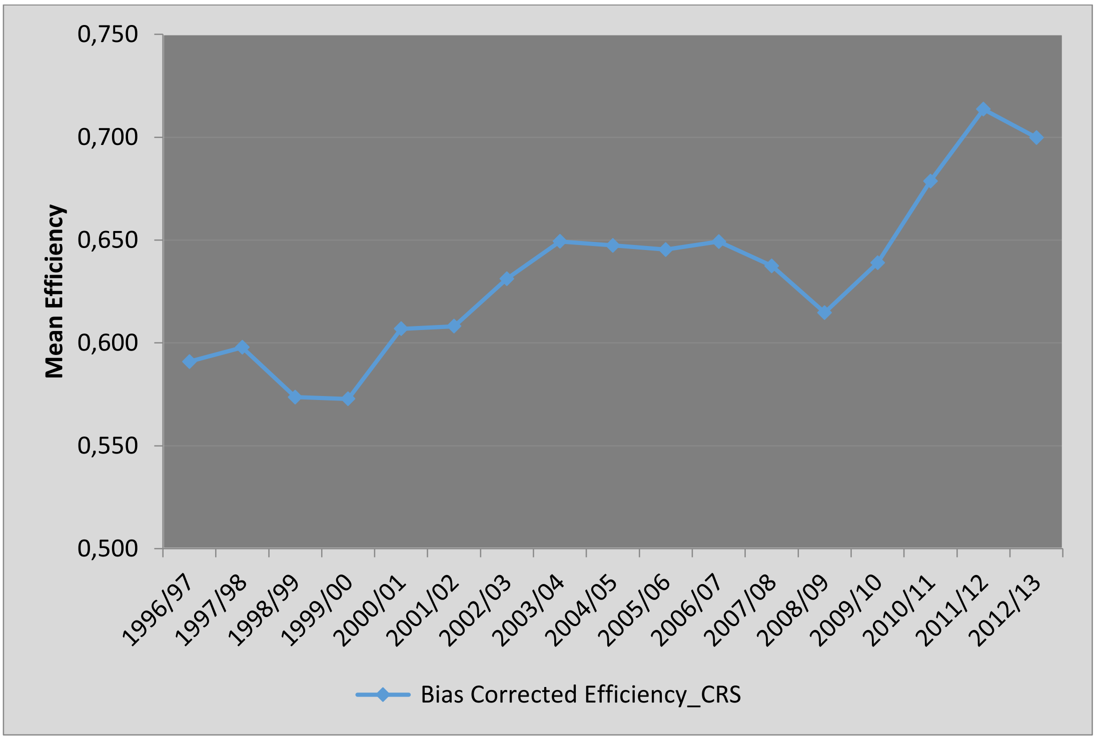
Overall, there has been an increased tendency towards efficiency improvements over the past 17 years, as revealed in the pooled-DEA estimates either under CRS or VRS, with the latter being indeed higher than the former. This is not the case for the within-year estimates, which, despite being higher than the pooled-DEA estimates, tend to vary significant year on year (see Table 5). This random pattern in within-year mean efficiency over time clarifies further why there is no obvious trend. This phenomenon may be attributed to the inevitable reduction in sample size, since, in the within-year estimation a new frontier is formed for each year of examination (technological change over time); hence, the estimates may be overestimated (Alirezaee et al., 1998).

Further scrutiny finds that pooled-DEA results between the first academic year in the sample (1996-97) and the last (2012-13) demonstrate obvious alleviation in the mean efficiency, since this rose from 0.6 to almost 0.7 under CRS and from 0.7 to nearly 0.8 under VRS. Since the VRS analysis is more flexible and envelops the data in a tighter way than the CRS analysis, we usually have VRS efficiency measures equal or greater than the CRS measures (Salerno, 2003).

The explanatory and interpretative nature of the pooled-DEA model implies that the best performance is now measured as the 'all-time' best performance (Rai, 2013). Thus, the pooled data offers an effective way for the best performance to emerge over a period of time. This allows the possibility that in a given year, no institution may have reached the efficient frontier, while the within-year methodology might control better for time period effects due to the isolation of the sample in each particular year. From a methodological point of view, pooled-DEA allows for more observations for DEA and, hence, a smoother envelopment. According to (Johnes, 2014) the pooled-DEA performed better due to the fact that it has a closely-related ranking with other methods widely applied such as the parametric methods; this, in turn, leads to better benchmarking. On average, the pooled-DEA mean efficiency for English HEIs over the examination period reached 0.658 under CRS and 0.792 under VRS scale assumption.

Over all HEIs and across all years, the mean overall TE (CRS) was at its lowest in 1999-2000 at 0.58 and peaked in 2011-12 at 0.71 (see Figure 3).[135] On average, therefore, there appears to be scope for some efficiency improvement in the sector. These results are a little lower than previous findings for the English HE sector, which also used DEA (Johnes, 2014); here, we use a different input output specification and, more importantly, cover a much longer time period. The strong upward trend in efficiency since 2008-09 (with the exception of the final year) is especially noteworthy and might reflect a response to the austerity measures following the global financial crisis; the drop in the final year might indicate that continuing to reap such efficiency gains is unsustainable.

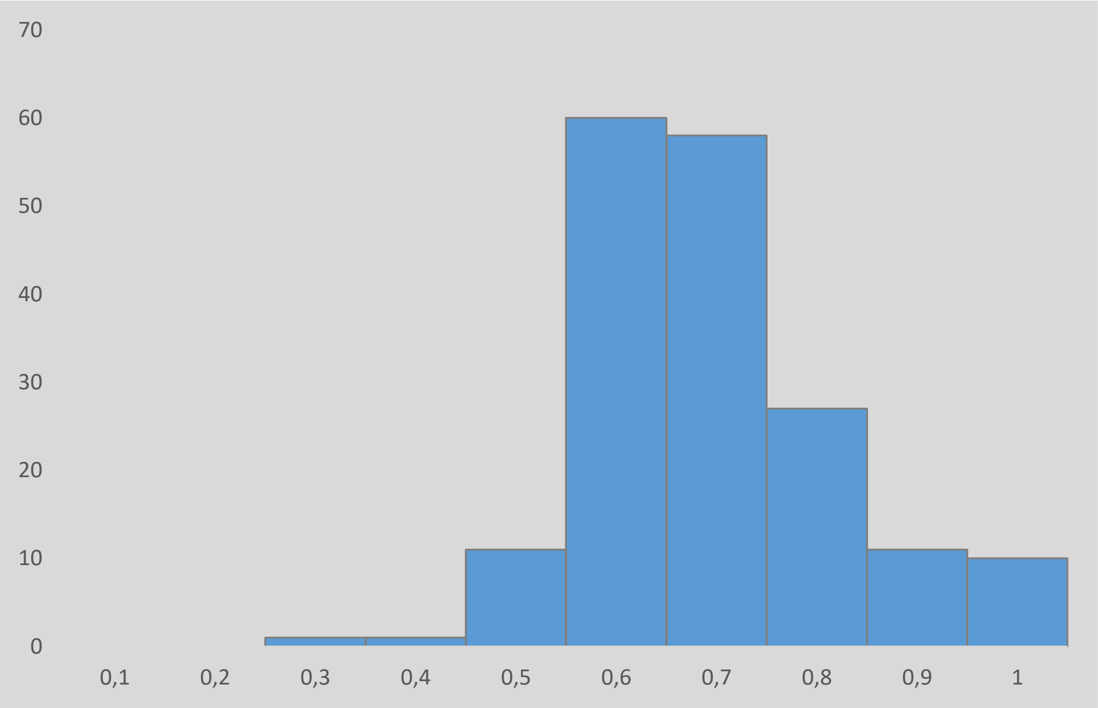
Figure 3: Mean DEA bias-corrected efficiency over time



If we calculate the mean efficiency over time for each HEI we find that, at the top end, there are two HEIs that are fully efficient with respect to overall TE over the period (see Figure 4). The worst-performing HEI, in contrast, has a mean overall TE of 0.214, and there are 13 HEIs with a mean overall TE below 0.5. At the bottom end, there are potentially large savings in efficiency to be made.

[135] Results are based on bias-corrected DEA efficiencies derived using bootstrapped estimation.

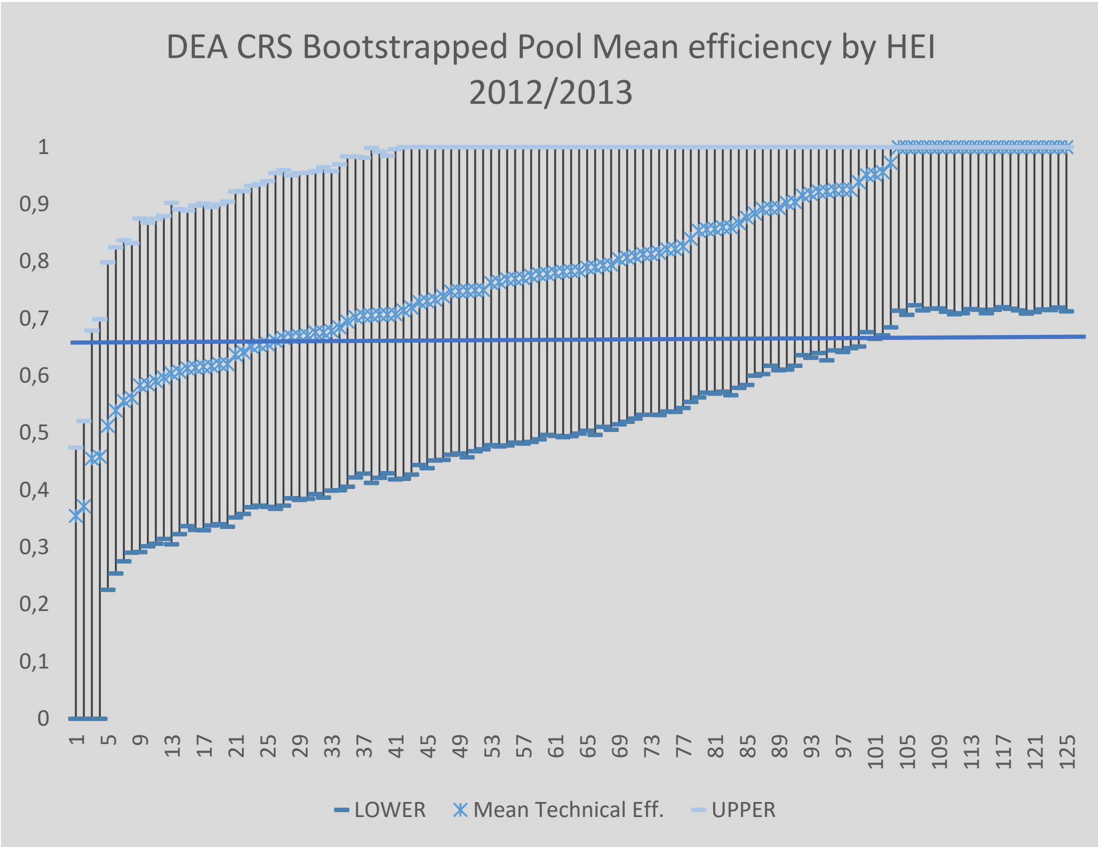
Figure 4: Histogram of mean HEI efficiencies (Bootstrapped DEA results) – overall technical efficiency



From a managerial and policy point of view we need to know whether these differences between HEIs represent real (or significant) differences. The bootstrapping estimation method allows us to calculate for each efficiency score a 95 percent confidence interval. This represents an interval for each HEI within which we are 95 percent confident that the true efficiency lies. We can plot the efficiency scores and accompanying confidence interval (see Figure 5 which plots the pure TE score and accompanying confidence interval for each HEI in 2012/13). The line of mean efficiency for that year (0.69) is superimposed onto the plot. It is clear that the intervals of all but four HEIs at the lower end of performance overlap the mean efficiency. This suggests that, on the whole, differences between HEIs in estimated efficiency are not significant: HEIs with the highest apparent efficiency scores are not significantly different from those with lower efficiency scores, with the exception of the lowest performing HEIs.

Previous findings for the English and UK HE sectors have also suggest considerable overlap in performance across HEIs, although there are apparent distinctions between both the best- and worst-performing HEIs (Johnes, 2014). It is worth noting that HEIs with apparently low levels of efficiency are characterised by being small and specialist and this result is similar to the findings derived from Johnes and Johnes's (2013) cost function study.

Figure 5: Mean pure technical efficiency score and associated 95 per cent confidence interval by HEI



Note: The mean here is 0.69, pool DEA CRS BOOT; only the 2012/2013 results have been retained

4.6.1.1 Efficiencies by Status of Institution

The main point this analysis aims to uncover is whether there are any significant differences in the mean values of technical efficiency obtained in the first stage for the universities taking the decision to merge. In this stage preliminary analysis to compare the mean values of efficiencies between the merged and non-merged universities (groups) is used. A one way ANOVA (Analysis of Variance) is used to compare two means from two independent (unrelated) groups using the F-distribution. The null hypothesis for the test is that the two means are equal. Therefore, a significant result means that the two means are unequal. The null hypothesis for identical MTE values derived by CRS bootstrapped estimates cannot be rejected between merged and non-merged DMUs since the p-value is (0.148) for 5 per cent significance level. However this is not the case when the test is performed with VRS bootstrapped estimates where the null is rejected with a p-value close to (0.000).

Meanwhile, a one way ANOVA inform the researcher that at least two groups were different from each other. But cannot be enlightening in terms of what groups were different. Since our test returns a significant $f-statistic$, we need to run an ad hoc test (like the Bonferroni Correction test, also known as Bonferroni type adjustment) so as to know exactly which groups had a difference in means. When conducting multiple analyses on the same dependent variable (MTE here), the chance of committing a Type I error increases, thus increasing the likelihood of coming about a significant result by pure chance. To correct for this, or protect from Type I error, a Bonferroni correction is conducted. In table 6 below, we can trace those differences by comparing the MTE of the different groups by two.

Table 6: Multiple comparisons Bonferroni correction

Dependent Variable: Bias_Vrs_ corrected Efficiency						
(I) merger	(J) merger	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Non-merged HEIs	Pre-merger HEIs	-0.004	0.012	1,000	-0.034	0.249
	merged HEIs	-0.106*	0.012	0.000	-0.136	-0.076
Pre-merger HEIs	Non-merged HEIs	0.004	0.012	1.000	-0.024	0.034
	merged HEIs	-0.101*	0.016	0.000	-0.141	-0.062
Merged HEIs	Non-merged HEIs	0.106*	0.012	0.000	0.076	0.136
	Pre-merger HEIs	0.101*	0.016	0.000	0.062	0.141

*The mean difference is significant at the 0.05 level.

When we access MTE between merged and non-merged DMUs this difference is significant at the 10 percent, 5 percent, and 1 percent significance levels respectively. This is the case for MTE differences between merged and pre-merger units as well. So there is an articulate hint of detected differences in the MTE values that should be further explored with more advanced and robust methods in later stages.

4.6.1.2 Efficiencies by Type of Institution

In the next section, the mean efficiency is calculated by taking into account the three different types of HEI in England (see Table 7). Overall, post-1992 HEIs achieve the highest efficiency scores regardless of the scale assumptions (CRS versus VRS). The least efficient type, on average, of the three explored, are the pre-1992 HEIs. Despite their large size, and the high levels of input and output volume, compared to their two counterparts, the efficiency estimates reveal that the pre- 1992 HEIs are deprived of an efficient allocation of their disposable resources. In the meantime, post-1992 HEIs stand out in terms of their high-ranked performance, since they obtain higher efficiency scores compared to the other two types, under both CRS and VRS.

The institutions involved in the current formulation of the English HE sector display fundamental differences, since they have diverse backgrounds and traditions. Consequently, this discrepancy in their common characteristics is reflected not only in their constitutional arrangements but also in the structure and powers of their governing bodies (Johnes, 2014).

Table 7: Mean DEA efficiency by HEI type

Mean Efficiencies of English Universities by Type		
Type of University	DEA (CRS)	DEA (VRS)
Post 1992 HEIs n= 527	0.805	0.896
Pre 1992 HEIs n=898	0.735	0.862
Colleges of Higher Education n=772	0.766	0.826

These findings are not verified by previous research by Johnes (2008), who suggests that TE is the highest among the former colleges of HE, followed by the post- and pre-1992 HEIs, however, they are in accordance with a later study by Johnes (2014), in which the DEA results generally suggest that both former colleges of HE and post-1992 HEIs are, on average, more efficient than pre-1992 institutions. A possible explanation for the poor performance of the pre-1992 HEIs is given by Johnes (2014). This fall in performance is attributed in the subject mix and the different kinds of operation of these institutions, which are more resource-intensive (e.g. teaching in medical and veterinary sciences).

4.6.1.3 Efficiencies by Time Cluster

In this section, the analysis is conducted in three different sub-periods (clusters). The first starts at the beginning of the examination period in 1996-97 and lasts until 2001-02. The second continues from 2002-03 to 2008-09, when the financial crisis broke. The last cluster reported here covers the period from 2009-10 to 2012-13. Here, three unique production technologies are assumed in each sub-period, making the attempt to skip out any overestimation or underestimation of the results that may be caused by technological shifts during those years. The results are summarised in table 8 and, on average, are consistent with the bootstrapped results from the pooled-DEA model, since there is only a minor difference on the MTE, which is slightly higher (0.806) in the cluster analysis case compared to the pooled-DEA analysis (0.769).

Table 8: Mean efficiency by time clusters

	Analysis by Time Clusters **Bootstrapped Results**				
	Mean Efficiency VRS	Std. Deviation	Mean Efficiency CRS	Std. Deviation	Number of Observations
First sub-period 1996-97-2001-02	0.845	0.122	0.693	0.154	794
Second sub-period 2002-03-2007-08	0.815	0.144	0.709	0.152	774
Third sub-period 2008-09-2012-13	0.846	0.126	0.784	0.133	629

A correlation matrix between the original bootstrapped efficiencies from the pool data and those produced in the cluster analysis are estimated. High correlation equal to 0.783 is identified in this stage between the two efficiency scores, confirming a coherence in the applied methods and signifying that technological advances have not the delivered results significantly.

4.6.1.4 Efficiencies in Russell Group Universities

The UK enjoys one of the most outstanding HE systems in the world. The quality of this high-performing sector is the determinant factor in the leading research and teaching excellence of most of the universities, which plays a chief role in the country's economic and educational competitiveness. In terms of the English HE synthesis, Russell Group universities are those that have invested more in facilities, staff, and support services some of the essential components of providing an outstanding teaching and learning experience (FHE, 2010). The status of these universities is testament to their quality and efficiency. Hence, they are constantly exploring innovative ways to improve their productivity and efficiency to secure leading positions in the public and private investment share. It has been speculated that Russell Group institutions are extremely efficient in international terms for a number of reasons (FHE, 2010) [136] and that they have identified areas in which cost savings and efficiency gains can be made. Therefore, it could be of extreme interest whether this promising task of increased efficiency and effective use of limited resources is indeed valid. According to the table below (see Table 9), Russell Group universities tend to perform better in terms of efficiency measures than other pre-1992 universities and non-Russell Group universities, validating the consensus of their being highly efficient bodies. This disciplinary-specific category of Russel Group institutions is on average, very close to an efficient operation, and distribution of their resources reaches 0.958 on the efficiency scale.

Table 9: Mean efficiency by type of HEIs

	Bootstrapped (replications=1,000)				
	Mean Efficiency VRS	Std. Deviation	Mean Efficiency CRS	Std. Deviation	Number of Observation
Russell Group Universities	0.958	0.051	0.941	0.056	339
Other Pre-1992	0.864	0.127	0.755	0.145	554
Non-Russell Group	0.799	0.145	0.716	0.146	1304

4.6.2 Second Stage Results

4.6.2.1 Simple Model Specification

[136] Three percent of global R&D investment stems from UK institutions, the UK publishes 14.4 percent of the world's most highly-cited publications, and it is the most efficient country in the G8 in terms of the ratio of citations to public funding for research.

The results obtained from the preliminary statistical analysis based on a random effects model are shown in table 10. This model presents the analysis of variations in mean technical efficiency (MTE) depending on six regressors. Two measures of TE depending on different scale assumptions (CRS vs VRS) obtained from the pooled-DEA model derived in the first stage are used in the analysis. In both cases, the interaction variable takes significant values, indicating that the MTE of the post-merger HEIs is greater in the post-merger period. The top half of the table shows some of the main characteristics of the breakdown of efficiency performance among the different types of HEI. On average, and in both model specifications, the post-1992 HEIs tend to be more efficient than the colleges of HE. If we now turn to the pre-1992 HEIs, the table reveals contradictory results depending on the scale assumptions. Under the VRS framework, the pre-1992 HEIs tend to perform better than colleges of HE. Unlike in the CRS case, the pre-1992 universities have no significant effect on the mean efficiency score. The intriguing element revealed in the table is that the MTE of the post-1992 HEIs outnumbers the pre-1992 HEIs. Finally, the size effect, although significant, has a minor negative effect in the MTE score.

In this simplest specification of the second stage model pre-and post-merging institutions are compared with non-merging HEIs. According to table 10 pre-merging HEIs do not differ significantly from non-merging HEIs in terms of efficiency (all else being equal); but that post-merging HEIs have significantly higher efficiency than non-merging HEIs by 0.052 points (or 5 percentage points if we consider efficiency in percentage terms), all else being equal. Merging therefore appears to have positive efficiency effects, even when all other factors underlying efficiency have been taken into account. This might be a consequence of more efficient administrative usage as Figure 2 has revealed a slowing of growth in this resource for the sector as a whole over time.

Table 10: Effects of HEIs merger on HEI efficiency

Random Effects Model	**Dependent Variable: Efficiency of HEIs**	
	Efficiency under CRS	Efficiency under VRS
Pre-1992	-0.003	0.039***
	(0.0078)	(0.0076)
Post-1992	0.096***	0.049***
	(0.0102)	(0.0099)
Pre-Merger	0.024***	0.021
	(0.0091)	(0.0087)
Post-Merger	0.052***	0.068***
	(0.0097)	(0.0094)

	-0.023***	-0.005***
Size	(0.0014)	(0.0013)
Size Square	0.0005***	0.0003***
	(0.5464D-04)	(0.5276D-04)
Constant	(0.765) ***	0.761
	(0.0123)	(0.0076)

*P<0.10; **P<0.05; ***P<0.01; CRS, constant returns of scale; VRS, variable returns of scale

One question that remains blurred and requires further clarification is whether the derived pooled-DEA efficiency estimations of the first stage are consistent and unbiased. However, a major problem with this kind of application is the consistency of the estimator. This is an essential property for any estimator, so the DEA efficiency estimator converges as the sample size increases, although at a slow rate. The practical use of this property is to confirm that the DEA estimator may be reasonable to use for efficiency estimation.

However, for applied research, more is needed; in particular, the applied researcher must have some knowledge of the sampling distributions in order to make inferences about the true levels of efficiency or inefficiency. Sampling distributions may also be approximated by bootstrap distributions in very general situations. The use of bootstrapping techniques allows a further transformation in the model and correction for the bias of the efficiency estimator and estimation of confidence intervals for the efficiency measures (Simar and Wilson, 2000). For the multivariate DEA case, at least so far, the bootstrap seems to offer the only approach to estimating the sampling variation of efficiency estimators. Therefore, bias-corrected DEA efficiency scores are derived in the first stage and used later in the second stage in order to gain further insight into the model.

Table 11: Effects of HEIs merger on HEIs efficiency-Bootstrapped results

Random Effects Model	**Dependent variable: Efficiency of HEIs**			
	Efficiency CRS	Efficiency VRS	Bootstrapped efficiency CRS	Bootstrapped efficiency VRS
Pre-1992	-0.003	0.039***	0.021**	0.044***
	(0.0078)	(0.0076)	(0.0106)	(0.0103)
Post-1992	0.096***	0.049***	0.129***	0.054***

	(0.0102)	(0.0099)	(0.0139)	(0.0134)
Pre-Merger	0.024***	0.021	0.036***	0.023**
	(0.0091)	(0.0087)	(0.0123)	(0.0117)
Post-Merger	0.052***	0.068***	0.065***	0.077***
	(0.0097)	(0.0094)	(0.0131)	(0.0127)
Size	-0.023***		-0.031***	
	(0.0014)		(0.0019)	
Size Square	0.0005***		0.0007***	
	(0.5464D-04)		(0.0000)	
Constant	(0.765) ***	0.761***	0.806***	0.727***
	(0.0123)	(0.0076)	0.0147)	(0.0091)

*P<0.10; **P<0.05; ***P<0.01; CRS, constant returns of scale; VRS, variable returns of scale

As indicated in table 11, the use of bootstrapped estimates does not cause significant discrepancy in the results. These findings are rather encouraging since not only the post-1992 universities but also the pre-1992 universities are shown to be more efficient than the colleges of HE. The present results are significant in at least major two respects since the merged institutions are prominently more efficient compared to their non-merged and pre-merged counterparts, regardless of the scale assumption. The results further enforce the claim of efficiency improvements in the HE sector in the post-merger period.

Turning now to the size effect in the MTE score from the partial derivative with respect to size, HEIs of sizes (measured as the total number of undergraduate and postgraduate students) over 10,850 students, may reap the benefits of greater returns of scale in their production side. The mean size across the dataset is 8,732 students and this fluctuates considerable across universities and throughout the years from 26 students up to 35,865. Therefore, on average, due to the U-shape relationship between MTE and university size, universities should increase their size of operation if they seek further efficiency gains more than 10,850 students. Around 70 universities operate at this size, exceeded 10,850 students; however, the years of operation vary significantly from one year[137] up to seventeen[138] consecutive years.

[137] Teesside University, the University of North London.

[138] University of Central England in Birmingham, the University of Birmingham, the University of Bristol, Cambridge University, Coventry University, De Montfort University, University of Hertfordshire, Leeds Metropolitan University, the University of Leeds, Liverpool John Moores University, the University of Liverpool, Manchester Metropolitan University, the University of Manchester, Middlesex University, the University of Newcastle-upon-Tyne, the University of Northumbria at Newcastle, the University of Nottingham, Nottingham Trent University, Oxford University, the University of Plymouth, the University of Portsmouth, Sheffield Hallam University, the University of Sheffield, the University of Southampton, the University of Warwick, the University of the West of England, Bristol, the University of Westminster, the University of Wolverhampton.

Additionally, there are universities that continue operating at this size of operation even after merger, such as the University of Birmingham, the University of Cambridge, the University of Central Lancashire, the University of Leeds, De Montfort University, Liverpool John Moores University, the University of Manchester, University College London, and the University of Northumbria at Newcastle.

4.6.2.2 Enhanced Model Specification

In the simplest specification of the second stage model (Model 1), pre- and post-merging institutions are compared with non-merging HEIs (see Table 12). We find that pre-merging HEIs do not differ significantly from non-merging HEIs in terms of efficiency (all else being equal), but that post-merging HEIs have significantly higher efficiency than non-merging HEIs by 0.046 points (or 5 percentage points, if we consider efficiency in percentage terms), all else being equal. Merging, therefore, appears to have positive efficiency effects, even when all other factors underlying efficiency have been taken into account.

Of course, mergers take place at different points in the study period so that the simple comparison potentially conceals differential effects from merging activity over time. In Models 2 to 4, therefore, we include post-merging dummies from one to four years after merger. These results reveal that efficiency benefits from merging occur in the first two years following merger; they have disappeared by the third year. There are no significant efficiency differences between merging and non-merging HEIs in the years leading up to merger.

With regard to HEI type, post-1992 universities are more efficient than colleges of HE by around 0.07 points (Models 1 to 4); pre-1992 HEIs, however, are less efficient than colleges of HE, and this result is largely driven by the non-Russell Group institutions.

The variable MEDICINE has a surprisingly large positive effect on efficiency (taking into account all other factors), and this is consistent across all four models. It appears, therefore, that the variable may be picking up a *quality* rather than a subject mix effect. As expected from the previous literature, the larger the proportion of income derived from government sources (GOVT), the greater the efficiency. HEI size has a negative (but non-linear) effect on efficiency (all else being equal) over the size range of most of the institutions in the dataset.

The year dummies suggest increasing efficiency from 2003–04 onwards relative to the base year 1996–97, although the years 2008–09 and 2009–10 are exceptions to this pattern and this is probably due to the shock of the global financial crisis on the sector. Increasing austerity and expansion of the sector over the period from 2003–04 are, therefore, accompanied by generally greater efficiency (see Table 12). These might be caused by improving technologies, which can affect positively both teaching and research production; they may also incorporate changes arising from increasing tuition fees (to a ceiling of £3,000 in 2006–07 and £9,000 in 2012–13, although the latter is unlikely to have any effect on these results).

Table 12: Possible determinants of university efficiency

Random Effects Model	**Bootstrapped Pooled DEA CRS Efficiencies**			
	Model 1	Model 2	Model 3	Model 4
PREMERGER	0.011	0.011	0.010	
	(0.010)	(0.010)	(0.010)	
POSTTMERGER	0.047***			
	(0.011)			
MERGERt-2				0.035
				(0.023)
MERGERt-1				0.007
				(0.021)
MERGERt		0.025	0.020	0.024
		(0.029)	(0.029)	(0.029)
MERGERt+1		0.071**	0.067**	0.067**
		(0.030)	(0.030)	(0.030)
MERGERt+2		0.052*	0.048	0.051*
		(0.030)	(0.030)	(0.029)
MERGERt+3		0.044	0.040	
		(0.031)	(0.031)	
MERGERt+4		0.039	0.037	
		(0.030)	(0.029)	
MEDICINE	0.096***	0.102***	0.097***	0.096***
	(0.018)	(0.018)	(0.018)	(0. 018)
PRE1992		0.005		
		(0.010)		

POST1992	0.089*** (0.012)	0.070*** (0.011)	0.085*** (0.012)	0.084*** (0.011)
RUSSELL	0.043** (0.017)		0.046*** (0.017)	0.046*** (0.017)
OTHERPRE1992	0.007 (0.010)		0.005 (0.010)	0.004 (0.010)
SIZE	-0.013*** (0.001)	-0.011*** (0.001)	-0.012*** (0.001)	-0.012*** (0.001)
SIZESQ	0.0002*** (0.63 D-04)	0.0002*** (0.63 D-04)	0.0002*** (0.63 D-04)	0.0002*** (0.63 D-04)
GOVT	0.269*** (0.050)	0.229*** (0.050)	0.259*** (0.050)	0.261*** (0.051)
YEAR97/98	0.011 (0.018)	0.111 (0.018)	0.011 (0.018)	0.011 (0.018)
YEAR98/99	-0.002 (0.018)	-0.003 (0.018)	-0.002 (0.018)	-0.005 (0.018)
YEAR99/00	-0.001 (0.018)	-0.003 (0.018)	-0.002 (0.018)	-0.004 (0.018)
YEAR00/01	0.033* (0.018)	0.032* (0.018)	0.033* (0.018)	0.0318 (0.018)
YEAR01/02	0.036 (0.018)	0.033* (0.018)	0.035* (0.018)	0.033* (0.018)
YEAR02/03	0.060*** (0.018)	0.057*** (0.018)	0.060*** (0.018)	0.060*** (0.018)
YEAR03/04	0.084*** (0.018)	0.081*** (0.018)	0.085*** (0.018)	0.085*** (0.018)
YEAR04/05	0.083*** (0.018)	0.080*** (0.018)	0.084*** (0.018)	0.084*** (0.018)

YEAR05/06	0.082*** (0.018)	0.080*** (0.019)	0.084*** (0.019)	0.083***	(0.018)
YEAR06/07	0.067*** (0.018)	0.067*** (0.018)	0.070*** (0.018)	0.068***	(0.018)
YEAR07/08	0.051*** (0.018)	0.052*** (0.018)	0.054*** (0.018)	0.054***	(0.018)
YEAR08/09	0.026 (0.019)	0.028 (0.018)	0.030 (0.019)	0.029	(0.018)
YEAR09/10	0.050*** (0.019)	0.052*** (0.018)	0.054*** (0.019)	0 .053***	(0.019)
YEAR10/11	0.089*** (0.019)	0.091*** (0.019)	0.093** (0.019)	0 .093***	(0.019)
YEAR11/12	0.121*** (0.019)	0.125*** (0.019)	0.127*** (0.019)	0.125***	(0.019)
YEAR12/13	0.085*** (0.020)	0.093*** (0.020)	0.091*** (0.020)	0.089***	(0.020)
CONSTANT	0.581*** (0.016)	0.584*** (0.017)	0.582*** (0.016)	0.584***	(0.016)

Note: Standard errors are shown in brackets. *, **, and *** signal that the coefficient is significantly different from zero at the 10 percent, 5 percent, and 1 percent significance levels, respectively.

We end this section with an exploratory analysis of the effects on efficiency of involvement in overseas campuses and third mission activities. This analysis can only be undertaken for the years since 2008–09 because of data availability and does not include any merger information. The results are presented in the next table (Table 13). These preliminary results suggest that a university's involvement in overseas campuses has no significant effect on its efficiency as defined in the first stage model. The number of branch campuses has nearly doubled in the past four years (Matthews, 2012), but there is no apparent effect of this activity on efficiency.

Table 13: Effects of third mission activities and overseas campuses on efficiency

Random Effects Model	Bias-Corrected Efficiency CRS Model 8
GOVT	0.082 (0.079)
MEDICINE	0.139*** (0.031)
OVERSEAS	-0.087 (0.152)
RUSSELL	-0.015 (0.034)
OTHERPRE1992	-0.079*** (0.020)
POST1992	0.073*** (0.022)
INCOMECPDCE	1.20*** (0.180)
INCOMETT	0.421** (0.177)
SIZE	-0.003 (0.002)
SIZESQ	0.252 (D-04) (0.92D-04)
YEAR09/10	0.029 (0.018)
YEAR10/11	0.068*** (0.018)
YEAR11/12	0.107*** (0.018)
YEAR12/13	0.098*** (0.021)
CONSTANT	0.572*** (0.043)

The two variables measuring third mission activity have a positive significant effect on efficiency. This is a surprising result: we expected that failure to capture a substantial activity of a HEI in the DEA model would underestimate the efficiency of those HEIs that produced more in terms of third mission than others, and, hence, that the effect of the third mission variables in the second stage would be negative. It should be noted that a third possible measure of third mission activity, namely the number of attendees taking part in free and chargeable events such as public lectures, performance arts (music, dance, drama, etc.), and exhibitions (galleries, museums, etc.) has no effect on efficiency. The significantly positive relationship between business interaction and efficiency suggests that involvement in (and income generation from) business education and research collaboration has positive effects on universities' efficiency, i.e. the activity has positive spill-over effects on the universities' business. It is unlikely that all third mission activity will have the same positive effects, however.

4.6.3 Discussion

Talking from a more comparable point of view, merger circumstances may vary excessively from case to case, however the main objectives are generally summarized to: greater efficiency, greater economies of scale and scope, improved competitiveness, continuing growth and in some cases the long-term survival of vulnerable institutions (Azziz et al., 2017).

Increasing returns to scale occur if physical input per unit of output falls as output rises. Thus administrative activity can be spread over larger output equirements (Fielden and Markham 1997; Patterson 2000; Kyvik 2002; Norgård and Skodvin 2002; Green and Johnes 2009; Ripoll-Soler and De-Miguel-Molina 2014), buildings and/or sites can be shed leading to lower maintenance (Fielden and Markham 1997; Teixeira and Amaral 2007), small duplicate programmes across separate HEIs can be eliminated from all but one HEI and concentrated in the remaining university (Skodvin 1999), and so teaching staff can be spread over more students (Fielden and Markham 1997). This is the case mainly in China where systematic mergers have been mainly driven by the belief that “bigger is better” in terms of efficiency, thus, larger, globally competitive, comprehensive universities have been created (Azziz et al., 2017). Similar policy have been also followed by other countries such as Australia, South Africa, Northern Europe, the Netherlands, Canada, and the UK.

Findings on economies of scale in the UK HE context vary depending on the underlying data. Studies which focus on pre-1992 universities find evidence of significant economies of scale for the typical university (Glass, et al., 1995a, b; Johnes 1996, 1998). Later studies using data across both pre- and post-1992 HEIs find that scale economies are close to constant or decreasing for the typical university (Johnes 1997; Izadi et al. 2002; Johnes et al. 2005; Johnes, et al., 2008; Johnes and Johnes 2009), but that increasing returns are observed in smaller HEIs (Johnes and Johnes 2016).

The relationship between size and efficiency is open to debate. On the one hand, increasing returns to scale would suggest that the relationship should be positive and there is evidence to support this in the context of European HE (Wolszczak-Derlacz and Parteka 2011; Wolszczak-Derlacz 2014). However, some studies find no relationship between size and efficiency (Bonaccorsi, et al., 2006); other studies (of British universities) find mixed results with size having a positive relationship with efficiency in some years, but the converse being observed in others (Flegg et al., 2004); and there is evidence of a negative relationship between size and efficiency in Swedish HE (Daghbashyan 2011). Cost function studies also suggest constant or decreasing returns to scale in English HE (Izadi et al. 2002; Johnes and Johnes 2009; Thanassoulis et al. 2011).

In this study according to the results derived in chapter 4, HEI size has a negative (but non-linear) effect on efficiency (all else being equal) over the size range of most of the institutions in the data set. There is therefore no evidence of increasing returns to scale here, which aligns with evidence from Sweden (Daghbashyan 2011), and from cost function studies for the UK, but not with studies based on European HE (Wolszczak-Derlacz and Parteka 2011; Wolszczak-Derlacz 2014). Given that merger inevitably increases university size this result appears somewhat perverse. There is clearly a tension between being merged (which has a positive effect on efficiency), and being large (which is pulling in the opposite direction). The positive effect of merger is therefore potentially a consequence of returns to scope rather than returns to scale.

Increasing returns to scope might arise from producing teaching and research jointly, or from producing teaching (or research) across disciplines. There might be additional benefits experienced from merging, for example, expanding a HEI's academic portfolio of programmes through merger can have benefits in terms of increasing student demand because of greater diversity and variety of degree programmes (Harman 2000; Harman and Meek 2002; Kyvik 2002; Harman and Harman 2003; Teixeira and Amaral 2007); this in turn might therefore lead to diversity in the student population of the institution (Harman and Harman 2003), and improve the scope of education for those students (Aarrevaara 2007). Evidence on economies of scope in UK HE is mixed but the studies which include the widest range of institutions consistently find global diseconomies of scope for the typical university (Johnes 1997; Izadi et al. 2002; Johnes et al. 2005; Johnes, Johnes, and Thanassoulis 2008; Johnes and Johnes 2009, 2016).

Some rare quantitative analyses of the efficiency effects of mergers can be found in the context of Chinese HE where there have been more than 400 mergers since the 1990s (Cai and Yang 2015). Merging has been found to have a positive effect on efficiency and productivity in the first year after merger but not in the subsequent year (Hu and Liang 2008; Mao, et al., 2009). This is in line with the results obtained here since the positive effect of merger tails off at the first year post merger. In addition, a recent study of English HE uses a panel of data from 1996/1997 to 2008/2009 to compare the mean technical efficiency (estimated using both SFA and DEA) of merged institutions (of which there are 19 instances) with mean efficiencies of pre- and non-merging institutions Johnes (2014).

There may, of course, be reasons why merger might cause greater inefficiency. First, there may be decreasing returns to scale caused by increased resource use if greater centralisation increases bureaucracy or if the merging institutions are geographically distant (Curri 2002). Second, a reduction in the number of HEIs in the sector inevitably leads to greater sector concentration which, by lowering competitive pressures, could lead to lower technical efficiency of HEIs (De Fraja and Valbonesi 2012). Third, there may be a reduction in efficiency if there is a loss of quality in the teaching experience enjoyed by students (and academics) (Tight 2011). Fourth, consolidation can lead to a reduction in diversity and choice between institutions – indeed a merger might intend to reduce duplication of programmes across HEIs in order to increase technical efficiency. Reducing this choice might be socially undesirable because of the negative impact on student access caused by imperfect geographical mobility amongst students

(Kelchtermans and Verboven 2010; De Fraja and Valbonesi 2012). It should be noted that the DEA model specification used in the analysis evaluates efficiency relating only to the first two of these points. Quality effects may be reflected by, for example, increased non-completion, and we use a production function approach to capture this. Further quality effects are accommodated in the second stage analysis.

Previous literature indicates that the average efficiency is considerably higher among merged than pre-merger and non-merging institutions: the null hypothesis of identical means in the three groups is rejected in all cases (Johnes 2014). These results should be treated with caution for a number of reasons.

First, examination of the average effects of merger conceal differences in the experiences of the different partners (Stewart 2003; Johnes 2014). In some cases, both partners enjoy unambiguous efficiency gains; in others, one partner gains while the other does not; and in still other cases, efficiency declines over time for both partners (Johnes 2014). Second, it is difficult to separate the effect of the act of merging from other underlying characteristics of the merged institutions. Universities which merge, especially if the decision to merge is a bottom-up institution-level one rather than a top-down directive, are likely to have different characteristics from those which do not merge, and these characteristics could themselves cause the observed differences in efficiency. Third, effects of merger may take time to be experienced and no study has looked at the evolution of efficiency over the periods following merger. Fourth, merger and efficiency might not exhibit a simple one-way causality (merger leads to greater efficiency), but may display a more complex, two-way relationship[139] whereby merger might lead to greater efficiency, but efficiency (or lack of it) might also be a motivation for merger. This has implications for the appropriate estimation approach. Finally, looking only at merger as the possible cause of changes in average efficiency may lead to omitted variable bias in the results if there are other factors which might also be affecting efficiency. Possible other factors are considered below.

Differences in the mission (such as subject mix or concentration on third mission activities) not taken into account in the DEA model may well affect the estimated efficiencies. In England, as earlier discussed in chapter 2 the HE sector is diverse and can be broadly split into 3 groups: pre-1992 universities – traditional institutions which offer degree programmes across the academic subject spectrum and an established research mission; post-1992 HEIs – institutions which have a balanced portfolio offering degree programmes across a range of academic and vocational subjects and a growing research mission; and former colleges of HE – institutions which have been awarded university status since 2003, which might be small and specialist, and often lack a strong research mission. Evidence from production studies suggest that the former colleges of HE have higher technical efficiency than the other two types (Johnes 2008, 2014); but this result is reversed in the context of cost function studies where they are found to be the least cost-efficient of the three types (Johnes et al. 2005; Johnes,

[139] Preliminary evidence does not in fact point to a two-way relationship (Johnes and Tsionas 2014).

Johnes, and Thanassoulis 2008; Johnes and Johnes 2009, 2013; Thanassoulis et al. 2011).

However the hypothesis that the composition of the student body might be expected to affect efficiency is not confirmed by the results: the proportion of total students from overseas has no significant effect on efficiency. Thus there is neither evidence of cross-subsidisation of research through international students nor of overseas students requiring greater resources. Also sources of funding might also affect efficiency; evidence from the USA suggests that reliance on tuition revenue has a negative effect on operating efficiency (Sav 2013) while government funding has a positive effect on both operating efficiency and cost efficiency in public universities (Sav 2012, 2013), but a negative effect on cost efficiency in private universities (Sav 2012). However, the funding systems for HE in the USA and England are different; mean proportion of income derived from government sources in the USA is 0.3 (Sav 2012, 2013).

It is therefore difficult to predict the direction of the effect in English HE. The increased competitive pressures caused by receiving a lower share of income from government sources (and hence a greater share from student fees, for example) might increase efficiency suggesting a negative relationship between the proportion of funding from government and efficiency as has already been observed in the wider European HE context (Wolszczak-Derlacz 2014). However according to the results obtained the effect of government funding on MTE is positive, a result with dubious policy implications that might be attributed to the fact that with a highly competitive funding and resource allocation system as it is the UK system we may expect positive impacts on the quality of research and publications as well as positive changes in efficiency and motivation (Liefner, 2003).

It is clear that there is little empirical evidence to date in the English HE context on the factors likely to affect technical efficiency, including merger activity, and this represents a gap in the literature which this study aimed to fill.

4.6.4 Conclusions

This chapter has reported the results of an empirical analysis of efficiency in English universities over 17 years. We find that mean efficiency for the sector is around 60 percent to 70 percent, suggesting that there is scope for increasing efficiency in the sector as a whole. The results also confirm previous findings that the efficiency levels of the vast majority of HEIs are not significantly different from each other. In other words, statistically, there is no difference in the performance of most of the HEIs. What is more, the small number of universities that have the lowest performance have specific characteristics (they are small and specialist) that are not captured in the DEA model. Thus, their apparent low level of efficiency should be treated with some caution.

The second stage of analysis of the efficiency scores suggests that, all else being equal, merger activity has a positive effect on efficiency. Merged HEIs have efficiency that is 5 percentage points higher post-merger than non-merging HEIs, holding all else constant. A further examination of the effects of merger over time suggests that the efficiency impact of merger comes soon after the merger takes place. These findings are clearly of interest to policymakers and HE managers. However, the context of the data analysed (i.e. English HE, which, generally, has not had a top-down policy of merger in the HE sector) should be remembered; whether or not such results would be forthcoming in a policy-led setting is still open to debate. Moreover, further work should be directed at establishing why the effects do not appear to continue beyond one year after merger.

Of the other factors included in the second stage of analysis, pre-1992 HEIs have lower efficiency than other types of institution. This may be a consequence of their size and specialty of provision. In addition, having a higher proportion of income from government sources is an incentive to greater efficiency – a result with clear policy implications. A preliminary exploration of the effect of third mission activity on efficiency suggests that interaction with business has positive spill over effects. As more data becomes available, this should be explored further.

These findings are clearly of interest to policy-makers who may claim that they justify a policy of merger in HE. Drawing such a hasty conclusion should be avoided for several reasons. First, the context of the data analysed (i.e. English HE which has generally not had a top-down policy of merger in the HE sector) should be remembered. Thus the merging institutions in this data set have made the decision to merge themselves – it has not been imposed from above, and it is important for HE managers to recognise this. As such, HEIs which are pre-disposed to merger may have distinct characteristics from those which do not. Indeed, there is some suggestion in the results that HEIs which subsequently merge are actually more efficient (by 2–3 percentage points) than those which do not merge. As a consequence, whether or not such results would be forthcoming in a policy-led setting is therefore still open to debate. Second, the results seem to suggest that there are decreasing returns to scale indicating that any positive merger effects are deriving from scope rather than scale. It is therefore important that merging HEIs have complementary operations where opportunities from economies of scope might be most likely. Third, the finding that two or more years after merger there is no significant difference in efficiency between merged and non-merging HEIs suggests that efficiency gains from merger may well be short-lived. Clearly this result requires further work to establish why the effects do not appear to continue into the longer term.

5. Chapter 5: Policy Evaluation-The Case of Propensity Score Matching

5.1 Evaluation Methods

The main task evaluation methods served is to estimate the impact of intervention in the presence of selection decisions by agents. In addition, they aim to correct empirical correlations to separate out the causal effect of the treatment from the confounding effect of the other factors influencing the outcome (Blundell et al., 2005). Implicitly, each method provides an alternative approach to constructing the counterfactual. Estimates from different methods may differ because they rely on different assumptions; they deal with different sources of bias and answer different questions.

In evaluation problems there are two types of studies based on the data used. Those split to those come from randomised trials (unlike case) or (non-randomised) observational studies. In the controlled experiments, assignment into treated and control groups is random, whereas in the observational studies, assignment into treated and control groups is not random. For the analysis of observational data, the structure depends on conceptualising the data as having arisen from an underlying regular assignment mechanism. Regular designs are like pure randomised experiments, except that the probabilities of treatment assignment are permitted to depend on covariates, and so can vary from unit to unit.

As suggested by Rubin (2008), we have to design observational studies to approximate randomised trials in order to obtain objective causal inference. In practice, simple comparisons or even regression-adjusted comparisons may provide misleading estimates of causal effects. This may reflect ambiguous potential differences between the two groups being compared due to omitted variable bias arising from unobserved and uncontrolled differences. In a more general framework, omitted variable bias (also known as selection bias) is the most serious econometric concern that arises in the estimation of treatment effects. Rosenbaum and Rubin (1983) proposed PSM as a method to reduce the bias in the estimation of treatment effects with observational datasets. The interaction between omitted variable bias, causality, and treatment effects can be seen most clearly using the potential outcomes framework. This was originally introduced by statisticians in the 1920s as a way to discuss treatment effects in randomised experiments.

5.2 Potential-Outcome Approach

According to Heckman and Vytlacil, (1999), treatment evaluation is the estimation of the causal effect of a programme or a treatment $(D = 1)$ relative to another treatment $(D = 0)$ on the outcome variable Y, experienced by units in the population of interest. $D_i \in \{0,1\}$ is the indicator of the treatment actually received. For each unit i, assume two potential outcomes (y_{0i}, y_{1i}) corresponding, respectively, to the potential outcomes in the untreated and treated states. Let $D_i = 1$ denote the receipt of treatment; $D_i = 0$ denotes non-receipt. Let y_i be the measured outcome variable so that:

$$y_i = D_i y_{1i} + (1 - D_i) y_{0i} \ (observed\ outcome\ for\ i)$$

$$y_i = y_{0i} + D_i (y_{1i} - y_{0i}) \equiv y_{0i} + \beta_i D_i$$

Where $\beta_i = \mathrm{y_{1i}} - \mathrm{y_{0i}}$ and $\mathrm{y_{0i}}, \mathrm{y_{1i}}, \mathrm{D_i}$ do not depend on $\mathrm{D_j}$, $\mathrm{j} \neq \mathrm{i}$ (SUTVA). The causal inference methods make the stable unit treatment value assumption (SUTVA) that one unit's outcomes are unaffected by another unit's treatment assignment. In other words, the treatment status of any unit does not affect the potential outcomes of the other units (non-interference) and the treatments for all units are comparable (no variation in treatment). One of a number of attributes, SUTVA excludes cross-effects, or general equilibrium effects, among potential treatment participants that could occur because of their actual participation decision (Lechner, 1999).

This is the Neyman-Fisher-Cox-Rubin model of potential outcomes. They are essentially the same as the econometric switching regression model of Quandt (1972), usually tied to a linear regression framework, the Roy model of income distribution (1951), the Heckman and Honore (1990) contribution to the classical Roy model, with further clarifications and extensions or the Heckman (1976; 1979) model in which a simple two-step estimator is used. Early influential contributions of this conceptual approach for experimental and non-experimental studies have been also made by Rubin, (1974; 1977),

The fundamental problem of causal inference is that only one of $Y_i(1)$ and $Y_i(0)$ is observed in any unit, so we can never find the true causal effect since it is impossible to observe the individual treatment effect β_i (Holland, 1986). So, estimation of missing data is needed from the data observed $(\mathrm{y_i}, \mathrm{x_i}, \mathrm{D_i})$ and assumptions are invoked to identify and estimate some characteristics (in particular means) of the distribution of β_i. It is central to define whether the parameter of interest has homogenous or heterogeneous impacts; hence, whether $y_{1i} - y_{0i} \equiv \beta \ \forall \ i$ or β_i, respectively.

5.3 Treatment Effects

It is impossible to measure causal effects at the individual level since they tend to be unobservable therefore, researchers focus on average causal effects. To make the idea of an average causal effect concrete, the causal estimands of interest are usually average treatment effects (ATE) on the whole population or on sub-populations. ATE is the difference between the outcomes of treated and control observations.

$$\Delta = y_1 - y_0$$

So, in other words, ATE is useful to evaluate the expected effect on the outcome Y if individuals in the population are randomly assigned to treatment.

$$ATE = E(y_1 - y_0)$$

This is equivalent to a simple t-test between the outcomes for the treated and the control groups. Consider the outcome $\mathrm{Y^{obs}}$, i.e. the observed efficiency scores for each institution; then:

$$\mathrm{E}\left(\mathrm{Y^{obs}}\middle|\mathrm{t}\right) = \alpha + \tau\mathrm{t}.$$

In this model, the treatment is the merger status t, denoted as a dummy variable. From the model we obtain, the average outcome for untreated units: $\mathrm{E}\left(\mathrm{Y^{obs}}\middle|\mathrm{t}=0\right) = \alpha$ and the average outcome for treated units: $\mathrm{E}\left(\mathrm{Y^{obs}}\middle|\mathrm{t}=1\right) = \alpha + \tau$.

The difference in mean's estimator is given by the slope of t, i.e. $\tau = \mathrm{E}\left(\mathrm{Y^{obs}}\middle|\mathrm{t}=1\right) - \mathrm{E}\left(\mathrm{Y^{obs}}\middle|\mathrm{t}=0\right)$.

It is likely that the expected average efficiency for the treated units is lower than for the control, so, mistakenly, the treatment may be considered dangerous. However, there are assumptions underlying the linear regression model that are not plausible so we need to adjust any difference in average outcomes for differences in pre-treatment characteristics (not being affected by the treatment). Adjusting for confounding variables, we can estimate the conditional average treatment effect (CATE).

$$E(Y_1 - Y_0|X = x)$$

We can adjust via specification of a conditional model for the potential outcome using regression models. In a standard regression approach, unconfoundedness (untestable) is implicitly assumed together with other functional or distributional assumptions.

$$\widehat{Y_i^{obs}} = \alpha + \tau D_i + \beta\, x_i + \varepsilon_i$$

With the usual exogeneity assumption that $\varepsilon_i \perp D_i$ and $\varepsilon_i \perp x_i$. The regression of Y_i^{obs} on a constant, D_i and x_i implicitly assumes the constant treatment effect and the slope τ of the treatment indicator will act as an estimator of the ATE.

$$\hat{\tau} = E\left(Y_i^{obs} \middle| X, D = 1\right) - E\left(Y_i^{obs} \middle| X, D = 0\right)$$

The model assumes a homogeneous treatment effect, i.e. the average change in the outcome variable due to treatment keeping x characteristics constant is equal to:

$$ATE = E(\Delta) = E((y_1|x, D = 1) - E(y_0|x, D = 0)$$

According to Heckman (1997), ATE might not be of relevance to policymakers because it includes the effect on persons for whom the programme was never intended. Hence, ATE is consistent in random experiments, but in observational studies it may be biased if treated and control observations are not similar. Therefore, what needs to be altered here is the calculation of the average treatment effects on the treated (ATT). The aforementioned method explicitly evaluates the effects on those for whom the programme is actually intended.

ATT is the difference between the outcomes of the treated observations if they had not been treated. Hence, it appears as a counterfactual situation, between the treated observations and the closest potential match of the control observations. This is a 'naïve' estimator, using (observed) outcomes of the non-participants. An individual's participation decision based on personal characteristics (observed and unobserved) is likely to affect Y.

$$ATT = E(Y_1 - Y_0|D = 1)$$

$$ATT = E(\Delta|x, D = 1) = E(y_1|x, D = 1) - E(y_0|x, D = 1)$$

This second term is the counterfactual element $E(y_0|x, D = 1)$ (i.e. the counterfactual mean) for those being treated, which is not observed, hence needs to be estimated. A potential substitute for it in order to estimate ATT is the mean outcome of untreated individuals $E(y_0|x, D = 0)$. In observational studies, this is not a good idea because it could be that covariates that determine the treatment decision also determine the outcome variable of interest. Therefore, the outcomes of individuals from the treatment and comparison groups (the difference between treated and non-treated outcomes in the absence of treatment) will differ even in the absence of treatment, resulting in so-called selection bias.

Using experimental data, Heckman, et al., (1998) provide a very useful breakdown of this bias term. In particular if we compare the outcomes by treatment status, we obtain a biased estimate of the ATT:

$$\begin{aligned} E(Y^{obs}|D=1) - E(Y^{obs}|D=0) &= \mathrm{E}(\mathrm{y}_1|\mathrm{D}=1) - \mathrm{E}(\mathrm{y}_0|\mathrm{D}=0) \\ &= \mathrm{E}(\mathrm{y}_1|\ \mathrm{D}=1) - \mathrm{E}(\mathrm{y}_0|\mathrm{D}=1) + \mathrm{E}(\mathrm{y}_0|\mathrm{D}=1) - \mathrm{E}(\mathrm{y}_0|\mathrm{D}=0) \\ &= ATT + bias \end{aligned}$$

The ATT can be identified only in the case in which the outcomes of individuals from the treatment and comparison groups would not differ in the absence of treatment. Consequently, $\mathrm{E}(\mathrm{y}_0|\mathrm{D}=1) - \mathrm{E}(\mathrm{y}_0|\mathrm{D}=0) = 0$.

As previously mentioned, in experimental studies, assignment to treatment is random; this is ensured and the treatment effect is directly identified. However, in observational studies, we must rely on some identifying assumptions to solve the selection problem (selectivity bias).

One source of selection bias may stem from the presence of units in one group that cannot find suitable comparison in the other, comparing non-comparable individuals (non-overlapping support of X in the treated and control group). Consequently, the first component of bias accounts for differences in the distribution of observed characteristics X between the two groups (non-overlapping support of the observables). Another potential reason for bias are the differences in the distribution of the observables between the two groups over the common support hypothesis (weighting comparable individuals incomparably). Hence, the second component of the error tries to cover inexpediencies due to differences in the resulting empirical distributions of observables even when restricted to the same support. Finally, selection of unobservables may lead to selection bias from sample differences that researchers can observe but fail to control (Blundell et al., 2004).

A properly designed randomised experiment would curtail or even eliminate the bias mentioned above, but such cases are very rare. In the vast majority of economic research, observational datasets are used. Therefore, in the absence of an experiment, researchers rely on a variety of statistical control strategies to reduce omitted variable bias. An effective way to confront is to conceptualise the data as having arisen from an underlying regular assignment mechanism (Grilli and Rampichini, 2011). A common practice is to adjust any difference in average outcomes for differences in pre-treatment characteristics (not being affected by the treatment). Hence, virtually non-experimental methods are used, each of which uses observed data combined with appropriate identifying assumptions to recover the missing counterfacts (Blundell et al., 2004). The sources of bias outlined above can be handled efficiently if the appropriate method from the alternatives is chosen. The *a priori* richness and nature of the available data, the postulated model for the outcome, and the selection processes are the determining factors in this process.

Non-experimental methods vary in terms of the appropriate identifying behavioural assumptions and the statistical methods, because of the differences in the observed microdata. Four broad classes of statistical method can be identified in the literature to recover missing counterfacts: model-based imputation methods (e.g. regression analysis[140]), instrumental variable methods, control function methods, and matching methods. By virtue of potential weaknesses and pitfalls may lurk in the above methods (e.g. sensitivity to linearity assumption, increased sensitivity to the model, and *a priori* assumptions) covariate choice (specification) and, more recently, non-parametric regression estimators. In the current study, in order to avoid model dependence, we focus mainly on matching techniques, which vary extensively in the literature, and, specifically, on propensity score methods of matching, which tend to have been well-established in the field.

There are two broad groups of matching estimators: the traditional matching estimator, i.e. individual neighbourhood, and smoothed weighted matching estimators. The former can be further divided into nearest neighbour matching (NNM) and/or caliper matching, and the latter on kernel-based matching. The aim of these matching techniques is to construct a suitable selected comparison group that is as similar as possible to the treatment group in terms of observable characteristics. There are two main assumptions postulated here: the conditional independence assumption and the common support assumption, which should hold simultaneously.

5.3.1 Assumptions

Partial equilibrium character (no general equilibrium effects)

Treatment does not indirectly affect the control observations (so there are no spill-over effects from the treated group to the control group).

Conditional independence assumption

The conditional independence assumption (CIA) states that the potential treatment outcomes are independent of the assignment mechanism for any given value of a vector of attributes (X). Hence, for random experiments the outcomes are independent of treatment.

This assumption is formalised in the following expression:

$$y_0, y_1 \perp D$$

For observational studies the outcomes are independent of the treatment, conditional on X. So treatment needs to be an exogenous variable, which ignores the outcomes. The

[140] Ordinary least square methods (simple or multiple linear regression).

intuition of this assumption implies that all the relevant differences between treated and non-treated observations are captured in their observed attributes

$$(y_0; y_1) \perp (D|x) \; for \; x \in S$$

Unconfoundedness assumption

A weaker version of the CIA is the unconfoundedness (ignorability) assumption or selection of observables. If the decision to take the treatment is purely random for individuals with similar values to the pre-treatment variables, then we can use the average outcomes of some similar individuals who were not exposed to the treatment. The underlying identifying assumption is called unconfoundedness (selection of observables or conditional independence).

Assignment to treatment is independent of the outcomes, conditional on the covariates. Hence the only source of omitted variables or selection bias is the set of observed covariates X_i. This is a weaker assumption than the conditional independence assumption (Lechner, 1999).The conditional independence assumption becomes:

$$y_0 \perp (D|x)$$

$$E(Y_0|x, D = 1) = E(Y_0|x, D = 0) \; for \; x \in S$$

This is the key assumption that facilitates causal inference (sometimes called an identifying assumption), which can be further developed as:

$$E\big(Y_{ji}\big|X_i, D_i\big) = E\big(Y_{ji}\big|X_i\big) \; for \; j = 0{,}1$$

$$E(Y_{1i} - Y_{0i}|D_i = 1) = E\{E[(Y_{1i}|X_i, D_i = 1)] - \langle Y_{0i}|X_i, D_i = 1\rangle|D_i = 1\}$$

$$= \; E\{E[(Y_{1i}|X_i, D_i = 1)] - \langle Y_{0i}|X_i, D_i = 0\rangle|D_i = 1\}$$

And likewise, $E(Y_{1i} - Y_{0i}) = E\{E[(Y_{1i}|X_i, D_i = 1)] - \langle Y_{0i}|X_i, D_i = 0\rangle\}$

Consequently, the ATT or ATE can be expressed by averaging X-specific treatment-control contrasts, and then reweighting these X-specific contrasts using the distribution of X_i for the treated (for ATT), or using the marginal distribution of X_i (for ATE). Since these expressions involve observable quantities, it is straightforward to construct consistent estimators from their sample.

Under unconfoundedness, the basic idea is to find units similar to the treated subjects in all relevant pre-treatment characteristics X from a large group of non-treated units. Since conditioning on all relevant covariates is limited in the case of a high dimensional vector X, Rosenbaum and Rubin, (1983a) suggest the use of the so-called balancing scores $b(X)$. These scores are functions of the relevant observed covariates X such that the conditional distribution of Xgiven $b(X)$ is independent of assignment into treatment:

$$X \perp (D|\mathrm{b(X)})$$

Of course, balancing scores are not unique. One possible type is the propensity score i.e. the probability of being treated given observed characteristics X. Matching procedures based on this balancing score are known as propensity score matching. The assignment to the treatment does not affect the outcome of the control group:

$$\mathrm{e(X)} = \mathrm{P_r}(D = 1|X = x) = \mathrm{E(D|X = x)}$$

$$\mathrm{P_r(D|x, y_0, y_1)} = \mathrm{P_r(D|x)}$$

According to Rosenbaum and Rubin (1983), at any value of a balancing score, the difference between the treatment and the control means is an unbiased estimate of the ATE at that value of the balancing score if treatment assignment is strongly ignorable. If treatment assignment is strongly ignorable given X characteristics, then it is strongly ignorable given any balancing score, i.e. D is independent of x given the propensity score. If unconfoundedness is a valid assumption then all biases due to observable covariates can be removed by conditioning solely on the propensity score:

$$\left((\mathrm{y_0, y_1}) \perp D\middle|\mathrm{e(X)}\right)$$

Usually, given a set of pre-treatment variables, unconfoundedness is viewed as a reasonable approximation of the actual assignment mechanism, with only vague *a priori* information about the dependence of the propensity score on the pre-treatment variables, (Grilli and Rampichini, 2011).

5.3.2 Matching or Overlap Assumption

The overlap or common support condition implies that the probability of assignment is bounded between zero and one. For each value of x, there are both treated and control observations, and for each treated observation, there is a matched control observation with similar x. We want the characteristics x to cover both the treated and the control group, (Rosenbaum and Rubin, 1983a). Therefore, the individual assignment possibilities (propensity scores) as a function of unit $i's$ value of the covariates, $p_i = \Pr(D = 1|x)$), are strictly between zero and one:

$$0 < prob(D = 1|x) < 1 \; for \; x \in S$$

The assignment mechanism can be interpreted as though we can analyse data from sub-samples with the same value of the covariates, as if they arose from a completely randomised experiment. The reduction to a paired comparison should only be applied if unconfoundedness is a plausibly assumption, based on the data and a detailed understanding of the institutional set-up by which selection into treatment takes place (Blundell et al., 2005).

According to Rosenbaum and Rubin (1983) findings, treatment is strongly ignorable when both unconfoundedness and overlap assumptions are valid. These two are the key assumptions that can identify the ATE. If the overlap assumption is violated at $X = x$, it is infeasible to estimate both $\mathrm{E}(\mathrm{y}_1|\mathrm{X} = \mathrm{x}, \mathrm{D} = 1)$ and $\mathrm{E}(\mathrm{y}_0|\mathrm{X} = \mathrm{x}, \mathrm{D} = 0)$.

5.3.3 Balancing Property

Assignment to treatment is independently conditional on specific functions of X, denoted as balancing score $e(X)$. This assumption fulfils the so-called balancing score property where the treatment decision is not reliant on the X characteristics given the same propensity score (testable condition):

$$X \perp (D|e(X))$$

And more specifically under the propensity score case:

$$X \perp (D|p(X))$$

Where $p(x) \equiv P(D = 1|X = x) = E(D|X = x)$

So, the balancing hypothesis implies that, observations with the same propensity score must have the same distribution of observable (and unobservable) characteristics independently of treatment status:

$$\mathrm{D} \perp (\mathrm{x}|\mathrm{p}(\mathrm{x}))$$

For a given propensity score, exposure to treatment is random and, therefore, treated and control units should be, on average, observationally identical. Rosenbaum and Rubin (1983) labelled the propensity score as particularly important for the set-up of the balancing scores since it reduces the dimension of the conditioning set to one. By avoiding model specification, we are coping with the curse of dimensionality and the problem of continuous covariates.

In order to ameliorate model dependence, matching methods have been extensively used. The underlying assumption of matching is that the only source of omitted variables or selection bias is the set of observed covariates X_i. In contrast to regression, treatment effects are constructed by matching individuals with the same covariates instead of through a linear model for the effect of covariates. Therefore, the goal here is balancing in covariates. This will further permit the generation of pairs of observations that have the exact same covariate values (perfect balance) and differ only on treatment assignment. Then, we have perfect conditional ignorability, assuming no omitted variable bias (Rubin, 2006). By tackling the data as matching pairs, we will achieve the same results regardless of the model. Matching is a method of trying to achieve better balance in covariates and to reduce model dependence.

5.4 Matching Strategies

Matching techniques have been applied in experimental work from the first half of the 20th century[141] and were advanced and developed in a set of papers by Rosenbaum and Rubin (1983a, 1983b, 1984, 1985). The standard matching strategy aims to pair each treated subject i with one or more comparable non-treated subjects. Then, the strategy intends to associate to the outcome Y_i^{obs} a matched outcome:

$$\widehat{Y_\iota}(0) = \sum_{j \in C^0(p_i)} w_{ij} Y_j^{obs}$$

Where $C^0(p_i)$ is the set of neighbours with $w =$ of the treated subject i and w_{ij} is the weight of non-treated j in forming a comparison with treated i and $w_{ij} \in [0,1]$ with $\sum_{j \in C(i)} w_{ij} = 1$.

Through the matching process we attempt to associate to the outcome y_i of treated unit i a matched outcome given by:

1. Traditional matching estimators.
   The outcome of the most observably similar control unit x (The general formula that formalizes one to one matching estimators is:

   $$C^0(p_i) = \left\{ j : |p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\} \right\}$$

   With $w_{ik} = 1(k = j)$

2. Simple smoothed matching. Two discrete categories exist here: K-nearest[142] neighbours matching and radius matching.

[141] See e.g. Rubin (1974) or Lechner (1998).
[142] **K-nearest neighbours:**
$C^0(p_i) = \{the\ K\ units\ in\ D = 0\ with\ p\ closest\ to\ p_i\}$
$w_{ij} = 1/K\ \ if\ j \in C^0(p_i)\ otherwise\ w_{ij} = 0$

3. A smoothed weighted matching[143] or kernel based matching.
   It is a weighted average of the outcomes of more (possibly all) non-treated (control) units where the weight given to non-treated unit j is in proportion to the closeness of the observables of i and j.

$$C^0(p_i) = \{D = 0\}$$

For Gaussian kernel $w_{ij} \propto K\left(\frac{p_i - p_j}{h}\right)$

An estimate of the propensity score is not enough to estimate the ATE. In fact, the probability of observing two units with exactly the same value of the propensity score is, in principle, zero since e(X) is a continuous variable. Typically, one treatment case is matched to several control cases, but one-to-one matching is also common and may be preferred (Glazerman et al. 2003). Several matching methods have been proposed in the literature, the most widely used are described in the following sections.

5.4.1 Nearest Neighbour Matching (NNM) (with or without caliper)

NNM matches treated and control units taking each treated units, and searching for the control unit with the closest propensity score, i.e. one-to-one matching. Although it is not necessary, the method is usually applied with a replacement,[144] in the sense that a control unit can be a best match for more than one treated unit. Once each treated unit is matched with a control unit, the difference between the outcome of the treated units and the outcome of the matched control units is computed. The ATE is then obtained by averaging these differences. All treated units find a match. However, it is obvious that some of these matches are fairly poor because, for some treated units, the nearest neighbour (NN) may have a very different propensity score; nevertheless, it would contribute to the estimation of the treatment effect independently of this difference.
If $e_i(x_i) = p_i$ is the propensity score of the i-th . Given a treated unit i , let $I_{m(i)}$ denote the index of the non-treated unit that is the m-th closest to unit i in terms of the distance measure based on the norm $\|.\|$

---

[143] Other matching methods in this category are:

I. Local linear regression matching: For each treated i, estimate $\hat{y}_i \equiv E(Y|D = 0, p(X) = p_i$non-parametrically: Fit a line estimated on a local neighbourhood of p_i and then apply a weighting scheme with weights K_{ij}, $\min_{\theta_0,\theta_1} \sum_{\kappa \in C^0(p_i)} \left(y_j - \theta_0 - \theta_1(p_i - p_j)\right)^2 K\left(\frac{p_j - p_i}{h}\right)$
II. Mahalanobis-metric matching: The weights $W$are combined into a distance measure and then match on the resulting scalar: $d(i,j) = (W_i - W_j)' \, V^{-1} \, (W_i - W_j)$ where V is the pooled within –sample covariance matrix of W. The weights are assigned to each w in proportion to the inverse of the variance of w. For further detail see Rubin, (1980).

[144] Matching with replacement keeps bias low at the cost of a larger variance, while matching without replacement keeps variance low at the cost of potential bias.

$$\sum_{j:W_j \neq W_i} \{\|p_j - p_i\| \leq \|p_I - p_i\|\} = m$$

Let $C(i)_M$ denote the set of indices for the first M matches for unit i:

$$C(i)_M = \{I_1(i), \dots \dots, I_M(i)\}$$

$$\widehat{Y_\iota}(0) = \sum_{j \in C(i)M} Y_j^{obs}$$

The individual NNM set for one to one matching (not for multiple nearest neighbours) is given by:

$$\mathrm{C(i)} = \min_{\mathrm{j}} \|\mathrm{p_i} - \mathrm{p_j}\|$$

Where $\mathrm{C(i)}$ is a set of control units matched to the treated unit i with an estimated value of the propensity score $\mathrm{p_i}$. Normally, the case of multiple neighbours should be not alike, in particular if the set of characteristics Xcontains continuous variables. In the implementation of this method, there is a common trade-off between bias and variance. In cases in which the matching is on just one NN the bias is minimised at the cost of a larger variance. However, if multiple nearest neighbours are used, a decrease in variance is feasible at the expense of a generated increase in bias. Another option that is commonly used here is the caliper option. NNM within a specified caliper distance is similar to the usual NNM embedded with some further restrictions that the absolute difference in the propensity scores of matched units must be below a pre-specified threshold which is the caliper distance (Austin, 2011).

5.4.2 Caliper Matching

When caliper matching is used, what differs from radius matching is that the nearest control is used as a match if a treated unit has no control units within radius r. Given $\delta > 0$, treated i is matched to non-treated j such that:

$$\delta > |p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

If no untreated unit is within δ from treated i, i is left unmatched. When the caliper is not applied, there are no *a priori* support restrictions (Cochran and Rubin, 1973).

5.4.3 Radius Matching

Each treated unit is matched only with the control units whose propensity score falls into a predefined neighbourhood of the propensity score of the treated unit. We use radius and we match the observations according to the equivalent radius. Each treated observation i, is matched with a control observation j that falls within a specified radius r:

$$C^0(p_i) = \{all\ j: \|p_i - p_j\| < r\}$$

$$w_{ij} = 1/N_i^C \ \ if\ j \in C^0(p_i)\ otherwise\ w_{ij} = 0$$

Where N_i^C is the number of controls matched with observation $i \in T$. The smaller the radius the better the quality of the matches and the higher the possibility that some treated units are not matched because the neighbourhood does not contain control units. The estimator's formula for both NN and radius matching can be expressed as follows (Becker and Ichino, 2002):

$$\begin{aligned}\tau^M &= \frac{1}{N^T}\sum_{\iota \in T}\left[Y_i^T - \sum_{j \in C(i)} w_{ij}Y_j^C\right] \\ &= \frac{1}{N^T}\sum_{\iota \in T}\left[Y_i^T - \sum_{i \in T}\sum_{j \in C(i)} w_{ij}Y_j^C\right] \\ &= \frac{1}{N^T}\sum_{\iota \in T} Y_i^T - \frac{1}{N^T}\sum_{j \in C} w_j Y_j^C\end{aligned}$$

Where M captures both NN and radius matching, N^T denotes the number of units in the treated group. For the variance formula of this estimator follow Becher and Ichino (2002).

5.4.4 Kernel Matching

For each observation in the treated group, we use all the observations in the control group where we are going to weight them. If the gap of the difference between the propensity scores is wide then the weights increase, otherwise, when the difference is narrow the arising weights decrease. Each treated observation i is matched with several control observations with weights inversely proportional to the distance between the propensity scores of treated and control observations. With matching based on propensity scores, the weights are defined as (Heckman et al., 1997, 1998):

$$w(i,j) = \frac{k\left(\frac{p_i - p_j}{h}\right)}{\sum_{j\in C^0(p_i)} k\left(\frac{p_i - p_j}{h}\right)}$$

Where h is a bandwidth parameter and $w(i,j)$ is the non-treated $j's$ outcome. $\sum_{j\in C^0(p_i)} w(i,j) = 1$. The higher the value h, the more 'tolerant' the matches in terms of closeness.

The outcome $y_i$of treated i is associated to a matched outcome given by a kernel-weighted average of the outcome of comparable non-treated, where the weight given to non-treated j is in proportion to the closeness between i and j:

$$\widehat{y_i} = \frac{\sum_{j\in C^0(p_i)} k\left(\frac{p_j - p_i}{h}\right) y_j}{\sum_{j\in C^0(p_i)} k\left(\frac{p_i - p_j}{h}\right)}$$

The choice of kernel function varies and it may be a Gaussian distribution where $K(u) \propto \exp(-u^2/2)$ for all the non-treated units $C^0(p_i) = \{D = 0\}$ or an Epanechnikov distribution with $K(u) \propto (1 - u^2)$ if $|u| < 1$ (=0 otherwise), moving window within the D=0 group.

5.4.5 Stratification Matching

This method consists of blocks to compare the outcomes within intervals/blocks of propensity score where blocks are defined by the same algorithm that estimates the propensity scores. Through the stratification method on the propensity score, the range of variation of the propensity score is classified in intervals such that, within each interval, treated and control units display, on average, roughly similar values of propensity score. Those stratified units are embedded into mutually exclusive subsets according to their estimated propensity score. Then, within each interval in which both treated and control units exist, the difference between the average outcomes of the treated and the controls is computed (Austin, 2011). Therefore, when the propensity score has been correctly specified, the distribution of measured baseline covariates will be approximately equivalent for both treated and untreated units within the same stratum.

A frequently used approach for stratification in the quintiles of the propensity score is suggested by Rosenbaum and Rubin (1984). This sub-classification of the propensity score is a popular method for estimating the (causal) difference of two treatment means used by Rosenbaum and Rubin (1984), where individuals are stratified based on estimated propensity scores and the difference is estimated as the average of within-stratum effects (Lunceford and Davidian, 2004). Through their approach, units are divided into five equal quintiles of the estimated propensity score, and they observed that 90 percent of the bias due to measured confounders when estimating a linear treatment effect is diminished. Increases in the number of strata used provoke a bias reduction, although the marginal reduction in bias decreases as the number of strata increases (Cochran, 1968; Hullsiek and Louis, 2002). In mathematical terms within each block, the program computes:

$$\tau_q^s = \frac{\sum_{i \in I(q)} Y_i^T}{N_q^T} - \frac{\sum_{j \in I(q)} Y_j^C}{N_q^C}$$

Where, q reflects the blocks defined over intervals of the propensity score, and $I(q)$ is the set of units in block q. N_q^T and $N_q^C$are the numbers of treated and control units in block q respectively.

The estimator of the ATT based on the stratification method is finally obtained as an average of the ATT of each block with weights given by the distribution of treated units across blocks. The ATT is then computed with the following formula:

$$\tau_q^s = \sum\nolimits_{q=1}^{Q} \tau_q^s \frac{\sum_{i \in I(q)} D_i}{\sum_{\forall i} D_i}$$

Where the weight for each block is given by the corresponding fraction of treated units and Q is the number of blocks. One of the pitfalls of the stratification method is that it discards observations in blocks where either treated or control units are absent or because no control is available in their block.

5.5 Propensity Score Matching Theoretical Framework

PSM uses an average of the outcomes of similar subjects who have the other treatment level to impute the missing potential outcome for each subject. Similarity between subjects is based on estimated treatment probabilities, known as propensity scores. The ATE is computed by taking the average of the difference between the observed and potential outcomes for each subject. These methods have become increasingly popular in medical trials and in the evaluation of economic policy interventions.

Suppose each observation has a true probability of receiving the treatment. The probability of receiving the treatment is the propensity score. We do not know the true propensity score, but we can estimate it for each observation with a regression of T (treatment) on X,[145] assuming we have the right set of X that went into the decision for assigning treatment. Since we do not have the true propensity scores, we need to check for balance in our covariates at the end.

[145] Baseline covariates that affect treatment assignment and/or potential confounders that affect the outcome variable.

The main element here is to assign the observations into two groups. The treated group (received the treatment) and the control group (did not receive the treatment). Treatment D is a binary variable that determines whether the observation has received the treatment or not. When $D = 1$ denotes treated observations and $D = 0$ for control observations.

After demonstrating the treatment and control groups, any standard probability model can be used to estimate the propensity score. Most software packages tend to use a probit or a logit model to define the propensity of observations given x variables that affect the likelihood of being assigned into the treated group. So, the propensity score model is a binary model with D serving as a dependent variable and x independent variables.

$$\mathrm{P(x)} = \Pr((D = 1|x) = E(D|x) = F(h(x_i)$$

Where: $h(x_i)$ is a function of covariates with linear and higher order terms, F (:) is a cumulative distribution, e.g. the logistic distribution, and $\mathrm{X_i}$ is a set of observed characteristics of individual i, Rosenbaum and Rubin (1983).

$$\Pr((D = 1|x) = \frac{\exp \mathrm{h}(x_i)}{1 + \exp\ h(x_i)}$$

The propensity score is the conditional (predicted) probability of receiving treatment given pre-treatment characteristics x. So, the balancing property should be satisfied in the obtained estimate of the propensity score. An appropriate specification of $\mathrm{h}(x_i)$ that satisfies the balancing property may contain even higher order terms in $\mathrm{h}(X)$. In general, in the received literature, more parsimonious rather than the full set of interactions are needed to match cases and controls on the basis of observables (Grilli and Rampichini, 2011). Therefore, after matching an observation that received treatment with an observation with a similar propensity score that received control, we can compare the outcomes of treated and control observations:

$$ATT = E(\Delta|p(x), D = 1) = E(y_1|p(x), D = 1) - E(y_0|p(x), D = 0)$$

The true propensity score is generally unknown, so that the propensity score needs to be estimated. If the potential non-treatment outcome is independent of the assignment mechanism conditional on $X = x$, then it is also independent of the assignment mechanism conditional on $P(X) \ = p(x)$, thus:

$$E\big(Y\big|D = 1, P(X) = p(x)\big) = E\big(Y\big|D = 0, P(X) = p(x)\big)$$

Hence, $E(Y|D = 1) = E\{E\langle Y|D = 0, P(X) = p(x)|D = 1\rangle\}$ can be used for estimation.

When the propensity score is known or can be N −consistently estimated with a parametric model, then the major advantage of this property is the reduction of the dimension of the estimation problem, which is especially important for non-parametric estimation techniques (Lechner, 1999).

The true propensity score is a balancing score, so the distribution of pre-treatment variables between treated and untreated units with the same propensity score should be the same and independent of treatment assignment. Therefore, appropriate methods (statistical tests) for assessing whether the propensity score model has been adequately well specified should be used (Heckman et al., 1998; Heckman and Smith, 1999; Black and Smith, 2003). In general, the choice of the baseline covariates included in the propensity score model should be based on theory or/and previous findings since the propensity score model does not entail a behavioural interpretation.

There is a lack of consensus in the applied literature regarding the inclusion (or exclusion) of covariates in the propensity score model. According to Austin et al., (2007), there is more than one possible set of covariates for inclusion in the propensity score model. Mainly, all measured baseline covariates (characteristics) should be included in the propensity score model safely, since, practically, the identification process of true confounders may be infeasible (Austin, 2011). Hence, the set of X covariates may consist of:

1. Baseline covariates that affect exposure to treatment, all covariates that affect the outcome (i.e. the potential confounders).
2. Only variables that are unaffected by treatment, but that affect the outcome so either fixed over time or measured before participation (Brookhart et al., 2006).
3. Only variables that affect simultaneously the treatment status and the outcome variable (i.e. the true confounders) (Sianesi, 2004; Smith and Todd, 2005).

In cases of uncertainty of the proper specification, the disputable choice is whether it is better to include too many rather than too few variables. Variables should only be excluded from analysis if there is strong evidence that is unrelated to the outcome or not appropriate covariates. If this is not the case, Rubin and Thomas (1996) recommend including the relevant variables in the propensity score estimation since omitting important variables can seriously increase bias in resulting estimates (Heckman et al., 1997; Dehejia and Wahba, 1999).

On the other hand, over-parameterisation steers on higher variance of the propensity score estimates and deteriorates the support problem (Bryson et al., 2002). In general, the plethora of available covariates varies and, according to the number of covariates used in the model, may result in bias increase for a wide bandwidth of covariates due to the weakness of the common support or to higher variance for a lower number of covariates due to the implausibility of the unconfoundedness assumption. This trade-off on the plausibility of the unconfoundedness assumption and the variance of the estimates in finite samples affects the estimated standard errors, which tend to be smaller for parsimonious specification where the common support condition poses no problem (Black and Smith, 2003).

The various methods considered above reach different points on the frontier of the trade-off between the quality and quantity of the matches, and neither method is considered *a priori* superior to the others. Their joint consideration, however, offers a way to assess the robustness of the estimates (Becker and Ichino, 2002). The quality of the matches may be further improved under the common support restriction. Under this scheme, observations whose propensity score belongs to the intersection of the supports of the propensity score of treated and controls are only considered. A caveat of this assumption suggests that imposing the common support restriction is not necessarily better (Lechner, 2001), since high-quality matches may be missed at the boundaries of the common support and the sample may be considerably downscaled. However, the imposition of this assumption provides better-quality matches.

5.5.1 Propensity Score Matching[146] Practical Exploration

For many years, the standard tool for performing PSM in Stata has been the psmatch2 command, written by Leuven and Sianesi (2003). Quite frequently, however, the p-score (Becker and Ichino, 2002) is also used. All the matching methods previously discussed[147] are available in the psmatch2 command. In addition, appropriate routines exist for the common support option, options for graphing, and covariate imbalance testing. Although the standard errors obtained using bootstrapping methods or variance approximation, it is not taken into account that the propensity score is an estimate, so the received output includes a caveat.

In a recent paper, Abadie and Imbens (2012) established how to take into account that propensity scores are estimated at an earlier stage, so there are estimates. Consequently, they launched the teffects psmatch command, and the recent development of this command that relies mainly in their work on psmatch2 is the fact that it takes into account that propensity scores are estimated rather than known when calculating standard errors. This often makes a significant difference, and sometimes in surprising ways. The underlying logic of the teffects psmatch command for estimating treatment effects from observational data includes PSM by determining how near subjects are to each other by using estimated treatment probabilities, known as propensity scores. A main advantage that PSM offers for the NNM estimator is that it does not need bias correction, because PSM matches on a single continuous covariate. In effect, the PSM estimator parameterises the bias-correction term in the treatment probability model, while the NN estimator constructs a bias correction term when matching on more than one continuous covariate.

[146] While methods based on PSM are the most common method of estimating treatment effects, there are other methods of estimating treatment effects. Other implementation methods include regression adjustment, i.e. model-based imputation methods (e.g., regression models), inverse probability weighting, augmented inverse probability weighting, inverse probability weighted regression adjustment etc.

[147] E.g. nearest neighbour, k-nearest neighbours, radius matching, kernel matching, local linear regression, Mahalanobis matching, etc.

In this study, we attempt to assess the effect of merger activity (treatment) on the average efficiency level performed by HEIs in England. Therefore, the ATE on universities that have been merged (treated) will be computed and compared to universities that have not experienced a merger. This analysis will be a sensitivity analysis check for previous findings using a two-step procedure that the typical merged HEI is significantly more efficient than either pre-merger or non-merging HEIs, suggesting that, on average, merging is a positive activity. Moreover, the merger effect tends to last for almost three periods after the merger; the year after the merger seems to concentrate higher efficiency levels compared to subsequent years.

Additional findings strengthen the claims of efficiency benefits through mergers that accrue from RTS, as a consequence of increased administrative, economic, and academic efficiency (Skodvin, 1999; Harman, 2000), or returns to scope if the merging institutions have complementary activities (Skodvin, 1999).

Matching estimators are rely mainly on the potential outcome model, in which each individual has a well-defined outcome for each treatment level. In a binary-treatment potential outcome model, y_1 is the potential outcome obtained by an individual if given treatment level 1 and y_0 is the potential outcome obtained by each individual i if given treatment-level 0. In this particular study, y_1 denotes the efficiency score of universities that have been merged and y_0 denotes the efficiency score of universities that have not been merged at all. The problem posed by the potential outcome model is that only y_{1i} or y_{0i} is observed, never both. Both y_{0i} and y_{1i} are realisations of the random variables y_0 and y_1. Throughout this document, i subscripts denote realisations of the corresponding, non-subscripted random variables.

Formally, the ATE is:

$$\tau_1 = E(y_1 - y_0)$$

And the ATET is:

$$\delta_1 = E(y_1 - y_0 | t = 1)$$

These expressions imply that we must have some solution to the missing-data problem that arises because we only observe either $\mathrm{y}_{1\mathrm{i}}$ or $\mathrm{y}_{0\mathrm{i}}$, not both. The teffects psmatch implements NNM on an estimated propensity score, which is a conditional probability of treatment.

For each individual, NNM uses an average of the individuals that are most similar, but get the other treatment level, to predict the unobserved potential outcome. NNM uses the covariates $\{\mathrm{x}_1, \mathrm{x}_2, \dots \dots, \mathrm{x}_\mathrm{p}\}$ to find the most similar individuals that get the other treatment level.

More formally, consider the vector of covariates $x_i = \{x_{i,1}, x_{i,2}, \dots \dots, x_{i,p}\}$ and frequency weight w_i for observation i. The distance between x_i andx_j is parameterised by the vector norm:

$$\|x_i - x_j\|_S = \{(x_i - x_j)' S^{-1} (x_i - x_j)\}^{1/2}$$

Where S is a given symmetric, positive-definite matrix.

The PSM estimator uses a treatment model, $p(x_i, t, \gamma)$ to model the conditional probability that observation i receives treatment t given covariates x_i. The literature calls $p(x_i, t, \gamma)$ a propensity score, and PSM matches on the estimated propensity scores. In this specific case, we use a probit model to compute these propensity scores using x_i covariates. Each institution is denoted by i and the set of covariates used in the probit model are: two dummy variables capturing the type of institution (i.e. pre-1992, post-1992), one variable named Size that quantifies the institution's size, one variable named Medicine that contains information about number of students enrolled in medicine subjects and the proportion of grant income plus tuition fees received in each institution.

Using this distance definition with matching on the estimated propensity score, we find that the set of NN indices for observation $i, i = 1, \dots, n$, is:

$$\Omega_m^P(i) = \{\langle j_1, j_2, \dots, j_{mi} | t_{jk} = 1 - t_i, |\hat{p}_i(t) - \hat{p}_{jk}(t)\rangle < |\hat{p}_i(t) - \hat{p}_I(t)|, t_I = t_i, I \neq j_k\}$$

Where $\hat{p}_i(t) = p(x_i, t, \hat{\gamma})$. As was the case with the NNM estimator, m_i is the smallest number such that the number of elements in each set, $m_i = |\Omega_m^P(i)| = \sum_{j \in \Omega_m^P(i)} w_j$ is at least m, the desired number of matches, set by the nneighbors (#) option. Once a matching set is computed for each observation, the potential-outcome mean, ATE, and ATET computations are derived and are identical to those of NNM. The ATE and ATET[148] standard errors, however, must be adjusted because the treatment model parameters were estimated; (see Abadie and Imbens, 2012). Hence the variances for $\widehat{\tau_1}$ and $\widehat{\delta_1}$ must be adjusted because we use $\hat{\gamma}$ instead of γ. Interestingly, the adjustment term for ATE is always negative, leading to smaller standard errors: matching based on estimated propensity scores turns out to be more efficient than matching based on true propensity scores. However, for ATT the adjustment term has two components and can be positive or negative, so the standard errors[149] may be too large or too small.

[148] For further clarification for the within treatment matches see Abadie and Imbens (2012).
[149] These are the standard errors reported by psmatch2 that need to be adjusted.

5.5.2 Results

Some summary statistics about the treated and control groups reveal that there are 266 observations in total within the sample that receive the treatment during the examination and 1,929 observations that can be used as control observations (see Table 14).

Table 14: Sample sizes for treated and control universities by year

Year	Treated Observations	Control Observations	Total
1996-97	1	137	138
1997-98	4	131	135
1998-99	6	126	132
1999-00	7	124	131
2000-01	10	118	128
2001-02	12	118	130
2002-03	13	118	131
2003-04	14	116	130
2004-05	18	111	129
2005-06	20	108	128
2006-07	22	107	129
2007-08	23	104	127
2008-09	24	102	126
2009-10	23	103	126
2010-11	24	102	126
2011-12	23	103	126
2012-13	24	101	125
Total	268	1,929	2,197

The number of treated observations varies from year to year since some universities have completely dropped from the sample due to closure, or some new universities have entered the HE system.

Table 15: Mean technical efficiency for treated and untreated universities

Groups	Treated		Untreated	
Variable	Mean	Std.dev	Mean	Std.dev
Bias_Crs Efficiency	0.658	0.138	0.629	0.164
Bias_Vrs Efficiency	0.796	0.172	0.690	0.190
pre1992	0.511	0.501	0.392	0.488
post1992	0.250	0.434	0.241	0.428
Size	14.363	7.035	7.951	6.167
Medicine	1,188.92	1,551.19	246.43	652.89
$\mathrm{Grants + Tuition/Total\ Income}$	0.669	0.183	0.721	0.150

Some summary statistics reveal that treatment observations have, on average, greater efficiency levels compared to non-treated observations (see Table 15). The result is more pronounced when we use a VRS assumption instead of a constant. However, this simple difference in the average efficiency level cannot be sufficient to verify our results that, indeed, those treated institutions tend to be more efficient. As we can observe from the table, there are some other characteristics that may account for those differences upon the mean efficiency between treated and control observations. So, only after matching our groups conditional on those characteristics can we extract a safe extrapolation of our result that the outcome variable (efficiency level) is higher for those treated. On average, treated institutions tend to be larger sized universities, they tend to have more students studying medicine subjects, and the received grants and tuition fee income as a proportion of the total income is lower than that of untreated institutions.

However, what remains meaningful in this stage is to create balanced samples in observational studies with a binary treatment where the control group data can be reweighted to match the covariate moments in the treatment group. So the estimation of causal effects in observational studies entails a pre-processing step of the data under the selection on observables assumption. In this way the covariate distribution of the control group data is adjusted by reweighting or discarding of units, so as to become more similar to the covariate distribution in the treatment group (Abadie and Imbens 2011). Popular techniques used in the literature to pre-process data prior to the estimation of causal effects are the nearest neighbour or the propensity score matching (Ho, et al., 2007; Sekhon 2009). Those methods produce a suitable weighting that balances the covariate distributions. However on the shortcoming of those matching techniques is their lack to jointly balance out all of the covariates (Diamond and Sekhon 2013; Iacus, et al., 2012).

Therefore a direct covariate balance is offered below (see Table 16 and 17) using entropy balancing, a method described in Hainmueller (2012). In this way we achieve a reweighting scheme that enables users to fit weights that satisfy a potentially large set of balance constraints. So a weight function that is used to adjust the control units is built that offers exact balance on the first, second, and possibly higher order moments of the covariate distributions in the treatment and the reweighted control group.

Table 16: Summary statistics without weighting before matching

Before matching	**Treated**			**Untreated**		
Variable	Mean	Variance	Skewness	Mean	Variance	Skewness
Bias_Crs Efficiency	0.657	0.019	-0.0005	0.628	0.026	0.115
Bias_Vrs Efficiency	0.796	0.029	-1.163	0.690	0.036	-0.639
pre1992	0.511	0.250	-0.044	0.392	0.238	0.440
post1992	0.25	0.188	1.155	0.241	0.183	1.211
Size	14.36	49.49	0.328	7.951	38.03	0.592
Medicine	1189	2406206	0.9143	246.4	426273	2.965
$\frac{Grants + Tuition}{Total\ Income}$	0.669	0.033	-0.191	0.721	0.022	-1.023

Table 17: Summary statistics after balancing weighting-matching

After matching	**Treated**			**Untreated**		
Variable	Mean	Variance	Skewness	Mean	Variance	Skewness
Bias_Crs Efficiency	0.657	0.019	-0.0005	0.657	0.019	0.001
Bias_Vrs Efficiency	0.796	0.029	-1.163	0.796	0.029	-1.161
pre1992	0.511	0.250	-0.044	0.511	0.250	-0.044
post1992	0.250	0.188	1.155	0.251	0.188	1.149
Size	14.36	49.49	0.328	14.36	49.48	0.329
Medicine	1189	2406206	0.9143	1189	2405782	0.914

$\frac{Grants + Tuition}{Total\ Income}$	0.669	0.033	-0.191	0.669	0.033	-0.189

In tables 16 and 17 summary statistics are offered for the covariates included in the reweighting. Specifically, in table 17, for each covariate a set of balance constraints is specified so as the moments of the covariate distribution between the treatment and the reweighted control group to be equal. The moment constraints here include the mean (first moment), the variance (second moment), and the skewness (third moment).

In table 16 we obtain differences in the covariate moments between treated and control universities, since, for each moment condition the results are different between treated and control units. In this preliminary analysis, the MTE estimate for those treated seems to be higher than the control units (0.796>0.690 under VRS and 0.657 >0.628 under CRS). However after matching in table 17 we adjust the covariate density of the reweighted control group. Given this two-group setup, the data from the control units are reweighted to match a set of moments computed from the data of the treated units. In this way we achieve exact covariate balance for all moments (here 1st, 2nd and 3rd), included in the reweighting scheme[150] so the specified moment conditions (mean, variance and skewness) of treated and control units are equal (Hainmueller and Xu, 2013).

In practice, the success of any propensity score method hinges on the quality of the estimated propensity score. Entropy balancing[151] is an alternative weighting method proposed by Hainmueller (2012) to estimate ATET for the population. The weighting methods (Robins et al., 1994; Hirano and Imbens, 2001) are usually more sensitive to model misspecification than matching (Rosenbaum and Rubin, 1985; Abadie and Imbens, 2006) and stratification (Rosenbaum and Rubin, 1984) so in what it follows we first utilize traditional statistical methods based on the propensity score to estimate the mean causal effect of mergers on MTE.

Table 18: Average treatment effect and the average treatment effect on the treated

Outcome Variable	Average effects	Coefficient	Std. Error	P>|z|	95% Conf. Interval	
VRS Efficiency	ATE	0.087***	0.021	0.000	0.459	0.128
	ATET	0.070***	0.012	0.000	0.046	0.094

[150] Here entropy balancing weights are used since we aim to avoid the time-consuming search over logistic or probit propensity score models to find a suitable balancing solution. After the pre-processing of the data any standard estimator for the subsequent analysis in the reweighted data can be easily applied.

[151] Entropy balancing is doubly robust weighting method that exactly matches the covariate moments for the different experimental groups in its optimization problem. Follow Zhao and Percival (2017) for the theoretical proof results and simulations that suggest the superiority of entropy balancing as an alternative to the conventional weighting estimators based on propensity score estimated by ML.

CRS Efficiency	ATE	0.038**	0.017	0.031	0.003	0.072
	ATET	0.049***	0.009	0.000	0.030	0.068

Regardless of the assumptions of scale returns used in the efficiency computation, the CATE on the outcome variable is higher (see Table 18). So, on average, the efficiency level in the treated group compared to the control group is greater after the treatment realisation, conditional on x_i characteristics described above. The ATE and the ATET are computed by matching first each subject to a single subject with the opposite treatment whose propensity score is closest. In this study, each subject was matched to multiple subjects and, more specifically, to three other subjects, since matching on more distant neighbours can reduce the variance of the estimator at a cost of an increase in bias. However, the ATE requires finding matches for both treated and control subjects. In contrast, estimating the ATET only requires finding matches for the treated subjects. One comparative advantage of the ATET is that, when we evaluate the policy implications of a particular intervention, in this case merger activity, we intend to access the effects of the treatment (merger) not just on the whole population but specifically for those to whom the treatment is administered.

There are two main caveats that should be discussed thoroughly. The first problem pertains to the fact that, when the software is searching for the appropriate matching pair on the aggregate level, the available control observations that can be served as potential matches contain universities that have been merged after a certain point. As a result, they may exist up to a certain time point (merger year) and after this they vanish from the sample since they are codified in a new entity (merged university). This may be problematic for the robustness of our results since it is counterfactual to match a university in the current year of examination with a pre-merger university that exists only in some earlier years, even if they perfectly matched on the x_i covariates. The most efficient way to tackle such a problem and reduce any bias is to fix the program to read the data per year. Therefore, the ATET is computed by year of examination (see Table19).

Table 19: Average treatment effect by year

Outcome variable	**CRS Efficiency**		**VRS Efficiency**	
Year	**ATET**	**AI Robust Std. error**	**ATET**	**AI Robust Std. error**
1997-98	0.002 (0.955)	0.037	0.007 (0.887)	0.049
1998-99	-0.017 (0.730)	0.049	-0.008 (0.851)	0.047

1999-00	0.058* (0.067)	0.031	0.113*** (0.000)	0.026
2000-01	0.061 (0.116)	0.039	0.109*** (0.006)	0.039
2001-02	0.071*** (0.004)	0.025	0.139*** (0.000)	0.028
2002-03	0.040 (0.110)	0.025	0.101 (0.008)	0.038
2003-04	0.085* (0.010)	0.032	0.146*** (0.000)	0.040
2004-05	0.032*** (0.000)	0.007	0.073** (0.031)	0.034
2005-06	0.097*** (0.000)	0.020	0.143*** (0.000)	0.033
2006-07	0.052 (0.013)	0.021	0.083*** (0.000)	0.023
2007-08	0.032 (0.390)	0.037	0.048 (0.314)	0.048
2008-09	0.048* (0.098)	0.028	0.093** (0.016)	0.038
2009-10	0.019 (0.314)	0.020	0.057** (0.050)	0.029
2010-11	-0.006 (0.807)	0.027	0.010 (0.797)	0.041
2011-12	0.032 (0.175)	0.024	0.063** (0.025)	0.028
2012-13	0.040* (0.068)	0.022	0.057* (0.074)	0.032

Note: Significance level *P<0.10; **P<0.05; ***P<0.01

The first year of examination, 1996–97, of the original sample is omitted since the low number of mergers has prevented the computational process. The most obvious finding to emerge from the analysis is that efficiency levels tend to be higher for those treated if they are treated compared to the counterfactual case that they had not been treated. Even though the results are not verified in every year, there is no strong evidence for the opposite case (see Table 19). Hence, there is no year in which the treatment effect results in a lower efficiency case. Any weaknesses in this model may be attributed in the huge reduction of the sample size and also to the available matching pairs per year.

A second pathogeny to tackle is the trustworthiness of our estimates. We have to guarantee that, once we control for observable characteristics, it is as if merged universities had been randomly assigned to control and treatment groups. Covariates are balanced in experimental data because treatment assignment is independent of the covariates due to the study design. In an experiment, the similarity of the characteristics between treatment and control groups is a straightforward process. Therefore, a simple glance at the data can provide a reliable source of how equivalent they are. Indeed, this is not the case with observational data. Therefore, for observational data, we model the treatment assignment process and, provided that our model is correct, the treatment assignment process is considered as good as random, conditional on the covariates of our model. In this case, covariates must be balanced by weighting or matching in observational data because treatment assignment is related to the covariates that may also have an impact on the outcome of interest. Then, it is perceptible that it is imperative to check the balancing property of the estimated propensity scores.

The estimators implemented in the majority of the available software used in the literature use a model or a matching method to make the outcome conditionally independent of the treatment by conditioning on covariates. A covariate is considered to be balanced when its distribution does not vary over treatment levels.[152] Hence, every time this model or matching method is well specified, it should balance the covariates. To this end, some balance diagnostic techniques have been developed, as well as formal tests implemented to check the specification of the conditioning methods used.

Imai and Ratkovic (2014) derived an official test for balance checking.[153] A weakness of this test is that it can be implemented only in cases in which the inverse probability-weighted (IPW) estimator, the augmented inverse probability-weighted (AIPW) estimator, or the IPW with regression adjustment estimator (IPWRA) are used.[154] So, when we use the teffects command for PSM, we cannot test the balancing property with an official test, only with exploratory diagnostic techniques. Such a test calculates the mean differences and the variance ratio between raw and weighted covariates, graphically, by plotting the distribution before fitting our model and the distribution after weighting; providing diagnostic box plots of the matched data can also indicate covariate balance.

[152]For further information on introductions to covariate balance, see Austin (2009, 2011) and Guo and Fraser (2015, sec. 5.52).

[153] For a formal test in STATA, there is an existing routine tebalance overid command.

[154] For further detail regarding the statistical methods used in these models and the relevant commands, see Brookhart et al. (2013) and Glynn and Quinn (2010).

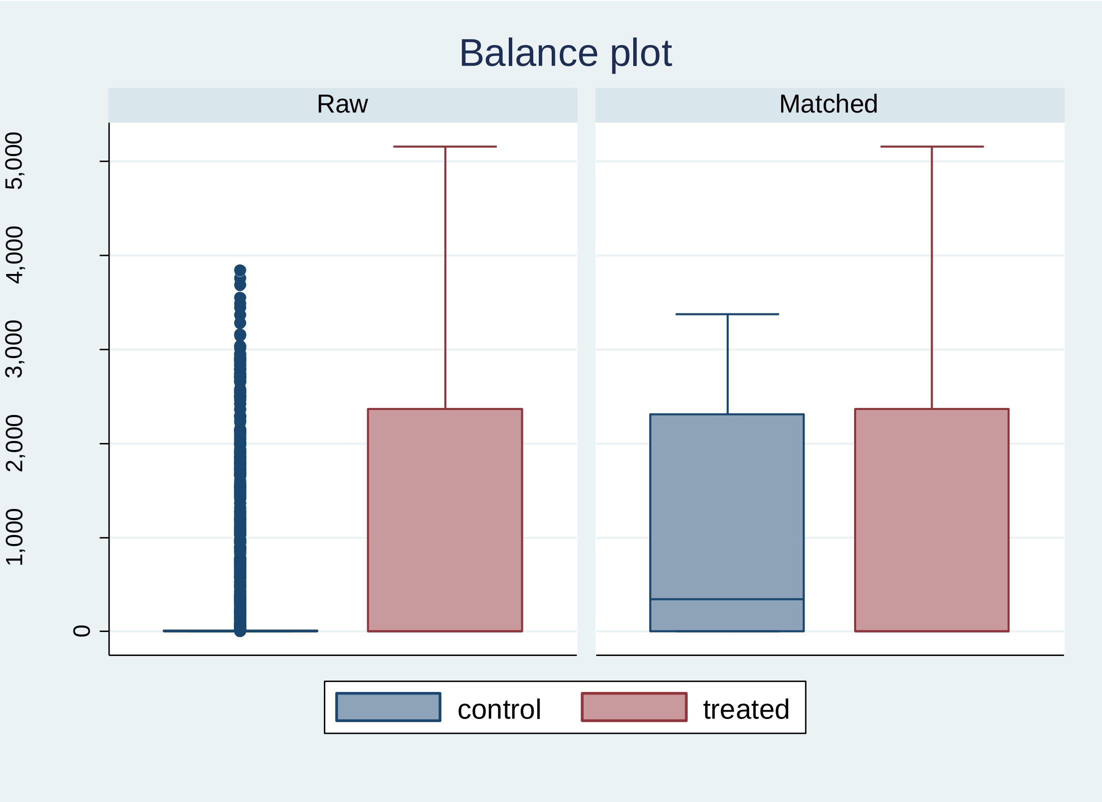
According to Austin (2009), an indication of the extent to which the covariates are unbalanced can be derived by summarising the covariates by group for the raw data and the weighted data. The differences can be explored further with standardised differences and variance ratios.

Table 20: Diagnostics for balancing property

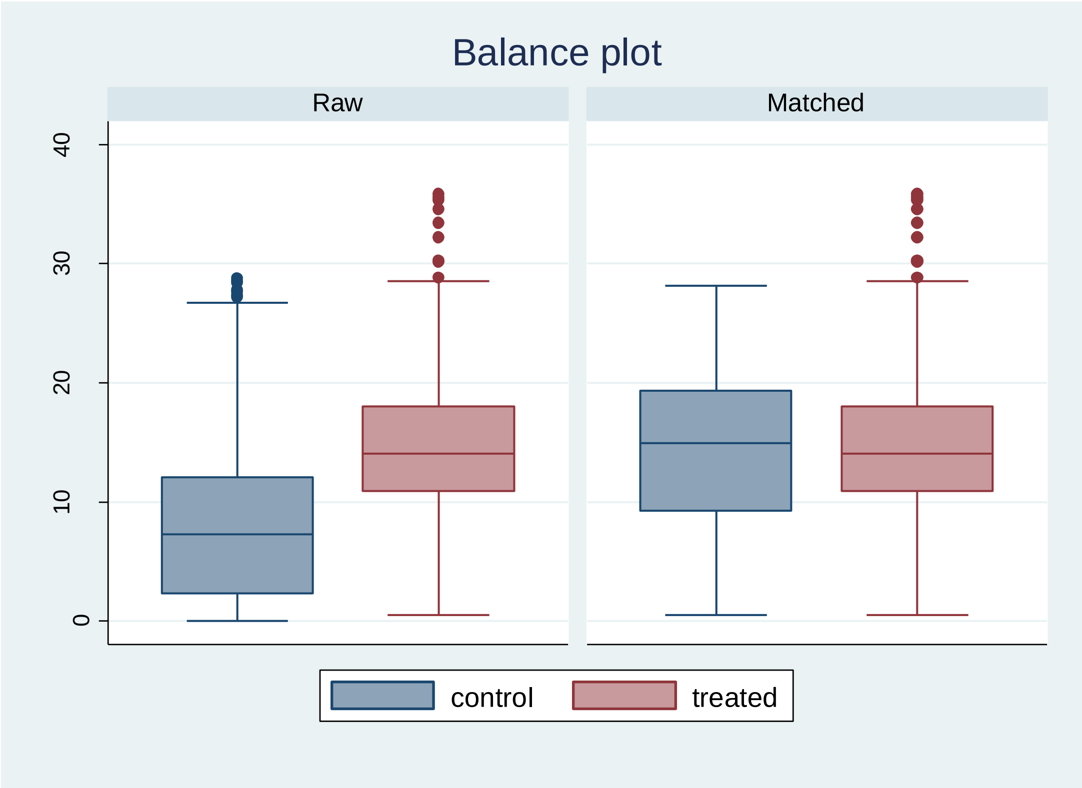
	Standardised Differences		**Variance Ratio**	
Covariates/Data	Raw	Matched	Raw	Matched
pre1992	0.240	0.094	1.051	1.004
post1992	0.020	-0.075	1.028	0.922
Size	0.969	-0.036	1.301	0.853
Medicine	0.791	0.123	5.644	1.454
(grants+tuition)/(total income)	-0.310	-0.140	1.486	1.698

Generally, in this type of diagnostic check, we need some benchmark values to compare our results. A perfectly balanced covariate has a standardised difference of zero and a variance ratio of one. There are no standard errors in these statistics, so inference is informal. The weighted standardised differences are all close to zero, mainly in the matched data. In particular, variable size seems to indicate that size may not be balanced by our model. The variance ratios are all close to one. However, output on the raw data for the variable Medicine tends to be markedly high. The weighted standardised difference is close to zero, but the weighted variance ratio still appears to be considerably less than one (see Table 20). Therefore, further exploration of the summary statistics permits the generation of diagnostic box plots (see Graph 1 and 2). The box plots reveal further information about the variability of the covariate between treated and control observations prior to and after the matching process.

Graph 1: variability of universities offering medicine subjects between treated and control observations prior and after the matching process



Graph 2: variability in universities' size between treated and control observations prior and after the matching process



The box plots of the raw data indicate huge variations between the control and treated groups. After matching, even if the situation improves and we obtain more balanced results, on average, both covariates seem to be imbalanced. However, in such a diagnostic process, we do not have standard errors in these statistics, so we cannot make any formal conclusions.

Therefore, a formal and valid answer on whether the balancing property is satisfied or not is been given, so only a formal test can provide a consistent answer. As mentioned earlier, an up-to-date test is offered for teffects by Imai and Ratkovic (2014). The only restriction here is that the estimation process should follow an IPW estimator rather than a propensity score estimator. IPTW using the propensity score is a method initially developed by Rosenbaum (1987), as a model-based standardisation process.[155] The method assigns weights based on the propensity score to eliminate the missing data problem stemming from the fact that each unit is observed in one of the two potential cases (treated or not). The desideratum here is the composition of a sample in which the distribution of the baseline covariates is independent of treatment assignment.

[155] IPW dates back to the 1950s (Horvitz and Thompson, 1952). Later work by econometricians extended the given framework to handle modern problems and estimation methods: see Robins and Rotnitzky (1995), Robins, Rotnitzky and Zhao (1994, 1995), and Wooldridge (2002, 2007).

This is a two-stage procedure since, at first, the parameters of the treatment model[156] are estimated and the inverse probability weights are computed. Next, the weights are used to compute the weighted averages of the observed outcomes for each treatment level to estimate the means of the potential outcomes. IPW estimators model the probability of treatment without any assumptions about the functional form for the outcome model. As mentioned earlier, $\mathrm{D_i}$ is an indicator variable denoting whether or not the $i-th$ unit was treated $D \in \{0,1\}$. In addition, let $\mathrm{p}(x_i)$ denote the propensity score for the $i-th$ unit, which is the probability that $\mathrm{D_i} = 1$ and is a function of covariates x_i. Weights are denoted by (Austin, 2011):

$$\mathrm{w_i} = \frac{\mathrm{D_i}}{\mathrm{p}(x_i)} + \frac{(1-\mathrm{D_i})}{1-\mathrm{p}(x_i)}$$

A unit's weight is equal to the inverse of the estimated probability of receiving the treatment that the unit actually received. Outcomes of units that are likely to receive treatment obtain a weight close to one, where outcomes of units that are unlikely receiving the treatment receive a weight larger than one[157], potentially much larger. According to Lunceford and Davidian[158] (2004, as cited in Austin, 2011), a variety of estimators for treatment effect exist based on IPTW. If the outcome variable is $\mathrm{Y_i}$, measured on the $i-th$ unit, a contrast of these weighted averages of the outcomes provides the estimates of the ATEs. An estimate of this:

$$\mathrm{ATE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \frac{\mathrm{D_i Y_i}}{\mathrm{p}(x_i)} - \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \frac{(1-\mathrm{D_i})\mathrm{Y_i}}{1-\mathrm{p}(x_i)}$$

Where n denotes the number of units. The first term is the potential outcome mean and for $\mathrm{D_i} = 1$ and potential outcomes y_0 and y_1 is equal to $E(y_1)$. If using the observed data $y_{1i} = D_i Y_i$ when $D_i = 1$ are always observed, the weights will all equal one. What should be noted here is that the weights may be inaccurate or unstable for units with a very low probability of receiving the treatment received, and the IPW estimator delivers higher weights for those observations for which y_{1i} is observed, even though its observation is not likely. However, if we want to calculate the ATT, then the weights assigned to the model are equal to:

$$w_{i,ATT} = D_i + \frac{(1-D_i)\mathrm{p}(x_i)}{1-\mathrm{p}(x_i)}$$

[156] Probit, logit, or heteroscedastic probit.
[157] A difficulty that can arise when using the weights described previously is that treated subjects with a very low propensity score can result in a very large weight (Austin and Stuart 2015).
[158] Lunceford and Davidian (2004) describe the theoretical properties of this estimator.

This method generates consistent estimates of the effect parameters because the treatment is considered to be independent of the potential outcomes after conditioning on the covariates. However IPW estimators model the probability of treatment without any assumptions about the functional form for the outcome model. In cases in which the overlap assumption is nearly violated, some of the inverse-probability weights become extremely large. Consequently, IPW estimators generate erratic estimates, and the large-sample distribution is a weak approximation of the finite-sample distribution of IPW estimators. The overlap assumption ensures that predicted inverse-probability weights do not become too large, but when the assumption is not valid even with a correct functional form for the treatment model, instability still occurs. IPW estimation methods have been used extensively in the modern treatment-effect estimation literature, (for further analysis see, Imbens, 2000; Hirano et al., 2003; Van der Laan and Robins, 2003; Curtis et al., 2007; Tan, 2010; Austin and Stuart, 2015).

In the next section, the results using an IPW estimator are presented. Once more, comparing merged (treated) with non-merged universities (untreated) reveals that the effect of merger on the outcome variable, which is the efficiency level that each university accrues, is higher in the merged than the non-merged. Also, our output reports that the average efficiency would be 0.743 (VRS) or 0.660 (under CRS) if all universities were merged and 0.694 (VRS) or 0.626 (under CRS) if none of the universities had been merged (see Table 21). So, again, the results imply a higher tendency of efficiency when merger activity occurs.

Table 21: Average efficiency by using inverse probability of treatment weighting (IPTW)

Outcome Variable	**Average Effects**	Coefficient	Robust Std. Error	P>|z|	95% Conf. Interval	
VRS Efficiency	ATE	0.048***	0.018	0.009	0.012	0.085
	ATET	0.071***	0.011	0.000	0.049	0.092
	POmean treated	0.743***	0.018	0.000	0.707	0.778
	POmean untreated	0.694***	0.004	0.000	0.686	0.702
CRS Efficiency	ATE	0.034**	0.015	0.029	0.003	0.066
	ATET	0.051***	0.008	0.000	0.033	0.068
	POmean treated	0.660***	0.015	0.000	0.630	0.691
	POmean untreated	0.626***	0.003	0.000	0.619	0.632

The estimation technique implemented by the IPW estimator is such that we do not need to correct the standard errors for any uncertainty associated with the predicted treatment probabilities. The results are presented only at the aggregate level since the analysis by year lacks validation for the implied overlap assumption.

In a model-based approach, the Imai and Ratkovic (2014) test checks whether the propensity score is correctly specified. We include all five covariates since they need to be balanced if the propensity-score model is correctly specified. So, the last diagnostic test performed here regards the assumption of balance in covariates used in the probit model. Intuitively, the score equations for the merged (treated) and non-merged (control) groups should be the same. We can test whether this is the case by using the score equations as moments in an over-identification test. The null hypothesis is that our covariates are balanced, so the propensity score model is correctly specified. Here, $\mathrm{Pr} > \mathrm{x}^2 = 0.1888$, so we cannot reject the null hypothesis. This implies that there is strong evidence that our covariates remain balanced. If this is not the case, powers and interaction terms or higher order terms of the base covariates are included to ensure a propensity-score is specified correctly.

5.5.3 Conclusion

The problem with the analysis in Chapter 4 is the use of empirical data rather than data from randomised trials; therefore, assignment to the group of merged universities is non-random. This can potentially cause selection bias or an endogeneity problem (Gerfin and Lechner, 2002), since either the decision to merge (or not to merge) may be caused by several factors that are usually not observed in the empirical data (Caliendo and Kopeinig, 2008), or there may be a two-way (endogenous) relationship between efficiency and merger activity,[159] both of which can bias the results. To assess the robustness of the results reported in the previous section, we redo the analysis using propensity score matching, a method proposed by Rosenbaum and Rubin (1983) to reduce the bias in the estimation of treatment effects (here, the effect of merger) with empirical datasets.

The findings in this chapter will doubtless be much scrutinised, but there are some immediately dependable conclusions from the comparison of merged HEIs with a sample of non-merged HEIs matched on the basis of characteristics such as type of HEI, size, number of students studying medicine, and source of funding. These results confirm the findings of Chapter 4 since the conditional effect of merger on efficiency is significantly higher in the merged than in the matched non-merging HEIs.

[159] Preliminary evidence does not, in fact, point to a two-way relationship (Johnes and Tsionas, 2014).

6. Chapter 6: Persistent and Transient Cost Inefficiency in the English HE Sector: A Generalised True Random Effects model

6.1 Introduction

Drivers of efficiency and value for money are core operational priorities and central concerns for the HE sector in the UK. In a climate in which public funding is becoming increasingly scarce and competition from other recipients of public funds such as healthcare is increasing, policymakers have found themselves asking how to ensure recurrent savings that will enable them to make necessary investments and, simultaneously, to define the extent to which HEIs productively should allocate their resources (Salerno, 2003).

The sector is adapting to a more competitive environment, and we must recognise and master the complexity of this turbulent landscape if there is substantial scope to achieve further cost savings and embed a continuous commitment to efficiency (UUK, 2011). However, significant progress has already been made, with the UK HE sector moving towards a ten-year track record of delivering efficiencies. According to the latest report by the HEFCE, efficiencies totalling more than £1 billion were delivered between 2011 and 2014. The UK HE sector has made significant progress in delivering efficiency and cost savings in both operational and academic areas over a sustained period of time (Jackson, 2013). Specifically, English universities tend consistently to have met efficiency targets that had been set in successive Comprehensive Spending Reviews (UUK, 2015).[160]

Despite the efficiency improvements and the cost-saving achievements due to the reduction in public funding since 2010, institutions that thrive in the new funding environment require decision making based on maintaining excellence and long-term sustainability of the increased surplus (UUK, 2015).[161] At the same time, controlling cost and maximising efficiency or eliminating inefficiency generates the concept of extrapolating further the well-established objective of cost efficiency. In light of the

[160] In total, £1.38 billion of efficiencies were reported against a cumulative target of £1.23 billion; universities in Scotland, Wales, and Northern Ireland have also had to attain efficiency targets and funding settlements over a similar period and have placed a similar emphasis on efficiency.

[161] Cuts in capital funding, plus the infrastructure expansion to remain competitive, have led universities to greater surpluses generated by efficiency savings in order to fund investment.

funding constraints of the UK government policy on the provision of the UK HE over the past decade, a rise has been seen in the number of studies seeking to assess productivity and cost efficiency. The development of parametric and non-parametric techniques for estimating efficiency in a more applied setting beyond theoretical frameworks has offered radical changes in the repository of empirical research. Researchers have gained both the ability and the flexibility to model complex production processes and cost structures within HEI (Salerno, 2003).

However, it might be kept in mind that HEIs are likely to produce using their inputs in a suboptimal way (Johnes, 2004b). Hence, the evaluation of their economic performance is harshening due to the presence of inefficiency. A common inference across studies, regardless of the underlying assumptions, the econometric technique, or the default educational system of examination, is the presence of inefficiency and, in some extensions, the determinants of inefficiency, since existing inefficiency has to be eliminated in the long run. The early work by Farell (1957) has been developed to identify the degree of sub-optimality along both non-parametric lines using established linear programming techniques and, more prominently, the method of DEA (Charnes et al., 1978), and parametric techniques, giving rise to SFA developed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977). The econometric assumptions underlying SFA impose that the error term can be disintegrated into two components with different distributions.[162] The first refers to ‘inefficiency’, and it is asymmetrically distributed most commonly as a semi-normal, while the second embodies the random ‘error’, and follows a white noise distribution.[163] In the classic cross-sectional SFA, all inefficiency is random and there is no explanation of its presence related to observable features of the DMU. Indeed, this is an extreme supposition since, in a more realistic framework, inefficiency can be determined through observables. Therefore, apart from the estimated level of inefficiency, researchers intend to define what other factors account for inefficiency (Parmeter and Kumbhakar, 2014).

Inefficiency can arise from different factors and not all of them can be controlled or curtailed by the institutions themselves. So, inefficiency can be an amalgam of different parameters, determined endogenously by the institutions or exogenously by policymakers and government officials. In this added layer of complexity, it is evident that factors such as the scale and scope of operations, the ownership structure, the use of information technology, the internal or external source of funding, the market share, and the geographic location, etc., may firmly define the inefficiency level of HEIs across adapting environments. In the meantime, stakeholders might be interested in whether regulations such as allowing institutions to merge improve universities' performance.

Some of the long-term factors that stimulate permanent differences among institutions, in many cases, cannot be modified by the institutions. As a result, institutions aim to be different from each other and institutional diversity is one of the main features of a

[162] See Kumbhakar and Lovell (2000) for analytical detail on SFA.

[163] It is imperative to assume that both components are uncorrelated (independent) to avoid distortions in the estimates.

healthy HE system (Ramsden, 2012). This diversity is desirable to an extent since the intention is to cover graded mental and learning abilities among students (students' quality),[164] different socio-economic statuses, academic orientations (research and/or teaching-intensive universities), subject variety, different funding and tuition fee regimes (Smith, 2007), further expansion in student enrolment, growing internationalisation in academic as well as in student recruitment, and considerable funding cuts, at least in countries most affected by austerity-related public expenditure cutbacks, including the UK HE (Lodge, 2015). Therefore, in the measurement of efficiency lurks the danger of being confounded with legitimate differences between institutions in cost and production structures.

Specifically, the English HE sector has remained highly diverse in many significant respects (Ramsden, 2012), so the position and shape of the cost frontier needs to be evaluated separately for each of a number of groups of institutions (Johnes and Johnes, 2013). The cost structure is altered upon the special characteristics that may vary widely. So, different groups of institutions circumscribe different challenges and missions and acquire different statuses. Therefore, the time-invariant and permanent characteristics should be eliminated from the efficiency term since individual specific heterogeneity among institutions can boost or undermine the efficiency estimates. In a simple SFA framework, there is no allowance for each HEI to have a different set of objectives or missions, unlike DEA and a random parameter SFA.[165] Thus, any efficiency estimations derived using the classical SFA should be interpreted with caution (Johnes and Johnes, 2013).

In short, the best evidence on HE efficiency is scattered among a diverse set of educational systems that are more prone to be different than similar in many aspects (Salerno, 2003). To this point, the accurate measurement of performance is based on the assumption that we know the 'true' technology among a range of production technologies. Although there are multiple available technologies, in the empirical research, there is a latent assumption that all institutions in the sample use the same technology, i.e. homogeneous production technology (Parmeter and Kumbhakar, 2014). Thus, if the unobserved technological differences are disregarded during the estimation process, their effects might be inappropriately labelled as inefficiency.[166]

Unlike the context of a cross-section of institutions, observed at a single point in time, a panel of data where HEIs can be observed over time allows us to assess inefficiency, and, at the same time, enables a rather greater overview of how HEIs are operating. Hence, with a panel, we may allow for latent heterogeneity where this unobserved heterogeneity could be inefficiency or individual specific heterogeneity. This heterogeneity, as it follows, distorts efficiency estimates when measured with the standard traditional approaches. Related to this, it stands to reason that a question that

[164] It is possible that different institutions differ in terms of the quality of students they can attract, and in the quality of graduates that they produce (Johnes and Johnes, 2013).
[165] RP SFA acknowledges that each university varies in its mission and faces distinct circumstances affecting its costs; for more detail, see Johnes and Johnes (2009).
[166] In the literature, *a priori* classification is used to split the sample so that homogeneity of technology and efficiency per group is retained. However, this is not an efficient solution; see Parmeter and Kumbhakar (2014, p. 64).

needs to be considered with regard to the time-invariant individual effects is whether the individual effects represent (persistent) inefficiency, or whether the effects are independent of the inefficiency and capture (persistent) long-term fixed factors labelled as unobserved heterogeneity. Hence, when persistent inefficiency is taken into account, a more accurate estimation is feasible since, in such a diverse environment as the English HE sector, some of the long-term factors are not fixed, so they can mistakenly be assigned to heterogeneity.

Generally, panel data studies that offer information on units over time using SFA do recognise that there are econometric models that produce efficiency indicators that vary over time or others that provide time-invariant efficiency indicators. As a result, panel data enables us to examine whether inefficiency has been persistent over time or whether the inefficiency is time-varying. Indeed, two types of inefficiency are verifiable, one component of inefficiency that has been stable (persistent) over time and another that varies over time, namely a varying short term (transient).

As we will see later in the analysis, each component conveys different types of information and yields different implications because they are relatively different in absolute value and not highly correlated. Remarkably, though, this provides a further step for better evaluation of the policy implications in the English HE sector, since the distinction between transient and persistent cost efficiency is novel and relatively little attention has been paid to the difference between these two components of cost efficiency. It is vital to decompose cost inefficiency into those two components in advance since the persistent part offers further insight into the presence of structural problems in the cost-minimising process of HEIs, factor misallocations that are difficult to change over time, and the presence of systematic shortfalls in managerial capabilities that may exist. So, it is an indication of operational reconstitution for both the institutions themselves and those beholden to educational policy implementation. At the other end of the spectrum, the transient part of inefficiency indicates singular non-systematic failures in management or spending strategies that can be solved or modified in the short term without operational changes in a firm.

In the operating framework for HE in England, in which funding and, therefore, central control is exercised by the HEFCE on the basis of certain criteria (Lodge, 2015), the decomposition of short- and long-term efficiency can be seen as a first step to uncover such influences. The mutative determined regulation of England's HE sector displayed many features,[167] specifically, a growing emphasis on markets and on audit, concurrently with an emphasis on internal university management in teaching and research aspects within institutions (Scott, 2004). As a result, this could cause short- or long-term repercussions in the efficiency levels of the HE sector. Therefore, the aim of this chapter is to control for unobserved heterogeneity (unchangeable factors) in the English HE system without pulling out either the time-varying or time-invariant inefficiency components.

[167] For further detail on the criteria, see Lodges (2015, p. 4).

In a systematic review of the literature, attempts have been made to isolate the time-varying and the time-invariant components even solely or concurrently. In the first sequence of papers (Kumbhakar, 1991; Kumbhakar and Hjalmarsson, 1993, 1998; Kumbhakar and Heshmati, 1995), all the time-constant effects (time-invariant component) can be viewed as persistent inefficiency, so they actually cover only the persistent part of efficiency. Thus, all time-constant variables are treated as persistent inefficiency, suffering from a deficiency in covering any time-constant firm heterogeneity. So, if the latter is true and unobserved firm heterogeneity is present, the model is mis-specified and an upward bias in inefficiency is expected.[168] Alternatively, the models[169] introduced by Greene (2005a, 2005b) build on an extension of the classical SFA model introduced by Aigner et al., (1977), which view firm effects as something other than inefficiency. These models (TFE and TRE) separate unobserved time-invariant effects from time-varying efficiency estimates. Thus, these models fail to capture persistent inefficiency, which is confounded with firm effects and provides information in the transient part exclusively. Therefore, these models are mis-specified as well and undermine the overall efficiency estimates with a downward bias.

It is not possible to identify more than a handful of empirical studies that utilise the decomposition of the inefficiency term into persistent and transient components. Recent papers on this topic (Colombi, 2010; Colombi et al., 2011, 2014; Kumbhakar and Tsionas, 2012; Kumbhakar et al., 2014; Fillipini and Greene, 2016) offer a theoretical platform that overcomes several of the limitations of the previously discussed models and separates persistent from transient inefficiency while controlling for heterogeneity effects. However, there are special features regarding the estimation procedure and the distributional assumptions in each case.

Specifically, Kumbhakar and Tsionas (2012) and Kumbhakar et al. (2014) chose a simple multi-step procedure based on least square methods imposing distributional assumptions in several steps, so overall the method is deemed as inefficient relative to the single-step MLE. Using a familiar tractable estimation technique Colombi (2010) and Colombi et al. (2011, 2014) extended the aforementioned framework by deriving the full ML function using closed skew normal distributional assumptions for both the time-variant and time-invariant random components. In this way identification is ensured since a distribution for each component is specified. The extreme complexity and lengthy process is eliminated by Fillipini and Greene (2016), who utilised a simulation-based optimisation routine and applied the Butler and Morffitt (1982) formulation. Hence, due to the reasons described so far, the appointed analysis of this chapter will follow the econometric framework established by Fillipini and Greene (2016).

This study aims to identify any possible inexpediency in the estimation of cost in the English HE sector, when standard traditional evaluation methods are used compared to the generalised true random effects model (GTRE) applied here. A more accurate

[168] Since all unobserved heterogeneity is viewed as inefficiency.

[169] Similar models that account for latent heterogeneity but are confounded with persistent inefficiency have been developed by Kumbhakar and Wang (2005), Wang and Ho (2010), and Chen et al. (2014).

conclusion is feasible on whether the chosen estimation method determines the policy inference. The novelty of this study indicates whether there are long-term (persistent) burdens that render English universities inefficient or if there are only short-term rigidities in the cost structure of each university. Specifically, in short panels, the level of persistent inefficiency is a determinant for the institutions, since it reflects the effects of inputs such as management (Mundlak, 1961), as well as other unobserved inputs that vary across institutions but not over time. As we will see later in this study, the distinction between time-varying and time-invariant components is crucial from a policy perspective since the level of persistent inefficiency is liable to change only if management practices, ownership, or government regulations alter.

To this end, in the English HE sector, despite the financial health of HEIs, which continue to show a sound position overall, according to data reported to the HEFCE, there are individual institutions across the sector whose financial performance continues to vary significantly (HEFCE, 2016).[170] For this purpose, and given that the HE sector continues to make a crucial contribution to the UK's development as a world-leading advanced economy, it is vital from a policy perspective to identify whether short- or long-term inefficiencies have a significant impact on the cost policy each institution adopts. This identification constitutes to play a vital role in institutional management and for underpinning public accountability throughout HEIs in England.

Evaluating efficiency in the HE sector in England is already having a significant impact in areas of the public sector and in public policy, so it is vital to ascertain whether the applied method or the model's special features (i.e. specification) overestimate or underestimate efficiency levels. In the next section, a review of the literature is attempted, mainly focus on the UK HE sector, followed by the data analysis. Next, the salient points of the newly introduced method by Fillipini and Greene (2016) will be discussed, focusing on a cost framework. The final section contains a summary of the findings, as well as concluding remarks that identify whether persistent or transient inefficiency is dominating the English HE sector and raising strategic issues for policymakers.

6.2 Literature Review

The received literature concerning the cost structure of HEIs has been widely expanded since the 1960s, examining different aspects of the cost configuration. This great expansion of cost studies is apparent not only in the UK case, but also beyond its borders[171]. The two state-of-the-art methods applied in the literature on costs and cost

[170] Further detail is available in the HEFCE 2016 report (http://www.hefce.ac.uk/pubs/year/2016/201620/) on the difference in the levels of surplus achieved by HEIs.

[171] An extensive summary of the literature for previous cost efficiency studies in the HE sector in UK and worldwide is laid out in appendix 18 chapter 6.

efficiency are the non-parametric approach of DEA and the parametric approach of SFA.

The multiple-input multiple-output framework of DEA is relatively less well researched in terms of cost analysis in HE. No more than a handful of studies can be traced in the literature, with five studies utilising UK data (Athanasopoulos and Shale, 1997; Johnes, 1998; Johnes et al., 2005; Casu and Thanassoulis, 2006; Thanassoulis et al., 2011), two in Spain (Gimenez and Martinez, 2006; Agasisti and Salerno, 2007), and one in Canada (McMillan and Datta, 1998). DEA is a frontier technique and, therefore, inefficient observations are omitted from the frontier estimation (Johnes et al., 2013). Furthermore, DEA is advantageous from the point of view that it allows different HEIs to have different objectives in terms of the mix of outputs produced. Therefore, DEA studies offer greater flexibility for analysing cost,[172] despite the severe critiques and the opposing views it receives for a number of reasons.[173] The mean efficiency for the HE sector, in the UK, under this methodological framework varies from 0.863 to 0.90 (Thanassoulis et al., 2011).

Turning to the statistical estimation of the cost, the well-established SFA method modifies the standard statistical tool of the ordinary least square (OLS) regression to account for efficiency variation across DMUs. Consequently, the cost function is fitted through the data points with a bias towards the data points that indicate lower costs for a given output level (Johnes et al., 2005), estimating a cost frontier rather than fitting a line in the middle of a scatter plot. SFA has been applied extensively in the literature as a standard tool to evaluate efficiency by splitting the regression residuals into two components:[174] one capturing random error and the other inefficiencies. Going one step further, Johnes (1998a) and Johnes et al. (2005) verify how the comparison of the two methods yields different rankings for HEIs in the UK according to the efficiency estimates occurring in each method.

The first attempt in the UK HE system to estimate cost functions originated in the 1960s and relates to data from 1968. In this primitive study by Verry and Layard (1975), identification of the impact of the multiplicity of different outputs on cost and technology of production is profound. They estimate six separate linear cost functions, one for each subject area, lacking the possibility of synergy formulation due to the joint production of the various outputs produced. This means that any benefits or significant cost savings stemming from such a joint production are disregarded, even in a later comparison study by Verry and Davies (1976). However, in terms of comparability, this study constitutes a benchmark for later developments.

In later years, researchers outlined and analysed more broadly the multi-product nature of HEIs, by defining teaching outputs by broad subject areas and incorporating in the analysis non-linear complex functional forms. The first cost efficiency empirical research for the UK HE sector was conducted in the 1990s by Johnes (1996), and later

[172] Cost definition may vary even in a single analysis.
[173] For more detail on these reasons, see Johnes et al. (2005).
[174] Each component follows a specified (assumed) distribution. The measurement error is a normal distribution while the inefficiency term is a semi-normal one; see Kumbhakar and Lovell (2000) for analytical detail on stochastic frontier analysis.

by other researchers (Glass et al, 1995a, 1995b; Johnes, 1998). The main common factor of these studies is an extrapolation of the analysis into identifying economies of scale and scope for British institutions. So, they aimed to control for: RRS,[175] which tend to be significant and unexhausted for the typical British university; PSRS,[176] which seem to be mixed and not in the same direction; and global economies of scope,[177]which have contradictory results according to the level of aggregation of the outputs. An earlier study on cost in US HE by Cohn (1989) specified economies of scope in the private HE sector up to large levels of output and more than double the mean output level of the public HE sector (Johnes et al., 2005).

Building on this sequence of studies Johnes (1997), Izadi et al. (2002), Stevens (2005), Johnes et al. (2005, 2008b), Johnes and Johnes (2009), and Thanassoulis et al. (2011) continued the use of panel datasets, which soon became customary. Another issue touched upon by these studies was the composition of the UK HE sector after 1992, which has been subject to major changes. Taking into account that polytechnics have gained university status and that colleges of HE are allowed to apply for university status, the aforementioned studies cover a subset of the English HE sector, namely the extended HE sector. They have controlled for economies of scale and scope, with ray economies of scale being close to constant or decreasing for the typical HEI, and global diseconomies of scope confirmed in all applications. The utilisation of panel data reveals relatively lower efficiency than in cross-sectional data; however, efficiencies vary also by type of university,[178] choice of data, and estimation method.[179] Another advancement here is the disaggregation of output by subject and type of HEI, and, indeed, cost divergence is verified when subjects split. Although there have been significant improvements, limited focus has been placed on the analysis of additional variables.[180]

A more articulate idea of the cost evaluation of HEIs in the UK can be found in subsequent studies by Johnes et al. (2005), Johnes et al. (2008b), Johnes and Johnes 2009, and Thanassoulis et al. (2011). These authors have developed the standard traditional approaches by embodying more sophisticated frontier estimation methods, by including measures of third mission activities and committing to more compatible datasets with the composition of the HE sector after 1992. This composition may account for further differences in the distribution of efficiencies since there is a considerable range in efficiency, with institutions lying at the very low end of the distribution and others displaying a firm high ranking in performance (Johnes and Johnes, 2013). Consequently, the special characteristics and features of an institution

[175] That is, RTS stemming from a simultaneous increase in all types of output being produced in the institution.

[176] That is, RTS associated with an increase in one output only.

[177] That is, economies arising from producing all outputs together rather than separately.

[178] Colleges of HE appear to be the least efficient, and post-1992 and some pre-1992 HEIs are typically the most efficient (Johnes and Johnes, 2013).

[179] SFA estimates tend to be lower than DEA since, in the latter, there is flexibility in the set of objectives and missions of each university.

[180] Student quality and location of HEI have limited importance in the cost determination.

suggest that, due to idiosyncrasies (that are not adequately captured by the data on outputs), cost inefficiencies may occur.

According to Ramsden (2012), the diversity of the UK HE sector remains high in many significant respects; despite the concerns about a wholesale diminution of diversity, only a modest amount of homogeneity exists. This diversity is fuelled by the major changes in subject provision,[181] in the composition of the student body,[182] and in the universities' statuses since there are research and/or teaching-intensive institutions. In addition, universities tend to evolve in a historic context, with new institutions joining the sector and others proceeding in merger compliances or complete closures. Furthermore, institutions are spread across different locations and regions and, especially in the UK, can be categorised in general versus specialist providers, going by the plethora of the formally recognised subject areas they offer. Further divergence in the sector can be accounted for by major or minor differences in teaching methods and conducive or discouraging governance structures.

It is manifest that heterogeneity across institutions due to these structural differences is an element that cannot be overlooked if we want to derive robust efficiency estimates and evaluate the sector consistently. Under the SFA approach, all DMUs are assumed to be directly comparable; however, this might be a false assumption since resource allocations and cost determination for a given level of production may vary, reflecting differences in the cost and production structures rather than differences in efficiency.

Therefore, researchers have used a better filtering of institutions *a priori* by dividing them into pre-specified sub-groups to accommodate their missions, so mission groups were often formed historically. These predefined groups are based on the preconception that, within the sub-groups, costs ought to vary when discharged from other differences. In the UK HE system, the first considerable attempt to anticipate the austere heterogeneity was made by Glass et al., (1998), who discerned three research quality groups of institution. Almost one decade later, Johnes et al. (2005) and Casu and Thanassoulis (2006) estimated cost functions specific to certain pre-specified sub-groups[183] of institutions. Following this pattern of classification, other researchers (Johnes et al., 2008, Johnes and Johnes, 2009; Thanassoulis et al., 2011) have taken into consideration the subject mode, extending this framework into further classes.[184] Beyond the UK border, researchers have attempted equally to control for heterogeneity with different scopes. In the US, Robst (2001) and Mensah and Werner (2003) incorporated in their analysis categorical variables denoting Carnegie classification,[185] and Laband and Lentz (2003) made a distinction between public and private ownership. In Germany, Kempkes and Pohl (2008) classified universities into three groups, dependent on to whether they operate under comparatively liberal or restrictive state

[181] HE institutions have matched the changes in demand, as evidenced by applicant choices.
[182] I.e qualifications, level and mode of study, widening participation, and balance of home and international students etc.
[183] Authors have divided the sample into pre-1992 universities (traditional universities such as Oxford and Cambridge), post-1992 universities (mainly formed by former polytechnics that gained university status), and colleges of HE.
[184] Here, researchers have split the pre-1992 HEIs into two groups, with or without medical schools.
[185] For further insight into the Carnegie Classification, see http://carnegieclassifications.iu.edu/classification_descriptions/basic.php

regulation.[186] However, the pre-specified sub-grouping is not necessarily satisfactory since ascribing main attributes for categorisation might be misleading, as universities that may once have had similar missions may not necessarily have similar outlooks today.

Researchers have realised the determinant role of heterogeneity, so they proceed with more advanced econometrical operations, controlling for structural differences promptly through the estimation process. Consequently, they developed Random Parameter SFA models (RP-SFA) which were mainly used in the literature by various researchers i.e. Johnes and Johnes (2009) UK data; Johnes and Scharzenberger (2011) German data; Agasisti and Johnes (2010) Italian data; Agasisti and Johnes (2015) US data. On the upsides of the method is the fact that HEIs are allowed to have different objectives and acknowledge that each institution varies in its mission. The random parameters frontier estimation framework enables coefficients on one output to vary by HEI and cost functions for HEIs to be entirely different from one another. The superiority of the method is due to the fact that the data determines in a single framework how the cost functions of each institution should look when released from any preconceptions and separate estimation for pre-determined groups (Johnes and Johnes, 2013). However, estimation might be demanding in terms of data availability since it can be difficult to fit the model and estimate the parameters. According to the results, HEIs are definitely heterogeneous in terms of both cost structure and efficiency (Johnes and Johnes, 2009).

Another caste of models developed in the literature are the Latent Class Stochastic Frontier Analysis Models (LC-SFA), initiated by Johnes and Johnes (2013) English data and Agasisti and Johnes (2015) US data. Consider here that efficiency may vary across time for each group of institutions. The structure of the cost is not expected to be the same across all institutions due to the special features they acquire. These models have been generated in order to ensure that the compared institutions in each cluster are comparable when evaluating efficiency, since different institutions face different challenges and missions. In addition, the models permit for variety in the parameters of the cost structures in each group[187] and to evaluate the efficiency of each institution in each group, rather than calculating cost functions by predefined groups. The first attempt to combine the SFA with latent class models was made by Orea and Kumbhakar (2004), controlling Spanish bank heterogeneity, and Greene (2005), examining the US banking industry. In the HE sector, Agasisti and Johnes (2009) mapped the latent class models (LCM) with SFA using data from the US, allowing objectives to vary by group suggested by the data. Turning to the English HE sector, the findings reveal that estimates of efficiency suggest that the sector is highly efficient when heterogeneity is accounted for using LCM, since even controlling for heterogeneity through a relatively

[186] Best law group (operate under a relatively liberal legal framework), medium law group, and worst law group.

[187] They use the LCM method to let the data suggest distinct groups, establishing which institution comprises the membership of each group (Johnes and Johnes, 2013).

unrefined latent class modelling significantly reduces the variation of efficiency across HEIs.

Until now, the nature of the analysis has suggested that inefficiency is calculated as an unexplained residual. Indeed, researchers have tried to obtain more information on the determinants of inefficiency, pointing out the crucial role of separating inefficiency and fixed individual effects inherited in each institution. On a methodological ground, efficiency scores suffer from the incidental parameters problem in the most recent literature, which deals with panel data (Agasisti et al., 2016). This forced all time-invariant individual heterogeneity into estimated inefficiency since the fixed parameters, often unobservable, distort the estimates. The emphasis of this distinction between unobserved individual heterogeneity and inefficiency is underlined by Greene (2005a, 2005b), who added to the classical SFA a firm-specific time-invariant component. Ergo, all the time-invariant individual heterogeneity stemming from fixed factors is disregarded from the estimated inefficiency. However, for a complete sheltered estimator of the incidental parameters problem, Wang and Ho (2010) adjusted the generalised Battese and Coelli (1995) formulation in order to incorporate heterogeneity into panel data in the stochastic frontier model. In this approach, they show how any fixed individual effect can be eliminated before estimation when transforming the model by either first-difference or within-transformation. The aforementioned approach was placed in an applied framework for the HE sector by Agasisti et al. (2016), utilising Italian data.

Despite the significant contribution of the TFE and TRE proposed by Greene (2005 a, 2005b), who included a term for time-invariant unmeasured unobserved heterogeneity, critiques within the econometric literature state that this time-invariant component, apart from heterogeneity, integrates another part that can be modified by the universities. This part that is mistakenly incorporated in the individual time-invariant effect is an adaptable part that can be under the control of the institution in the long term. Therefore, this adjustable part should be integrated in the (in) efficiency value since it can be a criterion of inefficiency in the long term. Following this argument, the time-invariant component can be split into an individual specific part capturing heterogeneity and another part, namely the persistent (long-term) part of inefficiency that accumulates all the long-term (stable) factors of inefficiency. Consequently, apart from the time-varying component providing information on the transient (short-term) part of efficiency developed by Greene (2005a, 2005b), another type of efficiency is ascertainable on the trajectory of efficiency that estimates a persistent term.

The meaningful role of persistent inefficiency is recognised in early studies (Kumbhakar, 1991; Kumbhakar and Heshmati, 1995; Kumbhakar and Hjalmarsson, 1993, 1995); however, these tend to treat firm effects as persistent inefficiency so they fail to separate from firm effects. In the context of the use of efficiency models for policy making, or managerial considerations, the truth is somewhere in between since firm effects might be persistent inefficiency and part of persistent inefficiency might include unobserved firm effects. Therefore, none of the above approaches is satisfactory; thus, the problem of separating those elements is crucial. Hence, there is interest in an econometric model that will provide the theoretical platform to distinguish and estimate the persistent and the transient parts of inefficiency simultaneously. Recent papers on this topic (Colombi, 2010; Colombi et al., 2011, 2014; Kumbhakar and Tsionas, 2012; Kumbhakar et al., 2014; Tsionas and Kumbhakar, 2014) attempt to develop a fully flexible error specification with a four-way error component SFA, namely the GTRE (Kumbhakar and Tsionas, 2012).

In the HE literature, there has been little interest into producing separate estimates for the persistent and transient parts of the productive or cost efficiency. Therefore, it is relatively novel in terms of HE efficiency to include both arguments in the analysis; only two studies can be traced so far. The first (Titus et al., 2016a) pertains to the US HE sector, and finishes with the finding that cost inefficiency tends to be persistent rather than short term. This study applied a slightly modified version of the recently developed multi-step SFA technique by Kumbhakar et al. (2014), taking into account serial and spatial correlation. On the same wavelength, a more recent study was performed in the German HE sector (Gralka, 2016) confirming that German universities are prone to comprehensive and structural changes, if they seek for improved efficiency levels. By analysing various models, it is confirmed that the specification developed by Kumbhakar et al. (2014) improves the accuracy regarding the heterogeneity assumption and reveals that inefficiency tends to be long term and persistent. However, university ranking is still hesitant since it is likely to vary by method.

Building upon the multi-step method presented by Kumbhakar (2014), a cross-country comparison of the Italian and German HE sector was attempted by Agasisti and Gralka (2017) who tried to relate the transient efficiency part of (in) efficiency with annual changes and therefore performance-in the institutional level, and the persistent part, with peculiarities in the state specific structures of the HE system. However, the method by Kumbhakar (2014) is deemed as complex and greatly dependent on the underlying distributional assumptions so as argued in Heshmati et al. (2016) the model is inefficient relative to a simulated maximum likelihood (SML) estimation method as proposed later by Fillipini and Greene (2016). Applications of the Fillipini and Greene’s model can be traced in the aviation industry (Heshmati et al., 2016), Swiss hydropower sector (Fillipini et al., 2018) and electricity sector in New Zealand (Fillipini et al., 2016).

A common handicap in the estimation process of the GTRE approach is the extreme complexity of the log likelihood (noted by Colombi, 2010; Colombi et al., 2011, 2014) in the practical implementation of the method. Even the partial Bayesian solution proposed by Kumbhakar and Tsionas (2012) is not sufficient for a full practical implementation of the MLE due to the sensitivity of the analysis for informative priors. Consequently, an alternative econometric approach has been developed in the literature for the estimation of the GTRE based on MSL function that transcends the limited information of the multi-step procedure based on least squares regression proposed by Kumbhakar et al. (2014). Following Filippini and Greene (2016), a relatively more transparent and effective application of the GTRE model is achieved, defeating most of the aforementioned limitations. Thus, an in-depth analysis of the novel specification by Filippini and Greene (2016) will be attempted for the English HE sector, including a comparison of the results from the most frequently used models in the literature, identifying whether long- or short-term inefficiencies dominate the sector.

6.3 Empirical Analysis

6.3.1 Preliminaries

The main assumption shared by almost all panel data stochastic frontier models introduced in the early 1980s (Battese and Coelli, 1988; Kumbhakar, 1987; Pitt and Lett, 1981; Schmidt and Sickles, 1984) was that technical inefficiency is individual-specific and time-invariant. This gives further validity to the understanding that inefficiency levels may vary for different producers, but they do not change over time, and that failure to change over time is the mark of an inefficient institution. In some situations, inefficiency is associated with managerial abilities; and no tractable changes during the study period are likely in management and production technologies for any of the DMUs. However, a static technology is an unrealistic framework, particularly when market competition exists (Heshmati et al., 2016). Another source of friction pinpointed in early studies is the inability to distinguish heterogeneity from inefficiency. Therefore, all time-invariant heterogeneity is confounded by inefficiency. In later studies, both time-invariant effects and time-varying inefficiencies are considered.

Despite the recent developments presented by Kumbhakar (1991), Kumbhakar and Hjalmarsson (1993, 1998), Kumbhakar and Heshmati (1995), and Greene (2005a, 2005b), a question remains as to whether time-invariant effects can be considered as persistent inefficiency or as firm heterogeneity that captures the effects of (unobserved) time-invariant covariates and, as such, is unrelated to inefficiency. Regularly, different SFA specifications produce different results, but the main goal here is not to evaluate all existing models, but rather to compare and contrast representative models for the econometric estimation of one of the two components of cost (or productive) efficiency with the newly introduced GTRE model developed by Fillipini and Greene (2016).

The research approach here follows the framework of a stochastic frontier model as first, and independently, proposed by Aigner et al. (1977) and Van den Broeck (1977). While early stochastic frontier models were mainly implemented with cross-sectional data, most of the frontier models using panel data[188] have been based on fixed and random effects models. In this study, some of the estimated SFA models provide time-invariant values of cost efficiency that tend to reflect the persistent part of the level of cost efficiency, i.e. the classic random effects (RE) by Pitt and Lee (1981). Others

[188] From the first attempts in the literature, the Battese and Coelli (1995) model was formulated for panel data (balanced and unbalanced).

estimate time-varying values of cost efficiency that tend to capture the transient component, i.e. the TRE model (Greene, 2005a, 2005b). As described earlier, HEIs in England are heterogeneous in many different aspects, so models by Kumbhakar and Heshmati (1995) and Kumbhakar and Hjalmarsson (1998), which tend to give a persistent inefficiency interpretation to heterogeneity and are bereft of heterogeneity among DMUs, will not be discussed.

6.3.2 Cost Function Estimation and Efficiency Measurement

The corresponding acceptation of technical efficiency when cost functions applied is the definition of cost efficiency. In a more generic framework, cost efficiency is the ratio of potential (ideal) cost to observed cost, or to express it differently, cost inefficiency is the percentage by which observed cost needs to decrease in order for the unit to become 100 percent cost-efficient i.e. produce observed output at minimum cost. The economic objective of the cost function is cost minimization[189] and, as such, information on relevant prices is required. Then a measure of cost efficiency, which depends on input prices, is provided by the ratio of minimum feasible cost to actual cost. This measure of cost efficiency attains a maximum value of unity if the DMU is cost-efficient, and a value less than unity indicates the degree of cost inefficiency (Fried et al., 2008). According to Fried et al., (2008) a measure of cost efficiency (CE) is provided by the ratio of minimum cost to actual cost:

$$CE(x, q, w) = \frac{c(q, w)}{w^T x}$$

Through the estimation of a cost function, appropriate measures of cost efficiency can be derived provided that enough information is available regarding costs, outputs, input prices and the estimation processes. The estimation of frontier functions entails the econometric exercise of making the empirical implementation consistent with the underlying theoretical proposition that no observed agent can exceed the ideal (Greene, 2008). Measurement of cost (in)efficiency is, then, the empirical estimation of the extent to which observed units in this case HEIs (fail to) achieve the theoretical ideal. The theory of cost minimization provides a description of the ultimate source of deviations from this theoretical ideal.

If we proceed to the practical implementation, the frontier model can be paralleled to a regression model in such a way that would be in line with the theoretical preconception

[189] There is an intense debate (Brinkman, 1990; Ehrenberg, 2000) within the academic community for the validity of the cost minimization assumption in the HE sector. The non-for profit character of HEIs circumvents rationality on their decision making process (Martin, 2005). However, recent budgetary constraints for public institutions have enforced HEIs to optimize the efficiency of resource usage, and thus minimize the costs related to educational production. As a result, the optimization assumptions from classical microeconomics are probably reasonably realistic in their application to HE (Johnes and Johnes, 2007).

that all observations lie within the theoretical extreme (Greene, 2008). Most of the early cost studies used the regression framework to estimate a line of best fit though the data (Johnes and Johnes, 2013). However, when deploying this approach only average estimates of cost were possible without any indication of how cheaply the output had been produced. So, by analysing cost efficiency it is possible to derive further information on whether the output can be produced in a cheaper way than the average. To this end, a cost curve most widely known as cost frontier delimit the position of the envelope below which it is technically infeasible for costs to go. Accordingly, for a cost function, all observations except those on the frontier must lie above the cost frontier.

A simple representation of the cost function according to Shephard (1953) and Nerlove (1963) is given below:

$$C(q, w) = \min\{w^T x : f(x) \geq q\}$$

Where w represents a vector of input prices determined exogenously. The cost function gives the minimum level of cost expenditure needed to produce a certain amount of output q. Any deficiency in the optimization process (technical or allocative) will show up as higher costs. As such, a unit that might be assessed as operating technically efficiently by a production function measure might still appear inefficient in relation to a cost function (Greene, 2008). Cost functions[190] can be very trivial in simple technologies as one type of output and a simple straight line can adequately demonstrate the relationship between cost and output. However, in most cases the complex nature of production with more than one type of output and the multi-product context as observed in HE where economies of scale and scope (from joint production) arises, more sophisticated cost functions should be used (Johnes et al, 2005). Generally, the cost function choice has to be committed to some criteria to derive competent functional forms. Those criteria include no predetermination of whether (dis)economies of scale or scope exist and to allow the data to decide on how to shield on zero output values effectively (Johnes et al., 2005).

In this particular chapter, the main interest is centred on the estimated models of cost which are means of achieving the objective of measuring cost inefficiency. To clarify this point, this chapter incorporates a formal analysis of the 'residuals' from cost models, since these aim to capture the gap between observed values for each data point

[190] Properties of the cost function Shephard (1953):

i. Non-negativity: $C(q, w) \geq 0 \; for \; w > 0$.
ii. Non-decreasing in w: $If \; w > w' \; then \; C(q, w) \geq C(q, w')$.
iii. Positively linearly homogenous in w: $C(q, \lambda w) = \lambda C(q, w), w > 0$.
iv. C is concave and continuous in w.
v. No fixed costs: $C(0, w) = 0, \forall \; w > 0$. We sometimes assume we have a long run problem.
vi. No free lunch: $C(q, w) > 0, w > 0, q > 0$.
vii. Non-decreasing in q (proportional): $C(\theta q, w) \leq C(q, w), for \; 0 \leq \theta \leq 1 \; and \; w > 0$.
viii. Non-decreasing in q: $C(q', w), = C(q, w), for \; q' \leq q \; and \; w > 0$.
ix. For any sequence q^ℓ such that $\|q^\ell\| \to \infty$ as $\ell \to \infty$ and $w > 0, C(q^\ell, w) \to \infty$ as $\ell \to \infty$.
x. $C(q, w)$ is lower semi-continuous in q, given $w > 0$.

and the line of best fit (in regression framework) or the differences in efficiency across the observed units (frontier methods). In the latter case, the magnitude of the non-normal residual associated with each observation reflects a measure of cost efficiency.

6.3.3 Data Source

This study utilises a set of panel data that cover a period of time from 2008-09 to 2013-14. The data provide a comprehensive view of the English HE landscape and embody 126 English HEIs for every academic year apart from the last two years (2012-13 and 2013-14) in which 125 institutions are presented after the School of Pharmacy merged with University College London, retaining the latter's name. Meanwhile, on 1st August 2011, Leeds College of Music left the HE sector and became a wholly owned subsidiary of Leeds City College, the largest further education establishment in the City of Leeds. In the same academic year (2011-12), Leeds College of Art joined the HE sector[191].

As previously mentioned in chapter 4 (p. 84) of this thesis, there are some institutions that have been deliberately removed from our sample. Those include the Open University, the University of London (institutes and activities), the University of Buckingham and Heythrop College for reasons previously described. Moreover, Liverpool School of Tropical Medicine has been excluded entirely for 2013-14 since its first separate data return for 'student and destinations of leavers' data was made only in 2015-16 when previously its data for these records were returned under the University of Liverpool. However, it has provided separate returns previously for staff, finance, business and community interactions, and estates management data from 2013-14. The last institution exempted from the sample is University Campus Suffolk since it was established only in 2007, and is one of the newest HEIs in the UK, building on the strengths of the Universities of East Anglia and Essex, its sponsoring institutions.

Table 22: Definition of variables used in the analysis

Variable Names	**Definition**
Cost: Total Expenditure (TExp)	Total expenditure including (cost of academic departments, administration and central services, premises, research and grants contracts and other expenditures) minus expenditure on residences and catering operations in £000s in December 2013 prices
Outputs	**Definition**
PGMed	Full-time-equivalent [FTE] postgraduate students (first degree and other) in medicine

191 For further details follow the HESA website: https://www.hesa.ac.uk/support/providers/mergers-changes

PGScience	FTE postgraduate students (first degree and other) in sciences other than medicine
PGArts	FTE postgraduate students (first degree and other) in all other subjects
UGMed	FTE Undergraduate students (first degree and other) in medicine
UGScience	FTE Undergraduate students (first degree and other) in sciences other than medicine
UGArts	FTE Undergraduate students (first degree and other) in all other subjects
RESEARCH	Income received from research grants and contract (in £000s) in December 2013 prices
INCTT	Total income from technology transfer and innovation (in £000s) in December 2013 prices
Dummy variables	
LONDON	Value equals to 1 if the institution's postcode is in the London territory
OXBRIDGE	Value equals to 1 if HEI is Oxford or Cambridge
YEAR 09/10... YEAR 13/14	Dummy variables reflecting potential year effects. There are 5 dummies in the model excluding the YEAR08/09 is the base year.

Source: HESA: Finances of Higher Education Providers; Students in HE; Business and Community Interaction Survey **Note: For further for FTE follow: https://www.hesa.ac.uk/support/definitions Subject definitions**: **Medicine**: Medicine and dentistry; Subjects allied to medicine. **Other science**: Biological sciences; Veterinary science; Agriculture and related subjects; Physical sciences; Mathematical sciences; Computer science; Engineering and technology; Architecture, building and planning. **Non-science**: Social studies; Law; Business and administrative studies; Mass communications and documentation; Languages; Historical and philosophical studies; Creative arts and design; Education. **Technology transfer and innovation contains income stemming from: Collaborative research involving public funding:** This includes research projects' public funding from at least one public body (in £ thousands); **Contract research**: This includes contract numbers and income(in £ thousands) identifiable by the institution as meeting the specific research needs of external partners, excluding any already returned in collaborative research involving public funding and excluding basic research council grants; **Consultancy contracts:** This includes contract income (in £ thousands) associated with consultancy, which are crucially dependent on a high degree of intellectual input from the institution to the client (commercial or non-commercial) without the creation of new knowledge.

The definitions of the variables used in this chapter and their respective units are displayed in table 22. The definition of cost varies from study to study. In most empirical studies in HE, total recurrent expenditure is recorded. In the UK, context expenditure can be decomposed further to various categories according to purpose, i.e. expenditure on academic departments, academic services, equipment, administration and central services, maintenance and running of premises, staff and student facilities.

Table 23: Descriptive statistics for the variables in the data set

Variable	Mean	Std. Deviation	Max	Min
COST				
TExp	175,663.70	203705.30	1499137.00	7205.45
OUTPUTS				
PGMed	86.18	249.52	1668.34	1
PGScience	804.61	924.97	5986.59	1
PGArts	1,388.96	1096.01	5056.01	1
UGMed	285.64	649.37	3032.48	1
UGScience	3,852.35	3181.52	12770.56	1
UGArts	5,240.95	3656.12	17787.00	1
RESEARCH	31,140.95	69581.30	471957.00	1
INCTT	15,770.74	28759.55	177329.00	1

Source: HESA. All values are in £ (000) deflated in December 2013 prices.

Descriptive statistics for total expenditure (minus expenditure on residences and catering operations), undergraduates and postgraduates split by subject subgroup and income from technology transfer, as a third mission activity proxy, are displayed for all HEIs in England in table 23. Two front-end activities of teaching and research are commonly, served in each HEI as primary activities (Abbott and Doucouliagos, 2009). The range of total expenditure across institutions is large, with an average of £175.5 million. This wide dispersion is reflected by the large standard deviation (more than £200 million).

On average, over the study period, HEIs produce more than 1,000 graduates from postgraduate degrees in arts and much fewer postgraduates in science (around 1,300) and medicine studies (around 100). When total undergraduate numbers are disaggregated by broad subject areas they reveal differences and trends, in the type of courses taken by undergraduate students. In the same period over 5,000 undergraduates in arts graduated from undergraduate degrees compared to more than 3,500 in science degrees and, on average around 300 medicine-related degrees. In addition, just over £30 million was received from research income and more than £15.5 million was received in activities associated with third mission aims. A notable feature of the descriptive statistics is that most of the variables' standard deviation is close to the mean or even higher, indicating the considerably high levels of heterogeneity among HEIs in England.

6.4 Methodology Review

This section is concerned with the methodology used for this study. Building upon the conventional SFA approach introduced by Aigner et al., (1977), various specifications and transformations of the original formulation have been developed in the literature. As described earlier in the literature review part different models can lead to varying efficiency scores and different results, which in some cases can be even in contradiction. The main objective is not to investigate all existing models or to produce one more study of efficiency estimates in HE. Instead, what needs to be covered is the gap in the existing English HE literature regarding the concurrent estimation of individual-specific effects (latent heterogeneity) and inefficiency which can be further decomposed into two parts: permanent (long-term) and transient (short-term).

Generally, SFA studies for panel data produce efficiency indicators which may vary over time or be time-invariant. However, it should be noted that when it comes to comparing between the different models it is misleading to compare values produced from time-invariant models with those obtained from time-varying models since they measure different aspects of inefficiency. Therefore, firstly in this study, following Fillipini and Greene (2016) an RE model as expressed by Pitt and Lee (1981) and a TRE model (Greene, 2005a, 2005b) will be evaluated and compared with the results obtained from the novel model introduced by Colombi (2010), Colombi et al. (2011, 2014) and Kumbhakar and Tsionas (2012). The latter provides a theoretical platform to disentangle persistent and transient inefficiency utilising maximum simulated likelihood estimation techniques. Finally, following Fillipini and Greene (2016), unobserved heterogeneity bias will also be controlled by incorporating in the analysis an auxiliary equation based upon the seminal work of Mundlak (1978). Table 24 refers to the main characteristics of the models applied in this chapter.

Table 24: Main elements of the cost frontier models

	RE	TRE	GTRE	MGTRE
Firm effect (heterogeneity)	No	Yes	Yes	Yes
Persistent inefficiency	Yes	No	Yes	Yes
Transient inefficiency	No	Yes	Yes	Yes

HE is more appropriately modelled as a multi-product organisation (Johnes and Johnes, 2007). Thus, multiple-output cost functions (Baumol et al., 1982; Mayo, 1984) are the main vehicle through which to study the cost structure of HE. Various functional forms have been proposed and applied in empirical applications. Among the most frequently used in the literature are the linear, the quadratic, the translog and the CES cost functions (see appendix 4 chapter 3). Here, to maintain simplicity a log-linear version is estimated to preserve degrees of freedom and avoid multicollinearity. However other specification (translog and quadratic) when possible will also be attempted as a matter of comparison on whether the results are sensitive to the choice of the specification of the cost function.

Model 1- *Random Effects Stochastic Frontier Analysis Model (RE-SFA)*
Among the mixed methodological tools available for panel data, random effects models have been widely applied. In this study, we adopt a random effects specification of the SFA model namely the RESFA methodology, since it was the first method developed in an SFA framework that provides persistent estimates of inefficiency. Unobserved differences are taken into account which are unrelated to the independent variables in the model and, unlike OLS and fixed effects models, can vary across time (Zhang, 2010). This RESFA framework allows for time-invariant variables (D_k) to be included in this study. Those variables include a binary variable listed as (LONDON) that covers possible influence on costs due to the location of an HEI, particularly location within the London conurbation where land prices and labour costs (differences in input prices) are higher than elsewhere (Lenton, 2006; Johnes et al., 2008; Johnes and Johnes, 2013), a binary variable named (OXBRIDGE), to reflect differences on costs of maintaining ancient buildings, and five year-dummies (YEAR) to capture sector-wide changes over time. The specification of the model is:

$$\ln(TExp_{kt}) = a_0 + \beta' x_{kt} + \gamma_l D_{london,k} + z_o D_{oxbridge,k} + \sum_{t=2009}^{t=2014} a_t D_t + v_{kt} - u_k$$

$$v_{kt} \sim N[0, \sigma_v^2]$$

$$u_k \sim |U_k|, U_k \sim N[0, \sigma_u^2]$$

$$e_{kt} = v_{kt} + u_k$$

Where $TExp_{kt}$ is total expenditure incurred by HEI k in year t, x_{kt} is a vector of outputs (and input prices) in logs that represents the number of postgraduates and undergraduates split according to subject, research income and income from technology transfer, β is a vector of parameters to be estimated. Regarding the symmetric disturbance v_{kt} this is a normally distributed error component that represents random noise, where u_k is a one-sided non-negative disturbance reflecting inefficiency. This is not a fixed parameter and once the model is estimated through GLS or ML estimation techniques, unit-specific estimated values of inefficiency can be obtained. If the estimation technique is the GLS, a free distributional assumption method, the efficiency is calculated as the difference between the estimated intercept of each unit minus the estimated maximum intercept (Titus and Eagan, 2016) and the estimated values of cost efficiency equal to:

$$CE_k = \exp(-u_k) \ with \ u_k \geq 0, k = 1, \dots \dots \dots . K$$

Larger values of u_k denote lower cost efficiency. If the parameters of the model are estimated through ML practises this means that explicit distributional assumptions are imposed on the random components of the model and that we cannot obtain a direct estimate of inefficiency since the inefficiency parameter is unobservable. This approach was originally articulated by Pitt and Lee (1981). However, it provides no sufficient information to generate an estimate of the conditional mean of inefficiency, i.e. $E(u|e), e_k = v_k + u_k$.

There are two[192] main conditional mean estimators used in the literature to obtain a unit-specific efficiency score (Kumbhakar, 1987), namely the Jondrow et al. (1982) formula (JLMS method) and the Battese and Coelli (1988) conditional mean estimator (BC method). A common flaw of RESFA models is the absence of possible other time–invariant effects (heterogeneity) that may improperly covered under the inefficiency term (Greene, 2005). In the latter case, cost inefficiency estimates are significantly distorted (Farsi et al., 2006; Greene, 2005, 2006) since they mistakenly also incorporate any unobserved time-invariant and individual-specific heterogeneity effects. Researchers of the frontier literature are rather neglectful when it comes to applying RESFA models due to their shortcomings. To be specific, conventional assumptions of independence between unit-specific heterogeneous characteristics and the variables included in the model may be seriously violated in the SFA framework where variables are related to unobserved unit-specific capital or costs variables (Titus and Eagan, 2016).

Model 2- *True Random Effects Stochastic Frontier Analysis Model [TRE-SFA]*

Most of the shortcomings of the frontier methods for panel data have been identified in the literature (Parmeter and Kumbhakar, 2014). Special efforts have been made to overcome those limitations, and Greene (2005) was from the pioneer researchers in this direction. Indeed, Green enhanced the classical SFA framework with a firm-specific time-invariant effect. Those models, later called the TFE and TRE models, allow for estimates of time-variant cost inefficiency. Those models accommodate both unobserved time-invariant effects and time-varying efficiency estimation (transient part of efficiency). The log-linear version of the model can be formed as:

[192] It has to be mentioned that both techniques produce statistically inconsistent estimates of inefficiency

$$ln\,(TExp_{kt}) = \alpha_k + \beta' x_{kt} + \gamma_l D_{london,k} + z_o D_{oxbridge,k} + \sum_{t=2009}^{t=2014} a_t D_t + \; v_{kt} - u_{kt}$$

$$\alpha_k = \alpha + w_k, w_k \sim N[0, \sigma_w^2]$$
$$v_{kt} \sim N[0, \sigma_v^2]$$
$$u_{kt} = |U_{kt}|, U_{kt} \sim N[0, \sigma_v^2]$$
$$e_{kt} = v_{kt} + u_{kt}$$

Where α_k is an individual-specific component that reflects heterogeneity among HEIs in terms of their structure, organisation, management, locations, missions etc. The inefficiency term is an *iid* random variable that can be distributed half normally, truncated normally or exponentially (non-normal distributions). The estimation technique proposed by Greene (2005) for those models is the maximum simulated likelihood and w_k is an *iid* random component in an RE framework and uncorrelated with all other terms.

One of the common weaknesses of those RE specifications is known in the literature as omitted variables bias whereby since there may be unobserved heterogeneity bias in the estimation of SFA models, this means unobservables may be correlated with the regressors. By modifying the model to incorporate unit (institutional) means of the explanatory variables, Mundlak (1978) tried to treat any existing unobserved unit heterogeneity. So, an auxiliary equation that takes into account correlation between this unobserved heterogeneity and the unit means of the explanatory variables was introduced in the literature, widely known as the CRE model (Wooldridge, 2010). The auxiliary equation is formulated as:

$$\alpha_k = \alpha + \phi' \, \overline{x_k} + w_k, w_k \sim iid[0, \sigma_w^2]$$

By incorporating this equation in the TRE model we obtain the MTRE model which incorporates the group means of the independent variables (time varying explanatory variables) $\overline{x_k}$ and it is estimated with SML techniques.

Model 3- *Generalised True Random Effects Stochastic Frontier Analysis Model [GTRE-SFA]*

In both models described so far there are dark angles regarding the underlying assumptions; RE-SFA models fail to control for time-varying inefficiency so any unit-specific effects or time-varying inefficiency effects are confounded with time-persistent inefficiency. On the other hand, TRE-SFA models fail to distinguish between fixed and adjustable long-term factors so long-term inefficiency (persistent) is suppressed. In order to remove ambiguity regarding the robustness of the cost efficiency estimates, there is an urgent need for a novel model that embodies both persistent and transient parts of inefficiency, while concurrently accounting for other exogenous unchangeable factors. Therefore, in a single step, it is aimed to separate persistent, transient inefficiency along with heterogeneity. The model, following Fillipini and Greene (2016), can be presented as:

$$\ln TExp_{kt} = a_0 + w_k + \beta' x_{kt} + \gamma_l D_{london,k} + z_o D_{oxbridge,k} + \sum_{t=2009}^{t=2014} a_t D_t + v_{kt} - h_{ko} - u_{kt}$$

Here, what has been added to the previous model is the term $h_{ko}\ with\ h_{k0} = |H_k|$ (half normal distribution) that aims to control for persistent or time-invariant inefficiency, since the TRE-SFA models fail to accommodate it, so all the time-invariant inefficiency (e.g. time-invariant ownership) is absorbed in the individual-specific constant term. Regarding the transient component, this is illustrated as u_{kt} and is considered as the residual (time-varying) component of cost inefficiency. Both parts of inefficiency are considered non-negative. The former is only unit-specific, while the latter is both unit- and time-specific. The model here has a fully flexible error specification with a four-way error component model, i.e. $e_{kt} = w_k + v_{kt} - h_{ko} - u_{kt}$ (Colombi, 2010; Kumbhakar et al., 2014; Kumbhakar and Tsionas, 2012; Tsionas and Kumbhakar 2014; Colombi et al., 2011, 2014). This model designation is GTRE since it is a simple TRE version in which the time-invariant effect has a skewed-normal distribution rather than a normal one. The GTRE model can be considered to contain a two-part disturbance: one time-varying i.e. $(v_{kt} - u_{kt})$, and one time-invariant $(w_k - h_{ko})$. Each part has its own skewed-normal distribution.

To obtain a tractable likelihood function, Colombi et al. (2014) used a skew normal distribution property for both the transient $v_{kt} - u_{kt}$ and time-invariant $w_k - h_{ko}$ random components. Therefore, for each component distributional assumptions are imposed as follows: v_{kt} is $i.i.d.$ standard normal and u_{kt} is $i.i.d.$ half normal, so the composed error $v_{kt} - u_{kt}$ has a skew normal distribution. The same set of assumption are imposed on the composite error $w_k - h_{ko}$, where w_k is normally distributed and h_{ko} has a half normal distribution, makes the composite error to follow a skew normal distribution. The GTRE model is simply a TRE model in which the time invariant effect has a skew normal distribution, rather than a normal distribution as assumed earlier.

Finally, the sum of the two independent skew normal random variables follow a closed skew normal distribution which is used to derive the likelihood function which can be maximised to obtain MLE of all parameters. Using the insights of Butler and Moffit's (1982) formulation Fillipini and Greene (2016) note that the density in Colombi et al. (2014) can be greatly simplified by conditioning on w_k and h_{ko} since the optimization problem itself is essentially the same as in the TRE models. In this case, the conditional density is simply the product over time of T univariate closed-skew normal densities. The simulation, itself, involves pairs of independent random draws from two standard normal populations. Thus, by simply taking the sum of a normal minus the absolute value of a normal draw is equivalent to simulate draws from a skew normal distribution and the model is estimated with SML[193].

The discretion of the two components of inefficiency is important and desirable since the measure of h_{ko} is an indication of the flexibility that DMUs or the policy officials have to introduce changes in management or new educational policies. However, h_{ko} does not fully reflect inefficiency because it does not account for learning over time since it is time-invariant, so the time-varying component aims to cover this aspect (Parmeter and Kumbhakar, 2014).

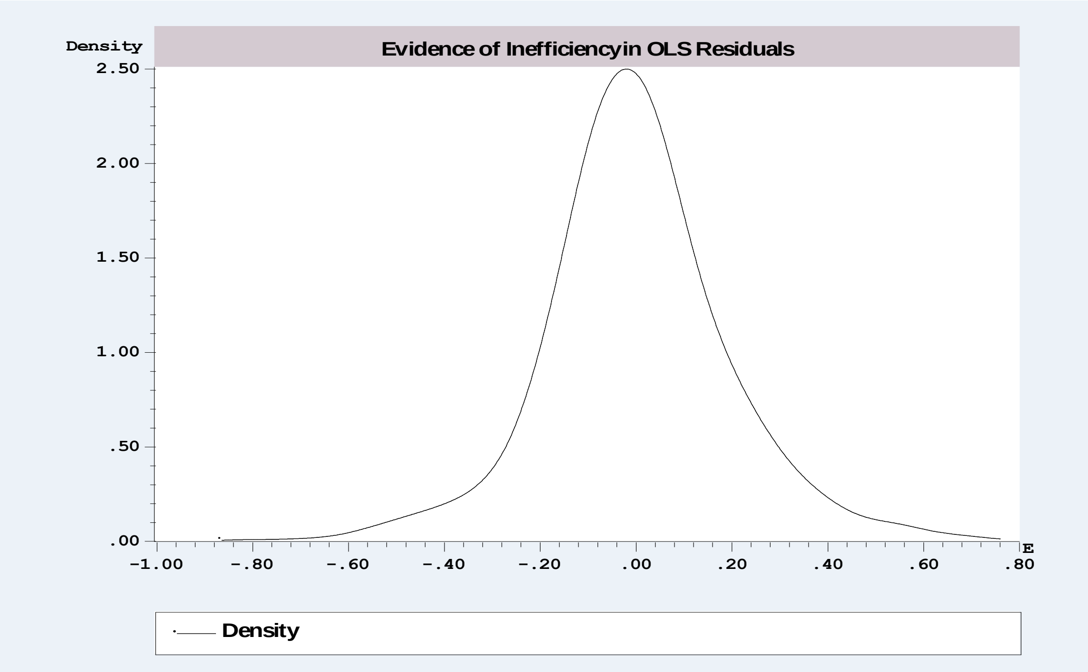
6.6 Results

Before proceeding to the discussion of the results, it is vital to trace whether any signs of inefficiency exist in the dataset. The skewness of the composed error term provides useful information in this direction since the direction of the kernel density[194] of the OLS residuals provides informative guidance on whether inefficiency can be distinguished from the stochastic error. The theory predicts that the OLS residuals have the right skewness when they are negatively skewed in production frontiers and positively skewed in cost frontiers, which is the case in this chapter. So, at this very first stage of this section, tracing any signs of inefficiency is an important indicator of the specification of the stochastic frontier model. In cases when the OLS residuals are skewed in the opposite direction from what theory predicts (i.e. ‘’wrong’’ direction[195]), in finite samples they tend to believe that the model is misspecified or that the data fail to conform to the model (Almanidis and Sickles, 2011). Therefore, before any maximum likelihood estimation begins, the skewness of the OLS residuals in the regression of c on x is checked. According to Waldman (1982) if the OLS residuals are skewed in the wrong direction, a solution for the maximum likelihood estimator for the stochastic frontier model is simply to use OLS.

[193] For the analytical approximation of the full log-likelihood function as well as the simulation version of the GTRE model, see Fillipini and Greene (2016).

[194] The kernel density estimator is a device used to describe the distribution of a variable non-parametrically, that is, without any assumption of the underlying distribution Greene (2015).

[195] If this condition is found, a reconsideration of the specification of the model is needed or changes in the modelling platform. Follow Chapter E62: Estimating Stochastic Frontier Models: http://people.stern.nyu.edu/wgreene/FrontierModels.htm



Picture 1: Kernel density of OLS residuals

Source: Own calculations

As predicted by the kernel density of the OLS residuals, the skewness value is around 0.117 positively skewed where the scores tend to cluster to the left, with the tail extending to the right, which is an indication of inefficiency in a cost framework (see Picture 1). In addition, the normality of the residuals has been tested by applying the Bowman and Shenton chi-squared statistic for testing against the null hypothesis of normality. The probability that the chi-squared variable with two degrees of freedom in this case (135.7) would exceed the 95 percent 'critical value' for chi-square (5.99), is zero so, based on this test, we reject 'normality'. Based on those outcomes we can further proceed to use frontier methods to estimate inefficiency.

6.6.1 Log Linear Specification

Having discussed the theoretical configuration of the four models in this part, we now turn to the empirical part of the cost efficiency assessment of HEIs in England. In this first step, a log-linear version of the cost function is used. According to table 25, the estimated output coefficients are relatively similar across the different model specifications apart from the MGTRE model specification. The output coefficients for postgraduates in medicine-related studies and for postgraduates in science and arts subjects, are positive and significant, which implies higher costs when the number of postgraduates increase. This is the also the case for undergraduates in science and non-science (arts) domains, however this positive impact is not confirmed for undergraduates in medicine-related subjects where an increase in their number leads to lower costs. More income obtained for research and third mission activities implies higher costs for HEIs. This, in turn, implies that research-related activities increase further the average cost of operations.

Table 25: Log-linear specification-estimation results of cost in English HEIs

	Dependent variable: Total expenditure			
	RE (Pitt and Lee)	TRE (Greene)	GTRE	MGTRE
PGMed	0.005	0.008***	0.054***	0.003
	(0.008)	(0.0003)	(0.007)	(0.006)
PGScience	0.028***	0.031***	0.063***	0.019**
	(0.007)	(0.0003)	(0.008)	(0.009)
PGArts	0.012	0.011***	0.055***	0.004
		(0.0001)	(0.008)	(0.010)
UGMed	-0.060***	-0.044***	-0.016***	-0.066***
	(0.010)	(0.0002)	(0.005)	(0.006)
UGScience	0.029	0.019***	0.089***	-0.002
	(0.019)	(0.0002)	(0.006)	(0.010)
UGArts	0.013	0.036***	0.128***	0.062***
	(0.019)	(0.0001)	(0.007)	(0.013)
RESEARCH	0.010**	0.012***	0.133***	0.005
	(0.004)	(0.0002)	(0.005)	(0.004)
INCTT	0.012***	0.010***	0.053***	0.008**
	(0.003)	(0.0001)	(0.006)	(0.003)
YEAR 0910	0.004	0.021***	0.033	0.019*
	(0.021)	(0.0006)	(0.080)	(0.011)
YEAR 1011	-0.019	0.008***	0.004	0.005
	(0.014)	(0.0006)	(0.069)	(0.010)
YEAR1112	-0.057***	-0.029***	-0.033	-0.031***
	(0.013)	(0.0006)	(0.049)	(0.009)
YEAR 1213	-0.040**	-0.019***	0.0001	-0.017

	(0.019)	(0.0006)	(0.057)	(0.010)
YEAR 1314	-0.019	-0.021***	0.009	-0.009
	(0.014)	(0.0006)	(0.055)	(0.007)
OXBRIDGE	4.422	2.256***	0.795***	0.581***
	(8.850)	(0.001)	(0.065)	(0.016)
LONDON	0.364**	0.666***	0.199***	0.161***
	(0.146)	(0.0006)	(0.021)	(0.005)
CONSTANT	8.758***	8.331***	7.433***	7.692***
	(0.076)	(0.033)	(0.062)	(0.015)

Note: Standard errors are shown in brackets. *, ** and *** signal that the coefficient is significantly different form zero at the 10%, 5% and 1% significance level, respectively. All time-varying variables in the models are expressed in log-forms. Mundlak terms are not shown.

Regarding the coefficients of the year dummy, they indicate a negative significant effect across the different specifications, implying that the total operation costs of English HEIs decreased over time. It could thus be implied, that this decreased tendency of suppressed cost is either an aftereffect of cuts in public funding, or the result of improved efficiency. The positive constant term coefficients suggest that fixed costs are positive and significant across the different model specifications.

In order to access other variables that influence costs, and more specifically to focus on any intra-country or-intra-regional effects due to higher land prices, the models are augmented with a LONDON dummy which signifies a positive effect on costs so universities in the London conurbation actually face higher costs. The significance of the OXBRIDGE dummy is also confirmed by all models indicating higher costs for the Universities of Oxford and Cambridge. These higher costs can be attributed to the expensive maintenance of upkeep of ancient buildings, but may also imply larger costs associated with the volume of operations or differences in teaching techniques/quality.

Lastly, it has to be mentioned that the Mundlak version of the GTRE model does not contradict the direction of the results. However, it deems some of the coefficients that were previously significant as insignificant. These results suggest the presence of unobserved heterogeneity bias since this modification actually aims to control and fix this aspect. For this model, the likelihood ratio test against the GTRE model gives chi-squared of 310.62[196]which far exceeds the critical value of $x^2_{(0.95;10)}$=18.30. So, on the basis of a likelihood ratio test, the null hypothesis is rejected that all Mundlak terms are equal to zero.

[196] The calculations made to produce this result are: $LR = 2\big((logL|unrestricted\ model) - (logL|restricted\ model)\big)$ = 2 [(649.40)- 494.09] =310.62

Hitherto, different model specifications have been compared which tend to deliver different results. Hence, we aim to check whether the outcomes are statistically similar or different along the different models, so as to confirm the superiority in terms of effectiveness of the GTRE model. To this end, we check the statistical validity and superiority of the MGTRE model against the RE model and the TRE model by performing two likelihood ratio tests.

On the basis of a likelihood ratio test, it is confirmed that we can reject the null hypothesis of statistical insignificance of the GTRE model against a RE model (likelihood ratio test is equal to 364.84 higher than the $x^2_{(0.95;12)}$=21.02). This is the case for the TRE model as well, since in statistical terms at the 0.05 significance level, the null (TRE model) is rejected in favour of the unrestricted GTRE alternative (likelihood ratio test is equal to 206.13 higher than the $x^2_{(0.95;11)}$=19.67).

As pointed out in Chapter 3, the choice of the functional form itself reveals information about the relationship between inputs and outputs. During the production process, more of the technical aspects, such as factor substitution, economies of scale/ scope or input demand elasticities, are implicitly embodied in the selection of the functional form. In the present analysis, the functional form choice is not of primary focus but, aside from the log-linear version of the model, we opt for a translog version in an attempt to gain further flexibility on the imposed restrictions on these features which may distort the cost efficiency estimates (Greene, 2008).

6.6.2 Translog Specification

The CD and translog models overwhelmingly dominate the applications literature in stochastic frontier and econometric inefficiency estimation. The translog is a flexible generalisation of the CD production function (Coelli et al., 2005). This increased flexibility permits for a more accurate representation of empirical production or cost function. The cost function is the dual of the production function which relates outputs to inputs and input prices, and specifications of the cost function therefore do not include inputs and input prices (Johnes et al., 2008). Here, we aim to adopt a translog multiple output cost function as specified by Tsionas and Greene (2003). In order to deal with zero values[197] we use a $\log(x+1)$ transformation following Wooldridge (2000). To be more accurate, this means that all variables are transformed by adding them a unit before taking the natural logarithm of their values, i.e. the zero values are transformed into one. The model specification can be written in general form as:

[197] Another alternative would be to employ a Box-Cox functional form in the translog model to accommodate zero values for some of the outputs Caves et al. (1980).

$$
\begin{aligned}
ln(TExp)_{kt} = \alpha \\
&+ \sum_{m=1}^{6} \delta_m ln(q_{mkt}) + \frac{1}{2}\sum_{m=1}^{6}\sum_{s=1}^{6} \phi_{ms} \ln(q_{mkt}q_{skt}) \\
&+ \sum_{m=1}^{6}\sum_{s=1}^{6} \theta_{ms} \ln(q_{mkt}q_{skt}) + \gamma_l D_{london,k} + z_o D_{oxbridge,k} \\
&+ \sum_{t=2009}^{t=2014} a_t D_t + e_{kt}
\end{aligned}
$$

Where e_{kt} is the error term which can be composed of different independent parts depending each time on the econometric specification chosen (see Table 26).

Table 26: Econometric specification of the error term

	Full random error e_{kt}
RE (Pitt and Lee)	$u_k + v_{kt}$
TRE	$w_k + u_{kt} + v_{kt}$
GTRE	$w_k + h_k + u_{kt} + v_{kt}$
MGTRE	$w_k + h_k + u_{kt} + v_{kt}$

Turning now to the experimental evidence in Table 27, there are no significantly observable differences in the direction or the significance of most of the coefficients apart from the points highlighted below. In this application, an aggregate approximation of the postgraduate students is used since statistical problems arose when a full version of disaggregation was attempted by the subject group. Also, in order to maintain degrees of freedom which are hard to find in translog specifications due to the inclusion of interaction and square terms, a single measure of postgraduates is used. The coefficients here can be interpreted as cost elasticities since all the continuous variables are in log-form.

Compared to the simple log-linear specification, the coefficients of undergraduates in medicine are still negative and significant indicating a decrease in costs due to a 1 percent increase in the number of postgraduates in medicine. Another important finding is that under both GTRE and MGTRE models the cost seems to have an increasing tendency due to a 1 percent increase in the number of postgraduate students. These findings further confirm previous evidence presented by Johnes et al. (2008) who highlighted an increase in the average incremental costs due to the fact that postgraduates, especially those involved in research projects, require intensive supervision on a one-to-one basis. The most costly output produced by HEIs is the research output since the implications of producing research are far-reaching, so HEIs which want to ensure a competitive position fit for the modern academic landscape invest more and more in new funding-heavy research projects.

There is a noticeable shift in the direction of significance in the coefficients accounting for income from third mission activities. Compared to the linear case model, in the translog specification the coefficients are negative, implying lower costs for institutions which engage more in third stream activities. This means that universities can build relationships to maximise engagement with third parties and new bodies that further reduce costs. However, those results are counterintuitive as they warn that statistical inference may vary according to minor changes in specification or modelling strategy.

Table 27: Translog specification-Estimation results of cost in English HEIs

	Dependent variable: Total expenditure			
	RE (Pitt and Lee)	TRE (Greene)	GTRE	MGTRE
PGs	-0.105***	-0.079***	0.137***	0.045***
	(0.029)	(0.008)	(0.015)	(0.008)
UGMed	-0.017	-0.070***	-0.143***	-0.218***
	(0.107)	(0.013)	(0.021)	(0.013)
UGScience	-0.296***	-0.299***	-0.018**	-0.181***
	(0.040)	(0.005)	(0.009)	(0.010)
UGArts	-0.240***	-0.170***	0.516***	0.026*
	(0.038)	(0.009)	(0.016)	(0.015)
RESEARCH	0.118***	0.118***	0.424***	0.134***
	(0.021)	(0.006)	(0.010)	(0.006)
INCTT	-0.007	-0.028***	-0.027***	-0.027***
	(0.015)	(0.005)	(0.009)	(0004)
YEAR 0910	-0.030*	-0.013	-0008	0.003
	(0.015)	(0.009)	(0.018)	(0004)
YEAR 1011	-0.062***	-0.036***	-0.034*	-0.020***
	(0.014)	(0.009)	(0.020)	(0.005)
YEAR1112	-0.110***	-0.086***	-0.083***	-0.059***
	(0.011)	(0.007)	(0.018)	(0.004)

YEAR 1213	-0.077***	-0.055***	-0.059***	-0.038***
	(0.013)	(0.008)	(0.019)	(0.005)
YEAR 1314	-0.045***	-0.030***	-0.032**	-0.020***
	(0.012)	(0.007)	(0.016)	(0.004)
OXBRIDGE	0.907***	0.828***	0.364***	0.817***
	(0.351)	(0.020)	(0.035)	(0.015)
LONDON	0163**	0.305***	0.142***	0.335***
	(0.070)	(0.006)	(0.010)	(0.004)
CONSTANT	9.302***	9.678***	4.483***	3.772***
	(0.178)	(0.056)	(0.103)	(0.101)

Note: Standard errors are shown in brackets. *, ** and *** signal that the coefficient is significantly different form zero at the 10%, 5% and 1% significance level, respectively. All time-varying variables in the models are expressed in log-forms. Mundlak terms are not shown.

The results obtained in this application are confirmed even when we use an aggregate metric for undergraduates in medicine and science subject groups. The only differential factor in this specification is the coefficient obtained regarding third mission activities which are positive, in line with the linear case. The full results of this application are given in appendix 19 chapter 6.

In the next table (Table 28), descriptive statistics for the cost efficiency estimates are provided for English HEIs for the period 2008-09-2013-14. Four different models have been examined, so for each model specification (i.e. linear and translog) six different estimates are available. It is expected that both GTRE and MGTRE models offer different estimates for the persistent and transient part of cost efficiency thus, the number of estimates exceeds the number of models used. As previously described, the RE model provides time-invariant cost efficiency estimates which further implies that it reflects persistent cost efficiency. Along with this explanation, the TRE model, since it incorporates a time-varying component of inefficiency, captures time-varying efficiency estimates.

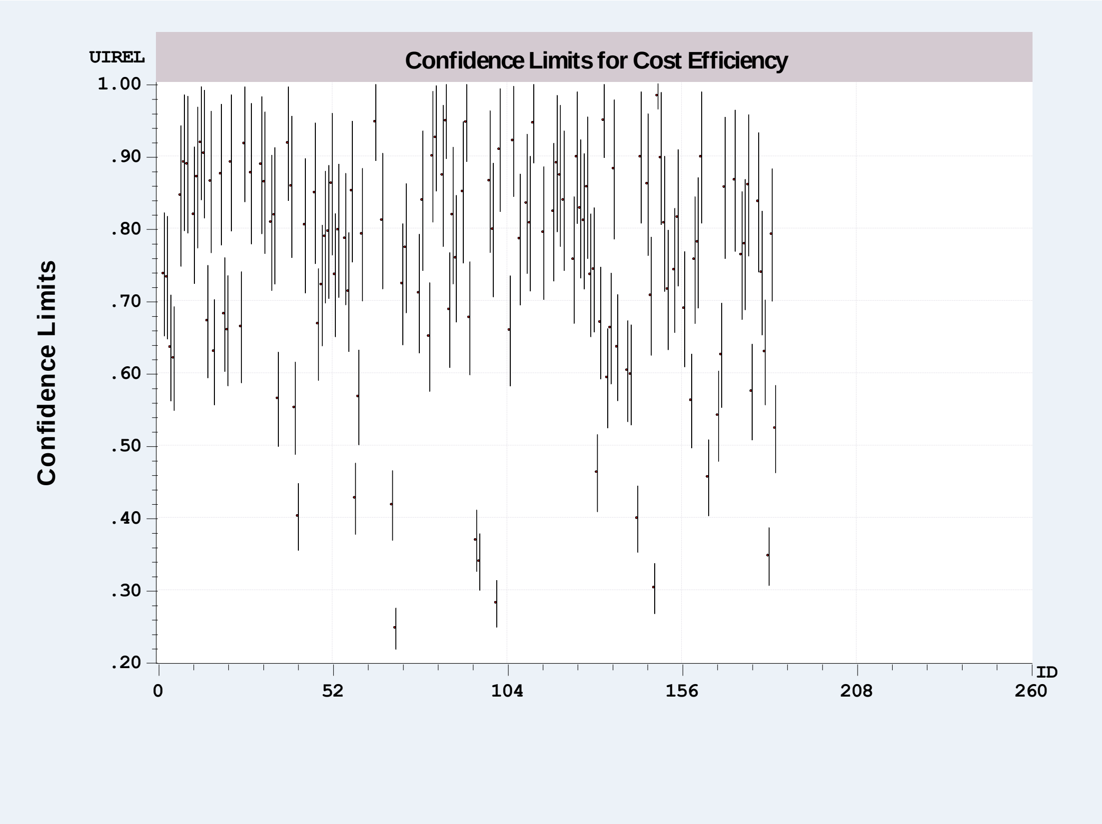
According to the findings of Table 28, the estimated average values of persistent efficiency vary from 21 percent in the RE model to 82 percent in the GTRE_LR and MGTRE_LR models that reflect the persistent part of cost efficiency. This is not the case for the translog specification, where the gap is certainly narrower since it varies from 74 percent to 85 percent. It is difficult to explain this result, but it might be related to the flexibility of the non-restricted specification (translog) since the nested linear case may lack informative interactions. Regarding the cost efficiency estimates of the transient component, those vary from 87 percent in the TRE case to 92 percent in the MGTRE_SR. This gap is again smaller in the translog specification case. What seems surprising is that under the GTRE and MGTRE models we trace fully efficient HEIs. Those are identified by having an efficiency score equal to 1.

Table 28: Cost efficiency estimates

Variable	Linear specification				Translog specification			
	Mean	SD	Min	Max	Mean	SD	Min	Max
RE	0.210	0.227	0.011	0.974	0.742	0.170	0.238	0.985
TRE	0.872	0.059	0.305	0.978	0.939	0.040	0.612	0.993
GTRE_SR	0.804	0.073	0.069	0.936	0.899	0.062	0.382	1
GTRE_LR	0.820	0.002	0.811	0.843	0.853	0.017	0.838	1
MGTRE_SR	0.925	0.046	0.539	1	0.923	0.072	0.351	1
MGTRE_LR	0.820	0.031	0.810	1	0.850	0.026	0.818	1

One of the issues that emerges from these findings is that the increased values of the standard deviation of the RE models under both specifications evokes further analysis. Therefore, we proceed with the construction of a 95 percent confidence limits since we aim to construct a confidence interval for the RE efficiency estimates. The centipede plot shows how the confidence limits vary from university to university (see Figure 6).

Figure 6: Confidence intervals for RE-translog case



The constructed confidence intervals now tell us the most likely range of the average cost efficiency estimates, providing both the location and precision of the estimate. Most of the institutions lie above the 50 percent efficiency level and there are only 13 institutions with mean values below that level. Those institutions include: the University for the Creative Arts; Guildhall School of Music and Drama; Imperial College of Science; Technology and Medicine; Institute of Education; University of the Arts, London; London Business School; London School of Economics and Political Science; University of Reading; Royal Veterinary College; School of Oriental and African Studies; Trinity Laban Conservatoire of Music and Dance; Writtle College and Falmouth University.

There are far more institutions that are either middle performing or highly performing in the efficiency scale. However, due to the time-invariant nature of the cost inefficiency, this inefficiency varies for each institution but does not change over time. This restricted framework implies that inefficient institutions never learn over time which might be true for certain periods of time or if the time dimension of the panel is particularly short, however inefficiency varies over time so the time dimension should indisputably be considered.

These findings will doubtless be much scrutinised, but there are some immediately dependable conclusions that further reveal the robustness of the GTRE and MGTRE models. The overall GTRE mean cost efficiency estimates (0.812; 0.876) lie in between the two values of the RE and TRE models as expected. The correlation coefficient matrix in the next table (Table 29) can adequately reflect the consistency of the persistent and transient parts obtained through the GTRE and MGTRE models.

Table 29: Correlation coefficients

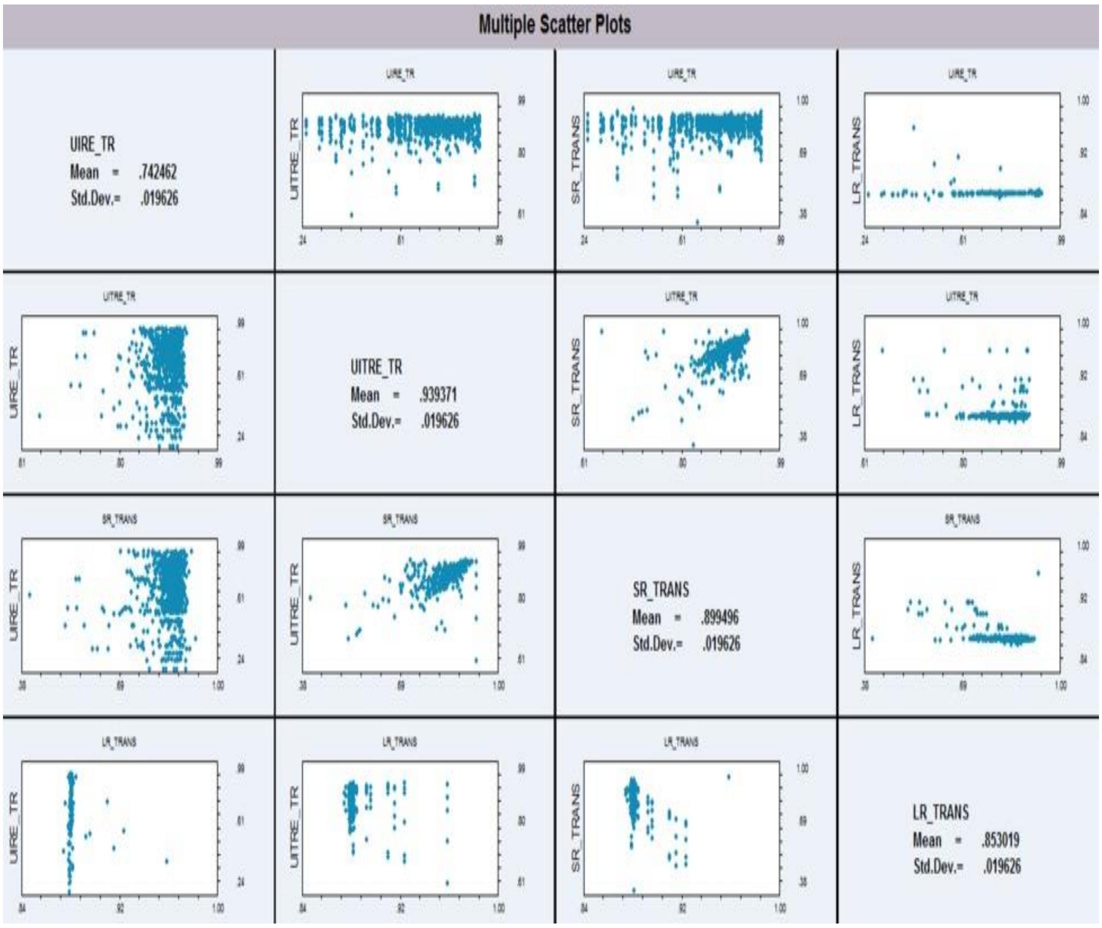
Correlation Matrix	**Linear Specification**					
	RE	TRE	GTRE_SR	GTRE_LR	MGTRE_SR	MGTRE_LR
RE	1					
TRE	-0.115	1				
GTRE_SR	-0.230	**0.661**	1			
GTRE_LR	**0.283**	-0.106	-0.441	1		
MGTRE_SR	-0.080	**0.387**	0.309	0.005	1	
MGTRE_LR	**0.043**	-0.255	-0.198	0.194	0.201	1

Source: Own calculations

Taken together, these results suggest that there is an association between the RE model and the persistent part of the GTRE_LR and MGTRE_LR models. This means that the value of the correlation coefficient between those models is positive at around 0.283 and 0.043. The same can be said between the correlation coefficient of the TRE model and the GTRE_SR and MGTRE_SR models which rise up to 0.661 and 0.387 respectively. The relatively low values of the correlation between the RE model and the persistent part of the GTRE models suggest that the results obtained with an RE model are not representative of the persistent efficiency part. These findings are in line with Greene (2005) who has pointed out the weaknesses of an RE where latent heterogeneity and persistent inefficiency both masqueraded as individual-specific effects. Interestingly, the persistent and transient parts of cost efficiency are relatively different in absolute values and not highly correlated, which is desirable since it further confirms that they account for different things. In appendix 20 chapter 6, a correlation matrix for the translog specification is available that confirms further the results reported for the linear case.

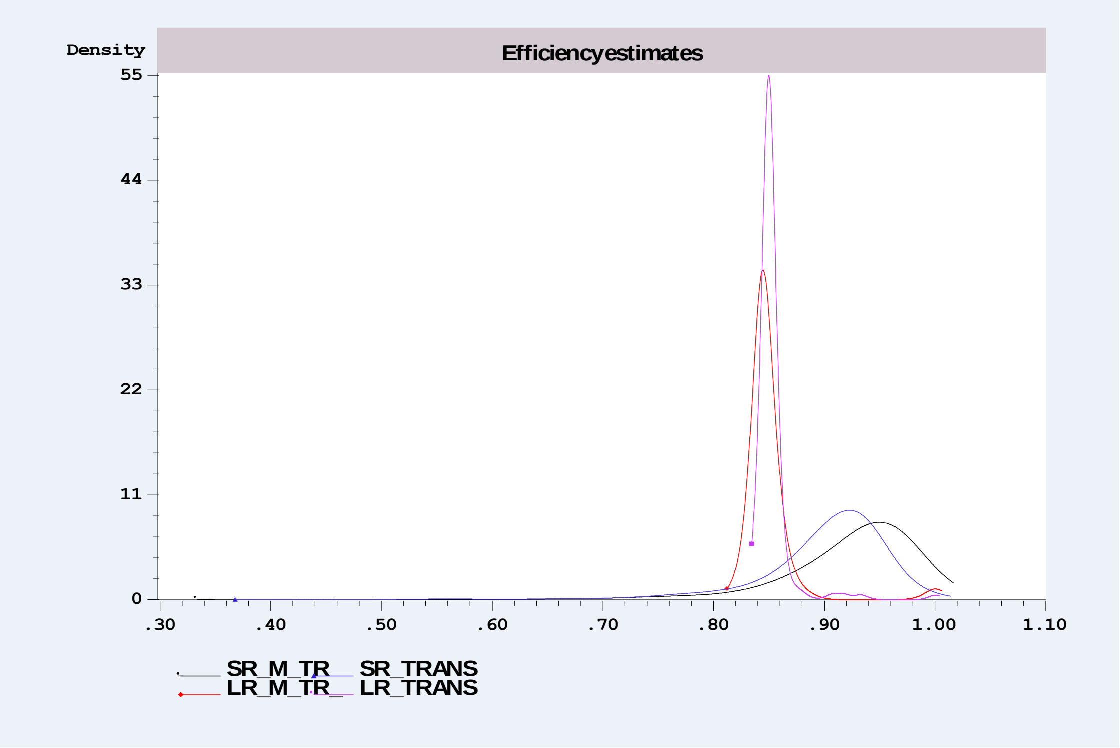
For a visual representation of the above relationship among the different models, we offer multiple scatter plots that validate the positive correlation between the TRE model since as one variable increases, the other variable increases too (see Figure 7). This seems to be the strongest relationship since the data points between the efficiency estimates of the two models are the most tightly clustered along an imaginary line. Regarding the association between the RE model and the persistent part of the GTRE model, their association is poor revealing the deficiencies of the RE models.

Figure 7: Scatterplot matrix-translog specification



For a thorough analysis of the results, apart from the absolute mean cost efficiency values, we turn on the kernel distribution of the estimated efficiency values for both GTRE and MGTRE. Under the linear specification of the GTRE model the long-run (persistent) cost efficiency (0.820) seems to be higher than the short-run (transient) (0.804). However, proceeding in the linear MGTRE as well as the quadratic versions of the models substantiate the opposite result. Consequently, persistent efficiency is lower than the transient efficiency suggesting that inefficiency is presumably not caused by something unexpected within the period covered but that there are rather persistent factors that drives it.

Figure 8: Kernel densities-GTRE and MGTRE models



Translog specification: persistent and transient cost efficiency estimates, source: own estimations

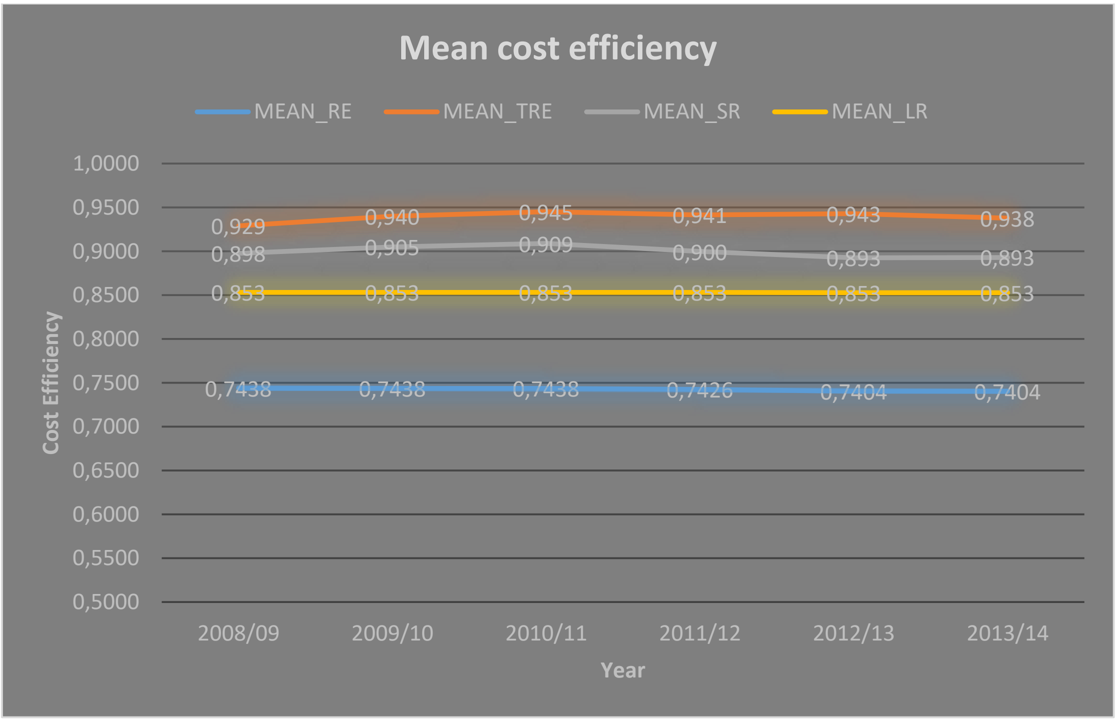
According to figure 8, the kernel distribution reveals that the short-run cost efficiencies disperse around the mean more than the long-run efficiencies. Long-run (persistent) efficiency rises to 0.850 whereas the short-run (transient) reaches even higher rates of around 89 percent to 92 percent. The kernel estimator suggests, however, that the differences in the long- and short-run estimates of efficiency are quite modest.

Descriptively, and without using a formal test, if we explore and compare the short-and long-run efficiencies for each institution we can extract useful information on whether each unit is dominated by one or the other. Hence, if the transient part of the inefficiency component for a unit is relatively large in a particular year then this is unlikely to be the case for the next year. However, if the persistent inefficiency component is large for a unit, then it is more than likely to be repeated over time, unless some changes in policy and/or management take place.

Moreover, there are cases where institutions tend to have more short- than long-run efficiency but after a period of time they manage to reverse this climate. Some examples of institutions that they finally attained greater long-run efficiency across the time are: Courtauld Institute of Art; University of Durham; Edge Hill University; University of Exeter; University of Surrey; Harper Adams University College; University of Kent; University of Lancaster; London School of Economics and Political Science; University College Plymouth St Mark and St John; University of Plymouth; Ravensbourne College of Design and Communication; Roehampton University; Royal Academy of Music; Royal Agricultural College; Royal Veterinary College and St Georges Hospital Medical School. However, this is a comparison that has not been tested statistically so we cannot confirm whether the difference between the short- and the long run efficiencies is statistically significant or not. Due to practical constraints, such as distributional assumptions of the efficiencies, this application is beyond the scope of this study.

Next, we explore cost efficiencies over time. According to figure 9 the persistent and the transient components of the GTRE model lie in between the two extreme values of the RE and the TRE models. This is likely to be the case since GTRE estimates calibrate and separate the short- and the long-run inefficiencies more effectively. The short-run estimates display a slight increase overtime however the persistent (long-run) estimates of efficiency are stable at a level of 0.853 on average lower than the short-run estimates across time. These findings further validate the initial hypothesis of predominance of persistent factors that hinder efficiency improvements in the long-run. What should be considered when those findings are interpreted is the fact that longer panels of data should perhaps be examined to firmly support the tolerance of the persistent inefficiencies in the English HE system.

Figure 9: Mean values of cost efficiency over time-translog specification



Lastly, in both extreme cases of RE and TRE the potential misspecification and the uncovered heterogeneity bias distorts the efficiency estimates either by overrating the estimated values (TRE case) or by undermining them (RE case).

6.7 Conclusions

Building upon the existing literature on the cost structure of English HEIs, this chapter provides estimates on cost functions for English HEIs and assesses whether inefficiency in the sector is dominated by persistent or transient factors. By using state-of-the-art parametric estimation methods as well as novel methods that account for time-varying and time-invariant inefficiency, the GTRE model allows policy-makers to identify those institutions operating in the most cost-effective manner in the long-run and encourage the sharing of best practices. The two types of cost efficiency are informative in terms of the short-run as well as the long-run cost saving university policy. Universities' management policies should respond with different improvement strategies; either by reallocating the available resources within the HE sector, or by taking more radical and structural interventions in the state specific regulation mechanisms i.e. funding, competition, grants, tuition fees etc.

There is a clear indication of lower costs across years and research has been deemed as the most costly output on average. The various findings provide some support for the conceptual premise that low values of persistent efficiency are of more concern from a long-term perspective since due to its persistent nature this adverse effect will tend to perpetuate unless some changes in policy and/or management take place. The findings suggest that this is the case for the English HE system where persistent efficiency is relatively lower compared to transient efficiency. There are clear implications for the need for structural changes/improvements in the English HE sector implying further possibilities to raise efficiency since the transient efficiency component is higher, thereby signifying repeated operational deficiencies in the long-run.

Thus, from a regulatory point of view, policy officials and various other university stakeholders[198] can capitalize on the results of this study for target-specific decisions. Those decisions can be associated with the administered amount of research grants and contracts coming from the state which can be curtailed for those HEIs showing a high level of cost inefficiency in the long run. The reciprocal contribution of short-and-long run rigidities in inefficiency mitigate a more cost efficient operational framework for HEIs. Consequently, policy formulation should be redefined in order for HEIs to prosper at an inter-and-intra country level.

The results suggest that HEIs tend to curtail part of their short-run inefficiency by removing some of the short-run rigidities (i.e. singular management mistakes), while some other sources of inefficiency tend to persist over time. Different approaches and specifications are compared which cannot yield specific conclusions, however through the GTRE and MGTRE models we can partially support the existence of two different parts of efficiency that differ in absolute value and are relatively different from previous

[198] Further discussion and definition on the main stakeholders in a HEI is offered by Mainardes et al. (2010).

approaches used in the literature. The valuable contribution of the GTRE and MGTRE models is apparent since, in a single step, any persistent undesirable factors and/or recurring identical management mistakes that hinder institutions from being efficient in the long-run, can be determined. However, with a relatively small sample size and short panel of data, the results should be interpreted with caution, as the findings might not be transferable and may not characterise the whole sector effectively. As a result an extension of the time scale of the panel could adequately validate or dispute the existing findings since there are factor misallocations that are difficult to rectify in a short period of time.

7. Chapter 7: Conclusions and Policy Implications

7.1 Summary

This thesis investigates efficiency measures of English HEIs and is summarised as follows. In chapter 1, an introduction to the motivation for and contribution toward the thesis is available. In addition, the main research objectives are outlined as well as the structure of each individual chapter. This chapter mainly focuses on the research gaps of previous studies regarding efficiency issues in HE and addresses two main research questions. Chapter 2 discusses the composition of the English HE sector. Specific details are offered regarding the history of the sector, the development phases, the structural financial reforms and the fair access and widening participation of the sector. Having identified some of the research gaps existing in the literature, a further review of the available methodological tools in the literature is offered in chapter 3. This chapter draws upon the entire literature, tying up the various theoretical and empirical strands of frontier estimation techniques in order to raise questions about both the theoretical configuration of efficiency measures as well as to identify most of the empirical weaknesses in the existing literature.

The first research objective of this thesis is developed in chapter 4 which adequately examines the effect of merger activity on the technical efficiency of HEIs in England. In this chapter, both parametric and non-parametric methods are utilised in an attempt of capture all the possible external factors that can drive technical efficiency towards an increasing or decreasing pathway. For comparative purposes, and to validate the proposed conclusions chapter 5 utilises the same dataset but implements a completely alternative approach to access the effect of mergers and to control for endogeneity issues that previous methods may have suffered from. Consequently, in chapter 5 propensity score matching and inverse weighting probability statistical techniques are applied to estimate the effect of a treatment, policy, or other intervention-in this case merger activity-by accounting for the covariates that predict how the treatment will be received.

The second research objective is analytically developed in chapter 6 and pertains to the existence of unobserved heterogeneity bias in most of the cost efficiency studies in HE. This chapter also pinpoints the critical step of separating two different types of cost inefficiency, one stemming from permanent and systematic shortfalls in the institutional or sector level (persistent part) and one that accounts for short-term changeable factors that can be adjustable in the short-term (transient part). As a result, two different measures of cost efficiency can be produced by utilising a GTRE model and both of them are consequential for policy makers and stakeholders when incentive-based regulation schemes are under consideration. Finally, the last chapter of this thesis binds the whole thesis by discussing the main findings, offering policy recommendations and discussing the implications of the findings for future research in this area.

7.2 Policy Recommendations and Conclusions

The English HE sector has undergone various changes in size and shape over time. HEIs in England have experienced unprecedented changes to their funding sources since extreme scrutiny has been put on public money spent in various commonwealth public services including the HE sector. Their overall trajectory has entailed practises and platforms of improved efficiency so as to produce an immune and efficient operational system to face any future uncertainty by diversifying their income and ensuring sufficient margins for reinvestment (UUK, 2014).

The arguments of this thesis intend to shed new light on existing topics within the HE efficiency literature and to unravel some of the main components contributing to institutional or sector inefficiency. The findings emerging from the statistical analysis presented in the previous chapters highlight the determinant role of merger activity as a policy intervention that improves efficiency since merged HEIs boast efficiency which is 5 percentage points higher post-merger than non-merging HEIs, holding all else constant. The positive spill-over effects on efficiency from the merger tend to vanish rapidly almost one year after the merger and this should be taken into account by policymakers not only for a policy as a whole but also with regard to the internal conformation of the new entity. Among the various changing factors that institutions have to comply with, such as changing public and student expectations, new tuition fee regimes, increased competition for international students, shifting patterns of enrolment, the findings of this thesis reveal that a higher proportion of income from government sources incentivises greater efficiency – a result with clear policy implications since generous public procurement has been vastly debated in the past. However, without proper public procurement based on transparency and accountability, the English HE system cannot be responsive to the current needs. So, for building an increasingly outward facing HE system with a research-led spectrum, public financial support is of substantial importance.

Universities have to develop activities beyond their traditional framework and, according to the indications of this thesis, the effect of third mission activity on efficiency suggests that interaction with business positively contribute to MTE. Despite the limited scope of the results due to the relatively small amount of data and its neglect of this third output in the efficiency determination of the first stage, the expectations were conducive for potential positive effects. Another intriguing point is the decreasing tendency for costs when universities increase their involvement with third mission activities. Nevertheless, this inference should be interpreted with caution since it is validated only when a translog specification is used. So, the result might be sensitive to the model specification since the log-linear case indicates the opposite effect. Hence, the generalisability of these results is subject to certain limitations.

An arguable point of interest is the effect of overseas campuses on the efficiency of HEIs. Despite the proliferation of courses offered outside the UK's borders, and the emergence of multi-campus institutions, according to our findings there is no significant effect in terms of efficiency from such an expansion of the English HE. However, this outcome may be misleading due to inappropriate or missing source of information. The insignificant effect of overseas campuses on efficiency leaves space for universities to build international partnerships to capitalise overseas opportunities even if there is no immediate significant effect in terms of efficiency.

This thesis touches upon another angle of the cost efficiency spectrum and, for the first time in the HE literature, identifies whether persistent (long-run) or transient (short-run) efficiency dominates the English HE sector. The precise mechanism of splitting the error component when, concurrently, latent heterogeneity is disentangled from inefficiency, has been previously used successfully elsewhere in the literature. Due to the large and diverse nature of the English HE sector this approach is adopted in this thesis as well.

Based on the findings, the English HE sector, on average, is more efficient in the short-run than the long-run. This supports the debate for systematic and predominant long-term deficiencies in management, operational and/or funding strategies or other factor misallocations, unfavourable to change over time. These results can function as an informative platform to policy-makers since long-run inefficiency stays with the unit (e.g. university) over time, while short-run inefficiency may alter from period to period. This is particularly important from a regulatory point of view since according to this study there appears to be an immense need for more long-term interventions and there should be a closer look into individual cases as to whether there are short-or long-term rigidities that hinder higher efficiency attainment. Building upon this framework, regulators should ensure a regime of government funding that rewards universities which attain long-run efficiency improvements but at the same time maintain quality compliance. Even short-run efficiency alleviation should be considered as desirable, but universities with systematic efficiency shortfalls in the long-run could be offered targeted efficiency programmes that would contribute in a uniform internal reorganisation of their sub-systems and in an external sector-wide transformation and development.

The crucial step of achieving efficiency savings in English HE, in a new territory for continuous efficiency improvements, is transparent according to the main findings of this thesis. Among the available policy interventions and actions that could potentially be put in place to ensure a continuous commitment to efficiency are shared services models or outsourcing[199]. The two main conclusions of this thesis imply that efficiency

[199] For a formal definition of the shared services follow Herbert and Rothwell (2015). The main conceptual sketch when shared services are applied is to provide consolidation of standardised processes across multiple organisations, within the same organisation, to achieve economies of scale.

savings are made through the merger mechanism and that there is a need for further cost efficiency improvements in the long-run. These results reinforce the understanding of shared services delivery as a policy tool that might offer a flexible and adjustable framework to suit a variety of requirements (UUK, 2011). Specifically, the identification of lower long-run cost efficiency in the English HE sector could fixed through the spectrum of shared services since this seems to increase management efficiency and effectiveness, in the areas of IT services, communication and management of financial and human resources. In general, the English HE sector should maintain its fundamental purpose of delivering high quality of teaching and research ingrained with the key principle of remarkable consistency with efficiency.

7.3 Considerations for Future Research

Despite its limitations and low reliability in discriminating between the middle-performing HEIs in terms of their level of efficiency, DEA can clearly discriminate between the worst- and best-performing HEIs (Johnes, 2006). While no differences emerge between HEI types in terms of the efficiency with which inputs are converted into outputs, the efficiency scores indicate that differences in efficiency between the worst- and best-performing English HEIs is significant.

The quantitative analysis conducted in Chapters 4 to 6 represents a 'top-down' approach to investigating HEIs' efficiency. There are high-performing universities in terms of efficiency so their efficiency score is high close to unit, and there are others with low performance, displaying low efficiency scores. Having explored the level of efficiency through the quantitative analysis, i.e. how much or to what extent is an institution efficient, we now investigate why universities behave in this way in terms of efficiency, how their decisions are formed, how HEIs are affected by the events that happen around them, how and why their policies have developed in certain ways, and what the differences are between groups of institutions in the same efficiency wavelength In that sense, rather than testing a hypothesis, qualitative research tends to engage in a much more dialectic process between the questions asked and the data observed. Hence, general forces play out in specific circumstances to ask questions that cannot easily be put into numbers (Greenhalgh and Taylor, 1997).

The explanatory nature of qualitative analysis empowers the method towards the quantitative analysis into stressing the research objectives labelled with the principles of breadth, precision, and accuracy (Becker, 1996). Instead of isolating variables to test any effects, qualitative analysis tries to look at a broad range of interconnected processes or causes. More specifically, we might be interested in undertaking a 'bottom-up' approach to try to gain further insight into the characteristics that contribute to a successful merger. To this end, a small number of mergers from the sample will be chosen for further exploration. These may be selected because of interesting features highlighted in the quantitative analysis or because of interest from policymakers. These chosen mergers will form case studies, and, using data collected from additional sources, as well as interviews, a much more detailed knowledge of the institutions involved in the merger, and the effects of the merger, is possible.

Therefore, the objective of the case study research is the development of a conceptual model that explains the adoption of the merger policy in low- to medium-performing universities so as to improve efficiency or to identify any other traits or patterns that can potentially lead to higher inefficiency. The objective of the study fits well with the philosophical nature of grounded theory. In this type of theory, the benefits are twofold, since it encompasses both induction, in which hypotheses are formulated from specific data, and deduction, in which specific conclusions are drawn from hypotheses.

Grounded theory originated in the work of Glaser and Strauss (1967) and is a method that has been used extensively across a variety of social science disciplines. A grounded theory is one that is discovered, developed, and provisionally verified through systematic data collection and analysis of data pertaining to a particular phenomenon (Strauss and Corbin, 1990). It also minimises influences from existing theories and the researcher's pre-conceived ideas since the method itself provides meanings, definitions, and interpretations that are made by the subjects of the study, and the researcher derives the categories from the field through in-depth examination and exposure to the phenomena (De Burca and McLoughlin, 1996).

Researchers state that using case studies in research enables them to answer questions such as 'how' and 'why', while also studying how a certain phenomenon is influenced by the context of its situation (Baxter and Jack, 2008). In particular, we aim to unfold research questions such as how and why some institutions tend to perform better than others in the efficiency scale and to identify common patterns or characteristics among the universities in each group (i.e. low-performing and high-performing) and see whether there are common practices that each group shares. To this end, case study research provides a suitable framework to understand the complex nature of efficiency and extend experience or add strength to what is already known from previous research (Soy, 1997). Also, it can be an additional tool to access merger effects since it emphasises detailed contextual analysis of a limited number of events or conditions (mergers) and their relationships. The case study definition is framed by Yin (1984, p. 23) as an empirical inquiry that investigates a contemporary phenomenon within its real-life context, when the boundaries between phenomenon and context are not clearly evident, and in which multiple sources of evidence are used.

A major caveat in the research design of case studies is the extent to which the research results are generalisable. This refers to whether the findings may be equally applicable to other research settings, such as other organisations. In cases in which the data collection emerges from one organisation, or a small number of organisations, or the organisation is markedly 'different'[200] in some way, we cannot claim that the results obtained are generalisable to all populations. That being the case, to improve the

[200] This is the case for English HEIs, which are different in many different ways. The diversity of the sector is high not only in terms of their mission and subjects, but also in the number of students recruited (Ramsden, 2012).

external validity of our study, we opt for multiple mini case studies rather than a single one as a mechanism to improve the generalisability and to test the robustness of our conclusions before universalising them (Saunders et al., 2009).

The sample criteria will be shaped according to purposeful sampling, which involves different methods amenable to different types of study, depending on the aims and conditions of the study. In particular, the sampling process will concentrate on a small number of institutions that are efficient, and a small number that are inefficient, which will be investigated further using data collection from additional sources, and using interviews. In addition, another criterion in the sampling process will be merger activity and the accessibility of the institution since, in most cases, a single interview may not be sufficient. The aim is to capture and describe the variations of the phenomenon, i.e. high or low efficiency levels in different contexts (such as merger activity). Hence, in a context of maximum variation in the sample, problems may arise in understanding the phenomenon if the sample is too small and too heterogeneous since individual cases might be different from one another. The main task, however, in the sample formulation will be to focus on finding information that highlights variations of the phenomena and significant common patterns within the varying sample (SBU, 2016).

During the data collection and data analysis process, focus group interviews with semi-structured questions and content analysis are considered appropriate. This semi-structured interview technique requires modest knowledge and allows the interviewer to follow up on ideas raised by the participants (Olson, 2011). This means that all interviews contain the same questions in the same order, but different follow-up questions are allowed (SBU, 2016). In order to maintain consistency between research questions asked and our research objectives, a pilot study is considered appropriate so as to confirm the validity of the research design of the questionnaire. Therefore, before administering the questionnaire, it is beneficial to complete all the revisions highlighted as necessary by the pilot testing (Sauders et al., 2009).

Once the data collection is completed, we may proceed with the data analysis process. The complexity of the analysis always depends on the current research question(s) and objectives. The main goal here is to identify potential emerging patterns. The amount of interactions may vary by the nature of the group or its activity, so patterns reflecting relationships between numbers of interaction categories may become evident (Sauders et al., 2009). The actual goal of this step is to generate proposals in order to advise policymakers on whether there are systematic and structural differences in the HE sector or if there are intriguing singular characteristics of each unit (university) that render them efficient or inefficient. Also, the data analysis can comprise a useful guideline for successful practices for institutions that lie at the bottom of the efficiency scale. This level of analysis is obviously more complex and will require computer software to calculate the cross-classifications. All interviews will be audio recorded. Then, they will be transcribed and analysed with the help of NVivo, so that themes and patterns might emerge.

An outline of the theoretical sketch of the qualitative research has been adequately described; however, the practical application of the interviews, the data analysis, and the suggested implications of this process will be left for future research.

Appendix

For compliance to the word length limit regulations of Lancaster University Management School [LUMS] the appendices part is cited in a separate external space. It is available online follow: https://sites.google.com/site/mariapapadimitr/phd_thesis_appendix

References

Aarrevaara, T. (2007). “Enhancing Universities in Finland.” Journal of Finance and Management in Colleges and Universities 4: 261–92.

Abadie, A. & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. Econometrica 74(1), 235-267.

Abadie, A., & Imbens, G.W. (2011). ‘Bias Corrected Matching Estimators for Average Treatment Effects’. Journal of Business and Economic Statistics, 29(1), 1-11.

Abadie, A., & Imbens, G.W. (2012). Matching on the estimated propensity score. Harvard University and National Bureau of Economic Research

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. (2001). Total factor productivity and efficiency in Australian colleges of advanced education. Journal of Educational Administration, 39(4), 384-393.

Abbott, M., & Doucouliagos, C. (2002). A data envelopment analysis of the efficiency of Victorian TAFE institutes. Australian Economic Review, 35(1), 55-69.

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

Abbott, M., & Doucouliagos, C. (2004). "Research output of Australian universities," Education Economics, Taylor & Francis Journals, vol. 12(3), pages 251-265.

Abbott, M., & Doucouliagos, C. (2009). Competition and efficiency: overseas students and technical efficiency in Australian and New Zealand universities. Education Economics, Taylor & Francis Journals, vol. 17(1), 31-57.

Abou-Warda, S. H. (2011). Quality in Egyptian Higher Education Context: DEA Approach to Enhancement of Internal Quality Assurance Systems in Technical Departments1. Arab Journal of Administrative Sciences, 18, 165-194.

Abramo, G., Cicero, T., & D’Angelo, C. A. (2011). A field-standardized application of DEA to national-scale research assessment of universities. Journal of Informetrics, 5(4), 618-628.

Abramo, G., D'angelo, C. A. & Pugini, F. (2008). The measurement of Italian Universities' research productivity by a non-parametric-bibliometric methodology, Scientometrics, 76, 225-244

Adams, R. M., Berger, A. N., & Sickles, R. C. (1999). Semiparametric approaches to stochastic panel frontiers with applications in the banking industry. Journal of Business & Economic Statistics, 17(3), 349-358.

Afonso, A., & Santos, M. (2008). A DEA approach to the relative efficiency of Portuguese public universities, Portuguese Journal of Management Studies, 67-87.

Afriat, S. N. (1972). Efficiency Estimation of Production Functions, International Economic Review, Vol. 13, No. 3, pp. 568-598.

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

Agasisti, T., & Dal Bianco, A. (2009). Reforming the university sector: effects on teaching efficiency-evidence from Italy. Higher education, 57(4), 477.

Agasisti, T., & Gralka, S. (2017). The transient and persistent efficiency of Italian and German universities: A stochastic frontier analysis (No. 14/17). CEPIE Working Paper.

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

Agasisti, T., & Johnes, G. (2010). Heterogeneity and the evaluation of efficiency: the case of Italian universities, Applied Economics, 42:11, 1365-1375, DOI: 10.1080/00036840701721463.

Agasisti, T., & Johnes, G. (2015). Efficiency, costs, rankings and heterogeneity: the case of US higher education, Studies in Higher Education, 40:1, 60-82, DOI: 10.1080/03075079.2013.818644.

Agasisti, T., & Pérez-Esparrells, C. (2010). Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities. Higher Education, 59(1), 85-103.

Agasisti, T., & Salerno, C. (2007). Assessing the Cost Efficiency of Italian Universities, Education Economics, 15:4, 455-471, DOI: 10.1080/09645290701273491.

Agasisti, T., Barra, C., & Zotti, R. (2016). Evaluating the efficiency of Italian public universities (2008–2011) in presence of (unobserved) heterogeneity, Socio-Economic Planning Sciences, Volume 55, September 2016, Pages 47–58.

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

Agha, S. R., Kuhail, I., Abdelnabi, N., Salem, M., & Ghanim, A. (2011). Assessment of academic department’s efficiency using data envelopment analysis, Journal of Industrial Engineering and management, 4, 301-325.

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 (ed) Creative and Innovative Approaches to the Science of Management Quorum Books, Westport, Connecticut, pp191-208

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 Non-profit Accounting, 5(2), 165-185.

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

Aigner, D. J., & Chu, S. F. (1968). On estimating the industry production function. The American Economic Review, 58(4), 826-839.

Aigner, D. J., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6 (1), 21-37.

Akhavein, J. D., Berger, A. N., & Humphrey, D. B. (1997). The effects of megamergers on efficiency and prices: Evidence from a bank profit function. Review of industrial organization, 12(1), 95-139.

Alirezaee, M. R., Howland, M., & Van de Panne, C. (1998). Sampling size and efficiency bias in Data Envelopment Analysis. Advances in Decision Sciences 2(1): 51-64.

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, 13-34.

Almanidis, P., & Sickles, R. C. (2011). The Skewness Issue in Stochastic Frontiers Models: Fact or Fiction?. In Exploring research frontiers in contemporary statistics and econometrics (pp. 201-227). Physica-Verlag HD.

Andersen, P., & Petersen, N.C. (1993). A procedure for ranking efficient units in data envelopment analysis, Management Science 39, 1261–64.

Anderson, C., Antellius, J., & Sund, K. (2010). Technical efficiency and productivity at higher education institutions–Some problems and some solutions. Journal of economics literature, 1.

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

Andor, M., & Hesse, F. (2014). 'The StoNED Age: The departure into a new era of efficiency analysis? A Monte Carlo comparison of StoNED and the \oldies" (SFA and DEA)', Journal of Productivity Analysis 41(1), 85{109.

Andrade, G., Mitchell, M., & Stafford, E. (2001). New evidence and perspectives on mergers. The Journal of Economic Perspectives 15 (2), 103–120.

Aragon, Y., Daouia, A., & Thomas-Agnan, C. (2005). Nonparametric Frontier Estimation: A Conditional Quantile-based Approach, Econometric Theory, 21, 358–389.

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

Archer, L., Hutchings, M., & Ross, A. (2002). Higher education and social class: Issues of exclusion and inclusion (London: Routledge Falmer).

Archibald, R. B., & Feldman, D. H. (2008). Graduation rates and accountability: Regressions versus production frontiers. Research in Higher Education, 49(1), 80-100.

Aristovnik, A., & Obadić, A. (2011). The funding and efficiency of higher education in Croatia and Slovenia: a non-parametric comparison.

Athanasoglou, P. P., & Brissimis, S. N. (2004). The effect of mergers and acquisitions on bank efficiency in Greece. Economic Bulletin 22, 7–31.

Athanassopulos, A. & Shale, E. A. (1997). 'Assessing the Comparative Efficiency of Higher Education Institutions in the UK by Means of Data Envelopment Analysis', Education Economics, Vol. 5, No.2, pp. 117–135.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in Medicine 28: 3083–3107.

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res., 46(3): 399–424, doi: 10.1080/00273171.2011.568786.

Austin, P. C., & Stuart, E.A. (2015). The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. Statistical methods in medical research, pii: 0962280215584401.

Austin, P. C., Grootendorst, P., & Anderson, G.M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study. Statistics in Medicine, 26, 734–753. doi:10.1002/sim.2580

Avkiran, N. K. (1999). The evidence on efficiency gains: The role of mergers and the benefits to the public. Journal of banking & finance, 23(7), 991-1013.

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

Azziz, R., Hentschke, G.C., Jacobs, B.C., Jacobs, L.A., & Ladd, H. (2017). Mergers in higher education: A proactive strategy to a better future? New York, NY: TIAA Institute.

Bădin, L., Daraio, C., & Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. European journal of operational research, 201(2), 633-640.

Bădin, L., Daraio, C., & Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. European Journal of Operational Research, 223(3), 818-833.

Bailey, T., & Xu, D. (2011). Input-Adjusted graduation rates and college accountability: What is known from twenty years of research? New York, NY: Columbia University, Teachers College, Community College Research Centre

Baker, S. (2011). 'School of Pharmacy to merge with UCL.' Times Higher Education 20th June 2014.

Balk, B. M., (2001). Scale Efficiency and Productivity Change, Journal of Productivity Analysis, 15, 159-83

Baltagi, B. H., (2013). Econometric Analysis of Panel Data, 5th edn, John Wiley & Sons, Great Britain.

Banker, R. D., & Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. Journal of Productivity Analysis, 3(4), 401-415.

Banker, R. D., & Morey, R.C. (1986). Efficiency analysis for exogenously fixed inputs and outputs. Operations research, 34(4), 513-521.

Banker, R. D., & Natarajan, R. (2004). Statistical tests based on DEA efficiency scores. Chapter 11. In: Cooper WW, Seiford L and Zhu J (eds). Handbook on Data Envelopment Analysis. Kluwer Academic Publishers: Norwell, MA, pp 299–321

Banker, R. D., Charnes A., & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci. 30:1078–92.

Barr, N., & Glennerster, H. (2014) Preface, in: N. Barr (Ed.) Shaping higher education: 50 years after Robbins (London: London School of Economics and Political Science).

Battese, G. E., & Coelli, T. J. (1988). "Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data," Journal of Econometrics, 38, pp. 387-399.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. In International applications of productivity and efficiency analysis (pp. 149-165). Springer Netherlands.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics, 20 (2), 325-332.

Baumol, W. J., Panzar, J. C., Willig, R. D., Bailey, E. E., Fischer, D., & Fischer, D. (1982). Contestable markets and the theory of industry structure. New York: Harcourt Brace Jovanovich, Inc.

Baxter, P. & Jack, S. (2008). "Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers", The Qualitative Report, Vol. 13 No. 4, pp. 544-559. Available at http://nsuworks.nova.edu/tqr/vol13/iss4/2 [Accessed 8 August 2017].

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

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

Becker, H. S., (1996). The epistemology of qualitative research. Ethnography and human development: Context and meaning in social inquiry, 53-71.

Becker, S. O. & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores - The stata journal, Volume 2 Number 4: pp. 358-377.

Berg, S. & Padowski, J. (2010). Overview of Water Utility Benchmarking Methodologies: From Indicators to Incentives. Public Utility Research Center, University of Florida, Gainesville.[]© AFD/The Regulation of Water and Sanitation Services in DCs/May, 102.

Bergantino, A. S., Capozza, C., & Porcelli, F. (2012). Measuring Teaching Efficiency of the Italian University System: the Role of Market Structure. A two-step DEA analysis at faculty level. Associazione Italiana di Scienze Regionali XXXIII Conferenza Scientifica Annuale Roma.

Berger, A. N. (1998). The efficiency effects of bank mergers and acquisition: A preliminary look at the 1990s data. In Bank Mergers & Acquisitions (pp. 79-111). Springer US.

Berger, A. N., & Humphrey, D. B. (1992). Megamergers in banking and the use of cost efficiency as an antitrust defense. The Antitrust Bulletin, 37(3), 541-600.

Berry, S. K., (2000). Excess returns in electric utility mergers during the transition to competition. Journal of Regulatory Economics 18 (2), 175–188.

Bisoffi F. G. B., Bonadia G. C., de Oliveira V. H. D. & Barbosa, S.R. (2017). Method for Estimating Confidence Intervals for DEA Efficiency Models Using Regression Models. In: Paradisi A., Godoy Souza Mello A., Lira Figueiredo F., Carvalho Figueiredo R. (eds) Cognitive Technologies. Telecommunications and Information Technology. Springer, Cham

Black, D., & Smith, J. (2003). “How Robust is the Evidence on the Effects of the College Quality? Evidence from Matching,” Working Paper, Syracuse University, University of Maryland, NBER, IZA.

Blundell, R., Dearden, L., & Sianesi, B. (2005). “Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS”, Journal of the Royal Statistical Society: Series A (Statistics in Society) volume 168, Issue 3, pages 473–512.

Blundell, R., Dias, M. C., Meghir, C., & Van Reenen, J. (2004). Evaluating the employment impact of a mandatory job search program. Journal of the European Economic Association, Volume 2, Issue 4, pages 569–606.

Bobe, B. (2009). Evaluating the efficiencies of university faculties: adjusted data envelopment analysis. In AFAANZ 2009: Accounting and Finance Association of Australia and New Zealand Annual Conference (pp. 1-26). AFAANZ.

Boliver, V. (2013). How fair is access to more prestigious UK universities?. The British journal of sociology, 64(2), 344-364.

Bolli, T., & Somogyi, F. (2011). Do competitively acquired funds induce universities to increase productivity?. Research Policy, 40(1), 136-147.

Bonaccorsi, A., Daraio, C., & Simar, L. (2006). Advanced indicators of productivity of universities. An application of robust nonparametric methods to Italian data. Scientometrics, 66(2), 389-410.

Botha, M. (2001). Models for mergers in South African Journal of Education, 21(4), 3-8.

Boxall, M., & Woodgates, P. (2014). Here be dragons: How universities are navigating the uncharted waters of higher education. London, PA Consulting Group.

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.

Breneman, D. W. (2001).The outputs of Higher Education. Ford Policy Forum. Cambridge, University of Virginia, - educause.edu

Breu, T. M., & Raab, R. L. (1994). Efficiency and perceived quality of the nation's “top 25” National Universities and National Liberal Arts Colleges: An application of data envelopment analysis to higher education. Socio-Economic Planning Sciences, 28(1), 33-45.

Brinkman, P. T. (1990). ‘Higher education cost functions’, in Stephen A. Hoenack and Eileen L. Collins (eds), The Economics of American Universities, Albany, NY: State University of New York Press, pp. 107–28.

Brockhoff, K. (1970). "On the Quantification of the Marginal Productivity of Industrial Research by Estimating a Production Function for a Single Firm", German Economic Review vol. 8, pp. 202–229.

Brookhart, A., Wyss, R., Layton, B., & Stürmer, T. (2013). Propensity Score Methods for Confounding Control in NonExperimental Research. Circ Cardiovasc Qual Outcomes, Vol 6(5): 604–611. doi:10.1161/CIRCOUTCOMES. 113.000359.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. American Journal of Epidemiology, 163, 1149– 1156.

Brown, R. (2010). Consolidation of the financial services sector: Implications for Australian credit unions. Economic Papers: A journal of applied economics and policy, 18(1) 72-82.

Browne, J. (2010). 'Securing a sustainable future for higher education: An independent review of higher education funding and student finance.' Department for Business, Innovation and Skills, London.

Bryson, A., Dorsett, R. & Purdon S. (2002). "The Use of Propensity Score Matching in the Evaluation of Labour Market Policies," Working Paper No. 4, Department for Work and Pensions.

Business and Higher Education Roundtable [BHERT], (2006). Universities' third mission Communities engagement. BHERT Position Paper. Fitzroy: Author.

Butler, J., & Moffitt, R. (1982) A computationally efficient quadrature procedure for the one factor multinomial probit model. Econometrica 50:761–764.

C.U.C. [Committee of University Chairs]. (2009). Guide for members of Higher Education Governing Bodies in the UK. Governance Code of Practice and General Principles, February 2009, 2009/02.

Cai, Y., & Yang, X. (2015). "Mergers in Chinese Higher Education: Lessons for Studies in a Global Context." European Journal of Higher Education 6 (1): 71–85

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of economic surveys, 22(1), 31-72.

Callender, C. (2010). Bursaries and institutional aid in higher education in England: do they safeguard and promote fair access?, Oxford Review of Education, 36:1, 45-62, DOI: 10.1080/03054980903518910.

Callender, C. (2011). Widening participation, social justice and injustice: part-time students in higher education in England, International Journal of Lifelong Education, 30:4, 469-487, DOI: 10.1080/02601370.2011.588462.

Callender, C. (2015). "The Cost-Cutting Reforms of Higher Education in England," International Higher Education, Number 63, Spring 2011, pp. 17-19.

Cameron, A. C., & Trivedi, P. K. (2005). MICROECONOMETRICS: Methods and Applications, Cambridge University Press, New York, May 2005.

Cameron, A. C., & Trivedi, P. K. (2010). Microeconometrics using stata (Vol. 2). College Station, TX: Stata press.

Carrington, R., Coelli, T., & Rao D.S. (2005). The performance of Australian universities: Conceptual issues and preliminary results. Economic Papers: A journal of applied economics and policy, 24(2), 145-163.

Castano, M. C. N., & Cabanda, E. C. (2007). Performance evaluation of the efficiency of Philippine Private Higher Educational Institutions: application of frontier approaches. International Transactions in operational research, 14(5), 431-444.

Casu, B., & Thanassoulis, E. (2006). Evaluating cost efficiency in central administrative services in UK universities. Omega, 34(5), 417-426.

Casu, B., Shaw, D., & Thanassoulis, E. (2005). Using a group support system to aid input–output identification in DEA. Journal of the Operational Research Society, 56(12), 1363-1372.

Caves, D., Christensen L., & Trethaway, M. (1980). "Flexible Cost Functions for Multiproduct Firms," Review of Economics and Statistics, 62, pp. 477-481.

Cazals, C., Florens, J. P., & Simar L. (2002). Nonparametric Frontier Estimation: a Robust Approach, in Journal of Econometrics, 106, 1–25

Chambers, R. G. (1988). Applied production analysis: a dual approach. Cambridge University Press.

Chang, D. F., Wu, C. T., Ching, G. S., & Tang, C. W. (2009). An evaluation of the dynamics of the plan to develop first-class universities and top-level research centers in Taiwan. Asia Pacific Education Review, 10(1), 47-57.

Chapple, W., Lockett, A., Siegel, D., & Wright, M. (2005). Assessing the relative performance of UK university technology transfer offices: parametric and non-parametric evidence. Research Policy, 34(3), 369-384.

Charnes, A., & Cooper, W. W. (1961) Management Models and the Industrial Applications of Linear Programming. John Wiley, New York.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). 'Measuring the Efficiency of Decision Making Units', European Journal of Operational Research, Vol. 2, pp.429–444

Charnes, A., Cooper, W. W., Lewin A. Y., & Seiford, L. M. (1995), Data Envelopment Analysis: Theory, Methodology and Applications, Kluwer

Chen, Y. Y., Schmidt, P., & Wang, H. J. (2014). Consistent estimation of the fixed effects stochastic frontier model. Journal of Econometrics, 181(2), 65-76.

Cherchye, L., De Witte, K., Ooghe, E., & Nicaise, I. (2010). Equity and efficiency in private and public education: a nonparametric comparison, European Journal of Operational Research, 202, 563–73.

Chu Ng, Y., & 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.

Cinar, Y. (2016). Research and Teaching Efficiencies of Turkish Universities with Heterogeneity Considerations: Application of Multi-Activity DEA and DEA by Sequential Exclusion of Alternatives Methods. arXiv preprint arXiv:1701.07318.

Clark, B. R. (1998). Creating Entrepreneurial Universities: Organization Pathways of Transformation. Guildford, UK: Pergamon

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics, 24, 295–313.

Cochran, W. G., & Rubin, D. B. (1973), “Controlling Bias in Observational Studies”, Sankyha, 35, 417-446.

Coelli, T. J. (1996a). A guide to DEAP version 2.1: a data envelopment analysis (computer) program. Centre for Efficiency and Productivity Analysis, University of New England, Australia.

Coelli, T. J., Perelman, S., & Romano, E. (1999). "Airlines, Environment and Technical Efficiency: An International Comparative Study", Journal of Productivity Analysis, 11, 251-273

Coelli, T. J., Rao, D. S. P., & Battese, G. E. (1998). An Introduction to Efficiency and Productivity Analysis, Boston: Kluwer Academic Publishers, Inc.

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

Cohn, E., & Cooper, S. T. (2004). Multi-product cost functions for universities: Economies of scale and scope. In G. Johnes & J. Johnes (Eds.), International handbook on the economics of education (pp. 579-612). Cheltenham, UK: Edward Elgar Publishing Ltd.

Cohn, E., Rhine S. L. W., & Santos, M. C. (1989). 'Institutions of higher education as multi-product firms: economies of scale and scope.' Review of Economics and Statistics 71(2): 284-290.

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.

Colombi, R. (2010). A skew normal stochastic frontier model for panel data. In: Proceedings of the 45-th scientific meeting of the Italian statistical society.

Colombi, R., Kumbhakar S., Martini G., & Vittadini G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. J Prod Anal (Published online).

Colombi, R., Martini, G., &Vittadini, G. (2011). A stochastic frontier model with short-run and long-run inefficiency random effects. In: Department of Economics and Technology Management, Economics, 43(1):44–56.

Considine, M. (2002). The end of the line? Accountable governance in the age of networks, partnerships and joined-up services. Governance, 15(1).

Cooper, W. W., Seiford, L. M., & Tone, K. (2000). Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Kluwer Academic Publishers, Boston.

Cooper, W. W., Seiford, L. M., & Zhu, J. (2011). Data envelopment analysis: history, models, and interpretations. In Handbook on data envelopment analysis (pp. 1-39). Springer US.

Cornwell, C., & Schmidt, P. (1996). Production frontiers and efficiency measurement. In The econometrics of panel data (pp. 845-878). Springer Netherlands.

Cornwell, C., Schmidt, P., & Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of econometrics, 46(1-2), 185-200.

Costa, E. M., de Sousa Ramos, F., & de Souza, H. R. (2012). Static versus dynamic DEA in Federal higher Education Institutions-IFES. XVII Encontro Regional de Economia. ANPEC Regional, Trabalho Completo.

Cubbin, J., & Tzanidakis, G. (1998). Techniques for Analysing Company Performance. Business Strategy Review 9(4): 37

Cuesta, R. A., & Orea L. (2002). 'Mergers and Technical Efficiency in Spanish Savings Banks: a Stochastic Distance Function Approach', Journal of Banking and Finance,A Vol. 26, No. 12, pp. 2231–2247

Cunha, M., & Rocha, V. (2012). On the efficiency of public higher education institutions in Portugal: an exploratory study. Universidade do Porto, Faculdade de economia do Porto, 486.

Curaj, A., Georghiou, L., Harper, J. H. & Egron-Polak, E. (2015). Mergers and Alliances in Higher Education, international Practice and Emerging Opportunities. Springer Open. DOI:10.1007/978-3-319-13135-1

Curri, G. (2002). "Reality Versus Perception: Restructuring Tertiary Education and Institutional Organisational Change – A Case Study." Higher Education 44 (1): 133–51.

Curtis, L. H., Hammill, B. J., Eisenstein, E. L., Kramer, J. M., & Anstrom, K. M. (2007). Using Inverse Probability-Weighted Estimators in Comparative Effectiveness Analyses With Observational Databases, Medical Care, Volume 45, Number 10 Suppl 2.

Daghbashyan, Z. (2011). The Economic Efficiency of Swedish Higher Education Institutions, CESIS Electronic Working Paper Series No. 245

Dalmau-Atarrodona, E. & Puig-Junoy, J. (1998). Market Structure and Hospital Efficiency: Evaluating Potential Effects of Deregulation in a National Health Service Review of Industrial Organization, 13(4): 447-466, doi:10.1023/A:1007775616593

Dan, M. C. (2012).The Third Mission of Universities in the Development Strategy of Vienna City. Informatica Economică vol. 16, no. 4/2012

Daneshvar Rouyendegh, B. (2011). The DEA and intuitionistic fuzzy TOPSIS approach to departments' performances: a pilot study. Journal of Applied Mathematics, 2011.

Dantzig, G. B. (1951a). "Maximization of a Linear Function of Variables Subject to Linear Inequalities," in T.C. Koopmans (ed.), Activity Analysis of Production and Allocation, John Wiley & Sons, New York, 339–347

Daouia, A., & Gijbels, I. (2011). Robustness and inference in nonparametric partial frontier modeling. Journal of Econometrics, 161(2), 147-165.

Daouia, A., & Ruiz-Gazen, A. (2006). Robust nonparametric frontier estimators: qualitative robustness and influence function. Statistica Sinica, 1233-1253.

Daouia, A., & Simar, L. (2005). Robust nonparametric estimators of monotone boundaries. Journal of Multivariate Analysis, 96(2), 311-331.

Daouia, A., & Simar, L. (2007). Nonparametric efficiency analysis: a multivariate conditional quantile approach, Journal of Econometrics, 140, 375–400

Daouia, A., Florens, J. P. & Simar, L. (2008). Functional Convergence of Quantile-type Frontiers with Application to Parametric Approximations, Journal of Statistical Planning and Inference, 138, 708–725.

Daouia, A., Florens, J. P. & Simar, L. (2009). Regularization of Non-parametric Frontier Estimators. Discussion paper 0922, Institut de Statistique, UCL.

Daouia, A., Florens, J. P. & Simar, L. (2010). Frontier estimation and extreme values theory. Bernoulli, 16(4), 1039–1063.

Daouia, A., Florens, J. P., & Simar, L. (2012). Regularization of nonparametric frontier estimators. Journal of Econometrics, 168(2), 285-299.

Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach, Journal of Productivity Analysis, vol 24, 1, 93–121.

Daraio, C., & Simar, L. (2006). A robust nonparametric approach to evaluate and explain the performance of mutual funds. European Journal of Operational Research, 175(1), 516-542.

Daraio, C., & Simar, L. (2007). Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. Springer Science & Business Media.

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

Daraio, C., & Simar, L. (2014). Directional distances and their robust versions: Computational and testing issues. European Journal of Operational Research, 237(1), 358-369.

Daraio, C., Bonaccorsi, A., Geuna, B., Lepori, L., Bach, P., Bogetoft, M. F., Cardoso, E., Castro-Martinez, G., Crespi, I., Fernandez de Lucio, H. O., Fried, A., García-Aracil, A., Inzelt, B., Jongbloed, G., Kempkes, P., Llerena, M., Matt, M. A., Olivares, C., Pohl, T., Raty, M. J., Rosa, C. S., Sarrico, L., Simar, S., Slipersaeter, P. Teixeira N. & Vanden Eeckaut P. (2011). 'The European university landscape: A micro characterization based on evidence from the Aquameth project.' Research Policy 40(1): 148-164.

Daraio, C., Simar, L., & Wilson, P. W. (2018). Central limit theorems for conditional efficiency measures and tests of the “separability” condition in nonparametric, two-stage models of production. The Econometrics Journal.

Das, S., & Das, M. (2014). Technical efficiency and performance of the higher educational institutions: a study of affiliated degree colleges at barak valley in assam. Journal of Social and Economic Policy, 11(1), 59-70.

De Burca, S., & McLoughlin, D. (1996). The Grounded Theory Alternative in Business Network Research, http://www.dcu.ie/business/research_interests.html

De Fraja, G., & Valbonesi, P. (2012). “The Design of the University System.” Journal of Public Economics 96 (3–4): 317–30

de Guzman, M. C. G. N., & Cabanda, E. (2011). Selected Private Higher Educational Institutions In Metro Manila: A DEA Efficiency Measurement. American Journal of Business Education (AJBE), 2(6).

de Mello, J. C. C. S., Gomes, E. G., Meza, L. A., de Mello, M. H. C. S., & de Mello, A. J. R. S. (2006). Engineering post-graduate programmes: A quality and productivity analysis. Studies in Educational Evaluation, 32(2), 136-152.

De Witte, K., & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables, Applied Economics, 45:17, 2401-2412, DOI:10.1080/00036846.2012.665602

De Witte, K., & Rogge, N. (2011). Accounting for exogenous influences in performance evaluations of teachers. Economics of Education Review, 30(4), 641-653.

Debreu, G. (1951). The coefficient of resource utilization. Econometrica: Journal of the Econometric Society, 273-292.

Dehejia, R. H., & Wahba, S. (1999). “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs,” Journal of the American Statistical Association, 94(448), 1053–1062.

Delmas, M., & Tokat, Y. (2005). Deregulation, governance structures and efficiency: the US electric utility sector. Strategic Management Journal 26, 441–460.

Denny, M. (1974). The Relationship between Functional Forms for the Production System. The Canadian Journal of Economics / Revue Canadienne D'Economique, 7(1), 21-31. doi:10.2307/134212

Department for Business Innovation and Skills [DBIS] (2011). Higher Education: Students at the Heart of the System, government White Paper, June 2011, http://www.bis.gov.uk/assets/biscore/higher-education/docs/h/11-944-higher-education-students-atheart-of-system.pdf (accessed 7 August 2017).

Deprins, D., Simar, L., & Tulkens, H. (1984). “Measuring Labor Efficiency in Post Offices.” In M. Marchand, P. Pestieau and H. Tulkens (eds.), The Performance of Public Enterprises: Concepts and Measurements. Amsterdam: North Holland, pp. 243-267.

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. Review of Economics and Statistics, 95(3), 932-945.

Diewert, W., & Wales, T. (1987). Flexible Functional Forms and Global Curvature Conditions. Econometrica, 55(1), 43-68. doi:10.2307/1911156

Domah, P. & Pollitt, M. (2001). "The Restructuring and Privatisation of Electricity Distribution and Supply Businesses in England and Wales: A Social Cost-Benefit Analysis." Fiscal Studies 22: 107-46.

Doyle, J. R., & Green, R. (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. Journal of the operational research society, 45(5), 567-578.

Doyle, J. R., & Green, R. H. (1994a). Self and peer appraisal in higher education. Higher education, 28(2), 241-264.

Doyle, J. R., Arthurs, A. J., Green, R. H., McAulay, L., Pitt, M. R., Bottomley, P. A., & Evans, W. (1995). The judge, the model of the judge, and the model of the judged as judge: analysis of the 1992 UK Research Assessment Exercise data for business and management studies (Working Paper, School of Management, University of Bath).

Duh, R. R., Chen, K. T., Lin, R. C., & Kuo, L. C. (2014). Do internal controls improve operating efficiency of universities?. Annals of Operations Research, 221(1), 173-195.

Dyckhoff, H., Dirksen, A., & Mbock, E. (2012). Measuring Balanced Efficiency with DEA: New Approach and Case Study of German Business Schools' Research Performance (August 26, 2012). Available at SSRN: https://ssrn.com/abstract=1990233 or http://dx.doi.org/10.2139/ssrn.1990233

E3M-Project. (2008). Application form 2008: Lifelong Learning Programme Sub-programmes – Multilateral Projects, Networks, Accompanying measures.

E3M-Project. (2012). Needs and constraints analysis of the three dimensions of third mission activities. Retrieved November 9, 2016, from http://www.e3mproject.eu/docs/Three-dim-third-mission-act.pdf

Eckles, J. E. (2010). Evaluating the efficiency of top liberal arts colleges. Research in Higher Education, 51(3), 266-293.

Ehrenberg, R. G. (2000). Tuition Rising: Why College Costs So Much, Cambridge, MA: Harvard University Press.

Emrouznejad, A., & Thanassoulis, E. (2005). A mathematical model for dynamic efficiency using data envelopment analysis. Applied mathematics and computation, 160(2), 363-378.

Enserink, M. (2007). Who Ranks the University Rankers? Science. 317, 1026–1028.

Etzkowitz, H. (2008). The Triple Helix model: University-industry-government innovation in action Routledge, London

Etzkowitz, H. et al. (2000). The Future of the University and the University of the Future: Evolution from Ivory Tower to Entrepreneurial Paradigm, Research Policy, 29, 313-330

EuroEducation [EED] (2017). HIGHER EDUCATION SYSTEM IN ENGLAND, THE UK. [online] Euroeducation.net. Available at: http://www.euroeducation.net/prof/ukco.htm [Accessed 8 August 2017].

European Commission. (2008). Needs and constraints analysis of the three dimensions of third mission activities. European indicators and ranking methodology for University Third Mission E3M. Retrieved on 1 March 2017 from http://www.e3mproject.eu/docs/Threedim-third-mission-act.pdf.

Eurydice (2009). National Summary Sheets on Education Systems in Europe and Ongoing Reforms United Kingdom (England, Wales and Northern Ireland). Brussels European Commission. Retrieved on 09/08/2017 from: http://www.voced.edu.au/content/ngv%3A13783

Evans, D., Tandon, A., Murray, C., & Lauer, J. (2000a). "The Comparative Efficiency of National Health Systems in Producing Health: An Analysis of 191 Countries," GPE Discussion Paper, No. 29, EIP/GPE/EQC, World Health Organization, Geneva.

Evans, D., Tandon, A., Murray, C., & Lauer, J. (2000b). "Measuring Overall Health System Performance for 191 Countries," GPE Discussion Paper, No. 30, EIP/GPE/EQC, World Health Organization, Geneva.

Fabrizio, K., Rose, N. L., & Wolfram, C. (2007). Do markets reduce costs? Assessing the impact of regulatory restructuring on US electric generation efficiency. American Economic Review 97 (4), 1250–1277.

Fan Y, Li Q., & Weersink, A. (1996). Semi-parametric estimation of stochastic production frontier. J Bus Econ Stat 14:460–468

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

Fare, R. (1988). Fundamental of production theory (Heidelberg, Springer-Verlag)

Fare, R., Grosskopf, S., & Logan, J. (1983). The Relative Efficiency of Illinois Electric Utilities, Resources and Energy, Vol. 5, No. 4, pp. 349-367

Fare, R., Grosskopf, S., & Lovell, C. A. K. (1994). Production Frontiers, Cambridge University Press, Cambridge

Fare, R., Grosskopf, S., & Roos, P. (1998). Malmquist Productivity In- dexes: A Survey of Theory and Practice, In R. Fare, S. Grosskopf and R. R. Russell (eds.) Index Numbers: Essays in Honour of Sten Malmquist. Kluwer Academic Publishers, Boston/London/Dordrecht.

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

Farsi, M., Filippini, M., & Kuenzle, M. (2005). Unobserved heterogeneity in stochastic cost frontier models: an application to Swiss nursing homes. Applied economics, 37(18), 2127-2141.

Fathi, B., Ahmadi, M., & dan Torabi, T. (2010). “The Measuring of the Efficiency of Islamic Azad University in Iran Using Data Envelopment Analysis (DEA) (Case study: Azad University Branches in Region Five).” Proceeding International Conference on Business and Economic Research (ICBER 2010).

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.

Ferrier G., & Valdmanis,V. (2004). Do Mergers Improve Hospital Productivity?” JORS, 55(10) 1071-1080.

Field, J., & Peck E. (2003). Mergers and Acquisitions in the Private Sector: What Are the Lessons for Health and Social Services? Social Policy and Administration 37(7):742-755

Fielden, J., & Markham, L. (1997). Learning Lessons from Mergers in Higher Education, CHEMS Paper No 17: 8pp. Hampshire: CHEMS Consulting.

Filippini, M., & Greene, W. H. (2016). Persistent and transient productive inefficiency: a maximum simulated likelihood approach. J Prod Anal (2016) 45:187–196 DOI 10.1007/s11123-015-0446-y

Filippini, M., Geissmann, T., & Greene, W. H. (2018). Persistent and transient cost efficiency—an application to the Swiss hydropower sector. Journal of Productivity Analysis, 49(1), 65-77.

Filippini, M., Greene, W. H., & Masiero, G. (2016, February). Persistent and transient productive inefficiency in a regulated industry: electricity distribution in New Zealand. In Energy: Expectations and Uncertainty, 39th IAEE International Conference, Jun 19-22, 2016. International Association for Energy Economics.

Flegg, A. T., & Allen, D. O. (2007a). Does expansion cause congestion? The case of the older British universities, 1994–2004. Education Economics, 15(1), 75-102.

Flegg, A. T., & Allen, D. O. (2007b). Using Cooper’s approach to explore the extent of congestion in the new British universities. Economic Issues, 12(Part 2), 47-81.

Flegg, A. T., & Allen, D. O. (2009). Congestion in the new British universities: A further analysis (< Special Issue> Operations Research for Performance Evaluation). Journal of the Operations Research Society of Japan, 52(2), 186-203.

Flegg, A. T., Allen, D. O., Field, K., & Thurlow, T. W. (2003). Measuring the efficiency and productivity of British universities: an application of DEA and the Malmquist approach. University of the West of England, Department of Economics, series Discussion Papers, 304.

Flegg, A. T., Allen, D. O., Field, K., & Thurlow, T. W. (2004). Measuring the efficiency of British universities: a multi-period data envelopment analysis. Education Economics, 12(3), 231-249.

Florens, J. P., & Simar, L. (2005). Parametric Approximations of Nonparametric Frontier, Journal of Econometrics, vol 124, 1, 91–116.

Florens, J. P., Simar, L., & Van Keilegom, I. (2014). Frontier estimation in nonparametric location-scale models. Journal of Econometrics, 178, 456-470.

Forsund, F. R., & Sarafoglou, N., (2002). "On the Origins of Data Envelopment Analysis." JPA 17(1-2): 23-40.

Førsund, F. R., Lovell, C. A. K. & Schmidt, P. (1980), 'A survey of frontier production functions and of their relationship to efficiency measurement', Journal of Econometrics, 13, 5–25.

Fozzard, A. (2001). 'The Basic Budgeting Problem: Approaches to Resource Allocation in the Public Sector and their Implications for Pro-Poor Budgeting', Centre for Aid and Public Expenditure, Overseas Development Institute, London

Fried, H. O., Lovell, C. A. K., & Schmidt, S. S. (Eds.). (2008). The measurement of productive efficiency and productivity growth. Oxford University Press, New York. DOI:10.1093/acprof:oso/9780195183528.001.0001

Fried, H. O., Schmidt, S. S., & Lovell, C. A. K. (Eds.). (1993). The measurement of productive efficiency: techniques and applications. Oxford university press.

Fu, T. T., & Huang, M. Y. (2006). Measuring the Performance and Relative Efficiency of Departments in the College of Commerce using DEA. In Forth workshop on knowledge economy and electronic commerce (pp. 42-62).

Fu, T. T., & Lu, Y. H. (2006). Heterogeneous Output Quality and Efficiency Evaluation of Polytechnic Higher Education Institutions via DEA. In 2006 Conference on Productivity and Efficiency (pp. 03-10).

Funding Higher Education [FDH], (2010). Staying on top: The challenge of sustaining world-class higher education in the UK, Russell Group Papers – Issue 2, 2010.

Gabrielsen, A. (1975). On estimating efficient production functions. In Working Paper no. A-85. Chr. Michelsen Institute, Department of Humanities and Social Sciences Bergen, Norway.

García-Aracil, A., & Palomares-Montero, D. (2010). Examining benchmark indicator systems for the evaluation of higher education institutions. Higher Education, 60(2), 217-234.

García-Aracil, A., López-Iñesta, E., & Palomares-Montero, D. (2009). An analysis of the Spanish public universities missions in efficiency terms. XVIII Jornadas de la Asociación Económica de la Educación.

Garniss, D. (2006). Measuring government education output quality in England: an overview of the issues and approaches developed by the Department for Education and Skills, OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health": London, October 3-5, 2006.

Gattoufi, S., Oral, M., & Reisman, A. (2004). Data envelopment analysis literature: A bibliography update (1951–2001). Journal of Socio-Economic Planning Sciences, 38(2-3), 159-229.

Gerfin, M., & Lechner, M. (2002). A microeconometric evaluation of the active labour market policy in Switzerland. The Economic Journal, 112(482), 854-893.

Geva-May, I. (2001), 'Higher education and attainment of policy goals: interpretations for efficiency indicators in Israel', Higher Education, 42(3), 265–305.

Gillard, D. (2011). Education in England: a brief history: www.educationengland.org.uk/history

Giménez, V. M., & Martínez, J. L. (2006). Cost efficiency in the university: A departmental evaluation model. Economics of Education Review, 25(5), 543-553.

Girod, O. A., & Triantis, K. P. (1999). The evaluation of productive efficiency using a fuzzy mathematical programming approach: the case of the newspaper preprint insertion process. IEEE Transactions on Engineering Management, 46(4), 429-443.

Gjirja, M. (2003). Assessing the efficiency effects of bank mergers in Sweden: A panel-based Stochastic Frontier Analysis. Goteborg University, Department of Economics.

Glaser, E. G., & Strauss, A. L. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research, Weidenfeld and Nicplson, London, England.

Glass, J. C., McCallion, G., McKillop, D. G., Rasaratnam, S., & Stringer, K. S. (2006). Implications of variant efficiency measures for policy evaluations in UK higher education. Socio-economic planning sciences, 40(2), 119-142.

Glass, J. C., McCallion, G., McKillop, D. G., Rasaratnam, S., & Stringer, K. S. (2009). Best-practice benchmarking in UK higher education: new nonparametric approaches using financial ratios and profit efficiency methodologies. Applied Economics, 41 (2). pp. 249-267

Glass, J. C., McKillop D. G. & Hyndman N. S. (1995a). 'The achievement of scale efficiency in UK universities: a multiple-input multiple-output analysis ' Education Economics 3(3): 249-263.

Glass, J. C., McKillop D. G. & Hyndman N. S. (1995b). 'Efficiency in the provision of university teaching and research: an empirical analysis of UK universities.' Journal of Applied Econometrics 10(1): 61-72.

Glass, J. C., McKillop, D. G., & O'Rourke, G. (1997). Productivity growth in UK accountancy departments 1989–96. Financial Accountability & Management, 13(4), 313-330.

Glass, J. C., McKillop, D. G., & O'Rourke, G. (1998). A Cost Indirect Evaluation of Productivity Change in UK Universities, Journal of Productivity Analysis Vol. 10, No. 2, pp. 153-175.

Glass, J. C., McKillop, D. G., & O'Rourke, G. (2002). Evaluating the productive performance of UK universities as cost-constrained revenue maximizers: an empirical analysis. Applied Economics, 34(9), 1097-1108.

Glazerman, S., Levy, D. M., & Myers D. (2003). Non-experimental versus experimental estimates of earnings impacts. The Annals of the American Academy of Political and Social Science 589(September): 63–93.

Glynn, A. N., & Quinn, K. M. (2010). An Introduction to the Augmented Inverse Propensity Weighted Estimator. Political Analysis, 18(1), 36-56. doi: 10.1093/pan/mpp036.

Görason, B., Maharajh, R., & Schmoch, U. (2009). New activities of universities in transfer and extension: multiple requirements and manifold solutions. Science and Public Policy, 36(2), 157-164

Goto, M., & Tsutsui M. (1998). "Comparison of Productive and Cost Efficiencies Among Japanese and U.S. Electric Utilities." Omega 26: 177-194.

Gozalo, P., & Linton, O. (2000). Local nonlinear least squares: Using parametric information in nonparametric regression. Journal of econometrics, 99(1), 63-106.

Gralka, S. (2016). "Persistent inefficiency in the higher education sector: Evidence from Germany", CEPIE Working Paper 06/16.

Great Britain. Department for Business, Innovation and Skills (BIS). (2009). Higher ambitions: the future of universities in a knowledge economy.

Green Paper Fostering and Measuring ´Third Mission´ in Higher Education Institutions. Retrieved from: http://www.e3mproject.eu/docs/Green%20paper-p.pdf accessed 20th June 2014.

Green, C., & Johnes, G. (2009). “Economies of Scale and Mergers in Higher Education.” In The Routledge International Handbook of Higher Education, edited by M. Tight, K.-H. Mok, J. Huisman, and C. Morphew, 369–380. New York: Routledge

Greene, W. (1980a). “Maximum Likelihood Estimation of Econometric Frontier Functions,” Journal of Econometrics, 13(1), pp. 27-56

Greene, W. (1980b). On the estimation of a flexible frontier production model. Journal of Econometrics, 13(1), 101-115.

Greene, W. (1990). “A Gamma Distributed Stochastic Frontier Model,” Journal of Econometrics, 46, pp. 141-163.

Greene, W. (2002). Alternative panel data estimators for stochastic frontier models. Unpublished manuscript (Septemebr 1, 2002), Department of Economics, New York University.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. The Econometrics Journal, 7(1), 98-119.

Greene, W. (2004b). "Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems." Health economics; 959-980.

Greene, W. (2005). 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model.' Journal of Econometrics 126: 269-303.

Greene, W. (2005a). Fixed and random effects in stochastic frontier models. J Prod Anal 23:7-32.

Greene, W. (2005b). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. J Econ 126:269–303.

Greene, W. (2007). The Econometric Approach to Efficiency Analysis. Oxford University Press, New York.

Greene, W. (2008). The econometric approach to efficiency analysis. The measurement of productive efficiency and productivity growth, 1, 92-250.

Greene, W. (2015). LIMDEP. Version 10. Econometric Modeling Guide by. William H. Greene. Econometric Software, Inc. Page 2. © 1986 - 2012 Econometric Software, Inc

Greenhalgh, T., & Taylor, R. (1997). Papers that go beyond numbers (qualitative research). BMJ: British Medical Journal, 315(7110), 740.

Griffin, R., Montgomery, J., & Rister, M. (1987). Selecting Functional Form in Production Function Analysis. Western Journal of Agricultural Economics, 12(2), 216-227. Retrieved from http://www.jstor.org/stable/40987872

Griffiths, S., (2010). 'Be afraid, parents'. The Sunday TImes. London Retrieved from http://www.timesonline.co.uk/tol/life_and_style/education/article6990896.ece assessed 10th June, 2014

Grilli, L., & Rampichini C. (2011). "Propensity scores for the estimation of average treatment effects in observational studies", Training Sessions on Causal Inference, Bristol, 28-29. Available at: http://www.bristol.ac.uk/cmm/software/support/workshops/materials/causalinference/prop-scores.pdf

Grosskopf, S. (1996). "Statistical Inference and Nonparametric Efficiency: A Selective Survey." JPA 7(2-3): 161-176

Guo, S., & Fraser, M. W. (2015). Propensity Score Analysis: Statistical Methods and Applications. Advanced Quantitative Techniques in the Social Sciences, 2nd ed. Thousand Oaks, CA: SAGE.

Hahn, J., & Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. Econometrica, 72(4), 1295-1319.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis, 20(1), 25-46.

Hainmueller, J., & Xu, Y. (2013). Ebalance: A Stata package for entropy balancing. Journal of Statistical Software, Volume 54, Issue 7.

Halkos, G., Tzeremes, N. & Kourtzidis, S. A. (2010). A DEA approach for measuring university departments' efficiency, Munich Personal RePEc Archive, No. 24029

Halkos, G., Tzeremes, N. G., & Kourtzidis, S. A. (2012). Measuring public owned university departments' efficiency: a bootstrapped DEA approach. Journal of Economics and Econometrics, 55(2), 1-24.

Hall, P., & Simar, L. (2002). Estimating a changepoint, boundary, or frontier in the presence of observation error. Journal of the American statistical Association, 97(458), 523-534.

Han, C., Orea, L., & Schmidt, P. (2005). Estimation of a panel data model with parametric temporal variation in individual effects. Journal of Econometrics, 126(2), 241-267.

Hanke, M., & Leopoldseder, T. (1998). Comparing the efficiency of Austrian universities: A data envelopment analysis application. Tertiary Education & Management, 4(3), 191-197.

Harman, G. (2000). Institutional Mergers in Australian Higher Educaiton since 1960. Higher Education Quarterly, 54(4), 343-366.

Harman, G., & Harman, K. (2003). Institutional mergers in higher education: lessons from international experience. Tertiary Education and Management, 9(1), 29-44.

Harman, K. (2002). "Merging Divergent Campus Cultures into Coherent Educational Communities: Challenges for Higher Education Leaders." Higher Education 44 (1): 91–114.

Harman, K., & Harman, G. (2008). Strategic Mergers of Strong Institutions to Enhance Competitive Advantage. Higher Education Policy, 99 (1), 99-121.

Harman, K., & Meek, V. L. (2002). Introduction to special issue: Merger revisited: international perspectives on mergers in higher education. Higher Education 44: 1–4. DOI: 10.1023/A:1015598525413

Harris, J., Ozgen, H., & Ozcan, Y. (2000). Do mergers enhance the performance of hospital efficiency?. Journal of the Operational Research Society, 51(7), 801-811.

Hattori, T. (2002a). "Regulatory Reform of the Electricity Industry in Japan: An Overview of the Process." In Berg, S., M.G. Pollitt and M. Tsuji, eds., Private Initiative in Infrastructure. Cheltenham: Edward Elgar.

Hattori, T., Jamasb, T., & Pollitt, M. (2005). Electricity Distribution in the UK and Japan: A Comparative Efficiency Analysis 1985-1998. The Energy Journal, 26(2), 23-47. Retrieved from http://www.jstor.org/stable/41323060

Haynes, M., & Thompson, S. (1999). The productivity effects of bank mergers: evidence from the UK Building societies. Journal of Banking and Finance 23, 825–846.

Hayton, A., & Paczuska, A. (Eds). (2002) Access, participation and higher education: Policy and practice (London: Routledge).

Hazelkorn, E. (2009). The problem with university rankings. Science and InnovationPolicy.Retrievedfrom:http://www.scidev.net/global/education/opinion/the-problem-with-university-rankings.html

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. Annals of Economic and Social Measurement, Volume 5, number 4, pp 475-492.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. Econometrica, Vol. 47, No. 1, pp. 153-161.

Heckman, J. J. (1997). Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. The Journal of Human Resources, 32(3), 441–462. http://doi.org/10.2307/146178

Heckman, J. J. (1998): “Matching as an Econometric Evaluation Estimator,” Review of Economic Studies, 65, 261–294.

Heckman, J. J., & Honore, B. (1990). The empirical content of the Roy Model. Econometrica 58, 1121–1149.

Heckman, J. J., & Smith, J. (1999): “The Pre-Program Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies,” Working Paper No. 6983, National Bureau of Economic Research.

Heckman, J. J., & Vytlacil, E. J. (1999). Economic Sciences Local instrumental variables and latent variable models for identifying and bounding treatment effects. Proc. Natl. Acad. Sci. USA Vol. 96, pp. 4730–4734.

Heckman, J. J., Ichimura H., & Todd P. E. (1997). “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme,” Review of Economic Studies, 64, 605–654.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1998). “Matching as an Econometric Evaluation Estimator”, Review of Economic Studies, 65, 261-294.

Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. E. (1998). Characterizing selection bias using experimental data. Econometrica, 66, 1017–1098.

HEFCE (2010). Literature review for the higher education collaborations, alliances and mergers project. Report to HEFCE by Oakleigh Consulting Limited (completed in May 2010).

HEFCE (2012). 'Collaborations, alliances and mergers in higher education: Consultation on lessons learned and guidance for institutions.' HEFCE 2012/06 Higher Education Funding Council for England, London.

HEFCE (2013). How should we measure higher education? A fundamental review of the Performance Indicators. Part one: Synthesis Report. Emma Pollard, Matthew Williams, Joy Williams, Christine Bertram, and Jonathan Buzzeo, IES, Emma Drever, Julia Griggs and Steve Coutinho, NatCen. Retrieved from: http://www.hefce.ac.uk/pubs/rereports/year/2013/ukpireview/#d.en.85232

HEFCE (2016). Higher education in England 2016 Key Facts. http://www.hefce.ac.uk/pubs/year/2016/201620/ Higher Education Funding Council for England, Bristol.

HEFCE (2017). Higher education providers. Retrieved July 18, from: http://www.hefce.ac.uk/analysis/HEinEngland/providers/colleges/

Henderson, D. J., & Simar, L. (2005). A fully nonparametric stochastic frontier model for panel data. Unpublished manuscript.

Herbert, I., & Rothwell, A. (2015). Collaboration and shared services in UK higher education: potential and possibilities. London: Efficiency Exchange, 50pp. https://dspace.lboro.ac.uk/2134/24450

HESA (2012). 'Performance Indicators in Higher Education in the UK', Retrieved 3 May 2014,from:http://www.hesa.ac.uk/index.php?option=com_content&task=view&id=2072&Itemid=141

HESA (2013). 'UK Performance Indicators in Higher Education'. Retrieved October 10, 2014 from: https://www.hesa.ac.uk/pis

HESA (2014). 'HE Business and Community Interaction Survey'. Resources Definitions 2008/09. Retrieved October 20, 2014 from: https://www.hesa.ac.uk/index.php?option=com_pubs&task=show_pub_detail&pubid=1718&Itemid=286

Heshmati, A., Kumbhakar, S. C., & Kim, J. (2016). *Persistent and Transient Efficiency of International Airlines* (No. 444). Royal Institute of Technology, CESIS-Centre of Excellence for Science and Innovation Studies.

Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Services and Outcomes Research Methodology 2, 259-278.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71: 1161–1189.

Ho, D.E., Imai, K., King, G., & Stuart, E.A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis, 15(3), 199.

Hogan, J. (2011). Is higher education spending more on administration and, if so, why?. Perspectives, 15(1), 7-13.

Holland, D., Liadze, I., Rienzo, C., & Wilkinson, D. (2013). The relationship between graduates and economic growth across countries. Department for Business Innovation & Skills: BIS Research Paper No. 110.

Holland, P.W. (1986). "Statistics and Causal Inference." Journal of the American Statistical Association 81(396): 945-960.

Hopkins, D. S. P. (1990). "The Higher Education Production Function: Theoretical Foundations and Empirical Findings," in Stephen A. Hoenack and Eileen L. Collins, eds., The Economics of American Universities. Albany, N.Y.: State University of New York Press.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47: 663–685.

Howard, J., & Sharma, A. (2006). Universities' Third Mission: Communities Engagement. B-HERT POSITION PAPER NO. 11, June 2006

Hsiao, H. C., Shen, C. H., Chou, C. M., Chen, S. C., & Ly, C. (2009). The relationship between programme management efficiency and evaluation results in an institute of technology: a Taiwanese case. World Transactions on Engineering and Technology Education, 7, 169-175.

Hu, Y., & Liang, W. (2008). Malmquist index analysis of the dynamic changes in scientific research productivity of some Chinese universities before and after merger. Frontiers of Education in China, 3(3), 429-447.

Huang, C. J., & Fu, T. T. (1999). An average derivative estimation of stochastic frontiers. Journal of Productivity Analysis, 12(1), 45-53.

Huang, T., & Wang, M. (2002), Comparison of Economic Efficiency Estimation Methods: Parametric and Non–parametric Techniques. The Manchester School, 70: 682–709. doi:10.1111/1467-9957.00320

Hughes, A. (1989). The impact of merger: a survey of empirical evidence for the UK. In: Fairburn, J., Kay, J. (Eds.), Mergers and Merger Policy. Oxford University Press, Oxford.

Hughes, A. (1993), Mergers and Economic Performance in the UK: A Survey ofthe Empirical Evidence 1950-1990, in Bishop, M. and J. Kay (eds.), European Mergers and Merger Policy, Oxford: Oxford University Press, pp. 9-95

Huiqian, Z. (2010). Evaluation and analysis of regional higher education performance based on DEA. In Proceedings of the 3rd International Conference on Education Management Science and Engineering, China [online] http://www. seiofbluemountain. com/(accessed 20 August 2011).

Hullsiek, K. H., & Louis, T. A. (2002). Propensity score modeling strategies for the causal analysis of observational data. Biostatistics, 3, 179–19.

Iacus, S., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. Political Analysis, 20(1), 1-24.

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. Journal of the Royal Statistical Society, Series B 76: 243–263.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. Biometrika 87: 706–710.

Independent Committee on Student Fees and Funding [Browne report]. (2010). Securing a Sustainable Future for Higher Education. London: ICSFF.

IPART. (1999). "Efficiency and Benchmarking Study of the NSW Distribution Businesses." Research Paper No. 13. Sydney. Prepared by London Economics for the Independent Pricing and Regulatory Tribunal (IPART), New South Wales.

Izadi, H., Johnes, G. Oskrochi R., & Crouchley R. (2002). 'Stochastic frontier estimation of a CES cost function: the case of higher education in Britain.' Economics of Education Review 21(1): 63-71.

Jackson S. (2013). Making the best better: UK Research and Innovation. More efficient and more effective for the global economy London: BIS, pp.9–11.

Jamasb, T., & Pollitt, M. (2003). "International Benchmarking and Regulation: An Application to European Electricity Distribution Utilities." Energy Policy 31 (December): 1609-1622.

Jenkins, L., & Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. European Journal of Operational Research, 147(1), 51-61.

Jeong, S. O., Park B. U., & Simar, L. (2010). Nonparametric conditional efficiency measures: asymptotic properties. Annals of Operations Research, 173, 105–122.

Jesenšek, M. (2006). Why the rankings are important?. Katedra (Maribor). 1(4), 6–7. (in Slovene).

Johnes, G. (1995). Scale and technical efficiency in the production of economic research. Applied Economics Letters, 2(1), 7-11.

Johnes, G. (1996). 'Multiproduct cost functions and the funding of tuition in UK universities.' Applied Economics Letters 3: 557-561.

Johnes, G. (1997). Costs and industrial structure in contemporary British higher education. The Economic Journal, 107 (442), 727-737.

Johnes, G. (1998). The costs of multi-product organizations and the heuristic evaluation of industrial structure. Socio-Economic Planning Sciences, 32(3), 199-209.

Johnes, G. (1999). The Management of Universities; Scottish Economic Society/Royal Bank of Scotland Annual Lecture, 1999. Scottish Journal of Political Economy, 46(5), 505-522.

Johnes, G. (2004a). A fourth desideratum: the CES cost function and the sustainable configuration of multiproduct firms. Bulletin of Economic Research, 56 (4), 329-332.

Johnes, G., & Johnes, J. (1992). ‘Apples and oranges: the aggregation problem in publications analysis’ Scientometrics, 25(2) pp353-365

Johnes, G., & Johnes, J. (1993). Measuring the Research Performance of UK Economics Departments: An Application of Data Envelopment Analysis, Oxford Economic Papers, New Series, 45, 332-347

Johnes, G., & Johnes, J. (2004). International Handbook on the Economics of Education. Edward Elgar Publishing Ltd Cheltenham, UK Northampton, MA, USA

Johnes, G., & Johnes, J. (2009). Higher education institutions' costs and efficiency: taking the decomposition a further step. Economics of Education Review, Vol. 28, No. 1, 02. p. 107-113.

Johnes, G., & Johnes, J. (2013). Efficiency in the higher education sector: technical exploration. Published by Department for Business, Innovation and Skills (BIS), London. Retrieved from: https://www.gov.uk/government/publications/efficiency-in-the-higher-education-sector-technical-exploration

Johnes, G., & Johnes, J. (2016). “Costs, Efficiency and Economies of Scale and Scope in the English Higher Education Sector.” Oxford Review of Economic Policy 32 (4): 596–614

Johnes, G., & Johnes, J. (Eds.), (2007). International handbook on the economics of education. Edward Elgar Publishing.

Johnes, G., & Salas-Velasco, M. (2007). The determinants of costs and efficiencies where producers are heterogeneous: the case of Spanish universities. Economics Bulletin, 4(15), 1-9.

Johnes, G., & Schwarzenberger, A. (2011). Differences in cost structure and the evaluation of efficiency: the case of German universities, Education Economics, 2011, 9, 487-499

Johnes, G., Johnes J., & Thanassoulis, E. (2008). 'An analysis of costs in institutions of higher education in England.' Studies in Higher Education 33(5): 527-549.

Johnes, G., Johnes, J., Thanassoulis, E., Lenton, P., & Emrouznejad, A. (2005). An Exploratory Analysis of the Cost Structure of Higher Education in England. London, Department for Education and Skills.

Johnes, J. (1996). ‘Performance assessment in higher education in Britain’ European Journal of Operational Research, 89 pp18-33

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 Department of Economics, Lancaster University

Johnes, J. (2004). Efficiency measurement. International handbook on the economics of education, 613.

Johnes, J. (2006). 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. (2006a). Measuring Efficiency: A Comparison of Multilevel Modelling and Data Envelopment Analysis in the Context of Higher Education. Bulletin of Economic Research, 58: 75–104. doi:10.1111/j.0307-3378.2006.00238.x

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

Johnes, J. (2008). Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/5. The Manchester School, 76(6), 653-674.

Johnes, J. (2010). Efficiency and Input-and Output-Substitutability in English Higher Education 1996/97 to 2007/08: A Parametric Distance Function Approach. Browser Download This Paper.

Johnes, J. (2012). Measuring efficiency and input substitutability: Evidence from the English higher education sector over a period of expansion. Paper to be presented at the Beijing Forum 2nd to 4th November 2012

Johnes, J. (2014). Efficiency and mergers in English higher education 1996/97 to 2008/09: parametric and non-parametric estimation of the multi-input multi-output distance function. The Manchester School, ISSN 1463-6786

Johnes, J. (2014). Efficiency and input substitutability in English higher education 1996/97 to 2008/09. In INFORMS conference, San Francisco. Available at: h ttp://eprints. hud. ac. uk/23955/1/INFORMS% 20San% 20Franciso% 20November (Vol. 202014).

Johnes, J. (2015). "Operational Research in education," European Journal of Operational Research, Elsevier, vol. 243(3), pages 683-696.

Johnes, J. (2016, November 15). Efficient universities: Are mergers really the answer? Retrieved from http://www.efficiencyexchange.ac.uk/10202/efficient-universities-mergers-really-answer/

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

Johnes, J., & Schwarzenberger, A. (2011). Differences in cost structure and the evaluation of efficiency: the case of German universities, Education Economics, 19:5, 487-499, DOI: 10.1080/09645291003726442

Johnes, J., & Taylor, J. (1987) 'Degree quality: an investigation into differences between universities' Higher Education, 16 pp581-602

Johnes, J., & Taylor, J. (1989a). 'The first destination of new graduates: comparisons between universities' Applied Economics, 21(3) pp357-374

Johnes, J., & Taylor, J. (1989b). 'Undergraduate non-completion rates: differences between UK universities' Higher Education, 18 pp209-225

Johnes, J., & Taylor, J. (1990). Undergraduate non-completion rates: a reply. Higher Education. 19, 3, 385-390

Johnes, J., & Taylor, J. (1992). 'The 1989 research selectivity exercise: a statistical analysis of differences in research rating between universities at the cost centre level' Higher Education Quarterly, 46(1) pp67-87

Johnson, A. L., & Kuosmanen T. (2015). An Introduction to CNLS and StoNED Methods for Efficiency Analysis: Economic Insights and Computational Aspects. In: Ray S., Kumbhakar S., Dua P. (eds) Benchmarking for Performance Evaluation. Springer, New Delhi

Johnstone, D. B., & Marcucci, P. (2007). Worldwide trends in higher education finance: Cost-sharing, student loans, and the support of academic research. Commissioned paper V. Lynn Meek and Dianne Davies, 81.

Jondrow, J., Lovell K., Materov, I., & Schmidt, P. (1982). "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," Journal of Econometrics, 19, pp. 233-238.

Jongbloed, B., Enders, J., & Salerno, C. (2008). Higher education and its communities: interconnections, interdependencies and a research agenda, Higher Education, 56, 303-324.

Kantabutra, S., & Tang, J. C. (2010). Efficiency analysis of public universities in Thailand. Tertiary Education and Management, 16(1), 15-33.

Kao, C., & Hung, H. T. (2008). Efficiency analysis of university departments: An empirical study. Omega, 36(4), 653-664.

Kao, C., & Lin, Y. C. (1999). Comparing university libraries of different university size. Libri, 49(3), 150-158.

Kao, C., & Lin, Y. C. (2004). Evaluation of the university libraries in Taiwan: total measure versus ratio measure. Journal of the Operational Research Society, 55(12), 1256-1265.

Kao, C., & Liu, S. T. (1999). Competitiveness of manufacturing firms: an application of fuzzy weighted average. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 29(6), 661-667.

Kao, C., & Liu, S. T. (2000). Data envelopment analysis with missing data: an application to university libraries in Taiwan. Journal of the Operational Research Society, 897-905.

Kao, C., Liu, S. T., & Pao, H. L. (2012). Assessing improvement in management research in Taiwan. Scientometrics, 92(1), 75-87.

Katharaki, M., & Katharakis, G. (2010). A comparative assessment of Greek universities' efficiency using quantitative analysis. International journal of educational research, 49(4), 115-128.

Kelchtermans, S., & Verboven, F. (2010). "Program Duplication in Higher Education is not Necessarily Bad." Journal of Public Economics 94 (5–6): 397–409.

Kempkes, G., & Pohl, C. (2008). Do Institutions Matter for University Cost Efficiency? Evidence from Germany. CESifo Economic Studies, 54 (2): 177-203. doi: 10.1093/cesifo/ifn009

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

Key, V. O. (1940). 'The lack of a budget theory' American Political Science Review 34 (6), 1137- 1144.

Khan, M. S., & Mahapatra, S. S. (2008). Service quality evaluation of technical institutions using data envelopment analysis. International Journal of Productivity and Quality Management, 3(1), 127-143.

Kim, Y., & Schmidt, P. (2000). "A Review and Empirical Comparison of Bayesian and Classical Approaches to Inference on Efficiency Levels in Stochastic Frontier Models with Panel Data," Journal of Productivity Analysis, 14, pp. 91-98.

Kjekshus, L. E., Bernstrøm, V. H., Dahl, E., & Lorentzen, T. (2014). The effect of hospital mergers on long-term sickness absence among hospital employees: a fixed effects multivariate regression analysis using panel data. BMC Health Services Research, 14, 50. http://doi.org/10.1186/1472-6963-14-50.

Kjekshus, L., & Hagen, T. (2007). Do hospital mergers increase hospital efficiency? Evidence from a National Health Service country. J Health Serv Res Policy, 12(4):230-5

Klein, J. (2002). More to third mission than counting pounds, Times Higher Education, Retrievedfrom:http://www.timeshighereducation.co.uk/story.asp?storyCode=170988 &sectioncode=26

Klumpp, M. (2012, September). European universities efficiency benchmarking. In U. o. University of Duisburg-Essen (Ed.), EAIR 34th Annual Forum.

Kneip, A., & Simar, L. (1996). A general framework for frontier estimation with panel data. Journal of Productivity Analysis, 7(2-3), 187-212.

Kneip, A., Simar, L., & Van Keilegom, I. (2012). Boundary estimation in the presence of measurement error with unknown variance. J. Econometrics.

Kneip, A., Simar, L., & Wilson, P. W. (2015). When bias kills the variance: central limit theorems for DEA and FDH efficiency scores. Econometric Theory 31, 394–422.

Kneip, A., Simar, L., & Wilson, P. W. (2016). Testing hypotheses in nonparametric models of production. Journal of Business and Economic Statistics 34, 435–56.

Kokkelenberg, E. C., Dillon, M., & Christy, S. M., (2008). The effects of class size on student grades at a public university. Econ Educ Rev 27, 221–233

Köksal, G., & Nalçaci, B. (2006). The relative efficiency of departments at a Turkish engineering college: A data envelopment analysis. Higher education, 51(2), 173-189.

Kong, W. H., & Fu, T. T. (2012). Assessing the performance of business colleges in Taiwan using data envelopment analysis and student based value-added performance indicators. Omega, 40(5), 541-549.

Kong, W. H., & Tsu-Tan, F. U. Investigating Output Weights of Peer Reviewers in the University Performance Assessment: A DEA Cross Efficiency Study.

Koop, T., Carslaw, K. S., & Peter, T. (1997). Thermodynamic stability and phase transitions of PSC particles. Geophysical research letters, 24(17), 2199-2202.

Koopmans, T. C. (Eds). (1951). Activity analysis of production and allocation (No. 13). New York: Wiley.

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

Koryakina, T., Sarrico, C. S., & Teixeir, P. N. (2015). Universities’ Third Mission Activities, Challenges to Extending Boundaries. The Transformation of University Institutional and Organizational Boundaries, Part of the series Higher Education Research in the 21st Century Series pp 63-82

Kosor, M. M. (2013). Efficiency Measurement in Higher Education: Concepts, Methods and Perspective. Procedia-Social and Behavioral Sciences, 106, 1031-1038. https://doi.org/10.1016/j.sbspro.2013.12.117

Kounetas, K., Anastasiou, A., Mitropoulos, P., & Mitropoulos, I. (2011). Departmental efficiency differences within a Greek university: An application of a DEA and Tobit analysis. International Transactions in Operational Research, 18(5), 545-559.

Kumbhakar, S. C. (1987). The specification of technical and allocative inefficiency in stochastic production and profit frontiers. Journal of econometrics, 34(3), 335-348.

Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. Journal of econometrics, 46(1-2), 201-211.

Kumbhakar, S. C. (1991). Estimation of technical inefficiency in panel data models with firm- and time-specific effects. Econ Lett 36:43–48.

Kumbhakar, S. C., & Heshmati A. (1995). "Efficiency Measurement in Swedish Dairy Farms: An Application of Rotating Panel Data, 1976–88." American Journal of Agricultural Economics 77 (3): 660–674.

Kumbhakar, S. C., & Hjalmarsson, L. (1993). Technical efficiency and technical progress in Swedish dairy farms. In: Fried HO, Lovell CAK, Schmidt SS (eds) The measurement of productive efficiency—techniques and applications. Oxford University Press, Oxford, pp 256–270.

Kumbhakar, S. C., & Hjalmarsson, L. (1995). Labour-use efficiency in Swedish social insurance offices. Journal of Applied Econometrics, 10(1), 33-47.

Kumbhakar, S. C., & Hjalmarsson, L. (1998). "Relative performance of public and private ownership under yardstick competition: electricity retail distribution," European Economic Review, Elsevier, vol. 42(1), pages 97-122, January.

Kumbhakar, S. C., & Lovell, C. K. (2003). Stochastic frontier analysis. Cambridge university press.

Kumbhakar, S. C., & Wang, H. J. (2005). Estimation of growth convergence using a stochastic production frontier approach. Economics Letters, 88(3), 300-305.

Kumbhakar, S. C., Lien, G., & Hardaker, J. B. (2014). "Technical Efficiency in Competing Panel Data Models: A Study of Norwegian Grain Farming." Journal of Productivity Analysis 41 (2): 321–337.

Kumbhakar, S. C., Park, B. U., Simar, L., & Tsionas, E. G. (2007). Nonparametric stochastic frontiers: a local maximum likelihood approach. Journal of Econometrics, 137(1), 1-27.

Kuosmanen, T. (2008). Representation theorem for convex nonparametric least squares. The Econometrics Journal, 11(2), 308-325.

Kuosmanen, T. (2012). 'Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model', Energy Economics 34, 2189-2199.

Kuosmanen, T., & Johnson, A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. Operations Research, 58(1), 149-160.

Kuosmanen, T., & Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. Journal of productivity analysis, 38(1), 11-28.

Kuosmanen, T., Johnson, A., & Saastamoinen, A. (2015). Stochastic nonparametric approach to efficiency analysis: A unified framework. In Data envelopment analysis (pp. 191-244). Springer, Boston, MA.

Kyvik, S. (2002). “The Merger of Non-university Colleges in Norway.” Higher Education 44 (1): 53–72.

Kwoka, J., & Pollitt, M. (2010). Do mergers improve efficiency? Evidence from restructuring the US electric power sector. International Journal of Industrial Organization 28, 645–656.

Laband, D. N. & Lentz, B. F. (2003). New Estimates of Economies of Scale and Scope in Higher Education, Southern Economic Journal, Vol. 70, No. 1 (Jul., 2003), pp. 172-183.

Lancaster, T. (2000). The incidental parameter problem since 1948. Journal of econometrics, 95(2), 391-413.

Land, K. C., Lovell, C. A. K., & Thore, S. (1993). Chance-constrained data envelopment analysis. Managerial and decision economics, 14(6), 541-554.

Laredo, P. (2007). Toward a third mission for Universities. ‘’ Main transformations, challenges and emerging patterns in Higher Education Systems’’. UNESCO research seminar for the regional Scientific Committee for Europe and North America, Paris 5-6 March 2007 UNESCO workshop

Lau, L. J. (1986). Functional forms in econometric model building. Handbook of econometrics, 3, 1515-1566.

Lechner, M. (1999). Identification and Estimation of Causal Effects of Multiple Treatments Under the Conditional Independence Assumption, IZA Discussion paper series, No. 91.

Lechner, M. (2001). “Identification and estimation of causal effects of multiple treatments under the conditional independence assumption,” in Econometric Evaluation of Labour Market Policies, ed. by M. Lechner, and F. Pfeiffer, pp. 1–18. Physica-Verlag, Heidelberg.

Lee, L. F., & Tyler, W. G. (1978). The stochastic frontier production function and average efficiency: An empirical analysis. Journal of econometrics, 7(3), 385-389.

Lee, Y. H., & Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical efficiency. The measurement of productive efficiency: Techniques and applications, 237-255.

Leggio, K. B., & Lien, D., (2000). Mergers in the electric utility industry in a deregulatory environment. Journal of Regulatory Economics 17 (2), 69–85.

Leitner, K. H., Prikoszovits, J., Schaffhauser-Linzatti, M., Stowasser, R., & Wagner, K. (2007). The impact of size and specialisation on universities' department performance: A DEA analysis applied to Austrian universities. Higher education, 53(4), 517-538.

Lenton, P. (2008). The cost structure of higher education in further education colleges in England. Economics of Education Review, 27(4), 471-482. doi:10.1016/j.econedurev.2007.05.003

Leuven, E., & Sianesi B. (2003). "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing," Software, http://ideas.repec.org/c/boc/bocode/s432001.html.

Li, L. (2013, March). Research of Universities' Input-Output Efficiency Basing on DEA. In Proceedings of the 2013 International Conference on Information, Business and Education Technology (ICIBET 2013). Atlantis Press.

Liefner, I. (2003). Funding, resource allocation, and performance in higher education systems. Higher education, 46(4), 469-489.

Lin, H. T. (2009). Efficiency measurement and ranking of the tutorial system using IDEA. Expert Systems with Applications, 36(8), 11233-11239.

Lissitsa, A., Coelli, T., & Rao, D. P. (2005). Agricultural economics education in Ukrainian Agricultural Universities: An efficiency analysis using data envelopment analysis. Halle (Saale) Germany.

Liu, L. C., Lee, C., & Tzeng, G. H. (2004). DEA approach for the current and the cross period efficiency for evaluating the vocational education. International Journal of Information Technology & Decision Making, 3(02), 353-374.

Lodge, M. (2015). Regulating higher education: A comparative perspective. Published by the Centre for Analysis of Risk and Regulation at the London School of Economics and Political Science, Discussion paper NO: 77.

Lovell, C. A. K. (1993). 'Production frontiers and productive efficiency', in H.O. Fried, C.A.K. Lovell and S.S. Schmidt (eds), The Measurement of Productive Efficiency, Oxford: Oxford University Press, pp. 3–67.

Lovell, C. A. K. (1994). Linear programming approaches to the measurement and analysis of productive efficiency, Top 2, 175-248

Lukman, R. (2010). Developing indicators for ranking universities from a “third mission” perspective. Knowledge Collaboration & Learning for Sustainable Innovation ERSCP-EMSU conference, Delft, The Netherlands, October 25-29, 2010

Lukman, R., Krajnc, D., & Glavič, P. (2010). University ranking using research, educational and environmental indicators. J. clean. prod. 18(7): 619–628

Lunceford, J. K., & Davidian, M. (2004). Stratification and Weighting Via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study Statistics in medicine, Volume 23, Issue 19, Pages 2937–2960.

Madden, G., Savage, S., & Kemp, S. (1997). Measuring public sector efficiency: A study of economics departments at Australian universities. Education Economics, 5(2), 153-168.

Mainardes, E. W., Alves, H., & Raposo, M. (2010). An exploratory research on the stakeholders of a university. Journal of Management and Strategy, 1(1), 76.

Malmquist, S. (1953). ‘‘Index Numbers and Indifference Surfaces.’’ Trabajos de Estadistica 4, 209–242

Mandelson, P. (2009). 'Higher Education Funding 2010-11'. Letter to HEFCE retrieved from: http://www.bis.gov.uk/wp-content/uploads/publications/Mandelson-Letter-to-HEFCEDec09.pdf accessed 10th June 2014

Mao, Y., Du, Y. & Liu, J. (2009). The effects of university mergers in China since 1990s from the perspective of knowledge production. International journal of educational management, 23 (1), 19-33

Marginson, S. (2007). Global University Rankings: Implications in General and for Australia. J. Higher Educ. Pol. Manage. 29 (2), 131–142

Martin, B. R., & Tang, P. (2007). The benefits from publicly funded research. Science Policy Research Unit, University of Sussex

Martín, 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.

Martin, J., & Samels, J. E. (1994). Merging colleges for mutual growth: A new strategy for academic managers. Baltimore, MD: The John Hopkins University Press.

Martin, R. E. (2005). Cost control, college access, and competition in higher education. Edward Elgar Publishing.

Martins-Filho, C., & Yao, F. (2015). Semiparametric stochastic frontier estimation via profile likelihood. Econometric Reviews, 34(4), 413-451.

Massy, W. (2011). Metrics for Efficiency and Effectiveness in Higher Education: Completing the Completion Agenda. Retrieved May 20, 2015, from: agb.org/sites/agb.org/files/u3/MetricsforEfficiency.pdf

Mastromarco, C. (2008). Stochastic frontier models. Department of Economics and Mathematics-Statistics, University of Salento.

Matthews, D. (2011). Funding council reaffirms case for Welsh mergers. Times higher education 13th June 2014

Mayo, A. (1984). The fast solution of Poisson's and the biharmonic equations on irregular regions. SIAM Journal on Numerical Analysis, 21(2), 285-299.

McBain, L. (2009). College and University Mergers: Recent Trends. A Higher Education Policy Brief. American Association of State Colleges and Universities

McManus, P. A. (2011). Introduction to regression models for panel data analysis, Workshop in methods, USA: Indiana Unviersity. Slides. 2011. http://www.indiana.edu/~wim/docs/10_7_2011_slides.pdf. Accessed 20 June 2017.

McMillan, M. L., & Chan, W. H. (2006). University efficiency: A Comparison and consolidation of results from stochastic and non-stochastic methods. Education economics, 14(1), 1-30.

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

Meeusen, W., & Van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production function with composed error. Int Econ Rev 8: 435–444

Mensah, Y. M. & Werner, R. (2003). Cost efficiency and financial flexibility in institutions of higher education Journal of Accounting and Public Policy, 22, 293–323.

Meschi, M. (1997). Analytical perspectives on mergers and Acquisitions: a survey. Research papers in international business, London South Bank University CIBS. Paper Number 5-97

Millett, J. D. (1976). Mergers in higher education: An analysis of ten case studies. Washington, DC: Academy for Educational Development.

Miranda, R., Gramani, M. C., & Andrade, E. (2012). Technical efficiency of business administration courses: a simultaneous analysis using DEA and SFA. International Transactions in Operational Research, 19(6), 847-862.

Molas-Gallart, J., & Castro-Martínez, E. (2007). Ambiguity and conflict in the development of 'Third Mission'indicators. Research Evaluation, 16(4), 321-330.

Mollas-Gallart, J., Salter, A., Patel, P., Scott, A., & Duran, X. (2002). Measuring third stream activities. Final report to the Russell group of Universities. SPRU, University of Sussex

Monaco, L. (2011). Measuring Italian university efficiency: a non-parametric approach.

Montesinos, P., Carot, J. M., Martinez, J. M., & Mora, F. (2008). Third mission ranking for world class universities: Beyond teaching and research. Higher education in Europe, 33(2-3), 259-271.

Mora, J-G., Detmer, A. & Vieira, M. J. (Eds). (2010). Good Practices in University-Enterprise Partnerships. Retrieved April 28, 2014, from: http://gooduep.eu/documents/GOODUEP-Final%20Report%20UEPS.pdf

Moreno, A. A., & Tadepalli, R. (2002). Assessing academic department efficiency at a public university. Managerial and decision economics, 23(7), 385-397.

Mugisha, S., Berg, S. V., & Muhairwe, W. T. (2007). Using internal incentive contracts to improve water utility performance: the case of Uganda's NWSC. Water Policy, 9(3), 271-284.

Mundel, D. S. (2008). What do we know about the impact of grants to college students. The effectiveness of student aid policies: What the research tells us, 9-38.

Mundlak, Y. (1961). Empirical production function free of management bias. Journal of Farm

Mylonidis, N., & Kelnikola, I. (2005). Merging activity in the Greek banking system: a financial accounting perspective. Southern Eastern Europe Journal of Economics 1, 121–144.

National Research Council [NRC]. (2012). Improving measurement of productivity in higher education. In T.A. Sullivan, C. Mackie, W. F. Massy, & E. Sinha (Eds.), Committee on National Statistics and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=13417

Nerlove, M. (1963). "Returns to Scale in Electricity Supply," in Christ, C., et. al. eds., Measurement in Economics, Stanford University Press, Stanford.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrica: Journal of the Econometric Society, 1-32.

Ng, Y. C., & Li, S. K. (2009). Efficiency and productivity growth in Chinese universities during the post-reform period. China economic review, 20(2), 183-192.

Ni, X., Zhang, D., & Zhang, H. (2010). Variable Selection for Semi-parametric Mixed Models in Longitudinal Models. Biometrics, 66:79–88. [PubMed: 19397585]

Norgård, J. D., & Skodvin, O.-J. (2002). "The Importance of Geography and Culture in Mergers: A Norwegian Institutional Case Study." Higher Edcuation 44 (1): 73–90.

Obadić, A., & Aristovnik, A. (2011). Relative efficiency of higher education in Croatia and Slovenia: an international comparison. Amfiteatru economic, 13(30), 362-376.

Office of Gas and Electricity Markets [OFGEM] (2002). Mergers in the electricity distribution sector Policy statement. Retrieved October 20, 2014 from: https://www.ofgem.gov.uk/ofgem-publications/37817/mergersandaquisitions-48.pdf

Office of Gas and Electricity Markets [OFGEM] (2002). The Review of the First Year of NETA: A Review Document, Vol. 1. Ofgem, London, p. 130.

Office of Gas and Electricity Markets [OFGEM] (2013). Oxera: Recommendations on cost assessment approaches for RIIO-ED1. Retrieved from: https://www.ofgem.gov.uk/publications-and-updates/oxera-recommendations-cost-assessment-approaches-riio-ed1

Olesen, O. B., & Petersen, N. C. (1995). Chance constrained efficiency evaluation. Management Science, 41(3), 442-457.

Olson, K. (2016). Essentials of qualitative interviewing. Routledge.

Ondrich, J., & Ruggiero, J. (2001), 'Efficiency measurement in the stochastic frontier model', European Journal of Operational Research, 129, 434–42.

Orea, L. (2002). Parametric decomposition of a generalized Malmquist productivity index. Journal of Productivity Analysis 18, 5–22.

Orea, L., & Kumbhakar, S. C. (2004). 'Efficiency measurement using a latent class stochastic frontier model.' Empirical Economics 29(1): 169-183.

Organisation for Economic Co-operation and Development [O.E.C.D.]. (2015). Education at a Glance 2015: OECD Indicators, OECD Publishing, Paris.

Park, B. U., & Simar, L. (1994). Efficient semi-parametric estimation in a stochastic frontier model. Journal of the American Statistical Association, 89(427), 929-936.

Park, B. U., Sickles, R. C., & Simar, L. (1998). Stochastic panel frontiers: A semi-parametric approach. Journal of Econometrics, 84(2), 273-301.

Park, B. U., Sickles, R. C., & Simar, L. (2003). Semi-parametric-efficient estimation of AR (1) panel data models. Journal of econometrics, 117(2), 279-309.

Park, B. U., Sickles, R. C., & Simar, L. (2007). Semi-parametric efficient estimation of dynamic panel data models. Journal of econometrics, 136(1), 281-301.

Park, B. U., Simar, L., & Zelenyuk, V. (2008). Local Likelihood Estimation of Truncated Regression and its Partial Derivatives: Theory and Application, Journal of Econometrics, 146, 185–198.

Parker, J. (2008). Comparing Research and Teaching in University Promotion Criteria, Higher Education Quarterly, 62, (3).

Parmeter, C. F., & Kumbhakar S. C. (2014), "Efficiency Analysis: A Primer on Recent Advances", Foundations and Trends® in Econometrics: Vol. 7: No. 3–4, pp 191-385. http://dx.doi.org/10.1561/0800000023

Parmeter, C. F., & Racine, J. S. (2013). Smooth constrained frontier analysis. In Recent advances and future directions in causality, prediction, and specification analysis (pp. 463-488). Springer, New York, NY.

Parteka, A., & Wolszczak-Derlacz, J. (2013). Dynamics of productivity in higher education: cross-european evidence based on bootstrapped Malmquist indices. Journal of Productivity Analysis, 40(1), 67-82.

Pastor, J. T., Ruiz J. L., & Sirvent, I. (2002). A statistical test for nested radial DEA models. Operations Research, 50 (4) (2002), pp. 728–735

Patterson, G. (2000). Findings on economies of scale in higher education: implications for strategies of merger and alliance. Tertiary Education & Management, 6(4), 259-269.

Peristiani, S. (1997). Do mergers improve the X-efficiency and scale efficiency of US banks? Evidence from the 1980s. Journal of Money, Credit, and Banking, 326-337.

Pinheiro, R., Geschwind, L., & Aarrevaara, T. (2016). Mergers in Higher Education, the Experience from Northern Europe. Higher Education Dynamics, Volume 46 2016, DOI: 10.1007/978-3-319-21918-9

Pitt, M. M., & Lee, L. F. (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. Journal of development economics, 9(1), 43-64.

Pollitt, M. (1995). Ownership and Performance in Electric Utilities. Oxford: Oxford University Press.

Porembski, M., Breitenstein, K., & Alpar, P. (2005). Visualizing efficiency and reference relations in data envelopment analysis with an application to the branches of a German bank. Journal of Productivity Analysis, 23(2), 203-221.

Pusztai, G., Hatos, A. & Ceglédi, T. (2012). Third Mission of Higher Education in a Cross-Border Region. Centre for Higher Education Research and Development – Hungary, University of Debrecen

Quandt, R. (1972). A New Approach to Estimating Switching Regressions. Journal of the American Statistical Association, 67(338), 306–310.

Raditloaneng, W. (2013). Impact of "Implementing the Third Mission of Universities in Africa (ITMUA) collaborative research project, 2010-2011". International Journal of Peace and Development Studies, 4(5), 90-99

Rai, A. (2013). Measurement of efficiency in the airline industry using data envelopment analysis. Investment Management and Financial Innovations, 10(1), 38-45.

Ramsden, B. (2012). Institutional Diversity in U.K. Higher Education. Higher Education Policy Institute (Great Britain).

Raty, T. (2002). Efficient facet based efficiency index: A variable returns to scale specification. Journal of Productivity Analysis, 17, 65–82.

Ray, D. J., & Thompson, H. E. (1990). Fifty in five: the prospects for merger in the electric utility industry. Journal of Regulatory Economics 2, 111–128.

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.

Rezitis, A. N. (2008). Efficiency and productivity effects of bank mergers: Evidence from the Greek banking industry. Economic Modelling, 25(2), 236-254.

Rhoades, S. A. (1993). Efficiency effects of horizontal (in-market) bank mergers. Journal of Banking & Finance, 17(2-3), 411-422.

Rhoades, S. A. (1998). The efficiency effects of bank mergers: An overview of case studies of nine mergers. Journal of Banking & Finance, 22(3), 273-291.

Richmond, J. (1974). “Estimating the Efficiency of Production,” International Economic Review, 15, pp. 515-521.

Ripoll-Soler, C., & De-Miguel-Molina, M. (2014). “Are Mergers a Win-win Strategic Model? A Content Analysis of Inter-institutional Collaboration between Higher Education Institutions.” Tertiary Education and Management 20 (1): 44–56.

Robbins, L. (1963). Higher education, report and appendices, CMND 2154 (London: HMSO).

Robins, J. M., & Rotnitzky, A. (1995). Semi-parametric efficiency in multivariate regression models with missing data. Journal of the American Statistical Association 90: 122–129.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association 89: 846–866.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of semi-parametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association 90: 106–121.

Robst, J. (2000). Do state appropriations influence cost efficiency in public higher education? Applied Economic Letters, 7, 715-719

Robst, J. (2001). Cost Efficiency in Public Higher Education Institutions. The Journal of Higher Education, Vol. 72, No. 6, pp. 730-750

Rosenbaum, P. R. (1987). Model-based direct adjustment. The Journal of the American Statistician, 82, 387–394.

Rosenbaum, P. R., & Rubin, D. B. (1983a). The central role of the propensity score in observational studies for causal effects. Biometrika, 70:41–55.

Rosenbaum, P. R., & Rubin, D. B. (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. Journal of the Royal Statistical Society, Series B, 45, 212–218.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using sub-classification on the propensity score. Journal of the American Statistical Association, 79:516–524.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician, 39, 33–38.

Rothaermel, F. T., Agung, S. D., & Jiang, L. (2007). “University entrepreneurship: a taxonomy of the literature” in Industrial and Corporate Change 16:4.

Roy, A. (1951). Some Thoughts on the Distribution of Earnings. Oxford Economic Papers 3, Vol. 3, No. 2, pp. 135-146.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66:688–701.

Rubin, D. B. (1977). Assignment to Treatment Group on the Basis of a Covariate. Journal of Educational Statistics Vol. 2, No. 1, pp. 1-26.

Rubin, D. B. (1980), “Bias Reduction Using Mahalanobis-Metric Matching”, Biometrics, 36, 293-298.

Rubin, D. B. (2006). Matched Sampling for Causal Effects, Cambridge University Press, Cambridge, UK.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. Annals of Applied Statistics 2: 808–840.

Rubin, D. B., & Thomas, N. (1996): “Matching Using Estimated Propensity Scores: Relating Theory to Practice,” Biometrics, 52, 249–264.

Salerno, C. S. (2003). What we know about the efficiency of higher education institutions: the best evidence, Vol 99, The Hague: Ministry of Education, Culture and Science. UT publications. Retrieved from: http://doc.utwente.nl/47097/

Sanchez, M. P., & Elena, S. (2006). Intellectual capital in universities: Improving transparency and internal management. Journal of Intellectual Capital, 7(4), 529-548

Sarafoglou, N., & Haynes, K. E. (1996). University productivity in Sweden: a demonstration and explanatory analysis for economics and business programs. The Annals of Regional Science, 30(3), 285-304.

Sarrico, C. S., & Dyson, R. G. (2000). Using DEA for planning in UK universities—an institutional perspective. Journal of the Operational Research Society, 51(7), 789-800.

Sarrico, C. S., & Dyson, R. G. (2004). Restricting virtual weights in data envelopment analysis. European Journal of Operational Research, 159(1), 17-34.

Sarrico, C. S., Hogan, S. M., Dyson, R. G., & Athanassopoulos A. D. (1997) ‘Data envelopment analysis and university selection’ Journal of the Operational Research Society, 48 pp1163-1177

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2009). Research methods for business students. Harlow, England: Financial Times/Prentice Hall 5th edition.

Sav, G. T. (2012a). Frontier and envelopment evaluations of university graduation efficiencies and productivities: elements for performance based funding. Problems and Perspectives in Management, 4.

Sav, G. T. (2012b). Managing operating efficiencies of publicly owned universities: American university stochastic frontier estimates using panel data. Advances in Management and Applied Economics, 2(1), 1.

Sav, G. T. (2013). Four-stage DEA efficiency evaluations: Financial reforms in public university funding. International Journal of Economics and Finance, 5(1), 24.

SBU. (2016). Evaluation and synthesis of studies using qualitative methods of analysis. Stockholm: Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU).

Scherer, F. M., Ross, D. (1990). Industrial Market Structure and Economic Performance, Third Edition. Houghton Mifflin Company, Boston

Schmidt, P. (1985–6). ‘Frontier production functions’, Econometric Reviews, 4(2), 289–328.

Schmidt, P., & Sickles, R. (1984). “Production Frontiers and Panel Data,” Journal of Business and Economic Statistics, 2, pp. 367-374.

Schoen, A., Laredo, P., Bellon, B., & Sanchez, P. (2007). Observatory of European University. PRIME Position Paper, version March 2007

Scott, C. (2004) ‘The UK: hyper-regulation and regulatory reform’, in C. Hood, O. James, B.G. Peters and C. Scott (eds), Controlling modern government, Cheltenham: Edward Elgar, pp. 124–9.

Scott, P. (1995). The meanings of mass higher education (Buckingham: Open University Press).

Scott, P. (2013) The Coalition Government’s reform of higher education: policy formation and political process. In C. Callender and P. Scott (eds.) Reforming English Higher Education: Browne and the White Paper. London: Institute of Education.

Scott, P. (2014). The reform of English higher education: universities in global, national and regional contexts, Cambridge Journal of Regions, Economy and Society, Volume 7, Issue 2, 1 July 2014, Pages 217–231, https://doi.org/10.1093/cjres/rst021.

Seaver, B. L., & Triantis, K. P. (1992). A fuzzy clustering approach used in evaluating technical efficiency measures in manufacturing. Journal of Productivity Analysis, 3(4), 337-363.

Seiford, L. M. (1994). “A DEA Bibliography.” In Charnes, Cooper, Lewin, and Seiford (eds.), Data Envelopment Analysis: Theory, Methodology and Applications. Boston: Kluwer Academic Publishers.

Seiford, L. M. (1996). “Data Envelopment Analysis: The Evolution of the State of the Art (1978–1995),” Journal of Productivity Analysis 7, 99–137.

Seiford, L. M., & Thrall, R. M. (1990), Recent Developments in DEA: The Mathematical Approach to Frontier Analysis, Journal of Econometris, 46, 7-38

Sekhon, J. S. (2009). Opiates for the Matches: Matching Methods for Causal Inference. Annual Review of Political Science, 12, 487-508.

Sellers-Rubio, R., Mas-Ruiz, F. J., & Casado-Díaz, A. B. (2010). University Efficiency: Complementariness versus Trade-off between Teaching, Research and Administrative Activities. Higher Education Quarterly, 64(4), 373-391.

Sengupta, J. K. (1992). A fuzzy systems approach in data envelopment analysis. Computers & Mathematics with Applications, 24(8-9), 259-266.

Sengupta, J. K. (1995). “A Dynamic View of Farrell Efficiency.” Int J Syst Sci 2: 1501-1509.

Sengupta, J. K. (1999). ‘The measurement of dynamic productive efficiency’, Bulletin of Economic Research, 51(2), 111–24.

Shephard, R. W. (1953). Cost and Production Functions. Princeton: Princeton University Press.

Shephard, R. W. (1970). Theory of Cost and Production Frontiers, Princeton, NJ, Princeton University Press

Sherman, H. D., & Zhu, J. (2006). Data Envelopment Analysis Explained. Service Productivity Management: Improving Service Performance using Data Envelopment Analysis (DEA), 49-89.

Shim, W. (2003). Applying DEA technique to library evaluation in academic research libraries. Library trends, 51(3), 312.

Shim, W., & Kantor, P. B. (1998, January). A novel economic approach to the evaluation of academic research libraries. In PROCEEDINGS OF THE ANNUAL MEETING-AMERICAN SOCIETY FOR INFORMATION SCIENCE (Vol. 35, pp. 400-410). LEARNED INFORMATION (EUROPE) LTD.

Sianesi, B. (2004). “An Evaluation of the Active Labour Market Programmes in Sweden,” The Review of Economics and Statistics, 86(1), 133–155.

Sickles, R. C. (2005). Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. Journal of econometrics, 126(2), 305-334.

Sickles, R. C., Good, D. H., & Getachew, L. (2002). Specification of distance functions using semi- and nonparametric methods with an application to the dynamic performance of Eastern and Western European air carriers. Journal of Productivity Analysis 17, 133–155.

Siegel, D. S., Waldman, D., & Link, A. (2003). Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: an exploratory study. Research policy, 32(1), 27-48.

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

Simar, L. (1996). Aspects of statistical analysis in DEA-type frontier models. Journal of Productivity Analysis, 7(2-3), 177-185.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. Journal of Productivity Analysis, 20(3), 391-424.

Simar, L. (2007), How to Improve the Performances of DEA/FDH Estimators in the Presence of Noise, Journal of Productivity Analysis, vol 28, 183–201.

Simar, L., & Vanhems, A. (2012). Probabilistic characterization of directional distances and their robust versions. Journal of Econometrics, 166(2), 342-354.

Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Management Science 44, 49-61

Simar, L., & Wilson, P. W. (1999). Performance of the Bootstrap for DEA Estimators and Iterating the Principle," Papers 0002, Catholique de Louvain - Institut de statistique

Simar, L., & Wilson, P. W. (2004). Estimation and inference in two-stage, semi-parametric models of production processes. Discussion Paper, Institut de Statistique, Universite´ Catholique de Louvain, Belgium.

Simar, L., & Wilson, P. W. (2006a). “Statistical Inference in Nonparametric Frontier Models: recent Developments and Perspectives”, forthcoming in The Measurement of

Productive Efficiency , 2nd Edition, Harold Fried, C.A.Knox Lovell and Shelton Schmidt, editors, Oxford University Press, 2006.

Simar, L., & Wilson, P. W. (2006b). “Efficiency Analysis: The Statistical Approach”, Manuscript, Institute of Statistics, UCL, Belgium.

Simar, L., & Wilson, P. W. (2007). Statistical inference in nonparametric frontier models: recent developments and perspectives.

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

Simar, L., & Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In: Fried H, Lovell CAK and Schmidt S (eds). The Measurement of Productive Efficiency and Productivity Growth. Oxford University Press: New York

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

Simar, L., & Wilson, P. W. (2011a). Inference by the m out of n bootstrap in nonparametric frontier models. Journal of Productivity Analysis, 36, 33–53

Simar, L., & Wilson, P. W. (2013). "Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives", Foundations and Trends® in Econometrics: Vol. 5: No. 3–4, pp 183-337. http://dx.doi.org/10.1561/0800000020

Simar, L., & Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. Journal of Productivity Analysis, 36(1), 1-20.

Simar, L., Van Keilegom, I., & Zelenyuk, V. (2017). Nonparametric least squares methods for stochastic frontier models. Journal of Productivity Analysis, 47(3), 189-204.

Simar, L., Vanhems, A., & Wilson, P. W. (2012). Statistical inference for DEA estimators of directional distances. European Journal of Operational Research, 220(3), 853-864.

Sinuany-Stern, Z., Mehrez, A., & Barboy, A. (1994). Academic departments’ efficiency via DEA. Computers & Operations Research, 21(5), 543-556.

Skodvin, O. (1999). Mergers in higher education – success or failure? Tertiary Education and Management, 5, 65-8dine, M. (2002). The end of the line? Accountable governance in the age of networks, partnerships and joined-up services. Governance, 15(1)

Smith, J. (2007). “Heterogeneity and Higher Education.” University of Michigan, mimeo. Universita Di Bergamo, Italy.

Smith, J., & Todd, P. (2005). “Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?,” Journal of Econometrics, 125(1-2), 305–353.

Sohn, S. Y., & Kim, Y. (2012). DEA based multi-period evaluation system for research in academia. Expert Systems with Applications, 39(9), 8274-8278.

Soy, S. K. (1997). The case study as a research method. Unpublished paper, University of Texas at Austin. Available at: http://www.ischool.utexas.edu/~ssoy/usesusers/l391d1b.htm [Accessed August 4, 2017].

Sreekumar, S., & Mahapatra, S. S. (2011). Performance modeling of Indian business schools: a DEA-neural network approach. Benchmarking: An International Journal, 18(2), 221-239.

Stevens, P. A. (2005). 'A stochastic frontier analysis of English and Welsh universities.' Education Economics 13(4): 355-374

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. Journal of econometrics, 13(1), 57-66.

Stewart, G. (2003). “College Mergers: Lessons to be Learned from Other Sectors.” Research in Post-Compulsory Education 8 (3): 305–28.

Strauss, A., & Corbin, J. M. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Sage Publications, Inc.

Sunil Dharmapala, P., & Saber, H. M. (2007). Resource allocation within academic units: DEA and goal programming approach with an application to an Arabian Gulf university. Benchmarking: An International Journal, 14(2), 202-210.

Tajnikar, M., & Debevec, J. (2008). Funding system of full-time higher education and technical efficiency: case of the University of Ljubljana. Education economics, 16(3), 289-303.

Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. Biometrika 97: 661–682.

Tauer, L. W., Fried, H. O., & Fry, W. E. (2007). Measuring efficiencies of academic departments within a college. Education economics, 15(4), 473-489.

Tavares, G. (2003). A bibliography of Data Envelopment Analysis (1978–2001). RUTCOR, Rutgers University, 2002; also available at:/http://rutcor.rutgers.edu/pub/rrr/reports2002/1_2002.pdfS.

Taylor, B., & Harris, G. (2004). Relative efficiency among South African universities: A data envelopment analysis. Higher Education, 47(1), 73-89.

Teixeira, P. N., & Amaral, A. (2007). “Waiting for the Tide to Change? Strategies for Survival of Portuguese Private HEIs.” Higher Education Quarterly 61 (2): 208–22.

Thanassoulis, E., (2001). Introduction to the Theory and Application of Data Envelopment Analysis: A foundation text with integrated software. Kluwer Academic Publishers, Boston, Hardbound, ISBN 0-7923-7429-0, 312 pp

Thanassoulis, E., de Witte, K., Johnes, J., Johnes, G., Karagiannis, G., & Portela, C. S. (2016). Applications of data envelopment analysis in education. In Data Envelopment Analysis; Springer: New York, NY, USA, 2016; pp. 367–438.

Thanassoulis, E., Kortelainen, M., Johnes, G., & Johnes, J. (2011). Costs and efficiency of higher education institutions in England: a DEA analysis☆. Journal of the Operational Research Society, 62(7), 1282-1297. ISSN 01605682

Thelin, J. R. (2011). A history of higher education. John Hopkins Press: Baltimore.

Thompson, J., & Bekhradnia, B. (2011). Higher Education: Students at the Heart of the System – an Analysis of the Higher Education White Paper, Higher Education Policy Institute, http://www.hepi.ac.uk/455-1987/Higher-Education-Students-at-the-Heart-ofthe-System.-An-nalysis-of-the-Higher-Education-White-Paper-.html (accessed 8 August 2017).

Thursby, J. G., & Kemp, S. (2002). Growth and productive efficiency of university intellectual property licensing. Research policy, 31(1), 109-124.

Tight, M. (2011). “How Many Universities Are There in the United Kingdom? How Many Should There Be?” Higher Edcuation 62 (5): 649–63.

Tirivayi, J. N., Maasen van den Brink, H., & Groot, W. (2014). Size and economies of scale in higher education and the implications for mergers.

Titus M. A., & Eagan, K. (2016). “Examining Production Efficiency in Higher Education: The Utility of Stochastic Frontier Analysis.” In Higher Education: Handbook of Theory and Research, 441–512. Springer.

Titus, M. A., Vamosiu, A., & McClure, K. R. (2016a). “Are Public Master’s Institutions Cost Efficient? A Stochastic Frontier and Spatial Analysis.” Research in Higher Education, 1–28.

Tomkins, C., & Green, R. (1988). An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting. Financial Accountability & Management, 4(2), 147-164.

Toth, R. (2009). Using DEA to evaluate efficiency of higher education. Applied Studies in Agribusiness and Commerce, 3(3-4), 79-82.

Trencher, G., Bai, X., Evans, J., McCormick, K., & Yarime, M. (2013). University partnerships for co-designing and co-producing urban sustainability. Global Environmental Change 28 (2014) 153–165

Triantis, K., & Eeckaut, P. V. (2000). Fuzzy pair-wise dominance and implications for technical efficiency performance assessment. Journal of Productivity Analysis, 13(3), 207-230

Triantis, K., & Girod, O. (1998). A mathematical programming approach for measuring technical efficiency in a fuzzy environment. Journal of Productivity Analysis, 10(1), 85-102.

Trippl, M., Sinozic, T., & Lawton Smith, H. (2012). The "third mission" of universities and the region: comparing the UK, Sweden and Austria. 52nd European Congress of the RSAI, 21st August - 25th August 2012

Trow, M. (1974). Problems in the transition from elite to mass higher education, in: Policies for higher education, from the general report on the Conference on Future structures of post-secondary education (Paris: OECD).

Tsionas, E. G. (2002). Stochastic frontier models with random coefficients. Journal of Applied Econometrics, 17(2), 127-147.

Tsionas, E. G., & Greene, W. (2003). Non-Gaussian Stochastic Frontier Models, Working Paper, Department of Economics, Athens University of Economics and Business, Athens.

Tsionas, E. G., & Kumbhakar, S. C. (2014). Firm Heterogeneity, Persistent And Transient Technical Inefficiency: A Generalized True Random-Effects model. Journal of Applied Econometrics, 29(1), 110-132.

Turner, D. (2005). Benchmarking in universities: league tables revisited. Oxford Review of Education, 31(3), 353-371.

Tuunainen, J. (2005). Hybrid practices? Contributions to the Debate on the Mutation of Science and University. Higher Education, 50(2), 275-298

Tyagi, P., Yadav, S. P., & Singh, S. P. (2009). Relative performance of academic departments using DEA with sensitivity analysis. Evaluation and Program Planning, 32(2), 168-177.

Tzeremes, N., & Halkos, G. (2010). A DEA approach for measuring university departments' efficiency, MPRA Paper, University Library of Munich, Germany.

Ulutas, B. H. (2011). Assessing the Relative Performance of University departments: teaching vs. Research. Ekonometri ve Istatistik Dergisi, (13), 125.

Universities UK. [UUK] (2005). Higher Education in Facts and Figures. Retrieved August 1, 2017, from: http://www.universitiesuk.ac.uk/facts-and-stats/data-and-analysis/Pages/higher-education-facts-and-figures-2011.aspx Universities UK, London

Universities UK. [UUK] (2006). Higher Education in Facts and Figures. Retrieved August 5, 2017, from: http://www.universitiesuk.ac.uk/facts-and-stats/data-and-analysis/Pages/higher-education-facts-and-figures-2011.aspx Universities UK, London

Universities UK. [UUK] (2011). Efficiency and effectiveness in higher education: A report by the Universities UK Efficiency and Modernisation Task Group. Retrieved January 20, 2016, from: http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2011/report-by-efficiency-and-modernisation-task-group.pdf Universities UK, London.

Universities UK. [UUK] (2014). Trends in undergraduate recruitment. Retrieved February 5, 2016, from: http://www.universitiesuk.ac.uk/Pages/Results.aspx?k=Trends%20in%20undergraduate%20recruitment%2E Universities UK, London.

Universities UK. [UUK] (2015). Efficiency, effectiveness and value for money. Retrieved March 20, 2016, from: http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2015/efficiency-effectiveness-value-for-money.pdf Universities UK, London.

Universities UK. [UUK] (2016). University Funding Explained. Retrieved March 30, 2016, from: http://www.universitiesuk.ac.uk/Pages/Results.aspx?k=UNIVERSITY%20FUNDING%20EXPLAINED Universities UK, London.

Universities UK. [UUK]. (2017). Patterns and trends, published annually by Universities UK, presents a comprehensive range of data and analysis on the changing size and shape of UK higher education. Retrieved September 5, 2017, from: http://www.universitiesuk.ac.uk/facts-and-stats/data-and-analysis/Pages/patterns-and-trends-2017.aspx. Universities UK, London.

Van der Laan, M. J., & Robins, J. M. (2003). Unified Methods for Censored Longitudinal Data and Causality. Springer, Berlin.

Vander Vennet, R. (1996). The effect of mergers and acquisitions on the efficiency and profitability of EC credit institutions. Journal of Banking & Finance, 20(9), 1531-1558.

Venditti, M., Reale, E., & Leydesdorff, L. (2011). The disclosure of university research for societal demand: A non-market perspective on the "third mission'', Science and Public Policy

Verry, D. W., & Davies, B. (1976). University Costs and Outputs. Amsterdam, Elsevier.

Verry, D. W., & Layard, P. R. G. (1975). 'Cost functions for university teaching and research.' Economic Journal 85: 55-74.

Vorley, T. & Nelles, J. (2008). (Re)Conceptualising the Academy: Institutional Development of and beyond the Third Mission, Higher Education Management and Policy, 20 (3), 119-135

Wagner, J. E. (2012). Forestry Economics: A Managerial Approach. Routledge Textbooks in Environmental and Agricultural Economics. Routledge Press, London (ISBN: 978-0-415-77476-5)

Waldman, D.M. (1982). A stationary point for the stochastic frontier likelihood. Journal of Econometrics 18, 275-279.

Wang, H. J. & Ho, C. W. (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. Journal of Econometrics, 157(2), 286-296.

Wang, W. K., & Huang, H. C. Measuring the Relative Efficiency of European MBA Programs: A Comparative analysis of DEA, SBM, and FDH Model. iji, 15(1), 0 Retrieved August 20, 2013, from: http://gebrc.nccu.edu.tw/proceedings/APDSI/2004/pdf/023.pdf.

Warning, S. (2004). Performance differences in German higher education: Empirical analysis of strategic groups. Review of Industrial Organization, 24(4), 393-408.

Watson, D. (2003). Universities and civic engagement: A critique and a prospectus. Paper for the 2nd biennial ''inside-out'' Conference, University of Queensland.

Wheelock, D. C., & Wilson, P. W. (2008). Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations. Journal of Econometrics, 145(1-2), 209-225.

Whitty, G., Hayton, A., & Tang, S. (2015). Who you know, what you know and knowing the ropes: A review of evidence about access to higher education institutions in England. Review of Education, 3(1), 27–67. doi:10.1002/rev3.3038

Wilson, P. W. (1993). Detecting Outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs. Journal of Business & Economic Statistics, 11(3), 319-323. doi:10.2307/1391956

Wilson, P. W. (1995). Detecting influential observations in data envelopment analysis. Journal of productivity analysis, 6(1), 27-45.

Wilson, P. W. (2011). Asymptotic properties of some non-parametric hyperbolic efficiency estimators, in I. van Keilegom and P. W. Wilson, eds., Exploring Research Frontiers in Contemporary Statistics and Econometrics, Berlin: Springer-Verlag.

Winsten, C. B. (1957). Discussion on Mr. Farrell's paper. Journal of the Royal Statistical Society, 120, 282-284.

Wolszczak-Derlacz, J. (2014). An Evaluation and Explanation of (In) efficiency in Higher Education Institutions in Europe and the U.S. with the Application of Two-stage Semi-parametric DEA. IRLE Working Paper No. 114-14. http://irle.berkeley.edu/workingpapers/114-14.pdf

Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: a two-stage multi-country approach. Scientometrics, 89(3), 887.

Wongchai, A., Liu, W. B., & Peng, K. C. (2012). DEA meta-frontier analysis on technical efficiency differences of national universities in Thailand. International Journal on New Trends in Education and Their Implications, 3(3), 3.

Wooldridge, J. M. (2000). Introductory econometrics: A modern approach. Mason, OH: South-Western Cengage Learning.

Wooldridge, J. M. (2002). Inverse probability weighted M-estimators for sample selection, attrition, and stratification. Portuguese Economic Journal 1: 117–139.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. The Review of Economics and Statistics, 87(2), 385-390.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. Journal of Econometrics 141: 1281–1301.

Worthington, A. C. (2001). An Empirical Survey of Frontier Efficiency Measurement Techniques in Education, Education Economics, (9), 245-268

Worthington, A. C., & Higgs, H. (2011). Economies of scale and scope in Australian higher Education, Higher Education, 61 pp. 387–414

Worthington, A. C., & Lee, B. (2008). Efficiency, technology and productivity change in Australian universities, 1998–2003. Economics of Education Review 27 (3), 285-298

Xiang-yang, S., Li, Z., & Qi-gang, L. (2012). An approach for evaluating journals of universities or colleges based upon multi-layer efficiency DEA. Physics Procedia, 25, 678-686.

Yin, R.K. (1984). Case Study Research: Design and Methods. Beverly Hills, Calif: Sage Publications.

Young Yoon, K. (2008). History of Higher Education in England and France, Korean Minjok Leadership Academy, International Program, Term Paper, AP European History Class, July 2008.

Zhang, L. (2010). The use of panel data models in higher education policy studies. In J. C. Smart (Ed.), Higher education: Handbook of theory and research (Vol. XXV, pp. 307–349). Dordrecht, The Netherlands: Springer.

Zhao, Q., & Percival, D. (2017). Entropy balancing is doubly robust. Journal of Causal Inference, 5(1).