Data Envelopment Analysis: Methodological Aspects and Empirical Applications

Dissertation
zur Erlangung des Doktorgrades
„Dr. rer. pol.“

der Fakultät für Wirtschaftswissenschaften der
Universität Duisburg-Essen

vorgelegt von

Gerald Fugger
aus Neunkirchen, Österreich
(Geburtsort, Geburtsland)

Betreuer:
Professor Dr. Andreas Behr
Lehrstuhl für Statistik

Essen, 12.8.2020

Erstgutachter: Prof. Dr. Andreas Behr

Zweitgutachter: Prof. Dr. Hendrik Schröder

Tag der mündlichen Prüfung: 16.12.2020

DOI: 10.17185/duepublico/73659
URN: urn:nbn:de:hbz:464-20210126-113231-4

Contents

List of Tables V

List of Figures VII

1 General Introduction 8

1.1 Motivation . . . 8

1.2 Efficiency Assessments using DEA . . . 11

1.3 Overview of the Five Studies . . . 13

2 Absolute and Relative Weight Restrictions in DEA - An Comparison 20

2.1 Introduction . . . 22

2.2 Method Origin and a Literature Review . . . 24

2.2.1 Method Development . . . 24

2.2.2 Applied Restrictions in Literature . . . 26

2.3 DEA Models and Weight Restrictions . . . 31

2.3.1 Basic Model . . . 31

2.3.2 Weight Restriction Types . . . 36

2.3.2.1 Absolute Restrictions . . . 36

2.3.2.2 Relative Restrictions . . . 38

2.4 Additional Weight Restriction Selection . . . 42

2.5 Conclusion . . . 44

3 Comparing the Slack-Based Maximum Measure of Efficiency: A Simulation Application 46

3.1 Radial and Non-radial DEA Models . . . 48

3.2 Literature Overview . . . 49

3.3 The DEA Models . . . 52

3.4 A Numerical Example . . . 56

3.5 A Simulation Application . . . 58

3.6 Conclusion . . . 61

4 PISA Performance of Natives and Immigrants: Selection versus Efficiency 63

4.1 Introduction . . . 65

4.2 Literature Overview . . . 67

4.3 Methodology . . . 70

4.3.1 Sets of Students 70
4.3.2 Students and Different Frontiers of Potential Scores 70
4.3.3 Benchmarking Individual Students 71
4.3.4 Benchmarking Sets of Students 72
4.3.5 Considering a Second Country 73
4.3.6 The DEA Model 74
4.4 The PISA study, Migration Regimes and Descriptive Results 75
4.4.1 The PISA Study and the ESCS 75
4.4.2 Migration Regimes 76
4.4.3 The Data and Descriptive Results 78
4.5 Efficiency Results and Efficiency Decomposition 82
4.5.1 National Frontiers 83
4.5.2 International Comparisons 85
4.5.3 Differences Between Immigrants and Natives 86
4.5.4 Selection Effects and Educational Efficiency 87
4.5.5 PISA Scores as Separate Outputs 89
4.6 Conclusion 90
A.1 Supplementary Results 92

5 Higher Education Institution Efficiency in Germany and the United Kingdom 105

5.1 Introduction 107
5.2 Literature Review 108
5.3 Operationalisation 110
5.3.1 Efficiency Assessments using DEA 110
5.3.2 Analysis Level 111
5.3.3 Inputs and Outputs 113
5.3.4 Models and Zero Weights 114
5.4 Methodology 115
5.4.1 DEA Models 115
5.4.2 Decomposition 118
5.4.3 An Illustration 118
5.5 The Efficiency of German and UK HEIs 120
5.5.1 Efficiency Results 123
5.5.2 Result Comparisons and Discussion 125
5.6 Conclusion and Suggestions for Future Research 129

B.1 Supplementary Results 131

6 An Encompassing Assessment of OECD Countries **145**

6.1 Introduction 147

6.2 Literature Overview 148

6.2.1 Economic Performance 149

6.2.2 Health 149

6.2.3 Environmental Studies 150

6.2.4 Well-being Studies 151

6.2.5 The Human Development Index and the Better Life Index . . . 152

6.2.6 Additional Weight Restrictions in DEA 153

6.3 Operationalisation and Variable Selection 154

6.4 Methodology 158

6.5 Results 159

6.5.1 Robustness Checks 162

6.6 Conclusion 163

C.1 Supplementary Results 168

7 Summary and General Conclusion **176**

References **183**

Attachments **206**

List of Tables

2.1 Weight restrictions in the literature 29
2.2 Example DMUs, efficiency scores, weights and intensity weights 34
2.3 Example DMUs using AWR, efficiency scores and weights 37
2.4 Example DMUs using ARI, efficiency scores and weights 39
2.5 Additional restrictions 40
2.6 Efficiency scores and weights for R.1 and R.2 41
2.7 Examples for weight restriction sources 43
3.1 Example DMUs and intensity weights 56
3.2 Efficiency scores 57
3.3 Descriptive results and correlation coefficients across all iterations . . . 59
3.4 Average efficiency scores comparisons 60
3.5 Computing times in seconds of all iterations 60
4.1 Average highest occupational status of parents 77
4.2 PISA scores correlations coefficients, overview 79
4.3 Descriptive results of the average PISA scores and the ESCS values . . 80
4.4 Decomposition within countries 83
4.5 International decomposition 85
4.6 Correlation coefficients between different outputs 90
A.1 Excluded outliers 93
A.2 Descriptive results of the PISA mathematics scores 94
A.3 Descriptive results of the PISA science scores 95
A.4 Descriptive results of the PISA reading scores 96
A.5 PISA scores correlation coefficients 97
A.6 Decomposition within countries, three outputs 103
A.7 International decomposition, three outputs 104
5.1 Efficiency scores given $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ 120
5.2 Descriptive results, all HEIs 123
5.3 Correlation coefficients, all HEIs 123
5.4 Efficiency results overview 124
5.5 Relative slacks 125
5.6 Student shares of the fields of education regression results 127
B.1 Inputs and outputs of German HEIs 131
B.2 Inputs and outputs of UK HEIs 132
B.3 Descriptive results for each country 133

B.4 Correlation coefficients for each country 133
B.5 Efficiency results of German HEIs 134
B.6 Efficiency results of UK HEIs 135
B.7 Slacks of German HEIs relative to $F^{\mathcal{U}_{GER}}$ 136
B.8 Slacks of German HEIs relative to $F^{\mathcal{U}_{ALL}}$ 137
B.9 Slacks of UK HEIs relative to $F^{\mathcal{U}_{UK}}$ 138
B.10 Slacks of UK HEIs relative to $F^{\mathcal{U}_{ALL}}$ 139
B.11 Descriptive results of the fields of education 140
B.12 Student shares of the fields of education 140
B.13 Literature review 141
6.1 Descriptive results 158
6.2 Correlation coefficients 158
6.3 Results of the BCC model 160
6.4 Results of the BCC model with weight restrictions 161
6.5 Efficiency scores for different thresholds 167
C.1 Country names and abbreviations 168
C.3 Overview of macroeconomic performance literature 168
C.4 Overview of health-related efficiency literature 169
C.5 Overview of environmental efficiency literature 171
C.6 Overview of human development efficiency literature 172
C.2 Efficiency results of the BCC model with AR1 175
C.7 Data sources 175

List of Figures

1.1 Production process . . . 12
2.1 Production possibility set of the example DMUs . . . 35
2.2 Example DMUs' weighting space . . . 35
2.3 Example DMUs' weighting space, including AWR . . . 38
2.4 Example DMUs' weighting space, including ARI . . . 40
2.5 Weighting spaces . . . 41
2.6 Additional weight restrictions implementation procedure . . . 42
3.1 Production possibility set of the example DMUs . . . 58
3.2 Sorted efficiency scores . . . 61
4.1 Relationship between PISA scores and ESCS values . . . 66
4.2 Benchmarking of a country with two student groups . . . 71
4.3 Benchmarking of multiple country with two student groups . . . 73
4.4 Average PISA scores distributions among natives and immigrants . . . 81
4.5 Efficiency differences between native and immigrant students . . . 86
4.6 Country rankings . . . 88
A.1 Country-level average PISA and ESCS scores . . . 92
A.2 Mathematics scores distributions . . . 98
A.3 Science scores distributions . . . 99
A.4 Read scores distributions . . . 100
A.5 Efficiency frontiers . . . 101
5.1 Efficiency assessment example . . . 111
5.2 Super-efficiency example . . . 117
5.3 Efficiency decomposition example . . . 119
5.4 Inputs-outputs relationships for German and the UK HEIs . . . 122
5.5 Average student shares of the fields of education . . . 126
B.1 Efficiency decomposition example . . . 144
6.1 Our operationalisation . . . 155
6.2 Normalised data . . . 165
6.3 Sorted efficiency scores . . . 166

1

General Introduction

1.1 Motivation

In this thesis, I apply a comparative deterministic and non-parametric method for calculating production efficiency, called Data Envelopment Analysis (DEA). DEA is based on the seminal article of Farrell (1957) and was introduced by Charnes et al. (1978) to assess the efficiency of decision-making units (DMUs). DEA has been continuously improved and adapted for new areas of application so that its various enhancements have become the most widely used efficiency assessment technique in economics and operations research (Emrouznejad et al., 2018).

The five studies in this thesis contribute to the existing literature by focusing on technical challenges and by emphasising methods and operational aspects of DEA. The first two studies are more technically focused and build the basis for the three empirical papers. The latter are efficiency evaluations of secondary and tertiary education institutions and an encompassing assessment of countries. Although DEA is one of the most widely used efficiency assessment techniques, operational and methodology aspects are often insufficiently emphasised in the majority of efficiency evaluations (Cook et al., 2009b).

In its most basic form, DEA models do not require information other than data on inputs consumed and outputs produced by the units compared. Methodology aspects in empirical efficiency assessment encompass the choice between radial and non-radial models, additional weight restrictions, and model orientation, among others. The relative productivity of the unit under consideration can be calculated relative to different reference sets. The variable selection, the classification in inputs and outputs, and the selected returns to scale assumption are analysis specific. Apart from the empirical results, the elaboration of operational and technical aspects and their implementation are the main contributions of this thesis to the existing efficiency literature.

In DEA, the unit of interest is benchmarked against similar units, and the most efficient ones are used to calculate a best-practice frontier against which the inefficient units are compared. The units are assumed to use inputs to produce outputs. Inputs and outputs are aggregated using weights to obtain a productivity measure (Cooper et al., 2009). The weights are calculated to fulfil the purpose of the target function, mainly to maximise the efficiency of the DMU under consideration given the restrictions of the respective model. These variable weights are associated with value judgments (Kong et al., 2012).

Traditionally, DEA models assume total flexibility in weight selection and do not need any previous assumptions about the underlying production process. Additional weight restrictions can implement prior knowledge on the interdependencies between specific inputs and outputs or reflect knowledge of individual factors. Additional restrictions can prevent DMUs from being assessed on only subsets of the data and unreasonable weights distributions, and thus, can improve discrimination between the DMUs (Atici et al., 2015). Additional weight restrictions can be absolute, relative, input- or output-specific, allow to link inputs and outputs, and limit the Production Possibility Sets (PPSs) of the DMUs. In my thesis, I outline how sources for additional weight restrictions can be identified, and how these restrictions can be formalised and implemented.

Another important decision in efficiency evaluations using DEA is whether to use radial or non-radial models. Radial DEA models maintain a specific input-to-output ratio and thus prevent any substitution among them (Khalili et al., 2010). These models are either units invariant or translation invariant, and they may overestimate technical efficiency by underestimating inefficiency, and they may fail to distinguish between input- and output-orientation (Dyson et al., 1988). Non-radial Slack-Based Measure (SBM) provides an efficiency measure based on relative input excess (input slacks) and output shortfalls (output slacks). SBM can easily be enhanced and is a more comprehensive alternative to the often-used radial models (Tone, 2001).

DEA is becoming increasingly important in the efficiency analysis of educational institutions and sectors (Thanassoulis et al., 2016). The Programme for International Student Assessment (PISA) data set is one of the most comprehensive and, therefore, most frequently used data sources for international efficiency analysis in secondary education. PISA is an international sample study that assesses the performance of 15-year-old students' in mathematics, science, and reading in most Organisation for International Co-operation and Development (OECD) countries (Teltemann et al., 2016). Interestingly, immigrant and native students perform differently in most countries. Here, DEA allows us to further decompose the students' efficiency scores. My co-author and I identify whether the performance gaps in PISA result from different immigration regimes or differences in the countries' education systems. The latter implies that countries without restrictive immigration regimes attract immigrants who have worse socio-economic endowments than their natives. Even if we account for the differences in the students' socio-economic endowments as inputs, differences in academic performance between natives and immigrants persist. Using DEA, we conduct an international student-level analysis of secondary education systems and assess how well they maximise the performance of immigrants and natives.

Apart from assessing the performance of secondary education institutions, one of DEA's recent main areas of application is the evaluation of higher education institutions (HEIs) (Liu et al., 2013). Various studies initially calculated the cost efficiency of HEIs. In recent years, studies have increasingly assessed the efficiency of HEIs in maximising their three main missions: research, teaching, and innovation (Frenken et al., 2017). Gawellek et al. (2016) and others have evaluated the efficiency of German HEIs in country-specific analyses or compared them in comprehensive evaluations together with HEIs from another country (Veiderpass et al., 2016). Country-specific and international efficiency analysis based on one common best practice frontier do not necessarily reveal country-specific input-output structures. Thus, the fourth study in this thesis uses super-efficient DEA models to identify and account for country-specific focus areas.

Although efficiency evaluations were originally conducted to analyse specific sectors or institutions, there is also a long tradition of measuring and comparing the performance of countries (Patrizii et al., 2017). Commonly used one-dimensional measures of economic prosperity, such as the gross domestic product, do not adequately reflect the standard of living and ignore the endowments of the respective countries (Lovell et al., 1995). Measures of subjective well-being better reflect the actual societal targets of countries (Mariano et al., 2015). The Human Development Index (HDI) is one of the most popular composite indices that measures human development. However, it insufficiently considers all areas of societal interest. Furthermore, the HDI does not account for countries' different endowments and neglects individual countries' preferences and policy targets (Greco et al., 2019). Therefore, we use DEA to evaluate the countries' strengths and weaknesses in providing their citizens with long and fulfilling lives, given their economic, environmental, and health endowments. Additional weight restrictions ensure that each country is assessed based on all variables and increases the discriminatory power of our analyses.

Due to its continuous improvements, DEA is a flexible and widely used technique to determine and compare the efficiency of DMUs. The actual production processes of DMUs are unknown and DEA assumes them based on the available data that must be by classified into inputs and outputs. It can be difficult to distinguish between inputs and outputs or to include all possible variables in efficiency assessment. One of the most important aspects of any DEA analysis is to operationalise the model so that it reflects the real production process as accurately as possible. Operationalisation comprises, among other things, the choice of model, the inclusion of additional weight restrictions, and the selection of variables.

The five studies in this thesis contribute to the existing operational research literature by analysing the impacts and implementation of additional weight restrictions, comparing radial and non-radial models, and performing multiple decompositions of efficiency results. Furthermore, the operationalisation used in this dissertation, the applied methodology, and the empirical analyses complement several of the most important current areas of efficiency assessment. In the field of secondary education, we find inefficiencies within several secondary school systems in relation to improving the performance of immigrants. In the area of higher education, an important contribution is made in revealing country-specific input-output structures due to different educational priorities. Finally, we conduct a comprehensive analysis of OECD countries. Compared to the composite indices commonly used, DEA is a more flexible, data-based approach that takes into account both the characteristics of countries and their endowments.

1.2 Efficiency Assessments using DEA

This section has three functions: it provides a basic understanding of efficiency evaluations using DEA, discusses differences between radial and non-radial models, and describes returns to scale assumptions. Unless otherwise stated, this section refers to the Charnes-Cooper-Rhodes (CCR) model introduced by Charnes et al. (1978) which can be regarded as a basic DEA model (Cooper et al., 2007).

In DEA, DMUs consume inputs in order to produce outputs. Figure 1.1 illustrates these input-to-output transformation, which is referred to as production process, for one DMU. The other DMUs use the same inputs and produce the same outputs but have most likely different input-output combinations and therefore different production

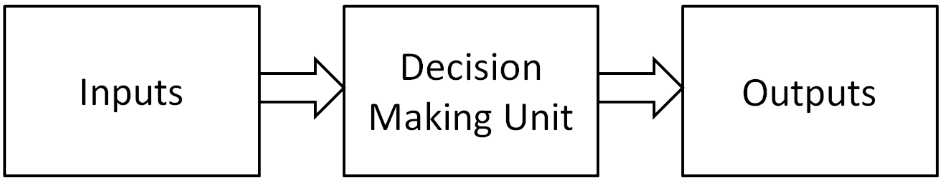


Figure 1.1 – Production process

processes. Productivity is defined as the ratio of produced output to consumed input. An efficiency measurement, or score, is obtained by comparing the output-to-input ratio of the DMU under consideration, hereafter referred to as DMU$_o$, with the maximum observed ratios of similar DMUs. The DMUs with which DMU$_o$ is compared form the reference set for DMU$_o$. The most productive units (highest output-to-input ratios) serve as benchmarks for their less productive counterparts. The efficient units span the efficiency frontier that envelops all data, hence the name DEA. If no efficient reference DMU exists to which DMU$_o$ can be compared, DEA computes a linear combination of the existing efficient DMUs that corresponds to the input-output structure of DMU$_o$. This linear combination, which does not exist in the data, is referred to as synthetic DMU. In a nutshell, inefficient DMUs are not on the efficiency frontier and are compared with best practice. Efficiency scores indicate how much inefficient DMUs must improve to become efficient (Golany et al., 1989).

In the one-input-one-output case, direct productivities can be calculated by dividing the amount of produced output by the amount of consumed input for each DMU. Even in this most basic form, DEA provides additional information by identifying appropriate reference units for DMU$_o$. Different reference sets can provide additional information about the relative productivity of DMU$_o$. In complex situations with several inputs and outputs, DEA uses weights to aggregate them. In the CCR model, DEA calculates the most favourable weights so that each DMU becomes as efficient as possible (Behr, 2015). The weights must not be chosen a priori, they must be non-negative, and the sum of weighted inputs must be not greater than the sum of weighted outputs. Without additional weight restrictions, the linear program may calculate specialised DMUs by assigning zero weights to some of their inputs and outputs. Specialised DMUs are assessed only on subsets of the data, which can result in unrealistic or unreasonable production processes (Doyle et al., 1994). Additional weight restrictions can prevent zero weights, can allow the implementation of prior knowledge on the interdependencies between the variables, and can improve discrimination between efficient and inefficient DMUs (Cooper et al., 2007).

Inefficient DMUs can become efficient by reducing their inputs or by increasing their outputs or a combination of both, depending on the underlying model assumptions. The CCR model is radial. It assumes proportional input reductions and proportional output increases. Non-radial models allow non-proportional input reductions and non-proportional output increases (Avkiran et al., 2008). The CCR model distinguishes between input- and output-orientation. The former assumes that DMUs minimise their inputs given their outputs and the latter implies output maximisation given the inputs. Most non-radial models can be calculated as un-oriented (they simultaneously consider output maximisation and input minimisation). Non-radial models account for output shortfalls and input excesses that are ignored in the CCR model (Tone, 2001).

DEA models can implement different returns to scale assumptions. Constant returns to scale (CRS) reflect the assumption that outputs will change by the same proportion as inputs are changed (i.e., a 100% increase of all inputs will increase all outputs by 100%). Increasing returns to scale allow over-proportional increases in outputs and decreasing returns to scale under-proportional increases (Fadeyi et al., 2019). Variable returns to scale (VRS) assume that an increase in inputs leads to disproportionate increases in outputs (Cooper et al., 2007). Banker et al. (1984) introduced the Banker-Charnes-Cooper (BCC) model, which includes the VRS assumption within the CCR model. Both models are radial and commonly used for empirical efficiency analysis across all disciplines (Suzuki et al., 2017).

Overall, DEA models have become the dominant efficiency assessment models in operations research as they allow, among other things, to distinguish between efficient and inefficient DMUs, to calculate their efficiency, and to identify sources of inefficiency (De Castro et al., 2017).

1.3 Overview of the Five Studies

The five studies in this dissertation deal with operational aspects and theoretical implications of efficiency assessments using DEA. In the first two studies, I discuss advantages, implementation, and consequences of additional weight restrictions and differences between radial and non-radial DEA models. The effects are demonstrated using small samples of artificial DMUs, as well as large-scale simulations with up to 1,000,000 DMUs. Papers three to five are empirical applications based on the insights of the former two studies. In the third study, the efficiency of the secondary education systems of 20 OECD countries in integrating immigrant students, given their socio-economic backgrounds, are evaluated. The results of 153,374 students are decomposed relative to national and international efficiency frontiers. After the assessment of secondary education systems, I evaluate 46 higher educational institutions (HEIs) in Germany and 45 in the United Kingdom using radial DEA models in the fourth paper. While the previous studies focus on specific sectors, the fifth paper is an encompassing assessment of 33 OECD countries. My co-authoer and I calculate how efficiently they enable their citizens to live as long and fulfilling lives as possible given their economic, environmental, and health resources. Additional weight restrictions ensure that each country is assessed based on all variables and allow a better distinction between efficient and inefficient countries. Whenever large amounts of data or a large number of linear programs are calculated, I reduce the computing time by parallelisation.

In the following, I provide a detailed overview of my five studies, including the research questions, the study's methodology, and the most important results.

Paper 1: Absolute and Relative Weight Restrictions in DEA - An Comparison

In the first paper, I focus on additional weight restrictions in DEA. I emphasise appropriate restriction sources and the motivations behind the restrictions, and I compare the implementation of different weight restriction types. Without additional weight restrictions, DEA may overestimate efficiency by assuming unrealistic production processes (Yang et al., 2019).

Thompson et al. (1990) introduced a new generalised approach to implementing weight restrictions in the DEA. These restrictions can be absolute for specific quantities or relative by linking inputs and outputs. Additional weight restrictions allow the implementation of prior knowledge on the interdependencies between inputs and outputs, increase discrimination among efficient and inefficient DMUs, and allow a more accurate representation of the production processes. Additional weight restrictions are mainly based on additional information such as descriptive results (Thompson et al., 1990), on expert opinions (Cooper et al., 2009), or on additional findings like linear regressions (Cook et al., 1991) or correlation coefficients (Mecit et al., 2013).

In this study, I present and discuss various DEA models and exemplify the impact of additional weight restrictions given five artificial DMUs. Using the DMUs' weighting space, my results show that additional weight restrictions limit the feasible weights region and can render the efficiency evaluation infeasible. The additional restrictions reduce the average efficiency and reveal specialised DMUs. Overall, the restrictions allow a more realistic representation of the targeted production processes (Wong et al., 1990). Robustness checks for different restriction thresholds are particularly necessary when the thresholds are chosen arbitrarily, since the restrictions directly affect the calculated efficiencies.

Paper 2: Comparing the Slack-Based Maximum Measure of Efficiency: A Simulation Application

In my second paper, I discuss differences between radial and non-radial DEA models. In radial models, DMUs can only become efficient by proportional contractions of their inputs or by proportional extensions of their outputs. The basic radial models distinguish between input- and output-orientation and ignore input excess (input slacks) and output shortfalls (output slacks). Non-radial DEA models allow non-proportional input reductions and non-proportional output increases. Tone (2001) introduced a slack-based measurement (in the following referred to as SBM-Min model) to enhance some of these drawbacks of radial models. The solutions of the SBM-Min do not depend on the units in which the inputs and outputs are measured (it is unit invariant), and its efficiency measure is monotone decreasing in each input and output slack. The SBM-Min has data-based lower weights restrictions, and its efficiency measure is as low as possible as the model maximises inefficiency. On the contrary, the SBM-Max is based on the SBM-Min and maximises efficiency by minimising inefficiency. The SBM-Max uses several linear programs to approximate the closest reference point at the efficiency frontier for each inefficient DMU. Exact efficiency scores cannot be computed as the SBM-Max would become too complex or too computationally demanding (Tone, 2017).

I first use five artificial DMUs to elaborate on the exact efficiency determinations and the differences between the radial and the non-radial models. These five DMUs are carefully selected to illustrate the characteristics of the different models but represent only one specific application. Consequently, I simulate two inputs and two outputs using truncated normal distributions for 1,000 DMUs 1,000 times to demonstrate the different interpretations of inefficiency between the models, as well as to calculate their computing efforts.

Correlation coefficients indicate strong positive relationships between the efficiency results of all models. If the efficiency scores are sorted in ascending order, the results of the radial models and the SBM-Max are more similar than the results of the SBM-Min. Apart from the efficient DMUs, the SBM-Min calculates the lowest efficiency scores for each quantile. The average computing times for the SBM-Max is about 70 times higher than for the other models due to the calculation of several linear programs.

This is the first study comparing the SBM-Max with other radial and non-radial DEA models on such a large scale. The SBM-Max offers an upper efficiency bound and the SBM-Min a lower efficiency bound. This study provides a guideline for future research by depicting the efficiency interpretation of the models and comparing their advantages and drawbacks.

Paper 3: PISA Performance of Natives and Immigrants: Selection versus Efficiency

In this study, my co-author and I analyse the performance of secondary education systems in 20 countries, given the socio-economic backgrounds of immigrant and native students. Performance gaps between immigrants and natives in PISA are due to different immigration regimes and different levels of success in integrating immigrants. We use DEA and conduct all analyses on the student level, the most disaggregated data available in PISA. Our sample comprises 153,374 students. The index of economic, social, and cultural status (ESCS) is our input because the socio-economic endowment is one of the most important determinants of students' educational successes (Hwang et al., 2018). We use the average of the mathematics, science, and reading scores on PISA as output. Aggregating the three strongly positively correlated PISA scores into one output allows a straightforward interpretation and decomposition of the efficiency frontiers. We calculate the efficiency of students relative to national and international efficiency frontiers.

In Denmark, Finland, and Sweden, native students perform substantially better than immigrants when compared to national efficiency frontiers. In contrast, immigrants are more efficient than their native peer groups in Australia, Canada, Israel, Singapore, and the United States. Relative to the international frontier consisting of all students and compared to their native peer groups, immigrants in Finland, Sweden, and Denmark perform relatively poorly. The opposite is true in Australia, Singapore, and the United States. The latter indicates that in countries with selective immigration regimes, immigrants outperform natives if the socio-economic background of the students is considered.

Our main findings are obtained by comparing the average PISA scores of immigrants of each country with their respective efficiency scores. The former indicates how well immigrants perform in absolute terms. The efficiency scores reveal how well the education systems maximise the efficiency of immigrants given their socio-economic endowments. Immigrants in Spain achieve relatively low PISA scores, but on average, they are among the top performing students regarding their efficiency. The education systems in France, Italy, and Portugal enable their students to perform relatively well given their socio-economic endowments. In contrast, the results of immigrants in Israel and the United Kingdom indicate relatively poorly performing education systems. Regardless of whether the ESCS is used as input, immigrants perform best in Singapore and worst in Denmark. We find that immigrants in countries with restrictive immigration regimes (Australia, Canada, the United Kingdom, and New Zealand) perform well in absolute PISA scores and are also quite efficient given their ESCS input levels.

Paper 4: Higher Education Institution Efficiency in Germany and the United Kingdom

In this paper, I address efficiency assessments in the context of higher education. Although there is sophisticated literature on the efficiency of HEIs, the results are often not further decomposed or are not compared with similar higher education systems or other studies. My study complements the literature by addressing the operationalisation of higher education efficiency assessments, outlining an applicable methodology, and evaluating the efficiency of German and UK HEIs.

Higher educational efficiency assessments are limited by the available data and thus mostly conducted on the departments of one specific HEI or the HEI-level. According to Wolszczak-Derlacz (2017), the department structure of the HEIs and the different focus areas of the various disciplines can be accounted for by selecting appropriate inputs and outputs that reflect the HEI's resources and missions. The HEIs' missions can be grouped into three outputs: research, teaching, and innovation (Frenken et al., 2017). The inputs represent personnel and capital resources (Rhaiem, 2017).

Without additional weight restrictions, radial DEA models calculate zero weights for most HEIs. Consequently, I use the output-oriented non-radial SBM-Min model, which prevents zero weights and has further favourable characteristics, which are discussed in detail in the second paper. In addition, I apply a multi-step super-efficient efficiency measurement to further decompose the efficiency results. In the super-efficient model, the HEIs are benchmarked against the efficiency frontier of the other country. The super-efficient model can calculate efficiency scores for HEIs that are located outside the feasible region. Such HEIs would render the SBM-Min infeasible.

The data are obtained from the Centre for Science and Technology Studies (CWTS) Leiden ranking and European Tertiary Education Register (ETER) data sets. 27 out of 46 German and 26 out of 45 UK HEIs are efficient when country-specific efficiency frontiers are used. In an international analysis that includes HEIs from both countries, 19 German and 24 UK HEIs are efficient. UK HEIs are on average more efficient, which indicates that German HEIs could produce more outputs given their inputs. Nearly all HEIs are super-efficient when benchmarked against the efficiency frontier of the respective other countries. These findings indicate vastly different input-output structures between the two countries which are also evident in the descriptive results. These country-specific differences are of particular importance for further international efficiency assessments when data for more countries become available.

Paper 5: An Encompassing Assessment of OECD Countries Using Weight Restricted DEA Models

In the last paper, my co-author and I conduct an encompassing assessment of 33 OECD countries. We evaluate how well these countries provide their citizens with long and fulfilling lives, given their economic, environmental, and health endowments. There is a long tradition in evaluating the performance of countries. First, studies such as Lovell et al. (1995) focused solely on measures of economic prosperity and did not account for the available resources. More recent evaluations focus on the standard of living and/or subjective well-being and account for a variety of socio-economic quantities as inputs (Mizobuchi, 2017).

Composite indices like the commonly used Human Development Index (HDI) often lack areas of societal interest, subjectively weigh the variables, and neglect countries' preferences and policy targets (Greco et al., 2019). Consequently, we use DEA as a data-driven technique to account for the countries strengths and weaknesses.

Our two outputs (healthy life expectancy and subjective well-being) are the primary goals of many societies (Ülengin et al., 2011). Our inputs represent the countries' economic, environmental, and health endowments. We include additional weight restrictions in our output-oriented DEA models. As shown in my first paper, such restrictions prevent fully specialised DMUs and increase the discriminatory power of our analyses. Without additional weight restrictions, zero weights are calculated at least for one input or one output in each country, and several countries are assessed based on only one input and one output. We use different weight thresholds and validate the robustness of our country ranking.

Our main findings indicate that the countries use heterogeneous strategies to maximise the outputs depending on their available inputs. Mexico and Russia perform well because they attain their relatively low outputs with relatively low inputs. In contrast, Japan, Switzerland, and Norway are efficient due to their high outputs. Australia, Canada, Denmark, Ireland, Israel, Latvia, Spain, and Turkey are efficient due to balanced inputs-outputs combinations or because they are partly specialised.

This study shows that DEA is more flexible in accounting for the countries' characteristics in maximising human development than an index with static weights. In addition, the inclusion of multiple inputs allows for the consideration of the countries' endowments, which is neglected in most commonly used composite indices assessing the countries' performance.

2

Absolute and Relative Weight Restrictions in DEA - An Comparison

Absolute and Relative Weight Restrictions in DEA - An Comparison

Gerald Fugger

University Duisburg-Essen, 45141 Essen, Germany.

Abstract

In Data Envelopment Analysis (DEA), the decision maker (DM) only selects inputs and outputs and does not have to make any previous assumptions about the underlying production process. Without additional weight restrictions, the linear program may assign zero weights to inputs and outputs of certain decision-making units (DMUs). Specialised DMUs can be calculated as efficient because only subsets of the data are included in their efficiency calculation. In most cases, however, it is not desirable to ignore inputs or outputs previously selected by the DM, so technical efficiency can overestimate true efficiency. Additional weight restrictions can avoid specialised DMUs, further improve discrimination between DMUs, and allow the implementation of prior knowledge of the production process. Depending on the application and the available information, the correct sources for additional weight restrictions must be identified, the restrictions formalised, and implemented. This publication provides a comprehensive overview of different additional weights restrictions in DEA, emphasising their motivation and the consequences of their implementation.

Keywords: Data Envelopment Analysis, Weight restrictions, Assurance regions
JEL Classification: C14 C52 C61

2.1 Introduction

The use of mathematical programming techniques for assessing the comparative efficiency of decision making units (DMUs) was proposed originally by Farrell (1957). Charnes et al. (1978) established Data Envelopment Analysis (DEA) as a tool for assessing relative efficiency. Their basic model (CCR-model) can calculate solutions for various kinds of production possibility sets (PPS) without the need for prior knowledge or assumptions that are not part of the inputs or outputs (Charnes et al., 1978). Banker et al. (1984) extended the CCR-model by introducing assumptions of variable scales in the so-called BCC-model (Banker et al., 1984). Since their introduction, the CCR- and BCC-models have become widely used for efficiency assessments and are referred to as DEA standard models in this publication. In these models, the DMUs produce outputs and consume inputs. Outputs and inputs are hereinafter referred to as factors. The linear program freely assigns the factor weights of the DMUs (as long as they remain non-negative) to maximise their efficiency. The selected weights can be associated with value judgements, and the resulting ratios of the inputs (outputs) can be interpreted as marginal rates of substitution (inputs) or transformation (outputs).

In DEA standard models, efficiency is calculated mathematically without additional assumptions, which may be subjective or can bias results. However, technical efficiency may differ from economic efficiency, which is obtained by including additional information (Yang et al., 2019). As one of the first publications, Thompson et al. (1986) introduced further weight restrictions to include information in the siting of nuclear physics facilities in Texas to enable discrimination between DMUs that are efficient in the standard DEA models (Thompson et al., 1986).

Since then, multiple studies have included additional weight restrictions for three main reasons:

Firstly, to implement prior knowledge on the interdependencies between specific inputs and outputs or to reflect knowledge of individual factors. If the imputed rates of substitution and transformation deviate from prior knowledge of the DM, additional information may help the model to adapt to its assumptions. For example, Kao et al. (2008) include weight restrictions to evaluate university departments. The restrictions are set based on prior knowledge provided by university administrators (Kao et al., 2008). In another educational efficiency study, Kong et al. (2012) implement output weight restrictions to take into account the marginal rates of transformations based on expert opinion (Kong et al., 2012).

Secondly, to prevent zero weights or unreasonable weight distributions. In standard DEA models, resulting weights can be zero or close to a non-Archimedean infinitesimal, resulting in marginal rates of substation and transformation that cannot always be defined (Angulo-Meza et al., 2002). Furthermore, zero weights imply that inputs (outputs) are neglected in the assumed underlying technology and the DMUs efficiency calculation, that may overestimate its' efficiency (Dyson et al., 1988; Joro et al., 2004). To demonstrate the consequences of the absolute specialisation of DMUs, Doyle and Green (1994) draw the following comparison:

> 'Best engine on the market-pity the car has no wheels'
> (Doyle et al., 1994, p. 569)

If only subsets of the outputs are considered, this can lead to unrealistic or unreasonable production processes. For example, to calculate a car without wheels as efficient because it has the best engine (Doyle et al., 1994). Allen et al. (1997) describe zero weights as a free lunch and point to conceptual problems of relating zero weights with economic theories such as resource allocation (Allen et al., 1997). Cook et al. (2009b) argue that weight restrictions based on assurance regions are a suitable instrument for limiting weights within boundaries to avoid too large weight discrepancies (Cook et al., 2009b).

Thirdly, to improve discrimination among efficient and between efficient and inefficient DMUs. When the number of DMUs is rather small in comparison to the number of factors, standard DEA models lack discriminatory power (Moshtaghi et al., 2018). This aspect motivated Thompson et al. (1986) to first restrict weights in DEA (Thompson et al., 1986). In the efficiency analysis of Atici et al. (2015), the number of factors is larger than the number of DMUs, and standard DEA models cannot discriminate between efficient and inefficient DMUs. Nevertheless, additional weight restrictions allow a meaningful calculation of efficiency (Atici et al., 2015).

The most common weight restrictions are absolute weight restrictions (AWR), assurance regions of type I (ARI) and type II (ARII). ARI and ARII impose restrictions on the ratios of the weights.

This paper describes the reasons for additional weight restrictions in the DEA and compares appropriate sources and implementations of different approaches. The structure of the paper is as follows. After the introduction, a literature overview of the origin of additional weight restrictions and application examples is provided. In the third section, the standard DEA models and weight restriction approaches are introduced and several artificial DMUs are used to illustrate the effects of the additional restrictions. Chapter four summarises the procedure for selecting additional restrictions, refers to appropriate literature, and provides guidance for the implementation of weight restrictions. The last section concludes.

2.2 Method Origin and a Literature Review

First, the early development of weight restrictions in DEA is presented to provide a better understanding of the original motivation. The chapter then outlines the broad scope of the additional restrictions in efficiency analysis by giving an overview of various publications that use additional weight restrictions. The publications were chosen either because they explain in detail their motivation for introducing weight restrictions, the methodology behind them, or because they are among the most influential in the field.

2.2.1 Method Development

Thompson et al. (1986) use DEA to support the siting of the Superconducting Super Collider (SSC) within nuclear physics facilities in Texas. Five out of six sites are efficient and the authors include additional weight restrictions to increase the discriminatory power of their analysis. Additional weight restrictions are calculated based on marginal rates of substitutions (to proxy environmental costs) to reveal which of the five technical efficient facilities is the most economically efficient. To allow weight variations, standard errors (based on cost estimations) determine upper and lower input weight bounds. Finally, the construction of the SSC was started in Waxahachie, the only remaining efficient site in the extended DEA analysis (Thompson et al., 1986).
Based on the problems in their previous study, Thompson et al. (1990) formulate and generalise their weighting approach by introducing relative restrictions on possible weights and calling them assurance regions. They distinguish between two types of assurance regions. Assurance region I (ARI) restrict input (output) weights independent of the output (input) weights and assurance region II (ARII) link input and output weights. In a case study, efficiency scores for Kansas farms are calculated. Using standard DEA models, the shares of efficient farms are relatively high. ARI based on various

information, e.g. cropland use patterns, historical cost and prices, crop yields,... decrease the number of efficient farms by up to 70%. The assurance regions allow the identification of extreme candidates among the efficient DMUs by successively tightening the restrictions. The weights assigned to the DMUs allow the DM to further distinguish between them (Thompson et al., 1990).

Dyson and Thanassoulis (1988) criticise that total weight flexibility in DEA can lead to some DMUs being assessed on only a small subset of the available inputs and outputs. In extreme cases, efficient DMUs can be efficient if the linear program assigns weights greater than zero to only one input and one output. Thus the efficiency scores would not reflect the DMUs' overall performance, but only individual aspects of their production processes. Additional weight restrictions are intended to decrease overestimated efficiency values, enable an economic interpretation of the weights, and reflect realistic production processes. The weights of the DMUs should still be able to deviate from their mean values within a certain range to maximise the efficiency of the DMUs as far as possible (Dyson et al., 1988).

Wong et al. (1990) describe the ignorance of some inputs or outputs by the DMUs caused by zero or close to zero weights as plainly unsatisfactory. Additional weight restrictions can avoid zero weights and enable the DEA to more realistically model the production processes. They propose restrictions for each input and output relative to the others, for example, based on a consensus between experts (Wong et al., 1990).

In most of the previous studies, the additional weights are derived from descriptive results or expert opinions. Cook et al. (1991) determine upper and lower restrictions using regressions to avoid inappropriately high or low weights. This approach is only feasible in single-input multi-output or single-output multi-input cases. The regression coefficients can serve as centres or as lower and upper bounds (Cook et al., 1991). However, the DM typically has to choose thresholds to decide how much weight variation they prefer, which is one of the biggest obstacles to the implementation of additional restrictions (Allen et al., 1997).

In a follow-up publication, Cook et al. (1992) distinguish between efficient DMUs by calculating absolute ranks of the efficient DMUs based on their ability to assign a balanced weights to their inputs and outputs. DMUs with relatively large or small weights are penalised. This approach favours balanced weight distributions. Thus, this is only meaningful if all input and output measures are roughly the same. Therefore inputs and outputs must be of a similar scale (Cook et al., 1992).

2.2.2 Applied Restrictions in Literature

Schaffnit et al. (1997) criticise that weights calculated with standard DEA models can be unreasonable (the weights do not equal the assumed production process), leading to an unrealistic frontier and an overestimation of efficiency. They argue that additional price information may assist in moving from technical to overall efficiency. The precise price information is seldom known, but weight bounds, weight interdependencies, or other managerial information may be available to tighten the PPS. As a case study, efficiency values are calculated for bank branches, and additional input restrictions are implemented based on average salaries. The tighter restrictions reduce the PPS and lower the average efficiency. An efficiency ranking is carried out based on various weight restriction thresholds and the resulting changes in efficiency (Schaffnit et al., 1997).

Olesen and Petersen (2002) also show that overall efficiency decreases when additional restrictions tighten the PPS. DMUs that remain efficient, independent of the selected thresholds, are identified as truly efficient, and they may serve as best practice benchmarks. DMUs that remain inefficient regardless of the thresholds are truly inefficient. DMUs showing efficiency score changes as a result of the selected thresholds should assess the reasons for their sensitivity to the thresholds (Olesen et al., 2002). An increase in discriminatory power is necessary if standard DEA models would not have a discriminatory effect due to a high number of outputs and a relatively low number of DMUs.

Sarrico and Dyson (2004) include VWR, ARI, and ARII based on descriptive results to calculate the institutional performance of universities. The restrictions enable a representation of preference structures and link the inputs to their corresponding outputs (Sarrico et al., 2004). Another efficiency analysis of universities is carried out by Kao and Hung (2008). They state that DMUs in standard DEA models often assign zero weights to unfavourable factors or select weights that may not reflect their real importance. Thus, additional weight restrictions, according to top university administrators, are incorporated in their analysis (Kao et al., 2008).

Amado and dos Santos (2009) calculate the efficiency of health care centres by including ARI and ARII based on their own opinions. These restrictions avoid unrealistic weights by restricting inputs (ARI) and improve the assumptions of the underlying production process by linking the output weights with their corresponding input weights (ARII) (Amado et al., 2009). Nevertheless, these restrictions may reflect simplified assumptions and unnecessarily restrict the PPS.

Cooper et al. (2009) improve the index of basketball players assessment used in the Spanish Basketball League by applying DEA on data of basketball players. Sets of ARI restrictions are implemented based on the opinions of the technical staff of the team Etosa Alicante. Basketball experts appreciated the efficiency evaluation approach of Cooper et al. (2009) as it allows flexible weights but within customary limitations and allows to further discriminate between players who are efficient in standard DEA models (Cooper et al., 2009).

Khalili et al. (2010) criticise the presence of zero weights in standard DEA models and propose an approach based on the determination of suitable restrictions due to a trade-off information approach. Nonlinear DEA models are used to calculate the efficiency of secondary schools. An expert is asked to provide values for variables of the changed schools to make the average school and the changed schools similar in performance. The expert opinions are incorporated as ARI and ARII in the DEA models. The approach of Khalili et al. (2010) is based on rather subtle differences between changes of the inputs and outputs. Only one expert has been consulted, her preferences cannot be reproduced, and opinions may vary between experts. However, this is one of the most comprehensible publications and provides an extensive overview of weight restrictions in DEA (Khalili et al., 2010).

Mecit and Alp (2013) prevent zero weights by incorporating weight restrictions based on correlation coefficients between and among inputs and outputs. The additional restrictions should resemble the production process and are described as ARIII. The authors claim that their approach is objective as no expert opinions are used. Thus, the results are more realistic, and a more balanced distribution of weights is obtained. In a case study, ARIII greatly reduces the number of zero weights, some are still present in each output, and yields a more balanced distribution of weights and avoidance of subjective information (Mecit et al., 2013). The ARIII approach more closely resembles the production process and reduces the number of zero weights. However, existing zero weights can still produce unreasonable results.

If only a few DMUs produce a large number of outputs, the efficiency scores determined by standard DEA models tend to be biased. Atici and Podinovski (2015) further refine an alternative efficiency measure first introduced by Podinovski (2004) to assess the efficiency of DMUs with output specialisations (Podinovski, 2004). Atici and Podinovski (2015) use production trade-offs between outputs based on expert opinions to increase the discriminatory power of their efficiency analysis. Their approach can discriminate between efficient and inefficient DMUs even if the number of DMUs is greater than the number of outputs (Atici et al., 2015).

Jain et al. (2015) reduce total weight flexibility in DEA to reflect management knowledge of inputs and outputs. Their multi-layer and multi-step approach incorporates different preferences (weighted based on the organisational hierarchy) and weight distributions of an unrestricted DEA model. The resulting economic efficiency discriminates more strongly between DMUs and reduces their weight variation (Jain et al., 2015).

In a case study, Ruiz et al. (2015) demonstrate that multiple DMUs that are efficient without additional weight restrictions are inefficient if expert opinions are included in the analysis. The authors conclude that the assurance regions provide additional insights into the efficiency calculation process by preventing DMUs from choosing an inappropriate weight distribution and reducing potential efficiency overestimates (Ruiz et al., 2015).

Theodoridis and Ragkos (2015) use additional weight restrictions to calculate the efficiency of dairy farms in Greece. They implement ARI based on a Cobb-Douglas production function. The authors strive to reduce unreasonable weight bandwidths and to neglect zero weights (Theodoridis et al., 2015). The model could underestimate the efficiency of the diary farms. The varying results of the inefficient DMUs for alternative thresholds are not reported.

Basso et al. (2018) measure the efficiency of the municipal museums of Venice. ARI restrictions based on expert opinions are included to increase the discriminatory power of the analysis and to avoid Zero weights. The restrictions are set rather strictly, which quite limits the number of efficient museums (Basso et al., 2018). By selecting different upper and lower limits, the impact analysis of the assurance regions allows the DMs to identify the strengths and weaknesses of each DMU.

Table 2.1 – Weight restrictions in the literature

Authors	Analysis	Restriction types	Sources	Motivation
Thompson et al. (1986)	Physics facilities	ARI	Additional data	Discrimination
Dyson et al. (1988)	Rates Departments	AWR	Regression coefficients	Zero weights
Charnes et al. (1990)	Commercial banks	Cone-ratio	Expert opinions	Discrimination
Thompson et al. (1990)	Farms	ARI	Additional data	Discrimination
Wong et al. (1990)	Hypothetical university departments	ARI	Own opinions	Zero weights
Cook et al. (1991)	Highway maintenance patrols	AWR	Regression coefficients	Unreasonable weights
Schaffnit et al. (1997)	Bank branches	ARI	Unrestricted weights	Unreasonable weights
Ray et al. (1998)	State-owned enterprises	ARI and ARII	Market prices	Allocative efficiency
Olesen et al. (2002)	Hospitals	ARI	Regressions	Allocative efficiency
Sarrico et al. (2004)	Universities	AWR, ARI and ARII	Own opinions	Unreasonable and zero weights
Camanho et al. (2005)	Bank branches	ARI	Price information	Unreasonable weights
Kao et al. (2008)	University departments	ARI	Expert opinions	Unreasonable and zero weights
Amado et al. (2009)	Health centres	ARI and ARII	Expert opinions	Unreasonable weights
Cooper et al. (2009)	Basketball players	ARI	Expert opinions	Zero weights and discrimination

Authors	Analysis	Restriction types	Sources	Motivation
Khalili et al. (2010)	Secondary schools	ARII	Expert opinions	Zero weights
Saen (2010)	Power plants	ARI	Expert opinion	Zero weights
Kong et al. (2012)	Business schools	ARI	Expert opinions	Unreasonable and zero weights
Mecit et al. (2013)	Robots	ARI, ARII, and ARIII	Regression coefficients	Unreasonable and zero weights
Atici et al. (2015)	Wheat farms	ARI	Expert opinions	Discrimination
Jain et al. (2015)	Supplier selection process	Weighted distance function	Expert opinions and unbounded weights	Discrimination
Ruiz et al. (2015)	Universities	ARI	Expert opinions	DMs preference
Theodoridis et al. (2015)	Dairy farms	ARI	Regression coefficients and cost shares	Unreasonable and zero weights
Castelo Gouveia et al. (2016)	Health centres	AWR and ARI	Expert opinions	Zero weights
Cook et al. (2017)	Power plants	ARI	Expert opinions	DMs preference
Basso et al. (2018)	Museums	ARI	Expert opinions	Discrimination and zero weights

Allocative efficiency: Technical efficiency is considered to be insufficient and allocative or economic efficiency is preferred.

Zero weights: Zero weights should be prevented so that all factors are considered.

DMs preference: The analysis should reflect the preferences of one or more DMs.

Discrimination: To increase the discriminatory power of the analysis.

Unreasonable weights: Total weight flexibility can result in a relative wide weight discrepancy and is not intended.

2.3 DEA Models and Weight Restrictions

The chapter begins with the introduction of standard DEA models. Afterwards, additional weight restrictions are introduced, and their implementation is demonstrated.

2.3.1 Basic Model

Productivity is defined as the output to input ratio. Efficient production as the ability to minimise input given a certain amount of output or to maximise output given a certain amount of input.
An efficiency measurement, or score, is obtained by comparing the output-to-input ratio of a decision making unit (DMU) with the maximum observed ratio of similar DMUs. The most efficient DMU serves as a benchmark for the others. This definition leads to the following problem for DMU$_o$ (o denotes a specific DMU under consideration):

$$\frac{\text{Sum of weighted outputs}}{\text{Sum of weighted inputs}} = \frac{\sum\limits_{r=1}^{s} y_{ro}u_{ro}}{\sum\limits_{i=1}^{m} x_{io}v_{io}}. \tag{1}$$

Output r is given by y_{ro} and is weighted by u_{ro} ($r = 1, ..., s$). s equals the number of outputs. x_{io} is input i and its weight is given by v_{io} ($i = 1, ..., m$). m is the number of inputs. u_r and v_i are derived from the data and most likely vary between DMUs. The choices of u_r and v_i are associated with value judgements and are found mathematically, by a linear program, or derived from theory.[1] In the standard DEA models, weights are only restricted to be non-negative, so that each DMU reaches its maximum efficiency (Behr, 2015; Cooper et al., 2007).

The input-oriented CCR-model in envelopment form and matrix notation is[2]:

$$\begin{aligned} &\min_{\theta,\boldsymbol{\lambda}} \theta \\ &s.t. \; \boldsymbol{X\lambda} \leq \theta \boldsymbol{x}_o \\ &\quad\;\; \boldsymbol{Y\lambda} \geq \boldsymbol{y}_o \\ &\quad\;\; \boldsymbol{\lambda} \geq 0. \end{aligned} \tag{2}$$

[1] Note that scaling u_r with the scalar c and scaling v_i with the scalar k does not alter the efficiency score.

[2] Bold lower case symbols indicate vectors and bold capitalized symbols matrices.

$\boldsymbol{x}_o$ $(m \times 1)$ and $\boldsymbol{y}_o$ $(s \times 1)$ are the input and output vectors of DMU_o. $\boldsymbol{\lambda}$ is a vector of dimension $(n \times 1)$ and contains the intensity weights. n equals the number of DMUs. θ is the efficiency score. The linear input and output combinations span the Production Possibility Set (PPS) such that the input produces the output:

$$\begin{aligned} \text{PPS} &= \{(\boldsymbol{X}, \boldsymbol{Y}) | \boldsymbol{x} \geq \boldsymbol{X\lambda}, \boldsymbol{y} \leq \boldsymbol{Y\lambda}, \boldsymbol{\lambda} \geq 0\} \\ \boldsymbol{X} &= (\boldsymbol{x}_1, ..., \boldsymbol{x}_n); \boldsymbol{Y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_n); \boldsymbol{\lambda} \in \mathbb{R}^n. \end{aligned} \tag{3}$$

The efficient DMUs span the efficiency frontier. The intensity weights are unit specific and show the composition of the frontier reference point for the respective DMU. This synthetic DMU uses $\boldsymbol{\lambda X}$ to produce $\boldsymbol{\lambda Y}$. $\boldsymbol{X}$, with dimension $(m \times n)$, consists of the input vectors $\boldsymbol{x}_j$ $(j = 1 \ldots n)$ and $\boldsymbol{Y}$, with dimension $(s \times n)$, of the output vectors $\boldsymbol{y}_j$ (Pannu et al., 2010).

Førsund (2013) mathematically proves that $\boldsymbol{\lambda}$ is zero for inefficient DMUs. These DMUs are not part of the efficiency frontier, and therefore are not suitable references. Furthermore, weights are zero if slacks are present. If a DMU is efficient by producing just one output, its other weights are zero. Visualised, these DMUs are corner solutions (Førsund, 2013).

The dual multiplier form of the linear program (2) is more closely related to the aggregation problem:

$$\begin{aligned} \underset{\boldsymbol{v},\boldsymbol{u}}{max}\ \theta &= \frac{\boldsymbol{u}\boldsymbol{y}_o}{\boldsymbol{v}\boldsymbol{x}_o} \\ s.t.\ \frac{\boldsymbol{uY}}{\boldsymbol{vX}} &\leq 1 \\ \boldsymbol{v}, \boldsymbol{u} &\geq 0. \end{aligned} \tag{4}$$

$\boldsymbol{v}$ $(1 \times m)$ and $\boldsymbol{u}$ $(1 \times s)$ can be interpreted as shadow prices, virtual prices, variable multipliers or weights. The input prices do not directly impact the efficiency scores. They influence θ indirectly through the solution for the output weights and vice versa for the output-oriented problem. The products of shadow prices and inputs (outputs) are called virtual inputs (outputs). $\boldsymbol{v}^*$ and $\boldsymbol{u}^*$ are the most favourable weights and θ^* the resulting (technical) efficiency score (Cooper et al., 2011). Cooper et al. (2007) use the term "technical" efficiency to distinguish between technical or mathematical and economic or real efficiency. The latter includes additional information (e.g. on prices) or value considerations that limit the available PPS (Cooper et al., 2007).

The restrictions in model (4) ensure, that for all DMUs the weighted outputs must not exceed the weighted inputs and that all weights must be non-negative. The primal (2) and dual (4) programs are referred to as the envelopment and the multiplier problems. The multiplier problem represents a value-space where the determinations of $\boldsymbol{v}$ and $\boldsymbol{u}$ imply value judgements (Thanassoulis et al., 2004). It is the focus of discussion in this paper.

Assuming non-negative input prices ($v \geq 0$) as well as positive amounts of inputs ($x > 0$), and normalising the output of DMU$_o$ to unity, converts model (4) towards:

$$\begin{aligned} &\max_{\boldsymbol{v},\boldsymbol{u}}\ \theta = \boldsymbol{u}\boldsymbol{y}_o \\ &s.t.\ \boldsymbol{v}\boldsymbol{x}_o = 1 \\ &\quad -\boldsymbol{v}\boldsymbol{X} + \boldsymbol{u}\boldsymbol{Y} \leq 0 \\ &\quad \boldsymbol{v},\boldsymbol{u} \geq 0. \end{aligned} \tag{5}$$

In the output-oriented model, DMUs maximise the output for constant input:

$$\begin{aligned} &\max_{\eta,\boldsymbol{\mu}}\ \eta \\ &s.t.\ \boldsymbol{x}_o - \boldsymbol{X}\boldsymbol{\mu} \geq 0 \\ &\quad \eta\boldsymbol{y}_o - \boldsymbol{Y}\boldsymbol{\mu} \leq 0 \\ &\quad \boldsymbol{\mu} \geq 0. \end{aligned} \tag{6}$$

The dual is

$$\begin{aligned} &\min_{\boldsymbol{p},\boldsymbol{q}}\ \boldsymbol{p}\boldsymbol{x_o} \\ &s.t.\ \boldsymbol{q}\boldsymbol{y}_o = 1 \\ &\quad -\boldsymbol{p}\boldsymbol{X} + \boldsymbol{q}\boldsymbol{Y} \leq 0 \\ &\quad \boldsymbol{p},\boldsymbol{q} \geq 0. \end{aligned} \tag{7}$$

The input-oriented model (5) is obtained from the output-oriented model via $\boldsymbol{\lambda} = \frac{\boldsymbol{\mu}}{\eta}$ and $\theta = \frac{1}{\eta}$. Although the notation of the input and output weights differs in (5) and (7) to clarify the differences between the in- and output-oriented models, in the remaining paper $\boldsymbol{v}$ indicates input weights and $\boldsymbol{u}$ indicates output weights.

The choice of number and type of inputs, outputs, and of the DMUs implicitly includes value judgements. Increasing the number of variables and reducing the number of DMUs decreases the discriminatory power of the DEA while increasing the potential number of zero weights. If the number of DMUs is relatively low, the chance that each DMU can find an input (output) mix on which it performs well, relative to the other DMUs, increases. This also enhances the possibility that its mix is not directly comparable with the combination of the other DMUs. The higher the number of inputs and outputs, the higher the likelihood that each DMU can find a factor to focus on while ignoring all other variables (Thanassoulis et al., 2004).

Table 3.1 presents five artificial DMUs to demonstrate the weight selection processes. The two outputs y_1 and y_2 are normalised to the sole input x.[3] This normalisation assumes constant returns to scale (CRS). VRS do not allow proportional input and output changes, preventing such a simplification.

Table 2.2 – Example DMUs, efficiency scores, weights and intensity weights

	x	y_1	y_2	η^*	v^*	u_1^*	u_2^*	λ_A	λ_B	λ_D
A	1.000	2.000	0.750	1.000	1.000	0.500	0.000	1.000	0.000	0.000
B	1.000	1.000	1.750	1.000	1.000	0.000	0.571	0.000	1.000	0.000
C	1.000	0.700	1.500	1.167	1.000	0.000	0.571	0.000	1.000	0.000
D	1.000	1.500	1.500	1.000	1.000	0.222	0.444	0.000	0.000	1.000
E	1.000	1.000	0.750	1.667	1.000	0.400	0.267	0.333	0.000	0.667

Using models (6) and (7), B, D and A span the efficiency frontier in Figure 3.1 and are also efficient in Table 3.1. C and E are inefficient. E can become efficient by radially increasing its outputs, represented by the synthetic DMU $\mathrm{E^s}$. $\mathrm{E^s}$ is a combination of A's (0.333%) and D's technologies (0.667%). B is C's sole peer. A, B and C focus on one output only by setting the other output weight to zero. They are corner solutions. The weights of the DMUs (Table 3.1) represent four hyperplanes (defined by the constraints $\sum\limits_{r=1}^{s} u_r^* y_{rj} - \sum\limits_{i=1}^{m} v_i^* x_{ij} = 0$, for DMU j), corresponding to different rates of transformation or substitution. B and C are part of the same hyperplane. A, D, and E are idiosyncratic in that they span their own. The amount of output k that is decreased if output w is increased by one unit is the marginal rate of transformation $\frac{dy_k}{dy_w} = -\frac{u_w^*}{u_k^*}$. The marginal rates can be directly derived from the hyperplane equations. E.g. for D's hyperplane: $0.4y_{1D} + 0.267y_{2D} - 1x_D = 0$ and D's marginal rate of transformation: $-\frac{0.267}{0.4} = -0.667$. An increase of the second output by one unit implies a reduction of 0.667 units of the first output at constant inputs. A's, B's, and C's marginal rates of transformation are

[3] To ensure the discriminatory power of the DEA, Cooper et al. (2007) recommend: $n \geq max\{m \cdot s, 3 \cdot (m+s)\}$. Otherwise, the number of degrees of freedom is too low, and efficiency discrimination among DMUs cannot be assured (Cooper et al., 2007). Although the five DMUs do not fulfil this rule of thumbs ($m + s = 3$), they are selected heterogeneously enough to ensure meaningful efficiency discrimination.

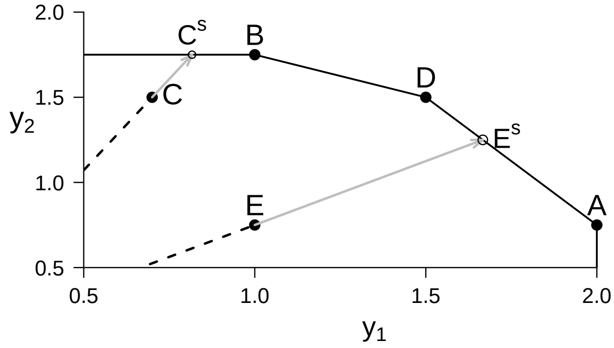


Figure 2.1 – Production possibility set of the example DMUs

undefined due to zero weights.

Figure 2.2 shows the DMUs' output weights space following the representations in Camanho et al. (2005) as well as Khalili et al. (2010). The restrictions $\sum_{r=1}^{s} u_r y_{rj} \leq \sum_{i=1}^{m} v_i x_{ij}$ determine the DMUs' lines. The capitalized letters are the DMUs positions given the weights calculated by the DEA. The slopes are the output ratios of the DMUs ($\frac{\boldsymbol{y_1}}{\boldsymbol{y_2}}$), and the resulting intercepts are $\frac{1}{\boldsymbol{y_1}}$ for the first output weight and $\frac{1}{\boldsymbol{y_2}}$ for the second output weight. Note that the amount of produced outputs directly links

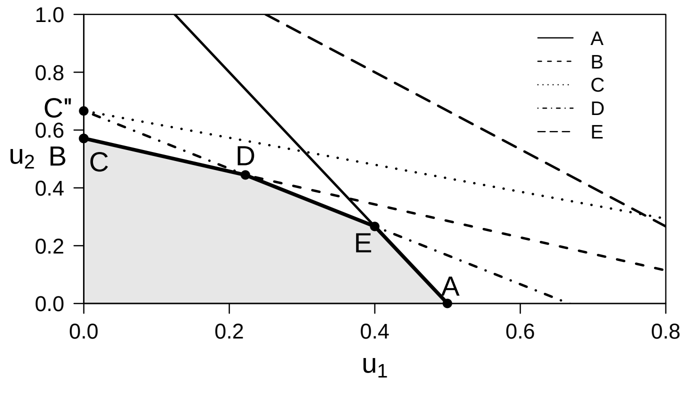


Figure 2.2 – Example DMUs' weighting space

to the weighting space through the first restriction of model (5) and that all outputs are normalised by $\boldsymbol{x}$. The feasible weights region (the grey area) of the minimisation problem (5) in the weighting space begins at the origin, is surrounded by the efficient boundary, and is limited to the non-negative space. From left to right, the efficiency frontier, represented by the thick line, is spanned by the hyperplane of B until it intersects with D's, by D's till its intersection with A's and by A's afterwards. A produces the highest amount of y_1 and is efficient for high values of u_1 and low values of u_2. B is efficient if its $u_1 \in [0, 0.222]$ and its u_2 can be set freely. D is efficient for every u_1 in the interval $[0.222, 0.4]$ if u_2 can be set accordingly. The intersection of A and D determines the upper bound in the interval (0.4). B dominates C and E. C and E are not part of the efficiency frontier. C'' is the closest point of C's hyperplane to the efficiency frontier but is not part of the feasible weights region. Thus, C's is inefficient, and the ratio of the distances ($\frac{\overline{0C''}}{\overline{0C}} = 1.167$) equals C's efficiency score (see Table 3.1). Since no further restrictions influence the weight selection of the DMUs, they can achieve their highest relative technical efficiency.

2.3.2 Weight Restriction Types

In basic DEA models, restrictions are applied to all DMUs so that they can only select weights that are feasible for all DMUs, and none can achieve an efficiency above the maximal efficiency threshold (typical one) (Wong et al., 1990). However, additional weight restrictions can be set for all DMUs or only for subsets. The latter is only feasible if the DMUs can be grouped according to additional information. This contradicts the DEA's underlying assumption that DMUs must be comparable. DMU-specific restrictions can result in infeasible weight combinations for the other DMUs. Therefore, restrictions that treat all DMUs equally are usually preferred (Sarrico et al., 2004).

2.3.2.1 Absolute Restrictions

AWR restrict the weights to vary within set boundaries. The restrictions can limit each variable individually or all variables together:

$$\begin{aligned} b_i^{lv} \leq v_i \leq b_i^{uv} \quad (i = 1, \dots m) \\ b_r^{lu} \leq u_r \leq b_r^{uu} \quad (r = 1, \dots s). \end{aligned} \tag{8}$$

Equation (8) provides examples of AWR that equally restrict the weights for all DMUs and may differ between inputs and outputs. The lower (b_i^{lv}) and upper bounds (b_i^{uv})

restrict the input weights and lower bound (b_r^{lu}) and the upper bound (b_r^{uu}) the output weights. Not all weights must be restricted, but all weights must remain non-negative. The inclusion of AWR for one or more input (output) weights affects all weights because the model interdependently links them through the restrictions. If the boundaries are too strict, the model becomes infeasible. Furthermore, the calculated efficiency scores may underestimate true technical efficiencies (Thanassoulis et al., 2004).

Table 2.3 shows the results of implementing the AWR $u_1 \leq 0.3$ in the linear program (7). Figure 2.3 provides the limited feasible weights region in the possible weighting space. The program including the weight restrictions is:

$$\begin{aligned} &\min_{\boldsymbol{v},\boldsymbol{u}} \ \boldsymbol{v}\boldsymbol{x_o} \\ &s.t.\ \boldsymbol{u}\boldsymbol{y}_o = 1 \\ &\quad -\boldsymbol{v}\boldsymbol{X} + \boldsymbol{q}\boldsymbol{Y} \leq 0 \\ &\quad u_1 \leq 0.3 \\ &\quad \boldsymbol{v}, \boldsymbol{u} \geq 0. \end{aligned} \tag{9}$$

Table 2.3 – Example DMUs using AWR, efficiency scores and weights

	η^*	v^*	u_1^*	u_2^*
A	1.250	1.000	0.240	0.427
B	1.000	1.000	0.000	0.571
C	1.167	1.000	0.000	0.571
D	1.000	1.000	0.222	0.444
E	1.933	1.000	0.155	0.483

Compared to the calculations without AWR, A's hyperplane is no longer part of the efficiency frontier, and E's preferred weight distribution is no longer feasible. Therefore, both DMUs have to choose different weights and become less efficient. The average inefficiency is higher than calculated by the unrestricted model (1.270 to 1.167), zero weights are excluded for u_1 and A's marginal rates of substitution and transformation are interpretable. B's and C's marginal rates of substitution and transformation are not interpretable.

AWR are easy to implement, mostly arbitrary set, rarely used in literature, and do not guarantee meaningful marginal rates of transformation and substitution for all DMUs (Olesen et al., 1996). Appropriate ARI can overcome the latter shortcoming.

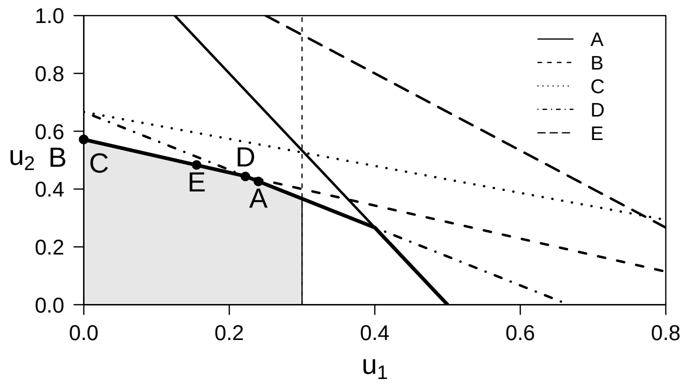


Figure 2.3 – Example DMUs' weighting space, including AWR

2.3.2.2 Relative Restrictions

ARI additively or relatively link the input or output weights. Since the ratios of input (output) weights reflect the marginal substitution rate (transformation), ARI should be preferred to AWR when the information is available.

ARI can correspond to any conceivable relationship and typically resemble relative linkages of input or output weights (Kao et al., 2008; Theodoridis et al., 2015):

$$\begin{aligned} b^{lv}_{i,k} &\leq \frac{v_i}{v_k} \leq b^{uv}_{i,k} \quad (i = 1, \dots, m), (k = 1, \dots, m), i \neq k \\ b^{lu}_{r,g} &\leq \frac{u_r}{u_g} \leq b^{uu}_{r,g} \quad (r = 1, \dots, s), (g = 1, \dots, s), r \neq g. \end{aligned} \tag{10}$$

Such ARI do not necessarily prevent zero weights as $b^{lv}_{i,k}v_k - v_i \leq 0$ and $v_i - b^{uv}_{i,k}v_k \leq 0$ are fulfilled for $v_i = 0$ and $v_k = 0$ and similarly for the outputs. Alternatively, AR1 can additively link the weights so that they are less or greater than a specific threshold (Cooper et al., 2009). For example:

$$\begin{aligned} \sum_{i=1}^{m} v_i &\geq b^{lv}_i \\ \sum_{i=1}^{m} v_i &\leq b^{uv}_i \\ \sum_{r=1}^{s} u_r &\geq b^{lu}_r \\ \sum_{r=1}^{s} u_r &\leq b^{uu}_r. \end{aligned} \tag{11}$$

Table 2.4 shows the implementation of the following ARI for the example DMUS:

$$0.2 \leq \frac{u_1}{u_2} \leq 1.2 \tag{12}$$

The inequalities can be rearranged to:

$$0.2u_2 \leq u_1 \leq 1.2u_2, \tag{13}$$

and further split into two inequalities:

$$0.2u_2 - u_1 \leq 0 \tag{14}$$

$$-1.2u_2 + u_1 \leq 0. \tag{15}$$

These inequalities can be implemented directly in the linear program.

Table 2.4 – Example DMUs using ARI, efficiency scores and weights

	η^*	v^*	u_1^*	u_2^*
A	1.048	1.000	0.364	0.303
B	1.000	1.000	0.103	0.513
C	1.189	1.000	0.103	0.513
D	1.000	1.000	0.222	0.444
E	1.692	1.000	0.364	0.303

The additional ARI relatively links both outputs, increases the average efficiency score compared to the unrestricted model (1.186 to 1.167), excludes zero weights, and provides interpretable marginal rates of substitution and transformation.

In contrast to ARI, ARII define relationships between input and output weights. They can resemble any relationship between two or more input and output weights. Amado et al. (2009) and Khalili et al. (2010) use the following additive ARII:

$$\sum_{r=1}^{s} u_r a_r + \sum_{i=1}^{m} v_i c_i \leq b^u. \tag{16}$$

At least one parameter of $\boldsymbol{a}$ $(s \times 1)$ and $\boldsymbol{c}$ $(m \times 1)$ must be non-zero. Otherwise, equation (16) would become an ARI restriction. The ARII can be directly included in

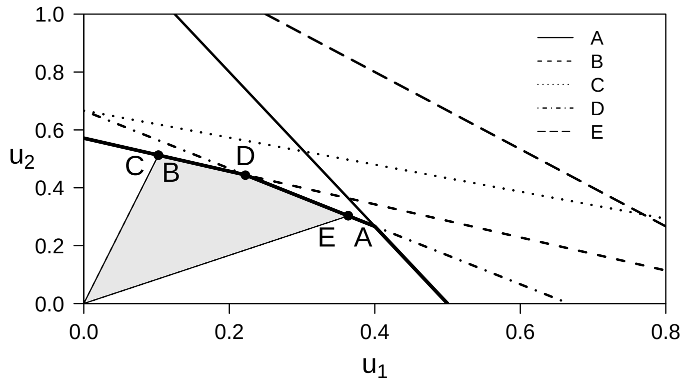


Figure 2.4 – Example DMUs' weighting space, including ARI

the output-oriented CCR model (5)

$$
\begin{aligned}
\max_{\boldsymbol{v},\boldsymbol{u}} \; & \theta = \boldsymbol{u}\boldsymbol{y}_o \\
s.t. \; & \boldsymbol{v}\boldsymbol{x}_o = 1 \\
& -\boldsymbol{v}\boldsymbol{X} + \boldsymbol{u}\boldsymbol{Y} \leq 0 \\
& \boldsymbol{u}\boldsymbol{a} - \boldsymbol{v}\boldsymbol{c} \leq \boldsymbol{b}^u \\
& \boldsymbol{v}, \boldsymbol{u} \geq 0.
\end{aligned}
\tag{17}
$$

Also, ARII may appear as trade-offs between input and output to indicate changes in outputs that would occur due to changes in inputs and vice versa (Thanassoulis et al., 2004):

$$
b_{i,r}^l \leq \frac{v_i}{u_r} \leq b_{i,r}^u \quad (i = 1, \ldots, m), (r = 1, \ldots, s). \tag{18}
$$

Table 2.5 – Additional restrictions

Example	Type	Restrictions
R.1	ARII	$0.15 \leq \frac{u_1}{v} \leq 0.5$ $0.3 \leq \frac{u_2}{v} \leq 0.55$
R.2	ARII	$0.1 \leq \frac{u_1}{v} \leq 0.3$ $0.1 \leq \frac{u_2}{v} \leq 0.35$
R.3	ARII	$0.4 \leq \frac{u_1}{v} \leq 0.5$ $0.4 \leq \frac{u_2}{v} \leq 0.6$

Table 2.5 contains several weight restrictions and Table 2.6 the resulting efficiency scores and weights. The results of R.3 are not reported as the additional restrictions determine possible weighting space outside the feasible weights region. This renders

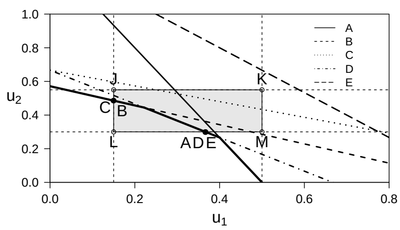
the linear program infeasible, and the relative efficiency cannot be calculated. Figures

Table 2.6 – Efficiency scores and weights for R.1 and R.2

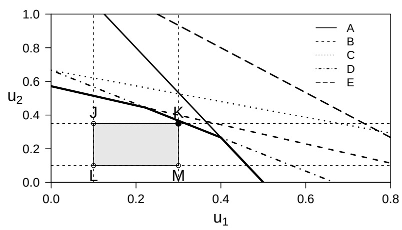
	R.1					R.2				
	θ^*	v	u_1	u_2	$\frac{u_1}{u_2}$	θ^*	v	u_1	u_2	$\frac{u_1}{u_2}$
A	1.043	1.000	0.367	0.300	1.222	1.159	1.000	0.300	0.350	0.857
B	1.000	1.000	0.150	0.486	0.309	1.096	1.000	0.300	0.350	0.857
C	1.200	1.000	0.150	0.486	0.309	1.361	1.000	0.300	0.350	0.857
D	1.000	1.000	0.367	0.300	1.222	1.026	1.000	0.300	0.350	0.857
E	1.690	1.000	0.367	0.300	1.222	1.778	1.000	0.300	0.350	0.857

2.5a to 2.5c elucidate the examples presented in Table 2.5. In Figure 2.5a, thin dashed lines represent the ARII of R.1 and span the reduced weighting space in rectangle JKLM.[4] Since the hyperplane of A is no longer feasible, A becomes inefficient. While

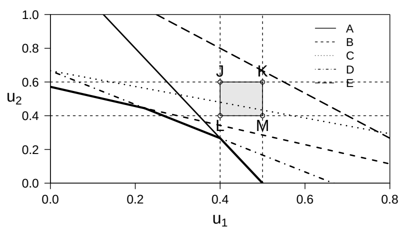
Figure 2.5 – Weighting spaces



(a) R.1



(b) R.2



(c) R.3

C and B decrease u_2 and increase u_1, A and E select a production process that yields a higher marginal rate of transformation. D selects the same hyperplane as A and E, adopting the same technology. The limited weight selection possibilities increase mean inefficiency from 1.167 to 1.187 compared to the unrestricted CCR model.

[4] In this example all inputs are normalized to unity so that the ARII are AWR. As AWR, the restrictions of R.1 are: $0.15 \leq u_1 \leq 0.5$ and $0.3 \leq u_2 \leq 0.55$.

In R.2 (Figure 2.5b) is the possible weighting space below the efficiency frontier and part of the feasible region. None of the DMUs is evaluated as efficient by the linear program, while D is calculated as the least inefficient DMU (its hyperplane is the closest to the assurance region). All DMUs are calculated so that they are in the position of K and for better clarity their letters are not included in Figure 2.5b. As efficiency is a relative concept, efficiency scores can be calculated relative to D (Podinovski, 1999). A's relative efficiency equals $\frac{1.159}{1.026} = 1.13$, B's $\frac{1.096}{1.026} = 1.068$, C's $\frac{1.361}{1.026} = 1.327$, D's $\frac{1.026}{1.026} = 1$, and E's $\frac{1.778}{1.026} = 1.733$. D is the only efficient DMU in R.2.

2.4 Additional Weight Restriction Selection

Data-driven additional weight restrictions allow accounting for additional information that is not input or output. However, market prices or other suitable supplementary information are rarely available. Most additional restrictions are based on expert opinions or are determined by statistical methods. Figure 2.6 illustrates the procedure for implementing additional restrictions. Most additional weight restrictions implement

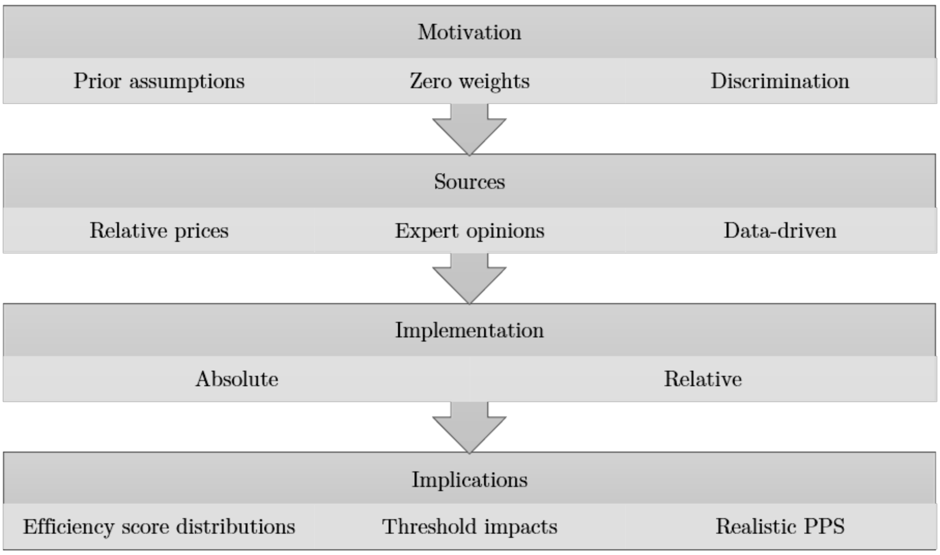


Figure 2.6 – Additional weight restrictions implementation procedure

prior assumptions (Kong et al., 2012), avoid total specification (Doyle et al., 1994) or should enable further discrimination between the DMUs (Thompson et al., 1990). If prior knowledge of the relationships between factors is available, this should be taken into account, as the weights of DMUs without limited PPS may be unreasonable (Kao et al., 2008). Standard DEA models maximise efficiency by assigning weights close to zero or zero to factors which would render the DMUs more inefficient. Additional weight restrictions can ensure that DMUs are evaluated with their full set of available inputs and outputs (Dyson et al., 1988). Table 2.7 outlines the most common sources

Table 2.7 – Examples for weight restriction sources

Data-driven	Expert opinions
Empirical results (Thompson et al., 1990)	AHP (Ruiz et al., 2015)
Correlation coefficients (Mecit et al., 2013)	Statements (Schaffnit et al., 1997)
Previous DEAs (Ramón et al., 2010)	Pairwise comparisons (Cooper et al., 2009)
Balanced weights (Dimitrov et al., 2010)	Relative importance (Kao et al., 2008)
Estimated regression coefficients (Theodoridis et al., 2015)	Swing method (Castelo Gouveia et al., 2016)

for additional weight restrictions. Expert opinions are the most frequently used source. Typically, their opinions are aggregated (Schaffnit et al., 1997) or captured through simple pairwise comparisons (Cooper et al., 2009). Alternatively, the expert opinions are implemented using a pairwise comparison technique like the AHP or swing method (Ruiz et al., 2015; Castelo Gouveia et al., 2016).[5] Databased weight restrictions are identified based on empirical findings (Sarrico et al., 2004), results of an unrestricted DEA (Ramón et al., 2010), stochastic models (Dyson et al., 1988) or alternative DEA models (Jain et al., 2015). Valued market prices are rarely available for factors but are well suited to impose additional restrictions (Joro et al., 2004).

The appropriate weight restriction type depends on the available data and the DM's motivation. Absolute restrictions, for example, can help to exclude zero weights by implementing an absolute lower bound for each weight (Castelo Gouveia et al., 2016). Absolute restrictions are rigid and can make the linear program infeasible (Dyson et al., 1988). Relative restrictions are more flexible as they reflect relationships between factors. They express marginal substitution or transformation rates (Atici et al., 2015) or trade-offs between inputs and outputs (Khalili et al., 2010). Additional weight restrictions limit the PPS and reduce weight flexibility, allowing for more realistic

[5] The AHP, introduced by Saaty (1980), helps the DM to structure and simplify complex decisions by reducing them to pairwise comparisons (Ruiz et al., 2015).

The swing method approach allows the implementation of restrictions based on expert opinions. The factors are ordered according to the preferences of an expert, and boundaries of their ratios are constructed. Additional limits are set for the first and last ranked weights to prevent zero weights (Castelo Gouveia et al., 2016).

weight distributions. Therefore, special attention must be devoted to the selection of the appropriate threshold values. In most publications, weight flexibility thresholds are set by the DM, to allow, e.g. 20% or 50% deviations from the value suggested by the source (Schaffnit et al., 1997). Theodoridis et al. (2015) state that their threshold (25%) is chosen because only a twice as high variation significantly changes the number of efficient DMUs. Most studies do not report detailed results for alternative thresholds that would be necessary to validate the impacts of the allowed deviations. The DM must choose appropriate weight restriction thresholds according to their motivation and the desired weight flexibility. Without weight flexibility, the linear program cannot maximise the efficiency of the DMUs and the model would become similar to that of the economic model of Joro et al. (2004).

The additional restrictions alter the efficiency values, reveal specified or extreme DMUs (Thompson et al., 1990) and limit the PPS to a more realistic representation of the targeted production processes (Wong et al., 1990). The results of different threshold values show to what extent balanced weights are assigned and the volatility of the efficiency scores (Ruiz et al., 2015). It is, therefore, necessary to compare and interpret the efficiency scores for a range of thresholds.

2.5 Conclusion

Additional weight restrictions increase the discriminatory power of the models, can implement prior knowledge of the interdependencies between factors and can prevent zero weights or unreasonable weight distributions. For this purpose, the correct sources (e.g. experts) for weight restrictions must be identified, information (e.g. from AHP) collected and restrictions implemented (e.g. as an insurance region).

Standard DEA models maximise the efficiency of DMUs as long as the weights are not negative. Any restriction of the PPS reduces technical efficiency. Applying weight restrictions is, therefore, an intervention in the weight calculation of the models. If the DM reduces weight flexibility (e.g. by applying thresholds to reduce the available weights range), the resulting weight distribution can be further located from technical efficiency but may be closer to that suggested, for example, by an expert (Atici et al., 2015). The increasing discriminatory power differently affects the efficiency scores and weight distributions of the DMUs. This allows distinguishing between technically and economically efficient DMUs (Mecit et al., 2013).

Market prices are rarely available, but they are an appropriate source of additional weight restrictions as they are objective information. If no market prices are available, experts can provide sufficient information to include additional restrictions. If neither of the above two data sources is available, restrictions may be based on the data. For example, a standard DEA may provide information about the number of zero weights, and constraints may be based on the calculated weights. Alternatively, correlation or regression relationships between inputs and outputs can be used to impose meaningful weight restrictions.

AWR are the most straightforward restrictions. Included as lower bounds, they exclude zero weights and allow the calculation of marginal rates of substitution and transformation. AWRs are usually set arbitrarily and rarely used in literature, as information about lower bounds is rarely available. The inclusion of absolute lower bounds does not necessarily guarantee meaningful marginal rates of transformation and substitution as upper bounds are missing (Olesen et al., 1996).

Assurance regions are the most common types of restrictions in the literature and restrict the ratio of inputs or outputs or combinations thereof (Allen et al., 1997). They are unit variant and, if their implementation changes the efficiency of certain DMUs, can provide additional information (e.g., an overpayment of certain types of factors). Their disadvantages are that they increase the computational effort, may lead to underestimation of true efficiency, and if they are set too strictly, they can make the program infeasible (Basso et al., 2018).

As no “one additional restriction fits all models” exists, the DM must choose the appropriate models and restrictions, matching their analysis and the available data. If all prices and costs of the factors are precisely known, exact productivities can be calculated. If not all information is available, DEA models with weight restrictions are well suited to calculate meaningful economic efficiency scores.

3

Comparing the Slack-Based Maximum Measure of Efficiency: A Simulation Application

Comparing the Slack-Based Maximum Measure of Efficiency: A Simulation Application

Gerald Fugger

University Duisburg-Essen, 45141 Essen, Germany.

Abstract

Radial Data Envelopment Analysis (DEA) models have been widely used in various applications. They are either units invariant or translation invariant, may overestimate technical efficiency by underestimating inefficiency, and distinguish between input- or output-orientation. The additive (ADD) model is non-radial and provides efficiency measures based on input and output slacks. Furthermore, the ADD model is unoriented and translation invariant. However, it is not units invariant and its inefficiency scores are not straightforward to interpret. Slack-based measurements (SBMs) are an extension of the ADD model. In the original SBM-Min model, relative slacks determine the efficiency measure. The SBM-Min is units invariant and allows substitution between different inputs and outputs. There are several extensions to the SBM-Min, including the recently introduced SBM-Max model that minimises inefficiency. This publication focuses on the SBM-Max, elucidates and compares it with radial and non-radial DEA models. The properties of the models are exemplified first for five artificial DMUs and then for 1,000,000 simulated DMUs.

Keywords: Data Envelopment Analysis, Slack-based measure, Model comparison
JEL Classification: C14 C52 C61
MSC2010 Classification: 46N10 62F07

3.1 Radial and Non-radial DEA Models

A variety of mathematical methods have been developed to calculate the efficiency of decision-making units (DMUs) that consume inputs and produce outputs. Radial Data Envelopment Analysis (DEA) models ignore input excess (input slacks) and output shortfalls (output slacks) and therefore can overestimate efficiency. Furthermore, the standard radial models pioneered by Charnes et al. (1978) (CCR model) and extended by Banker et al. (1984) to include the assumption of variable returns to scale (BCC model) distinguish between input- (minimising input and keeping output constant) and output-orientation (maximising output and keeping input constant) (Charnes et al., 1978; Banker et al., 1984). In radial models, DMUs can only become efficient by proportional contractions of their inputs or proportional extensions of their outputs (Avkiran et al., 2008).

Charnes et al. (1985) criticise that standard DEA models are either units invariant or translation invariant and that the slack component of radial DEA models is not units invariant.[1] Therefore, Charnes et al. (1985) introduced the non-radial additive (ADD) model. The ADD model considers input and output slacks simultaneously. Furthermore, unlike the radial models, the ADD model does not need any predefined input- or output-orientation.

Tone (2001) introduced the Slack-Based Measurement (in the following referred to as SBM-Min model) to improve the ADD model, which is not unit invariant and whose inefficiency results are not straightforward to interpret. The SBM-Min is units invariant and relative slacks determine the efficiency scores. The efficiency scores are monotone decreasing in each input and output slack and an efficiency score of one indicates an efficient DMU. By focusing on maximising or minimising inefficiency, slack-based models enable substitution within inputs or outputs. Their non-radial property allows non-proportional input reductions and non-proportional output increases (Avkiran et al., 2008). Furthermore, in standard radial models, weights can be set to zero or close to zero, resulting in marginal rates of substation and transformation that cannot always be defined (Angulo-Meza et al., 2002). Zero weights imply that certain inputs (outputs) are neglected in the efficiency calculation of the DMUs, which results in an overestimation of efficiency (Dyson et al., 1988; Joro et al., 2004). The SBM-Min and

[1] A DEA model is translation invariant if transforming the original data results in a new problem that has the same solution, see Ali et al. (1990, Theorem 4.6). A model is units invariant if the solutions do not depend on the units in which the in- and outputs are measured, see Coelli (1998).

its extensions overcome this problem by including data-based lower weight bounds to exclude zero weights.

Slack-based models differently interpret inefficiency. The SBM-Min maximises inefficiency. Its efficiency scores are the lower bound of slack-based efficiency. Tone (2017) developed an alternative model to calculate upper efficiency bounds, referred to as SBM-Max model. The SBM-Max model minimises inefficiencies while keeping the advantageous properties of the SBM-Min (Tone, 2017).

This publication focus on the SBM-Max, elucidates and compares the model with the most commonly used radial CCR and BCC models and the non-radial SBM-Min model. After the introduction, a literature overview provides the origin and the motivation for introducing slack-based measures and exemplifies applications. The models are formalised in the third section. Their differences are illustrated in the fourth section by five artificial DMUs to elaborate on the exact efficiency determinations. In the fifth section, inputs and outputs for 1,000 DMUs are generated 1,000 times to demonstrate the different interpretations of the inefficiency between the models on a larger scale as well as their computing effort. The last section concludes.

3.2 Literature Overview

In the following, some of the most relevant slack-based studies are presented. Slack-based approaches are quite popular in various efficiency fields, especially in transportation systems and environmental assessments (Bremberger et al., 2015; De Witte et al., 2017).

Sueyoshi et al. (2009) provide an overview of different radial- and non-radial DEA models. They apply the models to multiple data sets of artificial DMUs and state that the only disadvantage of the SBM-Min model is that it lacks the characteristic of homogeneity.[2] The ADD model is not unit invariant, or translation invariant, and the efficiency measure is not limited. Standard radial models are homogeneous, but their calculated efficiency scores do not react strictly monotonously to output or input changes. Furthermore, in radial-models, one input (output) cannot be substituted by another, as all inputs (outputs) can only be reduced (increased) to the same extent. Sueyoshi et al. (2009) conclude that no DEA model satisfies all desirable properties.

[2]Homogeneity implies that if all outputs are doubled, the output-based technical efficiency will double (Sueyoshi et al., 2009).

The SBM-Min is better than radial models in most applications because its efficiency measure is strict monotonic and its efficiency scores are between zero and one (Sueyoshi et al., 2009).[3]

Sueyoshi et al. (2012a) compare the efficiency scores of radial and non-radial DEA models of energy consumption, economic development, and environmental protection in Japanese prefectures. Their modified SBM-Min accounts for undesirable outputs and they prefer it over non-radial models because it prevents fully specialised DMUs by restricting zero weights (Sueyoshi et al., 2012a). Sueyoshi et al. (2012b) assess the environmental efficiency of U.S. coal-fired power plants. They apply multiple alternative scenarios (artificially decreasing and increasing the inputs and outputs) to calculate how the efficiency of the plants would change if they must decrease undesirable outputs while maximising their desirable outputs. Weak and strong disposable assumptions are implemented in an adjusted ADD model, which is quite similar to a SBM-Min.[4] The authors prefer a non-radial model to a radial model because they can more easily implement their extensions of desirable and undesirable outputs (Sueyoshi et al., 2012b).

Zhou et al. (2013) implement additional restrictions based on experts in an enhanced SBM-Min to calculate the efficiency of the power industry in Chinese provinces. The altered SBM-Min produces more reliable and reasonable results than any radial models (Zhou et al., 2013). Chang et al. (2014) extend the SBM-Min to include a weak disposability assumption into the efficiency evaluation of 27 global airlines. The weak disposability assumption increases the discriminatory power of the model and decreases overall inefficiency compared to an unaltered SBM-Min (Chang et al., 2014).

Overton et al. (2016) compare public school performance in highly unionised and less unionised states using radial- and non-radial models. Each group consists of eleven states. The combination of different models provides information about different input and output efficiency scores. The analysis reveals an adverse impact of unionisation on public education (Overton et al., 2016). The SBM-Min identifies more schools as inefficient than the radial models and together the models reveal which sources cause

[3] An efficiency measure is strict monotonic if it decreases monotonously in every input and output slack (Cooper et al., 2007).

[4] Weak and strong disposability: let $\boldsymbol{x}$ be a vector of inputs, $\boldsymbol{g}$ of the desirable outputs, $\boldsymbol{b}$ of the undesirable outputs, and $\boldsymbol{Z}$ is the production possibility set: $\boldsymbol{Z} = \{(\boldsymbol{x}, \boldsymbol{g}, \boldsymbol{b}) | \boldsymbol{x} \text{ can produce } (\boldsymbol{g}, \boldsymbol{b})\}$. Weak disposability of $\boldsymbol{g}$ and $\boldsymbol{b}$: if $(\boldsymbol{x}, \boldsymbol{g}, \boldsymbol{b}) \in \boldsymbol{Z}, 0 \leq \boldsymbol{g}' \leq \boldsymbol{g}$ and $\boldsymbol{x}' \geq \boldsymbol{x}$ then $(\boldsymbol{x}', \boldsymbol{g}', \boldsymbol{b}) \in \boldsymbol{Z}$. Strong disposability of $\boldsymbol{x}$ and $\boldsymbol{g}$: $(\boldsymbol{x}, \boldsymbol{g}, \boldsymbol{b}) \in \boldsymbol{Z}$ and $k \in [0, 1]$ then $(\boldsymbol{x}, k\boldsymbol{g}, k\boldsymbol{b}) \in \boldsymbol{Z}$ (Färe et al., 2004; Kuosmanen et al., 2009).

the inefficiencies.

The SBM-Min calculates the lowest efficiency scores by maximising inefficiencies and according to Zhu et al. (2018) unrealistically low efficiencies result in most applications. Andreu et al. (2014) propose several slack based measurement approaches that require additional statistical methods such as cluster analysis to approximate upper efficiency scores. Tone (2015) introduces the SBM-Max model that minimises inefficiency without additional statistical methods. In the first step, a SBM-Min is calculated. Afterwards, for each inefficient DMU local reference sets are defined. The local reference sets are based on distances between the efficient DMUs and the inefficient DMU under consideration. In the next step, two non-radial models are calculated for each inefficient DMU and each of its reference sets to obtain its optimal slacks and to project the solutions onto the efficiency frontier. The closest reference point on the efficiency frontier is identified. This final solution maximises efficiency by minimising inefficiency (Tone, 2015).

Johnes et al. (2017) use the CCR model, the SBM-Min, and the SBM-Max to measure the efficiency of 118 higher education institutions in England. The authors conclude that each model presents a different but reasonable way of measuring efficiency. Interestingly, SBM-Max efficiency scores are higher correlated with the CCR model results than with the SBM-Min efficiency scores. However, the correlation structure depends on the underlying data and varies depending on the application (Johnes et al., 2017).

[5] Bold lower case symbols indicate vectors and bold capitalized symbols matrices.

3.3 The DEA Models

In DEA, each DMU uses inputs ($\boldsymbol{x}$) to produce outputs ($\boldsymbol{y}$).[5] The output-oriented BCC model for DMU$_o$ (o denotes a specific DMU under consideration) is defined as:

$$\begin{aligned}
\min_{\eta,\boldsymbol{v},\boldsymbol{u}} \quad & \eta = \boldsymbol{v}\boldsymbol{x}_o - u_0 \\
\text{subject to } & \boldsymbol{u}\boldsymbol{y}_o = 1 \\
& \boldsymbol{v}\boldsymbol{X} - \boldsymbol{u}\boldsymbol{Y} - u_0 \geq 0 \\
& \boldsymbol{v} \geq 0 \\
& \boldsymbol{u} \geq 0 \\
& u_0 \text{ free in sign.}
\end{aligned} \tag{19}$$

η^* ($[1,\infty]$) denotes the solution to the minimisation problem. For convenience, we define $\theta^* = \frac{1}{\eta^*}$ ($[0,1]$). A DMU is efficient, only if $\eta^* = \theta^* = 1$, otherwise it is inefficient. Efficient DMUs are part of the reference set, span the efficiency frontier, and serve as benchmarks for the inefficient DMUs. $\boldsymbol{u}$ and $\boldsymbol{v}$ are the m input and s output weights. The scalar u_0 is free in sign and implements the assumption of variable returns to scale.

The non-radial ADD model considers input excess ($\boldsymbol{s}^-$) and output shortfalls ($\boldsymbol{s}^+$) simultaneously, combines the input- and output-orientations, and discriminates entirely between efficient and inefficient DMUs. The unoriented ADD model for DMU$_o$ is:

$$\begin{aligned}
\max_{\boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+} \quad & \psi = \boldsymbol{e}\boldsymbol{s}^- + \boldsymbol{e}\boldsymbol{s}^+ \\
s.t. \quad & \boldsymbol{X}\boldsymbol{\lambda} + \boldsymbol{s}^- = \boldsymbol{x}_o \\
& \boldsymbol{Y}\boldsymbol{\lambda} - \boldsymbol{s}^+ = \boldsymbol{y}_o \\
& \boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq 0 \\
& \boldsymbol{e}\boldsymbol{\lambda} = 1.
\end{aligned} \tag{20}$$

$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is the input matrix and $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ the output matrix, respectively. n is the number of DMUs and $\boldsymbol{e}$ is a vector of ones of an appropriate length. The DMU specific intensity weights $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$, with $\boldsymbol{\lambda} \in \mathbb{R}^n$, determine synthetic efficient DMUs. The restriction $\boldsymbol{e}\boldsymbol{\lambda} = 1$ implements the assumption of variable returns to scale. A synthetic DMU with the coordinates $(\hat{\boldsymbol{x}}_o, \hat{\boldsymbol{y}}_o)$ is the reference point for the inefficient DMU$_o$ on the efficiency frontier. Inefficient DMUs can improve efficiency by reducing slacks until they intersect with the efficiency frontier:

$$\begin{aligned} \hat{\boldsymbol{x}}_o &\leftarrow \boldsymbol{x}_o - \boldsymbol{s}^{-*} \\ \hat{\boldsymbol{y}}_o &\leftarrow \boldsymbol{y}_o + \boldsymbol{s}^{+*} \end{aligned} \tag{21}$$

(Banker et al., 1984; Cooper et al., 2007; Behr, 2015). A DMU is ADD-efficient ($\psi^* = 0$) if and only if it is radial efficient and does not have any slacks. The ADD model is translation invariant and can easily be extended with additional assumptions (Cooper et al., 2007). The ADD model's efficiency scores are the sum of the absolute slacks and have no upper bound and the ADD model is units variant.

While the SBM-Min is translation invariant too, it is units invariant as its measure is monotone decreasing in each input and output slack. Furthermore, the efficiency scores are bounded within zero and one.

By using a positive scalar variable t the unoriented SBM-Min for DMU$_o$ is given by

$$\begin{aligned} \underset{\boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+}{min} \quad & \tau = t - \frac{1}{m}\frac{t\boldsymbol{s}^-}{\boldsymbol{x}_o} \\ s.t. \quad & 1 = t + \frac{1}{s}\frac{t\boldsymbol{s}^+}{\boldsymbol{y}_o} \\ & \boldsymbol{X}(t\boldsymbol{\lambda}) + t\boldsymbol{s}^- = t\boldsymbol{x}_o \\ & \boldsymbol{Y}(t\boldsymbol{\lambda}) - t\boldsymbol{s}^+ = t\boldsymbol{y}_o \\ & \boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq 0, t > 0 \\ & \boldsymbol{e\lambda} = 1. \end{aligned} \tag{22}$$

A DMU is efficient if no input excess and no output shortfalls are present ($\tau^* = 1$). An inefficient DMU can become efficient by reducing its slacks. The SBM-Min accounts for technical inefficiencies and slacks, thus, $\theta^* \geq \tau^*$. Following Cooper et al. (2007), the dual of model (22) is

$$\begin{aligned}
\max_{\xi,\boldsymbol{v},\boldsymbol{u}} \quad & \xi \\
s.t. \quad & \xi + \boldsymbol{v}\boldsymbol{x}_o - \boldsymbol{u}\boldsymbol{y}_o - u_0 = 1 \\
& -\boldsymbol{v}\boldsymbol{X} + \boldsymbol{u}\boldsymbol{Y} - \boldsymbol{e}u_0 \leq 0 \\
& \boldsymbol{v} \geq \frac{1}{m}\left(\frac{1}{\boldsymbol{x}_o}\right) \\
& \boldsymbol{u} \geq \frac{\xi}{s}\left(\frac{1}{\boldsymbol{y}_o}\right) \\
& u_0 \text{ free in sign.}
\end{aligned} \tag{23}$$

The duality characteristics ($\tau^* = \xi^*$) apply, the lower bounds of the weights prevent zero weights. Increasing the number of variables does not necessarily decrease the discriminatory power of the SBM-Min, unlike in radial models. Contrary to the CCR model, all inputs and outputs are taken into consideration by the model. No DMU is evaluated as efficient by focusing on subsets of the inputs or outputs while ignoring the others (Thanassoulis et al., 2004). For some basic extensions see Cooper et al. (2007).

The SBM-Min measures the maximal distance of each DMU from the efficiency frontier. It calculates the lower bounds of slack-based efficiency scores. Tone (2015) introduces the SBM-Max model to approximate an upper slack-based efficiency bound. The objective functions in the following equations are not represented in matrix notation to improve intelligibility.

In the first step of the SBM-Max, model (22) is solved and its results are further used. R^{eff} consists of the efficient DMUs ($R^{eff} = \{j|\tau_j = 1, j = 1, ..., n\}$). The local reference set of an inefficient DMU$_o$ ($\tau_o < 1$) consists of all its efficient peers: $R_o^{loc} = \left\{j|\lambda_j^* > 0, j = 1, ..., n\right\}$. $\boldsymbol{\lambda}^*$ are the intensity weights obtained from the first step.

In the first iteration of the second step, the set R_o^{loc} comprises all peers of DMU$_o$. In the second iteration, R_o^{loc} consists of the closest to DMU$_o$ located efficient DMU. In the third iteration, the set consists of the two closest efficient DMUs, and so on. It should be noted that the closest DMU may not necessarily be a peer for DMU$_o$. The distances (d_{oz}) are the sum of the absolute differences between the inputs and outputs of the efficient DMU$_z$ and DMU$_o$ relative to the respective inputs and outputs of DMU$_o$:

$$d_{oz} = \sum_{i=1}^{m} \frac{|x_{iz}^{eff} - x_{io}|}{x_{io}} + \sum_{r=1}^{s} \frac{|y_{rz}^{eff} - y_{ro}|}{y_{ro}}. \tag{24}$$

The first program maximises the slacks:

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+} \quad & \frac{1 - \frac{1}{m}\sum\limits_{i=1}^{m} \frac{s_i^-}{x_{io}}}{1 + \frac{1}{s}\sum\limits_{r=1}^{s} \frac{s_r^+}{y_{ro}}} \\ s.t. \quad & \sum_{j \in R_o^{loc}} \boldsymbol{x_j}\boldsymbol{\lambda_j} + \boldsymbol{s}^- = \boldsymbol{x}_o \\ & \sum_{j \in R_o^{loc}} \boldsymbol{y_j}\boldsymbol{\lambda_j} - \boldsymbol{s}^+ = \boldsymbol{y}_o \\ & \boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq 0 \\ & \boldsymbol{e\lambda} = 1. \end{aligned} \tag{25}$$

$\boldsymbol{s}^{+*}$ and $\boldsymbol{s}^{-*}$ are derived as optimal solutions. The second program projects the solution onto the efficiency frontier:

$$\begin{aligned} \min_{\boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+} \quad & \frac{1 - \frac{1}{m}\sum\limits_{i=1}^{m} \frac{s_i^-}{x_{io} - s_i^{-*}}}{1 + \frac{1}{s}\sum\limits_{r=1}^{s} \frac{s_r^+}{y_{ro} + s_r^{+*}}} \\ s.t. \quad & \sum_{j \in R^{eff}} \boldsymbol{x}_j^{eff}\boldsymbol{\lambda}_j + \boldsymbol{s}^- = \boldsymbol{x}_o - \boldsymbol{s}^{-*} \\ & \sum_{j \in R^{eff}} \boldsymbol{y}_j^{eff}\boldsymbol{\lambda}_j - \boldsymbol{s}^+ = \boldsymbol{y}_o + \boldsymbol{s}^{+*} \\ & \boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq 0 \\ & \boldsymbol{e\lambda} = 1. \end{aligned} \tag{26}$$

$\boldsymbol{x}^{eff}$ and $\boldsymbol{y}^{eff}$ are the in- and outputs of the efficient DMUs. The optimal slacks of the second program are $\boldsymbol{s}^{-**}$ and $\boldsymbol{s}^{+**}$. Model (25) and (26) are calculated $n_{eff} + 1$ times for each inefficient DMU. n_{eff} is the number of efficient DMUs determined in the first step.

$$\rho_{oh} = \frac{1 - \frac{1}{m}\sum\limits_{i=1}^{m} \frac{s_{ih}^{-*} + s_{ih}^{-**}}{x_{io}}}{1 + \frac{1}{s}\sum\limits_{r=1}^{s} \frac{s_{rh}^{+*} + s_{rh}^{+**}}{y_{ro}}} \quad h = (1, \ldots, n_{eff} + 1). \tag{27}$$

For each DMU, the SBM-Max score (ρ^*_{max}) equals the highest ratio ($\boldsymbol{\rho_o}$) in equation (27) (Tone, 2015).

The SBM-Max approximates the closest reference point at the efficiency frontier. In contrast to the other DEA models, no exact efficiency scores are calculated. However, the SBM-Max leads to a sufficient approximation without being too complex or too computational demanding (Tone, 2015). The SBM-Min should be preferred if a worst case analysis is conducted. The SBM-Max calculates the smallest input and output changes that inefficient DMUs must make to become efficient.

Alternatively to the SBM-Max, Hadi-Vencheh et al. (2015) introduce a two-step model to exactly calculate the minimal distance to the frontier. However, their approach is based on the multiplier CCR model and uses fractional coefficients. This results in a high computational burden for large-scale problems (Hadi-Vencheh et al., 2015; Tone, 2015).

3.4 A Numerical Example

Table 3.1 lists five artificial DMUs and their intensity weights that provide the composition of their respective frontier reference point (Førsund, 2013). They produce two outputs y_1 and y_2. The input x of each DMU is one.[6]

Table 3.1 – Example DMUs and intensity weights

				BCC			SBM-Min		
	x	y_1	y_2	λ_A	λ_B	λ_D	λ_A	λ_B	λ_D
A	1.000	2.000	0.750	1.000	0.000	0.000	1.000	0.000	0.000
B	1.000	1.000	1.750	0.000	1.000	0.000	0.000	1.000	0.000
C	1.000	0.700	1.500	0.000	1.000	0.000	0.000	0.000	1.000
D	1.000	1.500	1.500	0.000	0.000	1.000	0.000	0.000	1.000
E	1.000	1.000	0.750	0.333	0.000	0.667	0.000	0.000	1.000

The intensity weights show that A, B, and D are technically efficient, C and E are inefficient. Using the BCC model, B is C's sole peer and E's reference DMU, projected onto the efficiency frontier, is located between A and D. The SBM-Min calculates D as

[6]Cooper et al. (2007) recommend: $n \geq max\{m \cdot s, 3 \cdot (m + s)\}$. If the number of degrees of freedom is too low, efficiency discrimination cannot be assured and the DEA lacks discriminatory power (Cooper et al., 2007). The five artificial DMUs are selected heterogeneously enough to ensure meaningful efficiency discrimination ($m + s = 3$).

the only peer of C and E. Figure 3.1 shows the different model approaches for measuring inefficiency.

Table 3.2 – Efficiency scores

	θ^*	τ^*	ρ^*_{max}	ρ^*_1	ρ^*_2	ρ^*_3	ρ^*_4
A	1.000	1.000	1.000	1.000	1.000	1.000	1.000
B	1.000	1.000	1.000	1.000	1.000	1.000	1.000
C	0.857	0.636	0.771	0.636	0.771	0.771	0.771
D	1.000	1.000	1.000	1.000	1.000	1.000	1.000
E	0.600	0.571	0.667	0.571	0.667	0.667	0.667

Table 3.2 provides the efficiency scores obtained from the BCC model (θ^*), the SBM-Min (τ^*), and the SBM-Max model (ρ^*_{max}). The SBM-Min considers technical inefficiencies and slacks. Thus, its efficiency scores must be equal or lower than the scores obtained from the BCC model ($\tau^* \leq \theta^*$). The SBM-Max calculates the maximum slack-based efficiency($\tau^* \leq \rho^*_{max}$). The radial BCC model and the non-radial SBM-Max interpret efficiency differently so that their results cannot be ranked in advance. For C, the SBM-Max score is lower than the score of the BCC model. The opposite holds for E.

Figure 3.1 elucidates C's and E's possibilities to become efficient by increasing their outputs. The radial model only allows proportional output increases. In the non-radial models, the inefficient DMUs can substitute one output for another, leading to non-proportional output increases. The optimal path for any inefficient DMU to increase efficiency depends on whether the efficiencies are to be maximised (SBM-Max) or minimised (SBM-Min).

In the following, the SBM-Max procedure is explained for DMU E. Its distances are: $d_{EA} = 1$, $d_{EB} = 1.333$, and $d_{ED} = 1.5$. In the first iteration of the second step, R^{loc}_E consists of only D, because D is the only peer of E in the SBM-Min. The location of D maximises the inefficiency of E, thus E$^{SBM-Min}$ equals D, and $\rho^*_{1E} = \tau^*_E$. In the second iteration, R^{loc}_E consists of the closest located DMU, which is A. The distance d_{EA} is smaller than d_{ED}, thus $\rho^*_{1E} < \rho^*_{2E}$. In the third and fourth iterations, first B and then D are additionally included in R^{loc}_E. E's scores remain unchanged ($\rho^*_{2E} = \rho^*_{3E} = \rho^*_{4E}$) because A's output composition already minimises the inefficiency of E.

Interestingly, the synthetic counterparts of C, C$^{SBM-Min}$ and C$^{SBM-Max}$, represent extreme paths (maximal substitution among outputs) as the DMU is not allowed to reduce (increase) overall output (input) in these models.

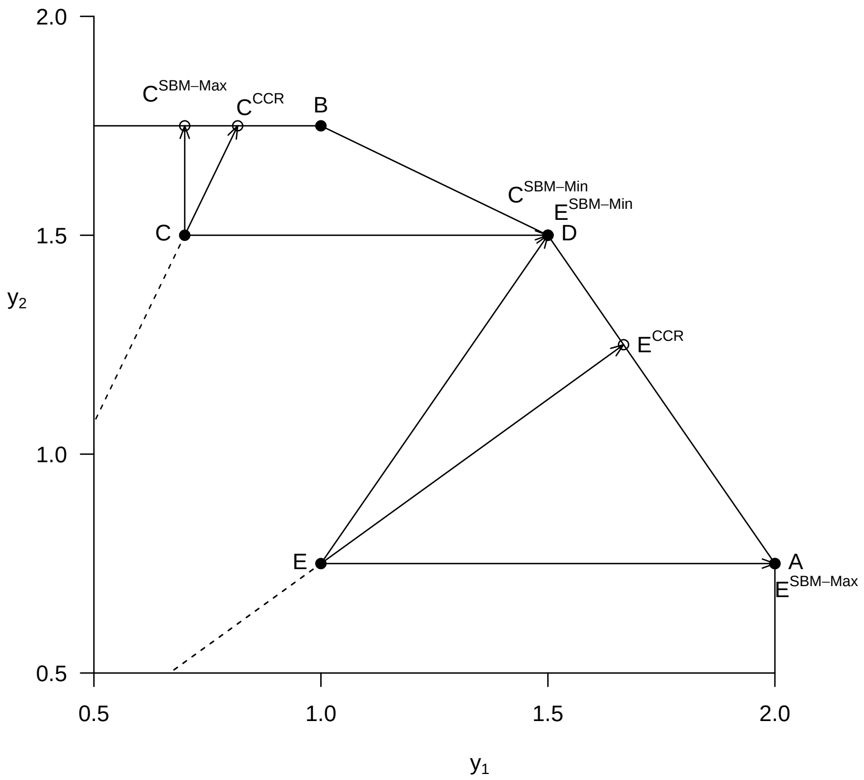


Figure 3.1 – Production possibility set of the example DMUs

3.5 A Simulation Application

The following simulation application shows the different interpretations of the inefficiency between the BCC, the SBM-Min, and the SBM-Max as well as their computing effort. 1,000 DMUs are simulated 1,000 times and the models are calculated for each iteration to improve the validity of the application. The multiple iterations allow to address a specific cases and deliver more reliable model comparisons. However, the results depend on the underlying data and may vary depending on the application. Over all iterations the models are calculated for 1,000,000 DMUs.

The inputs are generated using truncated normal distributions and are strictly positive. The outputs consists of both inputs (the relationship depends on a truncated normal distribution) plus a normally distributed random term to include more noise. Table 3.3 contains descriptive results and average correlation coefficients across all iterations.

Both outputs are positively correlated with both inputs, and all data is strictly positive in all iterations.

Table 3.3 – Descriptive results and correlation coefficients across all iterations

	x_1	x_2	y_1	y_2
Min	6.358	3.921	78.613	275.684
Median	2322.568	1501.578	1997.791	2284.598
Mean	2271.627	1501.098	1995.029	2280.837
Max	3997.744	2998.229	4018.157	4316.811
Standard deviation	776.816	636.437	603.445	607.243
Average correlation coefficients				
x_1	1.000	0.168	0.733	0.726
x_2	0.168	1.000	0.635	0.631
y_1	0.733	0.635	1.000	0.619
y_2	0.726	0.631	0.619	1.000

The SBM-Min and the SBM-Max are calculated under the assumption of variable returns to scale and combine the input- and output-orientations. For comparison, results of the BCC are derived using an input-orientation (BCC-in) and an output-orientation (BCC-out). Figure 3.2 shows the efficiency results for all models in ascending order across all iterations. On the left side all results are included and on the right side of five random iterations. Table 3.4 contains the average quantiles of the efficiency scores and their average correlation coefficients of all iterations. The full results and data can be provided upon request.

Across all iterations, the same 57,789 DMUs (of 1,000,000 DMUs) are efficient in all four models. If the efficiency scores are sorted in ascending order, the results of the BCC-out and the SBM-Max are quite similar and less dissimilar to the BCC-in scores than to the results of the SBM-Min. Apart from the efficient DMUs, the SBM-Min calculates the lowest efficiency scores for each quantile. The correlation coefficients of the unsorted DMUs indicate a strong positive relationship between the results in this application. The SBM-Max calculates 73.01% of the DMUs as more efficient than the BCC-out and 24.10% to be less efficient, and only the efficient DMUs are equally efficient in both models. No surprise is the low number of nearly efficient DMUs in the SBM-Min as it maximises inefficiency.

Using R (version 3.5.1) and the lpSolve package (5.6.13), the average computing times for the BCC-in, the BCC-out, and the SBM-Min are between 5.94 and 7.08 seconds. The average computing time for the SBM-Max is about 70 times higher (464.37 seconds)

Table 3.4 – Average efficiency scores comparisons

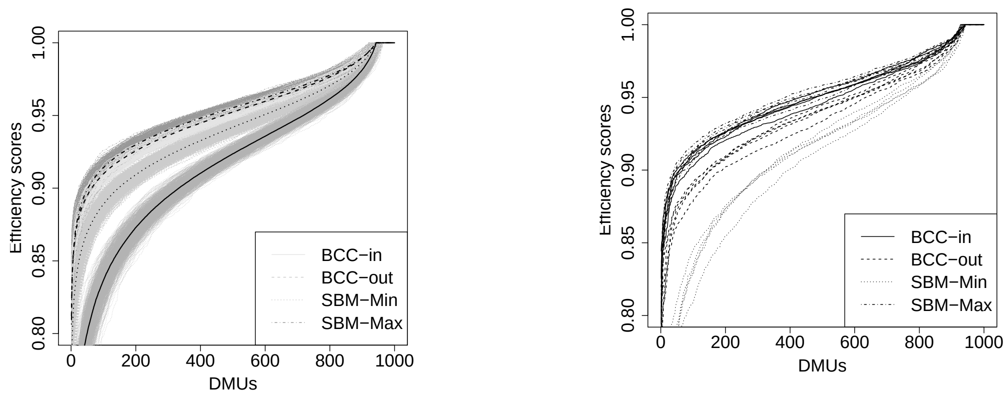
Quantile	BCC-in	BCC-out	SBM-Min	SBM-Max
0.0%	0.546	0.664	0.454	0.503
0.1%	0.889	0.910	0.836	0.914
0.2%	0.909	0.926	0.873	0.930
0.3%	0.922	0.936	0.894	0.940
0.4%	0.933	0.945	0.910	0.948
0.5%	0.942	0.952	0.924	0.955
0.6%	0.951	0.960	0.936	0.962
0.7%	0.960	0.968	0.948	0.969
0.8%	0.971	0.977	0.962	0.978
0.9%	0.987	0.990	0.982	0.990
1.0%	1.000	1.000	1.000	1.000
Average correlation coefficients				
BCC-in	1.000	0.993	0.887	0.957
BCC-out	0.993	1.000	0.884	0.960
SBM-Min	0.887	0.884	1.000	0.809
SBM-Max	0.957	0.960	0.809	1.000

due to the calculation of several linear programs. Table 3.5 contains an overview of the computing times. All calculations are performed on the same PC (using 64 GB of RAM, and an AMD R7 1800x CPU with up to 3.8 GHz).

Table 3.5 – Computing times in seconds of all iterations

	BCC-in	BCC-out	SBM-Min	SBM-Max
Min	5.507	6.609	6.123	279.390
Mean	5.939	7.083	6.416	464.366
Max	7.011	8.851	8.157	674.116
Standard deviation	0.222	0.263	0.224	63.825

The correlation coefficient structure, as well as the overall results, are in line with Johnes et al. (2017) who use the CCR model, the SBM-Min, and the SBM-Max to calculate the efficiency scores of 118 English universities. They include one input (total expenditure) and three outputs (research grants, research students, and taught students) in their analysis. The correlation coefficients of the efficiency scores vary between 0.443 and 0.812. Overall, the SBM-Max yields higher and the SBM-Min lower efficiency scores than the radial-model. However, the results of the individual universities differ strongly between the methods (Johnes et al., 2017).



(a) Sorted efficiency score of all iterations

(b) Sorted efficiency scores of five random iterations

Figure 3.2 – Sorted efficiency scores

3.6 Conclusion

Slack-based measures are better suited to calculate the efficiency of DMUs than radial models if slacks are present. Even if no slacks are present, the SBM-Min and the SBM-Max should be preferred if substitution among inputs and outputs is assumed. Based on the ADD model, the SBM-Min is units invariant, and its measure is monotone decreasing in each input and output slack, yielding comparable efficiency scores. Furthermore, it can be easily extended, e.g. to include weight restrictions (see Ruiz et al., 2015) or to consider undesirable outputs (see Barros et al., 2012).

The SBM-Max represents an alternative approach to measure efficiency based on slacks. Contrary to the SBM-Min, it minimises the distance to the efficiency frontier. The DM must understand the differences in assumptions between the DEA approaches to decide which model she prefers. The SBMs provide an efficiency corridor representing the maximal and minimal efficiency of inefficient DMUs. A closer look at the different efficiency scores of radial and non-radial models provides insights into the efficiency calculation and potential causes of inefficiency of the DMUs.

This is the first study to compare the SBM-Max with other radial and non-radial DEA models on this large scale. The simulation illustrates how the models interpret the inefficiency in different ways and the necessary computational effort. Overall, the SBM-Max offers an upper efficiency bound and the SBM-Min a lower efficiency bound. The efficiency scores of all methods are strongly positively correlated. The advantages of the non-radial slack-based models are that they consider slacks, allow substitution among inputs and outputs, prevent zero weights, and do not require assumptions about the model orientation.

If the DM does not prefer an explicit model, it is necessary to perform several radial and non-radial DEAS to account for the different interpretations of inefficiency. As the simulation shows, this leads to a considerably higher computing effort, especially due to the SBM-Max. If efficient DMUs that are not part of the reference set of the DMUs under consideration are excluded from the SBM-Max, its computational time could be reduced. In addition, parallel calculations of the linear programs are a further possibility to reduce the overall computational time. However, the optimisation of the SBM-Max remains a task for future research.

4

PISA Performance of Natives and Immigrants: Selection versus Efficiency

PISA Performance of Natives and Immigrants: Selection versus Efficiency

Prof. Andreas Behr, Gerald Fugger

University Duisburg-Essen, 45141 Essen, Germany.

Abstract

In most countries, immigrant and native students perform differently in Programme for International Student Assessment (PISA) due to two main reasons: different immigration regimes and differences in their home-country educational systems. While there is sophisticated literature on the reasons for these performance gaps, it is barely considered in the educational efficiency research. Our approach distinguishes between selection effects caused by immigration policies, and the efficiency of educational systems in integrating immigrant students, given their socio-economic background. Accordingly, we split our sample, which consists of 153,374 students in 20 countries, calculate various different efficient frontiers, and ultimately decompose and interpret the resulting efficiency values. We find large differences in educational system efficiency, when controlling for negative selection effects caused by immigration regimes.

Keywords: Data Envelopment Analysis, Migration, PISA, Efficiency decomposition
JEL Classification: C14 C61 I21

4.1 Introduction

The differences between natives and immigrants in the Programme for International Student Assessment (PISA), published by the Organisation for Economic Co-operation and Development (OECD), has gained considerable attention in the literature.[1] Apart from social, cultural, religious and historical reasons, different immigration policies and different levels of success in integrating immigrants are the two most important aspects (Kunz, 2016; Isphording et al., 2016). Countries attract different groups of immigrants with different socio-economic environments, due to country attractiveness, as well as their immigration policies (Entorf et al., 2005; Hochschild et al., 2010). In most countries, socio-economic endowment is one of the most important factors for the educational success of students (Parr et al., 2015; Rogiers et al., 2020). This is illustrated by the left panel of Figure 4.1, which shows a strong positive within-country correlation between the average reading, mathematics and science student scores in PISA, and their average ESCS values, the latter being an index of their socio-economic backgrounds, in 2015.[2] The index of economic, social and cultural status (ESCS) comprises several subcategories in the areas of parental education, highest parental employment and student housing. It is considered to be an appropriate measure of the students' socio-economic background (Hwang et al., 2018). The right side of Figure 4.1 shows the strong correlation between socio-economic endorsement gaps (ESCS gaps) and educational performance gaps between natives and immigrants across countries (Rogiers et al., 2020).[3]

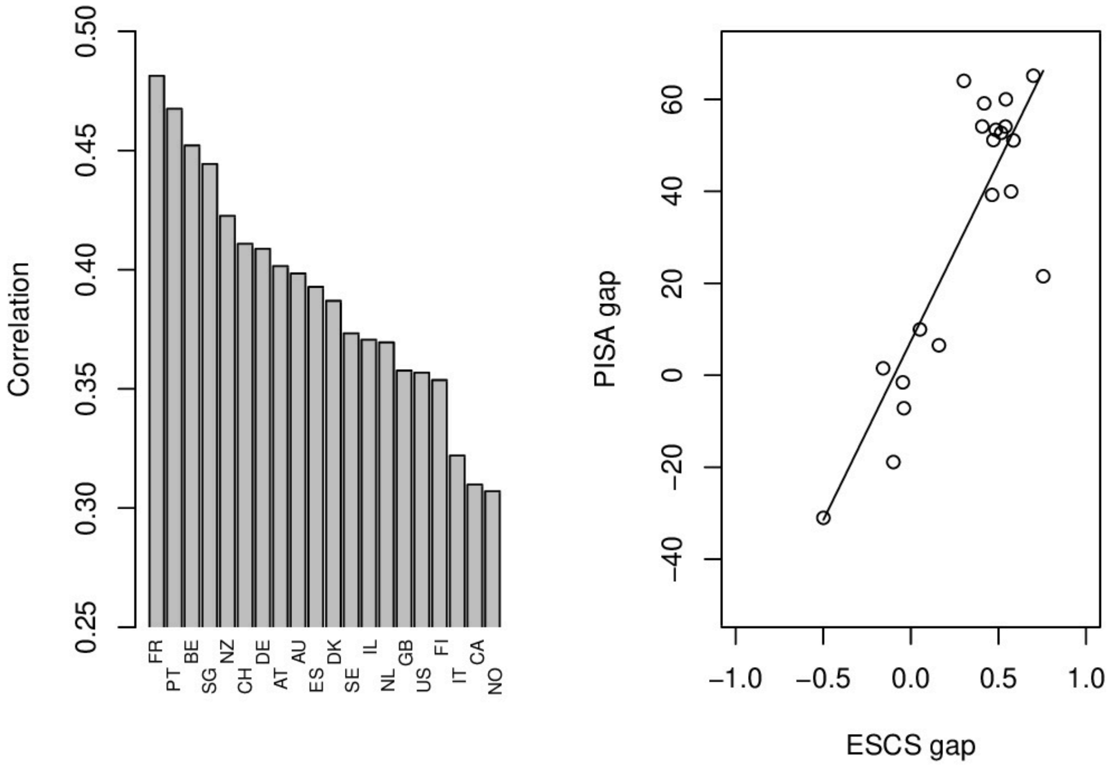
Our descriptive analysis reveals substantial educational (PISA) and socio-economic (ESCS) gaps between immigrants and natives and, that performance comparisons to a large extent implicitly reveal the students' different social and economic backgrounds. Without accounting for the students' backgrounds, studies run the risk of making implicit statements about immigration policy. We take this problem into account

[1]PISA is a worldwide study that assesses the 15-year-old students performance in mathematics, science, and reading. In addition, the individual backgrounds of the pupils and school data are collected. Following the PISA definition of immigration, an immigrant foreign-born in the second or first generation (OECD, 2017).

[2]Our included countries are: Australia (AU), Austria (AT), Belgium (BE), Canada (CA), Denmark (DK), Finland (FI), France (FR), Germany (DE), Israel (IL), Italy (IT), Netherlands (NL), New Zealand (NZ), Norway (NO), Portugal (PT), Singapore (SG), Spain (ES), Sweden (SE), Switzerland (CH), United Kingdom (GB), and the United States of America (US).

[3]Figure A.1 in the appendix shows the positive correlation between average PISA scores and the average ESCS scores on the country-level. The right panel elucidates that the average PISA scores are negatively correlated with the mean absolute deviations from the median ESCS scores.

Figure 4.1 – Relationship between PISA scores and ESCS values; left: within-country correlation, right: average gaps on country-level



explicitly, by analysing the performance of the educational system, given the varied social backgrounds of immigrant and native students.

An educational system can be integrating, despite a large educational gap, if it at least partially compensates for the gaps in socio-economic background. We use Data Envelopment Analysis (DEA) to examine the efficiency of educational systems. DEA models provide efficiency scores based on the students' performance relative to the performance of the best students comparable in their ESCS endowments. Our analysis is conducted at student-level, the most disaggregated data available in PISA. The students are evaluated according to their ability to maximise PISA scores given their socio-economic endowment. To account for the differences in socio-economic endowment between immigrants and natives, we split the PISA 2015 data into subsamples of natives and immigrants. Efficiency scores are calculated relative to various efficiency frontiers, which provides further insights and fosters our understanding of the relationship between selection effects in immigration, and the integrational abilities of the educational institutions in this context. Educational system performance is then obtained from the average efficiency scores of the students and further decomposed.

Our first efficiency analysis uses the average PISA score of the mathematics, science, and reading scores as output and the ESCS values as input. These three PISA scores are highly positively correlated. The aggregation into one output enables a straightforward interpretation and decomposition of the efficiency frontiers. In a further analysis we use the ESCS as input and include the three PISA scores (mathematics, science, and reading) as seperate outputs. DEA models allow the inclusion of several outputs, whereby all inputs and outputs are simultaneously included in the efficiency assessment by weighting them. The results of the second efficiency assessment confirm our main findings for average scores that in countries with restrictive immigration regimes, immigrants are not only performing relatively well but also use their endwoments rather efficiently. Some countries (e.g. Spain and France) perform considerably better according to their efficiency considering their ESCS endowments relative to their PISA ranking.

After this introduction, a literature overview of the performance gaps between natives and immigrants is provided. The third section outlines our methodology. In section four we explain the methodology of the ESCS and PISA scores, the differences between immigration regimes, and provides some initial results. The results of the efficiency analyses and their decomposition are discussed in section five preceding the conclusion.

4.2 Literature Overview

Differences in country attractiveness for immigrants, and different immigration policy regimes, attract different groups of immigrants, resulting in heterogeneous immigration populations between countries, and a wide range of challenges for the educational systems and societies in general (Entorf et al., 2005; Hochschild et al., 2010). While some countries attract immigrants whose socio-economic endowments are equal or even higher than those of the natives (Arabian oil-based economies, English speaking countries and Singapore), others, such as Central European countries, mainly attract immigrants who have a poorer socio-economic endowment than the natives (Jerrim, 2015). In Austria, Denmark, and Germany, for example, the differences between native and immigrant students are especially striking (Rindermann et al., 2016).

Besides their levels of educational, human capital, and wealth-related aspects (all part of PISA's ESCS index), native and immigration populations may also differ in cultural, religious, historical, and reputational aspects (Parr et al., 2015; Kunz, 2016). Immigrants may also face formal rights and legal status challenges, lack accumulated experiences as well as social connections that may result in educational information asymmetries, which can influence the educational performance of their children (Rindermann et al., 2016; Camehl et al., 2018).

Schneeweis (2011) decomposes the educational gap between immigrants and natives using the data of five international student assessment studies. Her results show that institutional characteristics of the education systems can increase differences between immigrants and natives. The results of Borgna et al. (2014) indicate that educational institutions and socio-economic backgrounds are mostly causing the gaps between immigrants and natives. Furthermore, PISA 2006 and 2009 data reveal that school attendance significantly reduces educational gaps. Dronkers et al. (2014) find that the countries' educational systems and the students' individual characteristics cause the differences between immigrants and natives. Harris et al. (2019) show that the access to certain areas of the curriculum depends at least in part on the socio-economic endowment of the students in the schools. Woessmann (2016) finds that educational institutions and family background have the highest explanatory power in determining educational achievements. Interestingly, the impact of school resources is much smaller than the students' social-economic endowment and institutional characteristics, which is also found by Falck et al. (2018).

Further empirical studies based on PISA data reveal that the different socio-economic backgrounds of immigrants and natives have the highest overall explanatory power regarding differences in educational attainment. Especially in European countries, nearly three-quarters of the performance gaps between natives and immigrants are accounted for primarily by differences in economic, social, and cultural status (Ammermueller, 2007; Levels et al., 2008; Arikan et al., 2017). Other factors, like linguistic barriers (previously considered the most important barrier for immigrants) only partially explain the performance gaps (Isphording et al., 2016; Rindermann et al., 2016).

Another important aspect in explaining performance gaps is the selection process among immigrants. Individual background factors vary between different immigrant groups which themselves vary between the countries (Schnepf, 2007; Arikan et al., 2017). In countries where immigrants are highly educated like Australia, they perform on average better in national and international comparisons than their native counterparts (Dustmann et al., 2012; Jerrim, 2015). The opposite holds for Central European countries in which a considerable share of the immigrants have on average a lower economic, social, and cultural status than the population of their immigration target countries and perform worse in PISA (Dustmann et al., 2012; Rindermann et al., 2016; Arikan et al., 2017).

Accordingly, heterogeneous immigrant populations provide specific challenges for educational systems that should be considered in efficiency analysis. Although the ESCS is an input (among others) in most educational efficiency analyses, regarding the importance of socio-economic backgrounds, international efficiency studies are deficiently, in how they consider the differences between immigrants and natives within and between countries.

Efficiency scores are based on the relationship between the sum of weighted *output* to the sum of weighted *input* of the students relative to the best students. As the socio-economic status is an environmental or non-discretionary input, it is not amenable to direct control by the educational system, and therefore cannot be regarded as a traditional input in efficiency analysis. But since it is found to have a significant impact in determining performance in PISA, socio-economic status is included in most efficiency analyses (Agasisti et al., 2018). For example, Sutherland et al. (2009) argue that student achievements depend on their social environment (family and peer-groups) and therefore must be included in student efficiency analysis. Similarly, Cordero et al. (2017b) argue that student socio-economic background is crucial for evaluating students according to their ability to make the most with their inputs (Cordero et al., 2017b). Aparicio et al. (2017a) refer to students as “raw material” that is transformed in schools and the impact of which is best reflected by the students’ socio-economic status (Aparicio et al., 2017a).

In the cross-country analyses of Sutherland et al. (2009), Aparicio et al. (2017a), and Agasisti et al. (2018), the students are not distinguished according to their country of origin. Moreover, the studies do not account for selection effects caused by immigration policies, that can lead to distinct immigrant groups with different socio-economic backgrounds. Aparicio et al. (2017b) proxy the socio-economic backgrounds of students by including the educational experience of their parents, which is only one aspect of the broader ESCS. As the performance gap determinants are manifold, a more comprehensive index should be preferred. De Witte et al. (2017) provide a broad overview of recent educational efficiency studies.

A considerable number of publications have been published in both the efficiency strand and the literature strand, focusing on the determination of the performance gaps between immigrants and natives. However, no international educational efficiency study so far accounts for the different challenges arising from different immigration policy regimes.

4.3 Methodology

In this section we explain our notation and our methodological approach in detail using a small artificial data set. As our decomposition approach regards several countries and the two subsets of students with or without immigration background, we introduce index sets denoted in calligraphic characters to facilitate referencing to specific groups of students.

4.3.1 Sets of Students

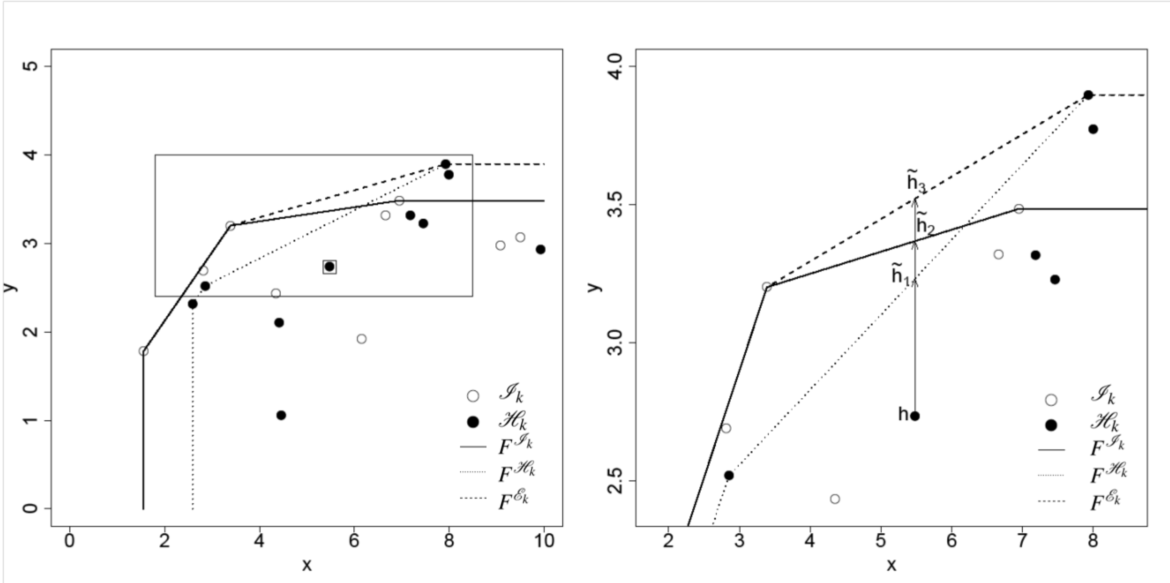
The set of all countries is denoted $\mathcal{K}$ and individual countries are referred to with index k $(k = 1, ..., K)$. In each country k we have two sets of students. The set of students in country k having an immigration background is denoted with $\mathcal{I}_k$. Immigrant students in country k are referred to using the index i $(i = 1, ..., I_k)$. Native students (home) in country k build the index set $\mathcal{H}_k$ and are indexed with h $(h = 1, ..., H_k)$. All students in country k, that is students with and without immigration background are referred to with $\mathcal{E}_k = \{\mathcal{I}_k, \mathcal{H}_k\}$.

Calligraphic characters without an index refer to the set combining the subsets from all K countries. I.e. $\mathcal{E} = \{\mathcal{E}_1, ..., \mathcal{E}_k, ..., \mathcal{E}_K\}$ is the set of all students from all K countries and $\mathcal{I}$ =$\{\mathcal{I}_1, ..., \mathcal{I}_k, ..., \mathcal{I}_K\}$ is the set of all students with immigration background from all K countries. We also have $\mathcal{E} = \{\mathcal{I},\ \mathcal{H}\}$ with $\mathcal{H}$ =$\{\mathcal{H}_1, ..., \mathcal{H}_k, ..., \mathcal{H}_K\}$.

4.3.2 Students and Different Frontiers of Potential Scores

In our illustrating example we only consider two countries, that is k and k'. First, we consider country k and the two subsets $\mathcal{I}_k$ (immigrant students) and $\mathcal{H}_k$ (native students). For each we observe their input x (ESCS-score) and their output y (PISA-score). We represent in Figure 4.2 native students ($\mathcal{H}_k$) by closed circles and students with immigration background ($\mathcal{I}_k$) with open circles.

Figure 4.2 – Benchmarking of a country with two student groups. The left panel shows the three frontiers, the right panel shows the rectangle enlarged.



We observe that some students with rather similar inputs reach quite different outputs. The observations of the 'best students', subsequently named efficient students, are joined with linear junctions and the resulting frontier is used as a yardstick to benchmark the remaining students. How we identify the best students is explained in more detail below (see model 34). As we have three different subgroups, natives ($\mathcal{H}_k$), immigrants ($\mathcal{I}_k$) and all students combined ($\mathcal{E}_k$), we can obtain three different frontiers. These frontiers we denote in general by F and the superscript indicates based on which subset of students the frontier is obtained, accordingly we have drawn the three different frontiers $F^{\mathcal{I}_k}$, $F^{\mathcal{H}_k}$ and $F^{\mathcal{E}_k}$ in Figure 4.2.

4.3.3 Benchmarking Individual Students

The performance of a specific student h, we pick for illustration the one indicated with the square, can now be assessed using three different benchmarks. To ease the readability, the right panel of Figure 4.2 displays a part of the left panel enlarged.

A benchmark student denoted by $\tilde{h}_1$ is a synthetic student on frontier $F^{\mathcal{H}_k}$. This benchmark student is a linear combination of two efficient native students located at the frontier $F^{\mathcal{H}_k}$ (dotted line). If we compare the obtained score of student h

with the score of $\tilde{h}_1$ on the frontier $F^{\mathcal{H}_k}$, we calculate a relative efficiency score of $D^{\mathcal{H}_k}(h) = 2.740/3.230 = 0.850$. We use D for the efficiency score and the superscript indicates on which set of students the frontier is obtained, here we use frontier $F^{\mathcal{H}_k}$. Hence, the student h only obtained 85% of the score that is regarded as being possible given his input amount. Or, equivalently, he could increase his output by 17.6% if he would be as efficient as his benchmark fellow students.

If we compare our native student h with an efficient synthetic student with immigration background $\tilde{h}_2$, which is located at the frontier $F^{\mathcal{I}_k}$ (solid line) obtained from immigration students $\mathcal{I}_k$, we obtain students h score as $D^{\mathcal{I}_k}(h) = 2.740/3.370 = 0.810$, hence, in this comparison he is underperforming by 19%.

And finally we can benchmark student h with synthetic student $\tilde{h}_3$ located at the frontier $F^{\mathcal{E}_k}$ (dashed line) which is based on all students in country k. As this hypothetical benchmark student $\tilde{h}_3$ performs even better than $\tilde{h}_1$ and $\tilde{h}_2$, we find that according to this yardstick, student h underperforms by $D^{\mathcal{E}_k}(h) = 2.740/3.510 = 0.780$, i.e. 22%. Note that in this last comparison the benchmark student $\tilde{h}_3$ is a hypothetical student obtained as a linear combination of an efficient native and an efficient immigrant student.

4.3.4 Benchmarking Sets of Students

To obtain a measure of the performance of a complete set of students we use the arithmetic mean of individual scores. E.g. to obtain the average performance of immigrant students $\mathcal{I}_k$ using the frontier $F^{\mathcal{I}_k}$ obtained based on this set of students, we calculate

$$M^{\mathcal{I}_k}\left(\mathcal{I}_k\right) = \frac{1}{I_k}\sum_{i=1}^{I_k} D_i^{\mathcal{I}_k}\left(\mathcal{I}_k\right) \tag{28}$$

I_k is the number of students benchmarked, here the students with immigrant background in country k. We use M for arithmetic mean, the superscript $\mathcal{I}_k$ to indicate that we use the frontier $F^{\mathcal{I}_k}$ and the argument in parentheses indicates which group of students is benchmarked.

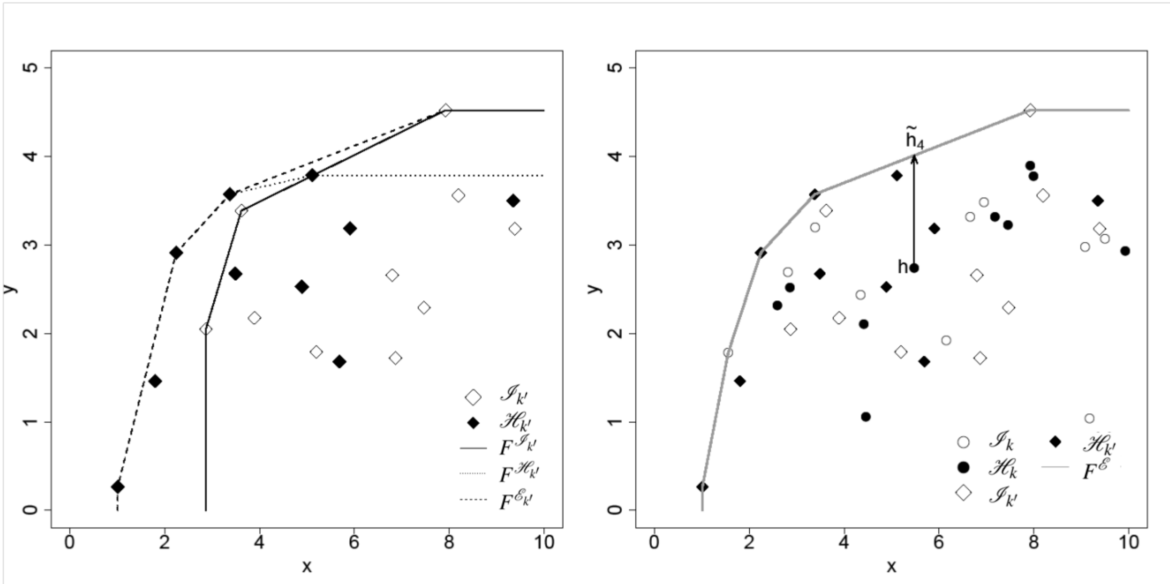
In our illustrative example considered in Figure 4.2 we obtain for immigrants $M^{\mathcal{I}_k}\left(\mathcal{I}_k\right) = 0.827$ and for natives $M^{\mathcal{H}_k}\left(\mathcal{H}_k\right) = 0.839$. For comparing the performance of immigrants and natives, one may like to use the frontier $F^{\mathcal{E}_k}$ obtained considering all students $\mathcal{E}_k$ in country k. In this example we obtain for immigrants $M^{\mathcal{E}_k}\left(\mathcal{I}_k\right) = 0.788$

and for natives $M^{\mathcal{E}_k}(\mathcal{H}_k) = 0.796$ as average efficiencies.

4.3.5 Considering a Second Country

We now consider a second country k'. We use filled diamonds for native students and open diamonds for immigrant students. The left panel of Figure 4.3 contains the situation for country k', again with the three different national frontiers indicated by dotted, dashed and solid lines. The right panel combines the students of both countries and allows us to obtain an international frontier $F^{\mathcal{E}}$ collated from all students of all (here: two) countries.

Figure 4.3 – Benchmarking of another country with two student groups and for both countries together



This allows the benchmarking of the immigrant students of county k ($\mathcal{I}_k$) and of the native students of country k ($\mathcal{H}_k$) using the international frontier. E.g. our student h of country k is now benchmarked based on the score of a synthetic student $\tilde{h}_4$ located at the international frontier $F^{\mathcal{E}}$. Accordingly in this comparison her efficiency score $D^{\mathcal{E}}(h) = 2.740/4.010 = 0.680$ is the lowest obtained in the comparisons and hints for a potential increase in her score of 47%.

Using the international frontier $F^{\mathcal{E}}$ for benchmarking all native students in country k results in an average score $M^{\mathcal{E}}(\mathcal{H}_k) = 0.686$. The immigrants of country k obtain an average score $M^{\mathcal{E}}(\mathcal{I}_k) = 0.698$.

4.3.6 The DEA Model

We use the output-oriented BCC model, introduced by Banker et al. (1984). The output orientation implies that students maximise their output given their inputs. For student o the model is defined as:

$$\begin{aligned} \min_{\eta,\boldsymbol{v},\boldsymbol{u}} \quad & \eta = \sum_i v_i x_{io} - u_0 \\ \text{subject to} \quad & \sum_r u_r y_{ro} = 1 \\ & \sum_i v_i x_{ij} - \sum_r u_r y_{rj} - u_0 \geq 0 \quad (j = 1, \ldots, n) \\ & v_i \geq 0 \quad (i = 1, \ldots, m) \\ & u_r \geq 0 \quad (r = 1, \ldots, s) \\ & u_0 \text{ free in sign.} \end{aligned} \tag{29}$$

Output r of student o is given by y_{ro} and is weighted by u_r $(r = 1, ..., s)$. s equals the number of outputs. Her input i (x_{io}) is weighted by v_i $(i = 1, \ldots, m)$. m is the number of inputs and n is the number of all students under analysis. The weights are restricted to be non-negative, derived from the data, and most likely vary between students. The weights are not chosen a priori but determined when solving the linear program. The most favourable composition of weights to make student o as efficient as possible are chosen given the restrictions. The linear program is set up and solved for each student under analysis individually (Behr, 2015; Cooper et al., 2007).

η^* denotes the solution to the minimisation problem. For convenience, we define $D^* = \frac{1}{\eta^*}$. If $\eta^* = D^* = 1$ student o is efficient. The limits of η^* and D^* depend on whether the student o belongs to the group of students she is compared to. If she belongs to the group of students she is compared to, η^* is equal to or greater than one and D^* is equal to or less than one. If student o does not belong to the group of students she is compared to, η^* may be smaller than one (the student is super-efficient). In this case the student o is above the efficiency frontier of the students she is compared to, and D^* is greater than one (Chen, 2005).

The scalar u_0 is free in sign and implements the assumption of variable returns to scale (VRS). VRS allow non-proportional output changes when the inputs change. The input and output tuples of students are neither allowed to be scaled up (increasing returns to scale) nor down (decreasing returns to scale) in the BCC model.

4.4 The PISA study, Migration Regimes and Descriptive Results

We use students' socio-economic status as the input and the average PISA score of the students in reading, mathematics and science, as output in the first efficiency analysis. If necessary, the data are transformed to obtain positive values as DEA can only handle positive inputs and outputs. Outliers are excluded.

4.4.1 The PISA Study and the ESCS

PISA is a worldwide stratified two-stage sample study conducted by the OECD, to measure 15-year-old students' performance in mathematics, science, and reading. It was conceived to offer insights into sources of performance variation within and between countries. It was first performed in 2000 and then repeated every three years. The PISA assessment in 2015 focused on science, and was published in December 2016 (OECD, 2016). Student performance is reported as the corresponding mathematics, science, and reading scores.[4]

A minimum of 150 schools must be selected in each country to ensure quality standards. If a participating country has fewer than 150 schools, all schools are selected. Within each participating school, a predetermined number of 15-years-old students, usually 42 students, is randomly chosen with equal probability. In schools with fewer students, all students are selected. If the response rate is too low, the sample size of schools is increased beyond 150 to ensure a minimum student sample size. A response rate of 85% is required for initially selected schools. If the initial school response rate falls between 65% and 85%, an acceptable school response rate can still be achieved by using replacement schools. Schools are classified into similar groups according to selected variables (region, private or public school, funding,...). A minimum student response rate of 50% within each school is required for a school to be regarded as participating (OECD, 2016).

Since its publication, the results of the PISA study have influenced the design of the education systems of the participating countries. For example, Ho (2016) shows how the insights resulting from PISA were used in Hong Kong, Damiani (2016) in Italy, and Ababneh et al. (2016) in Jordan. Tobin et al. (2016) provide a world wide overview of how large-scale educational assessments influence education policy and most studies

[4]Now data for 2018 have become available but the preliminary version of 2018 is still incomplete and lacks for example individual scores in of the three subjects for spanish students.

find significant effects of secondary education on the economic development of countries (Aduand et al., 2017; Karatheodoros, 2017).

The index of economic, social and cultural status (ESCS) comprises three main categories: parental education, highest parental occupation, and home possessions. The latter combines five indices: family wealth, household possessions, cultural possessions, home educational resources, and information and communication technology resources. These indices are derived from the availability of 16 household items at home, including three country-specific household items. The ESCS's three main components are standardized with a mean of zero and a standard deviation of one, over the full sample. Finally, a principal component analysis (PCA) of the three main components is conducted, and the ESCS is defined as the first principal component score (OECD, 2017).[5] For first-generation immigrants, parental education and partly the highest parental occupation may result from the educational institutions of their country of origin, rather than from integration results or the educational system of their target country, in whose educational efficiency we are interested. However, both the home possession measures and the success of the second-generation immigrants depend on the integration and education quality in their target country (Reparaz et al., 2019). The ESCS covers a wide range of different economic, social and cultural topics, enabling an approximation of possible determinants of education performance gaps between immigrants and natives. Furthermore, through the use of PCA, the ESCS is a construct that is well suited for capturing and comparing the whole students' socio-economic status (Hwang et al., 2018).

4.4.2 Migration Regimes

When examining the efficiency of educational systems in terms of the immigrant performance, the respective immigration regimes of the countries must be taken into account. Bjerre et al. (2015) and Bonjour et al. (2018) provide an overview of a large number of definitions and distinctions in the literature.

In addition to limiting official immigration policies (strict ones are mainly based on points systems), another important aspect is how many people enter the country through unofficial channels. For example, a comparison between Germany and Australia shows that the proportion of immigrants in Australia for family and humanitarian

[5] The common ESCS component weights across cycles are 0.79 (parental occupation), 0.82 (parental education), and 0.74 (home possessions) (**OECD2014**).

reasons is far lower and the percentage who do so for economic reasons is higher (Beine et al., 2016). Based on their selective immigration policy and low proportion of non-economic immigration, Australia, Canada, New Zealand, and the United Kingdom can be regarded as having rather restrictive immigration regimes. The United States of America also has a restrictive immigration policy, but unlike the remaining countries in this group, it does not succeed in attracting immigrants who perform on average at least as well as their native peer group, as shown below (see also Camarota et al. (2016)). The European Union introduced a points-based system in 2009, but it is far less strict than in the other countries with a selective immigration policy, and the share of immigrants for family and humanitarian reasons is relatively high. Therefore, we do not regard the members of the European Union as being restrictive (Bertoli et al., 2016).

Table 4.1 – Average highest occupational status of parents

	Immigrants	Natives	Gaps	t-Tests	p-Values
AU	57.870	56.465	-1.405	0.380	0.704
AT	42.294	53.350	11.056	-2.991	0.003***
BE	45.995	54.497	8.502	-2.300	0.021**
CA	57.564	56.987	-0.576	0.156	0.876
DK	42.244	56.849	14.605	-3.951	0.000***
FI	45.511	53.386	7.875	-2.130	0.033**
FR	40.722	53.296	12.574	-3.401	0.001***
DE	42.235	53.775	11.540	-3.121	0.002***
IL	58.358	60.347	1.989	-0.538	0.591
IT	37.215	50.929	13.714	-3.710	0.000***
NL	45.231	55.830	10.599	-2.867	0.004***
NZ	57.923	57.638	-0.285	0.077	0.939
NO	51.832	63.253	11.421	-3.089	0.002***
PT	46.032	45.672	-0.361	0.098	0.922
SG	67.916	59.240	-8.677	2.347	0.019**
ES	39.428	50.396	10.968	-2.967	0.003***
SE	50.573	59.201	8.628	-2.334	0.020**
CH	43.227	56.933	13.705	-3.707	0.000***
GB	55.359	56.346	0.987	-0.267	0.789
US	43.005	57.112	14.107	-3.816	0.000***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

We use the average occupational status of parents, which is available in PISA (higher values stand for better status) to substantiate our country classification. The occupational status of parents is an important determinant of the educational attainment of immigrants, as the educational mobility of immigrants is generally lower than that of natives (Schneebaum et al., 2016; Reparaz et al., 2019). Descriptive results show that in most countries, the occupational status of parents of natives is higher than that of immigrants. Only in countries with a selective immigration regime, are the gaps close to zero or even negative. Singapore attracts immigrants whose parents have the highest level of education.[6] These results can be provided upon request.

4.4.3 The Data and Descriptive Results

Our sample comprises 153,374 students in 20 industrialized countries for PISA 2015.[7] We combine first- and second-generation immigrants, otherwise several countries would have too few data points in at least one group (e.g. Finland and the Netherlands), and both groups have similar performance differences (relative to the natives), which are determined to a similar extent by the ESCS (Rangvid, 2007).

As a frontier based non-parametric technique, DEA is sensitive to outliers. We exclude outliers based on their influence, measured by Cook's distance. We define outliers to have a Cooks' distance of at least eight times the average distance for each country and each regression, which is a reasonable threshold according to Cook (1979).[8] Table A.1 shows the number (between 44 and 207) and the percentages (ranging from 0.759% to 1.186%) of excluded outliers per country.

PISA reading, mathematics, and science scores are constructed to have an international mean of 500 and a standard deviation of 100. The standardization provides student results that are directly comparable between countries. Table 4.2 summarizes within-country correlations between the scores. All scores are highly positively correlated, and the correlations vary between 0.743 for the mathematics and reading results in Italy,

[6] In Singapore, the recruitment of skilled workers is systematically promoted and part of the official government strategy, as the following quote from prime minister Goh Chok Tong's speech at the national day rally 2001 shows:

> "[...] some Singaporeans may again question the need for more global talent. I urge you to understand that this is a matter of life and death for us in the long term. [...] If we do not top up our talent pool from the outside, in ten years time, many of the high-valued jobs we do now will immigrate to China and elsewhere, for lack of sufficient talent here." (Tong, 2001)

Singapore is the most successful of all countries in attracting highly qualified and top performing immigrants. In our analysis, immigrants in Singapore are on average the most efficient.

[7] Japan, Korea, and Poland are excluded because of having too few immigrants.

[8] The results are robust for alternative thresholds (e.g. from two times up to 20 times the average distance) and can be provided upon request.

and 0.908 for the reading and science results in Singapore. Table A.5 in the appendix depicts the correlation coefficients for each country.

Table 4.2 – PISA scores correlations coefficients, overview

Scores	Min	Country	Max	Country	Mean
Mathematics-Reading	0.743	Italy	0.860	Netherlands	0.795
Mathematics-Science	0.849	Italy	0.899	France	0.883
Reading-Science	0.828	Sweden	0.908	Singapore	0.868

We use the students' average PISA scores as output y, to enable comprehensible visual and contextual illustrations. After discussing the results for the average PISA score as output, we also present the results for the three PISA scores in mathematics, science, and reading as outputs.

Figure 4.4 presents the average PISA score distributions, using a Gaussian kernel with a bandwidth of 70% of Silverman's "rule of thumb" to disclose more details for immigrant and native students separately for each country (Silverman, 1986). In countries with selective immigration policies, as well as in Israel and Portugal, immigrants and natives perform similarly well. In Singapore, the immigrants perform even better than the natives. In the other countries and especially in most European countries, natives perform better. The differences between natives and immigrants between countries further indicate that the prevailing immigration regime influences the selection among immigrants. However, Figure 4.4 focuses only on our output and does not distinguish between the selection effects and the efficiency of educational systems. Figures A.2 to A.4 in the appendix provide the distributions of the three PISA scores. They are rather similar to the distributions of the average PISA scores and the same distinctions between countries with and without restrictive regimes can be made.

The index of economic, social and cultural status of each student (ESCS) is regarded as input (x'). x' is internationally comparable, has a mean of zero and a standard deviation of one. Radial DEA models can only handle strictly positive variables. Therefore, x' is transformed: $x' - min(x') + 0.01 = x$. x is the input used in our efficiency analysis and is not further transformed.

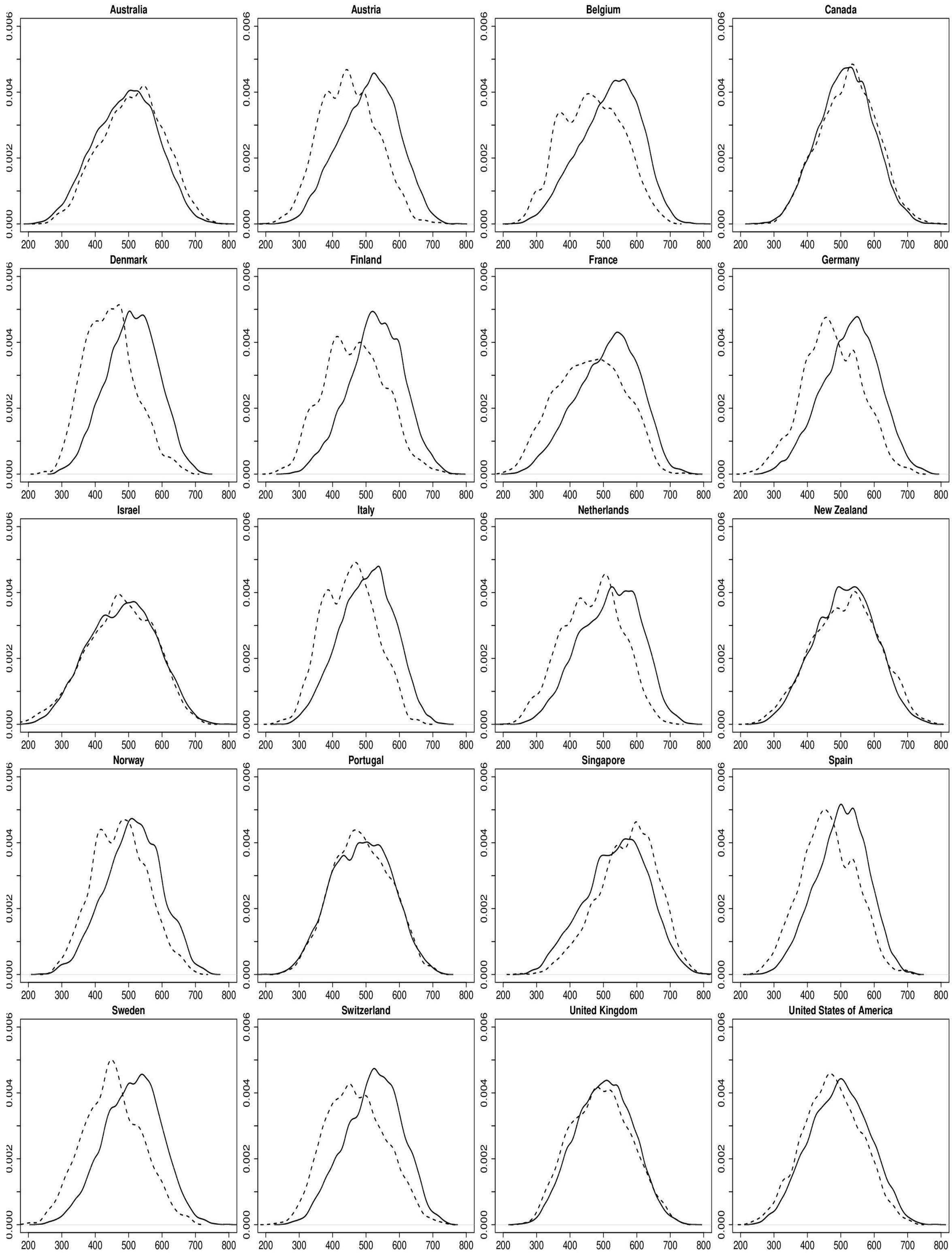
Table 4.3 provides descriptive results and correlation coefficients between the average PISA scores and the ESCS values for each country at the student level, for students with and without an immigration background. In most countries, natives perform better and have a better average socio-economic background. In Australia, Canada, and New

Table 4.3 – Descriptive results of the average PISA scores and the ESCS values

Countries	Group	PISA	Mean diff.	ESCS	Mean diff.	Corr	n
Australia (AU)	Nat	492	-18.872	0.194	-0.100	0.403	10744
	Mig	511	(2.012)	0.294	(0.017)	0.369	2651
Austria (AT)	Nat	508	60.012	0.207	0.542	0.373	5533
	Mig	448	(2.626)	-0.335	(0.026)	0.299	1242
Belgium (BE)	Nat	519	54.107	0.272	0.407	0.446	7684
	Mig	465	(2.573)	-0.135	(0.026)	0.358	1445
Canada (CA)	Nat	514	-7.163	0.487	-0.040	0.319	14555
	Mig	521	(1.481)	0.527	(0.014)	0.281	4057
Denmark (DK)	Nat	510	65.141	0.630	0.699	0.362	5224
	Mig	445	(2.177)	-0.069	(0.027)	0.170	1567
Finland (FI)	Nat	529	64.021	0.281	0.303	0.349	5495
	Mig	465	(6.391)	-0.022	(0.054)	0.289	200
France (FR)	Nat	512	52.708	-0.038	0.515	0.486	5089
	Mig	459	(3.921)	-0.553	(0.032)	0.290	706
Germany (DE)	Nat	528	54.087	0.238	0.539	0.409	4614
	Mig	474	(3.156)	-0.301	(0.032)	0.213	881
Israel (IL)	Nat	481	6.512	0.227	0.160	0.388	5223
	Mig	475	(3.426)	0.067	(0.029)	0.318	1023
Italy (IT)	Nat	502	51.145	-0.006	0.471	0.316	10199
	Mig	451	(2.737)	-0.477	(0.031)	0.196	867
Netherlands (NL)	Nat	519	53.374	0.245	0.485	0.359	4587
	Mig	466	(4.102)	-0.240	(0.035)	0.225	504
New Zealand (NZ)	Nat	512	-1.539	0.173	-0.046	0.417	3031
	Mig	514	(3.403)	0.219	(0.026)	0.457	1075
Norway (NO)	Nat	513	39.224	0.550	0.463	0.294	4535
	Mig	474	(3.465)	0.087	(0.033)	0.202	616
Portugal (PT)	Nat	487	1.512	-0.570	-0.158	0.468	6647
	Mig	486	(4.325)	-0.412	(0.055)	0.456	416
Singapore (SG)	Nat	539	-31.006	-0.120	-0.499	0.450	4734
	Mig	570	(2.913)	0.379	(0.027)	0.342	1164
Spain (ES)	Nat	502	39.959	-0.371	0.572	0.380	5808
	Mig	462	(3.268)	-0.943	(0.044)	0.333	664
Sweden (SE)	Nat	510	59.129	0.425	0.418	0.374	4311
	Mig	451	(3.298)	0.007	(0.031)	0.192	819
Switzerland (CH)	Nat	522	51.076	0.323	0.585	0.356	3907
	Mig	471	(2.528)	-0.262	(0.026)	0.357	1711
United Kingdom (GB)	Nat	502	9.967	0.232	0.052	0.365	11329
	Mig	492	(2.360)	0.181	(0.023)	0.334	1607
United States (US)	Nat	496	21.527	0.280	0.755	0.366	4153
	Mig	474	(2.849)	-0.475	(0.034)	0.302	1215

PISA and ESCS: group-specific country averages; Mean diff.: Differences between the means of natives and immigrants; the values in brackets are a variance measure: $\sqrt{\frac{var(v_{\mathcal{I}})}{n_{\mathcal{I}}} + \frac{var(v_{\mathcal{H}})}{n_{\mathcal{H}}}}$ where v represents the students' PISA and ESCS values and n their respective numbers.

Figure 4.4 – Average PISA scores distributions among natives (straight line) and immigrants (dashed line)



Zealand (all countries with selective immigration systems), immigrants achieve higher average PISA scores and have higher ESCS endowments. On average, immigrants in Singapore have ESCS values that are above the PISA average, and the values of the natives are lower (Becker, 2012; Facchini et al., 2014). In comparison, both Canadian population groups have above-average ESCS averages and the smallest gap. This hints for the selectivity of the Canadian immigration system, so that the average immigrant in Canada has a socio-economic background similar to that of the average native. The United States has the largest ESCS gap between the two groups. Although the United States has a selective immigration system, it attracts immigrants with relatively poorer socio-economic backgrounds. However, the differences in performance are smaller in the United States than in Germany and Norway, for example. Spanish immigrants have, on average, the lowest ESCS values, and Portugal is the only country in which the natives achieve higher PISA values despite worse socio-economic backgrounds, although the gap is not significantly different from zero. Such specific challenges must be taken into account in an international efficiency analysis of educational systems. Tables A.2 to A.4 in the appendix provide descriptive results for the individual PISA scores. All scores are greater than zero and students with missing values are excluded from our analyses.

We use regressions to gauge the relationship between students' average educational performance and their socio-economic endowments for each country separately. The regressions include both a dummy for immigrant background and an interaction term. The results indicate that performance gaps between immigrants and natives are determined strongly by their respective ESCS endowments. Increasing ESCS values have the highest positive impact in France and lowest in Spain and Italy. The results indicate a significantly better performance of immigrants in Australia, Canada and Singapore and a positive but insignificant relationship in Israel and the United States of America. In all other countries, immigrants perform significantly worse than natives. All results can be provided upon request.

4.5 Efficiency Results and Efficiency Decomposition

All results are obtained using R (version 3.6) and, unless otherwise stated, the average PISA results are used as output. The efficiency scores indicate how relatively well the students perform, given their socio-economic backgrounds. First, the results are decomposed relative to national and then international frontiers, followed by a comparison of the performance of natives and immigrants, and finally, the impact of the selection processes and the efficiency of educational systems are evaluated.

4.5.1 National Frontiers

Table 4.4 – Decomposition, national students and national frontiers, average PISA scores as output

	(1)	(2)	(3)	(4)	(5)
	$M^{\mathcal{H}_k}(\mathcal{H}_k)$	$M^{\mathcal{I}_k}(\mathcal{I}_k)$	$M^{\mathcal{E}_k}(\mathcal{H}_k)$	$M^{\mathcal{E}_k}(\mathcal{I}_k)$	$M^{\mathcal{E}_k}(\mathcal{E}_k)$
AU	0.665	0.707	0.665	0.684	0.669
AT	0.694	0.673	0.694	0.649	0.686
BE	0.704	0.706	0.704	0.661	0.697
CA	0.681	0.702	0.680	0.688	0.682
DK	0.727	0.690	0.726	0.668	0.712
FI	0.724	0.701	0.724	0.652	0.721
FR	0.703	0.703	0.702	0.669	0.698
DE	0.716	0.708	0.716	0.669	0.708
IL	0.645	0.699	0.645	0.651	0.646
IT	0.707	0.714	0.707	0.657	0.703
NL	0.707	0.706	0.707	0.664	0.703
NZ	0.696	0.706	0.693	0.692	0.693
NO	0.707	0.726	0.707	0.682	0.704
PT	0.709	0.733	0.709	0.699	0.708
SG	0.699	0.753	0.699	0.711	0.701
ES	0.729	0.727	0.729	0.690	0.725
SE	0.684	0.700	0.684	0.636	0.676
CH	0.725	0.691	0.724	0.684	0.712
GB	0.693	0.701	0.693	0.683	0.692
US	0.671	0.721	0.670	0.691	0.675
Mean	0.699	0.708	0.699	0.674	0.696

Table 4.4 provides country-specific arithmetic mean efficiency scores for all students, for the student groups relative to both national frontiers, and comparisons between the groups. The column numbers are given above the formal terms to simplify the interpretation.

The initial descriptive results showed that natives have higher average PISA scores (see Table 4.3 and Figure 4.4), but they disregard the socio-economic backgrounds of the students, that are taken into account in the efficiency analysis. Column (1) and (2) of Table 4.4 contain the results of natives and immigrants relative to their respective frontiers for each country. Across all countries, both groups of students are on average almost equally efficient (0.699 in column (1) to 0.708 in column (2)) if compared to their benchmark students from their group.

Columns (3) and (4) of Table 4.4 show the average efficiency scores when using the national frontier based on both subsets of students. We observe that there are hardly any changes among the natives, if immigrants are also taken into account when calculating the efficient frontier. In contrast, the performance of immigrants decreases when natives are taken into account as revealed by the comparison of column (2) and (4).

Natives outperform immigrants on average by $(M^{\mathcal{E}_k}(\mathcal{H}_k) - M^{\mathcal{E}_k}(\mathcal{I}_k)) \cdot 100 = 5.741\%$ in Denmark, by 7.188% in Finland, and by 5.009% in Italy. Natives also perform better in most countries, but the gaps are not as large as in the previous countries and range from 0.100% in New Zealand to 4.774% in Sweden. In all these countries, immigrants perform far worse, according to their efficiency scores, taking into account their socio-economic endowment. The educational systems do not succeed in fostering both groups equally, which leads to inequalities in educational performance beyond the differences due to their endowments.

In Australia, Canada, Israel, Singapore, and the United States, immigrants perform on average better than their native peer group, considering their efficiency based on ESCS endowments. In the United States, immigrants perform best relative to the natives. The performance difference is 2.066%. In Israel both groups perform similarly, immigrants being slightly better (0.613%).

Column 5 of Table 4.4 provides the mean efficiency scores for all students, based on their own frontiers for each country. Israel achieves the lowest (0.646) and Spain the highest (0.725) mean. Since the efficiency frontiers are country- and group-specific, they are rather a measure of inequality than a means of comparing efficiency between countries. Table 4.4 does not provide any information on which students form the efficiency frontiers, and how efficient the national educational systems are.

Figure A.5 in the appendix displays the frontiers for each student group within the countries and the international frontier, calculated for all students. In several countries, the best-performing students are immigrants for low ESCS values and natives for higher ESCS values (e.g. in Austria, France, Germany, and the United States). In the remaining countries, only natives constitute the efficiency frontier, as is the case in Finland, Portugal, Singapore, Spain, and the United Kingdom. It is striking that the students in Portugal, Singapore, and Spain have input-output combinations that are on average far less distant from the international efficiency frontier than in the other countries. Therefore, these countries are among the top performers in our analysis.

4.5.2 International Comparisons

Table 4.5 – Decomposition, national students and international frontier, average PISA scores as output

	(1)	(2)	(3)
	$M^{\mathcal{E}}(\mathcal{H}_k)$	$M^{\mathcal{E}}(\mathcal{I}_k)$	$M^{\mathcal{E}}(\mathcal{E}_k)$
AU	0.619	0.638	0.623
AT	0.639	0.590	0.630
BE	0.652	0.603	0.644
CA	0.637	0.645	0.639
DK	0.630	0.577	0.617
FI	0.660	0.594	0.658
FR	0.652	0.615	0.647
DE	0.667	0.624	0.660
IL	0.602	0.604	0.603
IT	0.644	0.603	0.641
NL	0.648	0.606	0.644
NZ	0.643	0.643	0.643
NO	0.631	0.602	0.627
PT	0.663	0.649	0.662
SG	0.696	0.710	0.698
ES	0.672	0.652	0.669
SE	0.632	0.578	0.624
CH	0.652	0.619	0.642
GB	0.632	0.622	0.630
US	0.623	0.638	0.626
Mean	0.645	0.621	0.641

Including all students, Figure A.5 shows that the international efficiency frontier for low ESCS values consists of three Spanish native speakers (one of whom has the lowest ESCS value in the sample), followed by one Portuguese and one Singaporean native speaker (with the highest average PISA value).[9]

Table 4.5 provides further within and between-country comparisons. $M^{\mathcal{E}}(\mathcal{H}_k)$ is the average score of the native students of country k, $M^{\mathcal{E}}(\mathcal{I}_k)$ is the average efficiency of its immigrant students, and $M^{\mathcal{E}}(\mathcal{E}_k)$ is the mean efficiency of all of students from country k with respect to the international frontier of all students.

Columns 1 and 2 show how well each group performs within each country, and allows within-country comparisons relative to the international frontier consisting of all students. Compared to their native peer group, immigrants perform best in Australia

[9]Using an output-oriented BBC-model with one input and one output, and variable returns to scale, the student with the highest output value must be efficient by construction.

(on average 1.977% better), followed by the United States (1.558%), and Singapore (1.415%). The countries where natives perform best compared to immigrants are Finland (on average 6.646% better), Sweden (5.476%), and Denmark (5.277%).

The results so far have been group-specific. Column 3, on the other hand, provides a comparisons of the efficiencies of the national educational systems. The values result from an international frontier and do not differentiate between natives and immigrants within countries. The mean inefficiencies show how much the average PISA scores of a country could be increased, if its educational system were to enable students to perform similarly to the most efficient international students with comparable ESCS endowments. The average inefficiencies over the entire sample are 35.9%. The country with the highest mean efficiency is Singapore. In Finland, Germany, Portugal, and Spain, the mean efficiency scores are also relatively large. The highest inefficiencies exist in Israel and Denmark, given the ESCS backgrounds of their respective students.

4.5.3 Differences Between Immigrants and Natives

Figure 4.5 – Arithmetic mean efficiency differences between native and immigrant students in each country relative to the international frontier

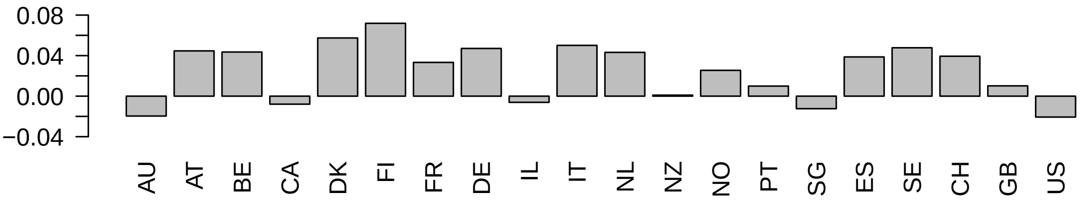


Figure 4.5 shows the differences between the arithmetic means of students with and without immigration background, relative to the countries' frontiers, providing an overview of the within-country differences. By including the ESCS as input, our analysis takes into account the socio-economic endowment of the students. Selection effects that result in high or low ESCS scores should therefore not influence the efficiency scores, given the ESCS input levels.

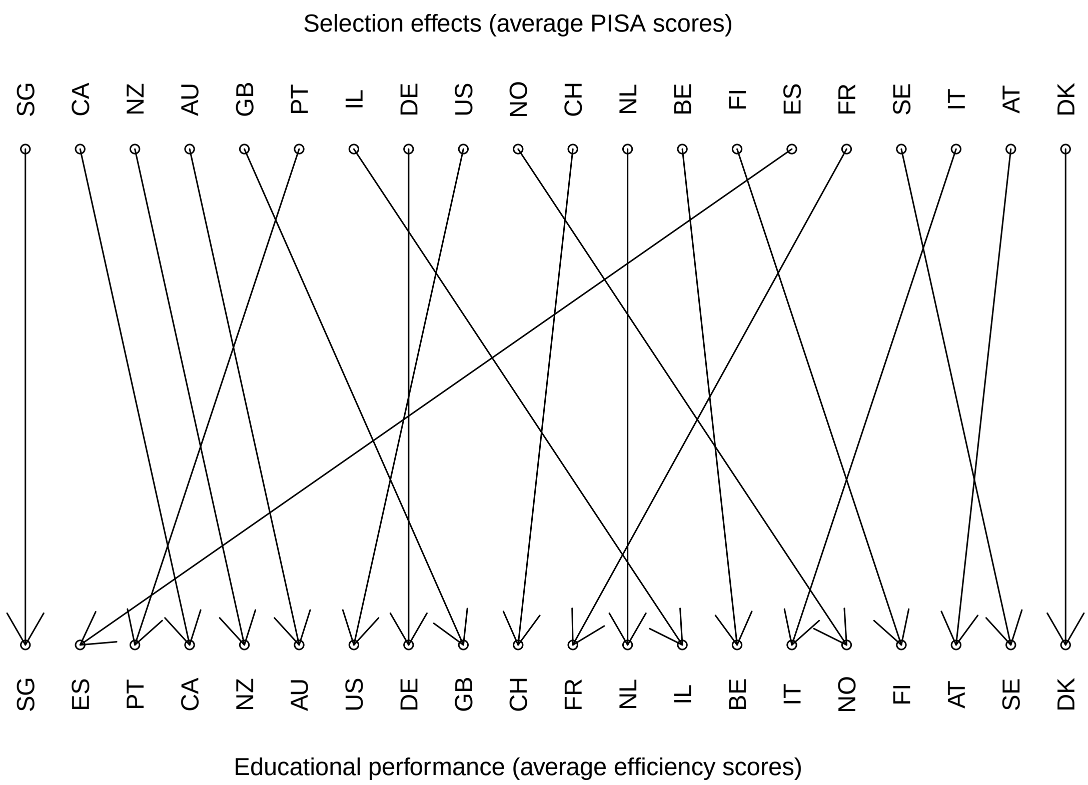
The efficiency gaps between the groups are smallest in Canada (−0.008), Israel (−0.006), New Zealand (0.001), and Portugal (0.010). In the other countries, the differences are greater than 1%. In all European countries and especially in Sweden (0.048), Denmark (0.057), and Finland (0.072), the immigrant students perform on average considerably worse than their native counterparts given their ESCS backgrounds. Finland is often regarded as a country with a superior educational system and integration success, but according to the efficiency scores the educational system in Finland is highly inefficient in closing the gap between natives and immigrants. Recent literature confirms these performance deficits of immigrants in Finland, taking into account background factors such as gender, grades, socio-economic background, home language and age of arrival in Finland (Kirjavainen (2015); Yeasmin et al. (2018)). However, these results have not yet attracted much attention in recent literature. Arikan et al. (2017), for example, claim that reducing the ESCS gap would close the performance gap between natives and immigrants in Finland, but our results indicate that especially an efficient use of the ESCS endowment is more important than the low ESCS levels (Arikan et al., 2017). We argue that the sole use of PISA results in native immigrant comparisons mainly reflects selection effects due to different immigration policies, rather than an analysis of the educational systems. Given the social structure of immigrants (and natives) we evaluate the educational systems according their ability to transform social endowments into good PISA results.

4.5.4 Selection Effects and Educational Efficiency

In the upper line of Figure 4.6, the countries are arranged in descending order according to their immigrants' average PISA scores. The order is solely based on the absolute performance of immigrants in PISA. Here, the efficient countries are characterised by a strict immigration policy, selecting immigrants who achieve the highest PISA levels. In the lower line, the countries are ordered according to their immigrants' average efficiency relative to the international frontier ($M^{\mathcal{E}}(\mathcal{I}_k)$). Thus, the countries are ranked according to their students' performance, given their ESCS endowments. Therefore, the impact of selection procedures is to a large extent controlled for, and the ranking reveals how successfully educational systems use the ESCS endowment.

The arrows indicate the rank changes. In both analyses, students perform best in Singapore and worst in Denmark. The ranks of all other countries change due to taking the ESCS endowment into account. Without regarding the ESCS endowment (upper ranking), countries with strongly selective immigration systems rank second to fifth. Taking into account the socio-economic backgrounds of their students (lower ranking), their ranks deteriorate to four, five, six and nine. This indicates that simple PISA score comparisons examine rather immigration policy and less so the efficiency of educational systems.

Figure 4.6 – Country rankings based on mean PISA scores of immigrants and of mean efficiency



Austria, Denmark and Sweden are the countries where immigrants perform worst according to their average DEA scores. Given the socio-economic background of their students, these countries could achieve much higher PISA scores, if they were to adapt their educational systems to those of the efficient countries.

Without including the ESCS as input, immigrants in Spain perform relatively poorly, but on average they perform very well regarding their efficiency. France (five ranks), Italy (four), and Portugal (three) are also countries which improve their rankings compared to the simple PISA score comparison. Regarding their socio-economic backgrounds, these three countries have relatively less favourable immigrant compositions, but their educational systems are relatively more efficient than in most other countries.

Our analysis shows that, despite a very large educational gap (see Table 4.3), the French school system performs well on average, because it at least partially compensates for the large differences in the socio-economic background of immigrants. While most countries lose up to three ranks, Israel (nine), and the United Kingdom (four) are far worse ranked, indicating relatively poorly performing educational systems. Immigrants in Denmark perform worst both when their ESCS endowment is considered and when it is not considered.

4.5.5 PISA Scores as Separate Outputs

The DEA allows the inclusion of separate outputs that are simultaneously included in the efficiency assessment. In this section, students are assessed on the basis of their ability to maximize the three PISA scores, given their ESCS endpoints. Model (34) allows for specialisation so that the efficiency of students focusing on a subset of the three abilities is adequately taken into account. Tables A.6 and A.7 in the appendix contain the decomposition of the efficiency results for national and international frontiers.

The efficiency scores of the average PISA score and three separated PISA scores as outputs are highly positively correlated. The Pearson correlation coefficient between the DEAs are 0.984 for $M^{\mathcal{H}}(\mathcal{H})$, 0.981 for $M^{\mathcal{I}}(\mathcal{I})$, and 0.984 for $M^{\mathcal{E}}(\mathcal{E})$. Table 4.6 provides the correlation coefficients of the efficiency scores based on the aggregated output and that of the three outputs for each country.

The inclusion of the separated PISA scores as outputs allows the DEA model to weight the outputs separately and thus to calculate overall higher efficiency scores. The similarity of the results to those of the previous analysis shows that students who perform well on average also perform quite well in the individual PISA subjects. These results confirm that immigrants in countries with restrictive immigration regime perform relatively better than in other countries and that immigrants in Spain, Portugal, and Singapore perform relatively best given their socio-economic endowments.

Table 4.6 – Correlation coefficients between the efficiency scores for the average PISA score and the three PISA scores as outputs

	$M^{\mathcal{H}_k}(\mathcal{H}_k)$	$M^{\mathcal{I}_k}(\mathcal{I}_k)$	$M^{\mathcal{E}_k}(\mathcal{H}_k)$	$M^{\mathcal{E}_k}(\mathcal{I}_k)$	$M^{\mathcal{E}_k}(\mathcal{E}_k)$	$M^{\mathcal{H}_k}(\mathcal{H})$	$M^{\mathcal{I}_k}(\mathcal{I})$	$M^{\mathcal{E}_k}(\mathcal{E})$
AU	0.985	0.983	0.985	0.984	0.985	0.986	0.986	0.986
AT	0.981	0.978	0.981	0.980	0.981	0.987	0.985	0.987
BE	0.988	0.990	0.988	0.990	0.989	0.988	0.989	0.988
CA	0.983	0.984	0.984	0.983	0.984	0.986	0.986	0.986
DK	0.985	0.974	0.985	0.979	0.984	0.983	0.980	0.983
FI	0.981	0.961	0.981	0.972	0.981	0.986	0.979	0.986
FR	0.985	0.985	0.985	0.988	0.985	0.988	0.991	0.988
DE	0.984	0.981	0.984	0.981	0.983	0.987	0.985	0.986
IL	0.981	0.985	0.981	0.981	0.981	0.986	0.987	0.986
IT	0.982	0.972	0.982	0.980	0.982	0.983	0.979	0.983
NL	0.987	0.983	0.987	0.986	0.987	0.990	0.989	0.990
NZ	0.985	0.981	0.984	0.980	0.983	0.986	0.986	0.986
NO	0.984	0.974	0.984	0.979	0.983	0.984	0.978	0.983
PT	0.989	0.975	0.990	0.986	0.990	0.989	0.986	0.989
SG	0.983	0.983	0.985	0.978	0.984	0.988	0.981	0.986
ES	0.987	0.977	0.987	0.986	0.987	0.988	0.989	0.989
SE	0.975	0.979	0.974	0.978	0.976	0.983	0.980	0.983
CH	0.986	0.985	0.987	0.988	0.988	0.983	0.985	0.984
GB	0.986	0.978	0.986	0.982	0.985	0.987	0.985	0.987
US	0.982	0.986	0.982	0.982	0.982	0.989	0.989	0.989

4.6 Conclusion

Our analysis focuses on the abilities of the national educational systems to integrate immigrants, given their socio-economic backgrounds. Country-specific means of efficiency scores based on national frontiers reveal that in Denmark, Finland, and Sweden, native students perform substantially better than immigrants. In Australia, Canada, Israel, Singapore, and the United States, immigrants are more efficient than their native peer group.

Relative to the international frontier consisting of all students and compared to their native peer groups, immigrants in Australia, Singapore, and the United States perform relatively best. The opposite is true in Finland, Sweden, and Denmark.

Even if the differences in the socio-economic endowment of the students are taken into account, differences between natives and immigrants persist. According to PISA scores, as well as the efficiency scores, in most countries with more selective immigration regimes, immigrants perform on average similar or even better than natives. The persistent differences are somewhat surprising, as the broad ESCS should capture the most relevant socio-economic factors.

We find that the Spanish educational system is relatively best in increasing immigrants' performance, and Israel's system is worst, given the respective socio-economic backgrounds of their immigrants. Australia, Canada, the United Kingdom, and New Zealand are countries with selective immigration policies, which attract immigrants who perform relatively better or almost as well as their natives. If, however, the socio-economic backgrounds are taken into account, the immigrants in these countries perform on average worse than in Spain and Portugal. The latter have low PISA values, but highly efficient education systems.

The result that countries with relatively selective immigrant policies perform not only well in absolute PISA scores, but are also quite efficient given their ESCS input levels, is truly astonishing. This result implies that the selection process not only affects ESCS levels, but also the immigrant capacity to use their endowments efficiently.

A.1 Supplementary Results

Figure A.1 – Country-level average PISA scores relative to the average ESCS scores and mean absolute deviations from the median ESCS scores; left: average scores, right: average PISA scores and mean absolute deviations from the median ESCS scores

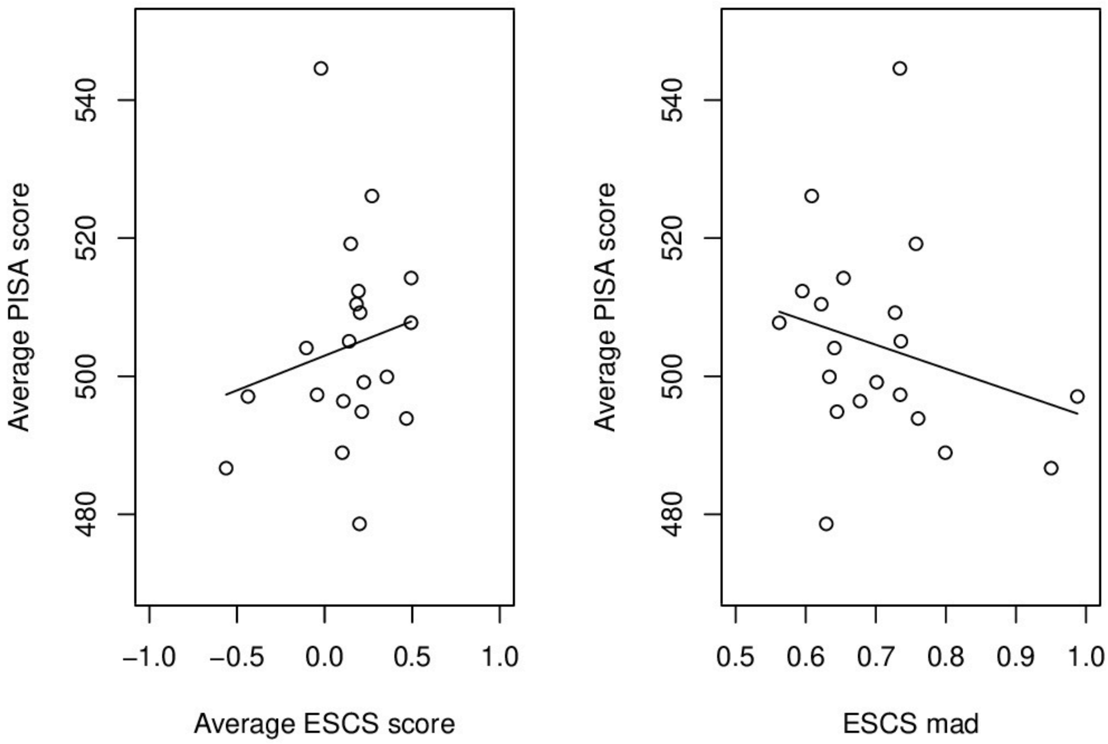


Table A.1 – Excluded outliers

	Amount	Percentage
Australia	151	1.079
Austria	67	0.966
Belgium	77	0.815
Canada	207	1.066
Denmark	67	0.959
Finland	67	1.153
France	52	0.875
Germany	57	1.012
Israel	72	1.108
Italy	110	0.971
Netherlands	59	1.108
New Zealand	49	1.131
Norway	56	1.059
Portugal	65	0.900
Singapore	60	0.985
Spain	53	0.794
Sweden	63	1.186
Switzerland	44	0.759
United Kingdom	115	0.851
United States of America	51	0.905

Table A.2 – Descriptive results of the PISA mathematics scores

Countries	Group	Min	Max	Median	Mean	Sd.	n
AU	Nat	166.883	800.542	484.679	483.382	91.113	10744
	Mig	225.436	757.799	505.180	504.928	91.018	2651
AT	Nat	193.875	797.841	516.602	512.318	88.728	5533
	Mig	170.358	731.659	446.329	448.960	84.998	1242
BE	Nat	197.540	818.559	529.075	523.481	90.932	7684
	Mig	237.039	727.384	468.027	468.323	90.944	1445
CA	Nat	218.370	807.652	504.818	504.624	81.876	14555
	Mig	261.416	810.875	516.563	517.336	83.974	4057
DK	Nat	266.826	751.205	518.059	516.057	76.720	5224
	Mig	213.486	700.433	455.982	457.803	74.074	1567
FI	Nat	245.949	751.693	516.953	515.419	78.857	5495
	Mig	208.933	695.899	457.889	466.733	88.732	200
FR	Nat	197.621	765.411	515.105	507.942	91.081	5089
	Mig	200.976	723.209	455.736	456.625	96.992	706
DE	Nat	227.368	803.717	524.496	522.885	85.158	4614
	Mig	218.142	715.613	475.107	475.467	84.294	881
IL	Nat	147.251	776.097	479.573	476.417	98.822	5223
	Mig	122.084	728.409	476.589	471.336	103.668	1023
IT	Nat	171.668	792.029	505.564	503.999	87.043	10199
	Mig	224.551	674.787	459.160	458.088	82.293	867
NL	Nat	203.188	783.104	528.884	523.470	87.652	4587
	Mig	229.374	689.471	473.280	470.753	87.335	504
NZ	Nat	248.467	768.730	497.402	497.696	86.419	3031
	Mig	249.591	766.447	511.344	508.374	94.368	1075
NO	Nat	240.418	748.189	507.671	507.450	82.138	4535
	Mig	264.560	702.677	469.598	471.186	77.707	616
PT	Nat	157.556	783.224	485.353	483.819	95.255	6647
	Mig	214.974	702.556	470.405	474.209	93.780	416
SG	Nat	242.225	847.230	555.887	552.496	93.070	4734
	Mig	293.152	842.615	591.471	584.756	88.716	1164
ES	Nat	244.435	752.071	499.491	496.916	79.837	5808
	Mig	226.047	690.859	454.489	455.834	79.728	664
SE	Nat	206.226	771.176	507.830	507.146	82.513	4311
	Mig	221.587	712.981	451.156	452.016	84.530	819
CH	Nat	254.399	800.594	539.532	536.024	87.237	3907
	Mig	183.867	779.058	486.424	488.177	90.664	1711
GB	Nat	185.785	769.712	495.911	494.442	84.650	11329
	Mig	205.712	729.779	492.467	488.871	91.280	1607
US	Nat	204.268	766.785	477.317	477.480	85.992	4153
	Mig	227.198	720.594	458.297	457.413	84.894	1215

Table A.3 – Descriptive results of the PISA science scores

Countries	Group	Min	Max	Median	Mean	Sd.	n
AU	Nat	191.045	833.478	504.560	501.608	100.573	10744
	Mig	227.272	803.415	519.730	515.728	100.702	2651
AT	Nat	227.032	826.725	513.994	511.576	91.553	5533
	Mig	204.623	741.154	443.153	447.522	86.996	1242
BE	Nat	231.127	813.512	525.989	518.910	94.424	7684
	Mig	210.139	711.469	459.965	461.836	96.402	1445
CA	Nat	213.671	821.825	521.184	519.625	88.046	14555
	Mig	250.632	828.142	526.060	523.003	91.824	4057
DK	Nat	202.866	758.113	507.808	507.843	86.583	5224
	Mig	219.308	731.502	430.812	432.864	85.046	1567
FI	Nat	232.233	852.902	540.267	537.439	92.260	5495
	Mig	239.698	735.533	462.255	462.091	96.635	200
FR	Nat	226.574	784.065	517.705	511.625	95.790	5089
	Mig	189.502	746.084	451.341	453.775	100.996	706
DE	Nat	253.093	810.494	536.333	531.437	92.732	4614
	Mig	188.335	781.910	464.048	467.684	92.284	881
IL	Nat	181.542	825.603	476.171	476.777	102.558	5223
	Mig	116.428	742.017	471.508	470.807	105.217	1023
IT	Nat	209.030	772.320	502.760	498.819	86.628	10199
	Mig	219.710	711.228	451.098	449.931	82.727	867
NL	Nat	233.621	786.983	525.554	520.158	96.996	4587
	Mig	225.921	685.571	468.732	461.069	93.907	504
NZ	Nat	245.909	795.682	525.200	521.620	97.949	3031
	Mig	185.943	800.925	521.780	518.146	108.474	1075
NO	Nat	204.571	819.329	509.873	508.423	92.275	4535
	Mig	247.333	699.818	458.451	460.458	87.591	616
PT	Nat	195.810	781.855	491.015	490.120	91.883	6647
	Mig	252.490	726.540	482.707	488.197	87.464	416
SG	Nat	248.752	870.020	549.887	542.634	101.790	4734
	Mig	265.541	835.631	580.107	574.010	97.278	1164
ES	Nat	230.400	740.724	506.647	503.965	82.564	5808
	Mig	226.049	724.153	457.635	460.772	85.888	664
SE	Nat	166.987	845.611	511.242	508.842	95.297	4311
	Mig	165.070	739.310	437.695	440.494	96.810	819
CH	Nat	242.234	771.501	525.921	522.781	90.129	3907
	Mig	219.737	750.943	459.952	465.410	94.428	1711
GB	Nat	214.389	807.431	511.242	509.452	93.702	11329
	Mig	271.842	792.647	495.884	497.483	97.426	1607
US	Nat	238.079	806.788	504.314	505.552	95.449	4153
	Mig	239.927	737.466	475.062	477.650	91.942	1215

Table A.4 – Descriptive results of the PISA reading scores

Countries	Group	Min	Max	Median	Mean	Sd.	n
AU	Nat	96.374	832.034	497.103	491.976	101.709	10744
	Mig	169.719	800.578	519.464	512.925	102.203	2651
AT	Nat	131.917	762.012	506.123	499.912	94.060	5533
	Mig	183.376	712.661	447.364	447.287	94.479	1242
BE	Nat	205.263	807.437	524.693	515.790	93.568	7684
	Mig	184.550	720.464	468.211	465.703	96.005	1445
CA	Nat	223.017	798.814	520.883	517.968	87.553	14555
	Mig	219.633	812.414	526.823	523.367	91.836	4057
DK	Nat	213.575	781.762	509.405	505.788	83.225	5224
	Mig	192.824	713.168	440.661	443.598	83.058	1567
FI	Nat	200.763	778.198	538.810	533.374	86.657	5495
	Mig	194.514	716.262	474.596	465.347	99.653	200
FR	Nat	169.224	850.750	524.294	515.017	103.111	5089
	Mig	181.602	753.531	474.086	466.058	110.194	706
DE	Nat	201.098	808.879	538.702	530.944	92.075	4614
	Mig	171.798	749.074	481.730	479.852	97.112	881
IL	Nat	123.358	861.854	496.734	490.971	109.097	5223
	Mig	161.619	781.152	491.460	482.487	107.601	1023
IT	Nat	218.244	766.463	507.069	502.664	85.794	10199
	Mig	185.012	676.707	448.731	444.028	86.843	867
NL	Nat	151.081	778.437	521.125	513.710	96.334	4587
	Mig	176.822	741.653	470.805	465.394	93.320	504
NZ	Nat	169.027	810.103	520.348	516.821	99.665	3031
	Mig	231.686	811.821	520.667	514.233	107.019	1075
NO	Nat	183.062	807.994	528.604	523.727	94.026	4535
	Mig	239.430	777.681	494.664	490.283	93.814	616
PT	Nat	159.732	773.769	492.554	488.197	91.653	6647
	Mig	216.284	723.842	500.284	495.196	91.901	416
SG	Nat	230.626	818.352	529.974	522.078	97.500	4734
	Mig	162.940	782.729	560.674	551.459	92.967	1164
ES	Nat	161.767	767.832	511.244	505.802	81.443	5808
	Mig	162.794	708.353	475.377	470.201	90.464	664
SE	Nat	181.227	826.607	520.679	515.129	94.084	4311
	Mig	127.785	756.546	465.677	461.219	99.536	819
CH	Nat	204.781	771.608	510.955	506.492	90.407	3907
	Mig	192.295	747.702	456.481	458.481	94.613	1711
GB	Nat	213.622	846.678	503.169	501.540	89.369	11329
	Mig	186.130	794.240	487.340	489.179	94.946	1607
US	Nat	198.408	772.617	508.019	503.818	95.449	4153
	Mig	181.668	742.142	491.332	487.206	98.431	1215

Table A.5 – PISA scores correlation coefficients

	Mathematics-Reading	Mathematics-Science	Reading-Science
AU	0.789	0.879	0.872
AT	0.794	0.886	0.864
BE	0.834	0.891	0.897
CA	0.766	0.878	0.865
DK	0.769	0.874	0.863
FI	0.783	0.863	0.861
FR	0.828	0.899	0.892
DE	0.796	0.885	0.856
IL	0.823	0.887	0.892
IT	0.743	0.849	0.829
NL	0.860	0.899	0.891
NZ	0.772	0.884	0.866
NO	0.778	0.885	0.836
PT	0.806	0.889	0.862
SG	0.829	0.890	0.908
ES	0.756	0.888	0.847
SE	0.756	0.881	0.828
CH	0.801	0.882	0.871
GB	0.783	0.879	0.869
US	0.826	0.890	0.889

Figure A.2 – Mathematics scores distributions among natives (straight line) and immigrants (dashed line)

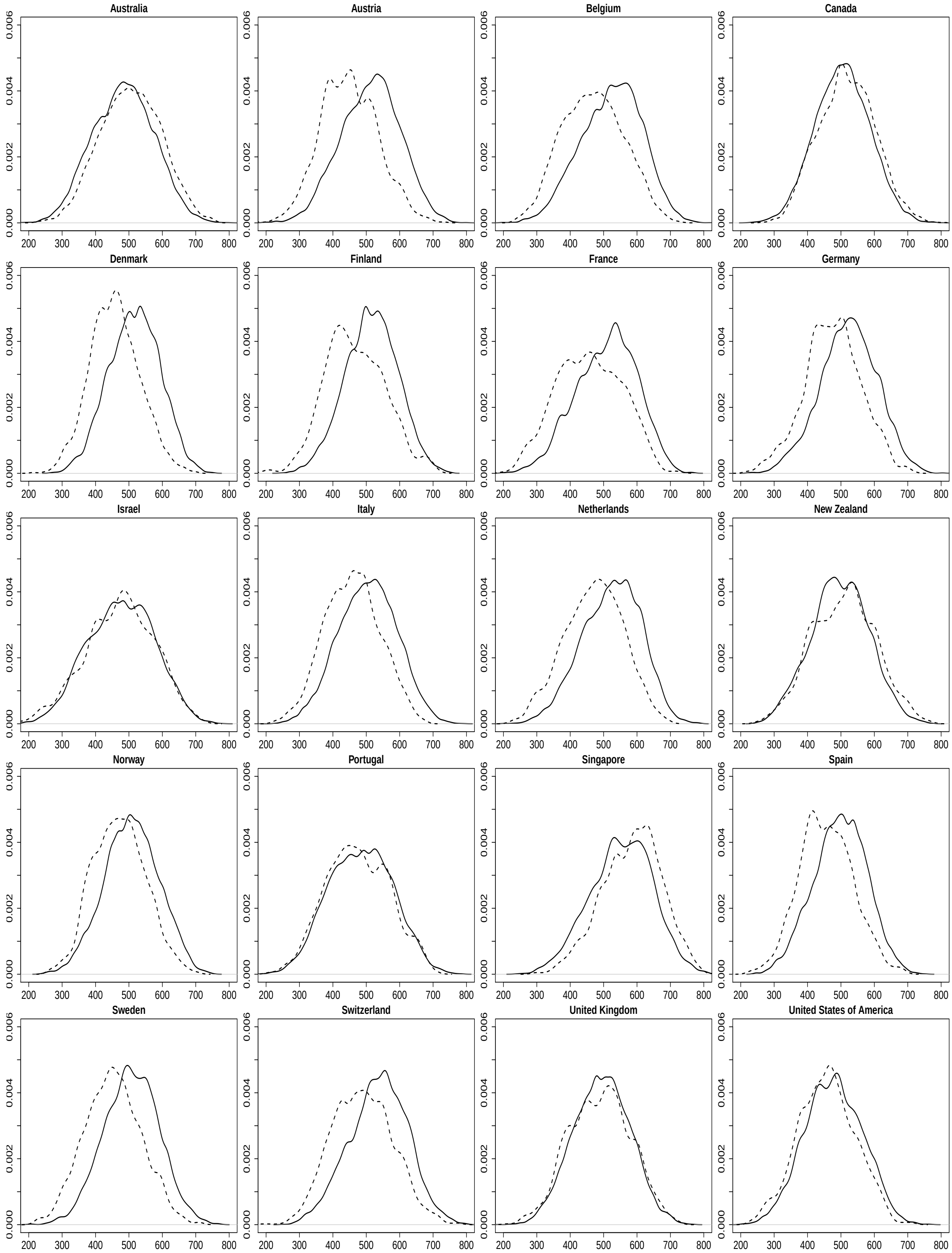


Figure A.3 – Science scores distributions among natives (straight line) and immigrants (dashed line)

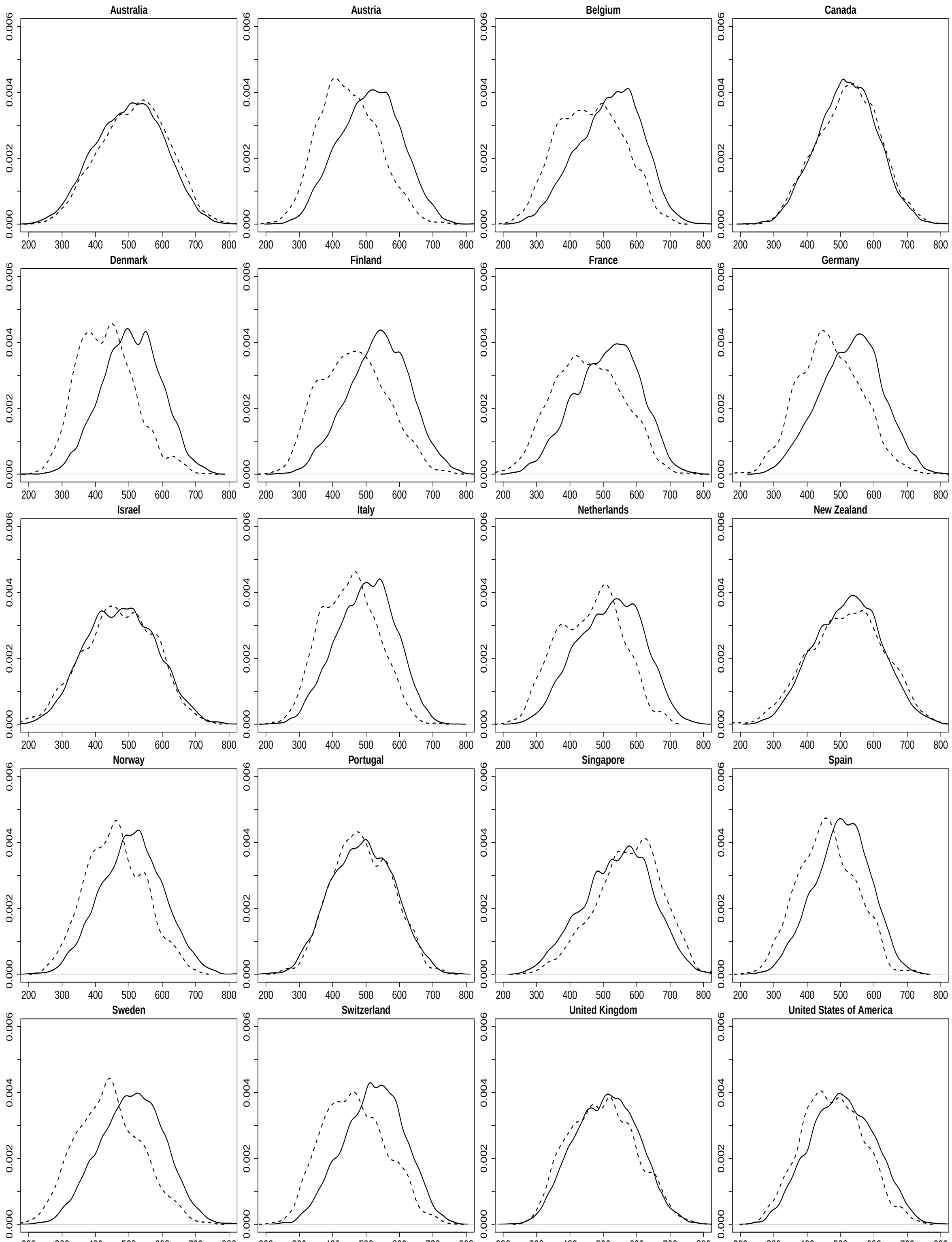


Figure A.4 – Read scores distributions among natives (straight line) and immigrants (dashed line)

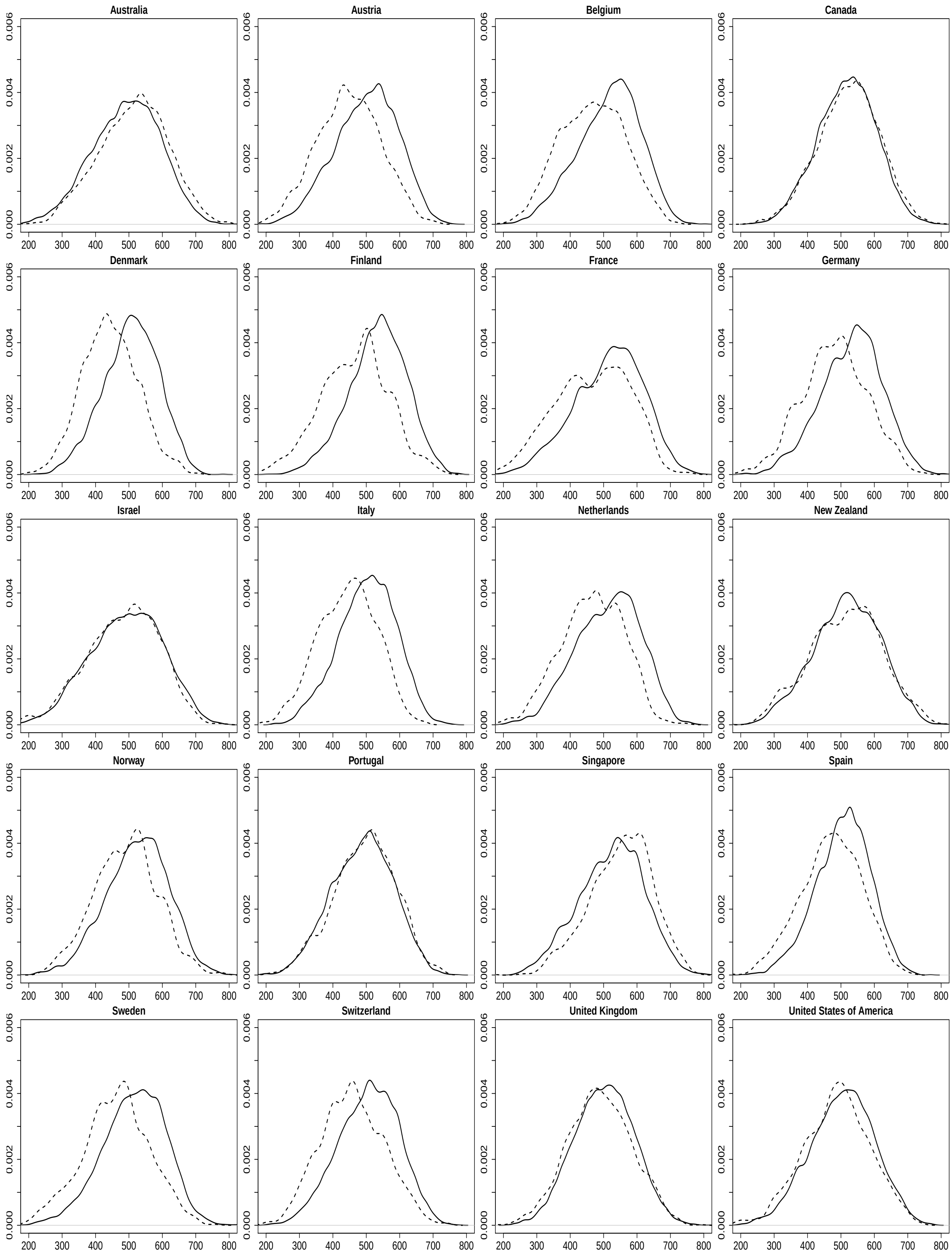
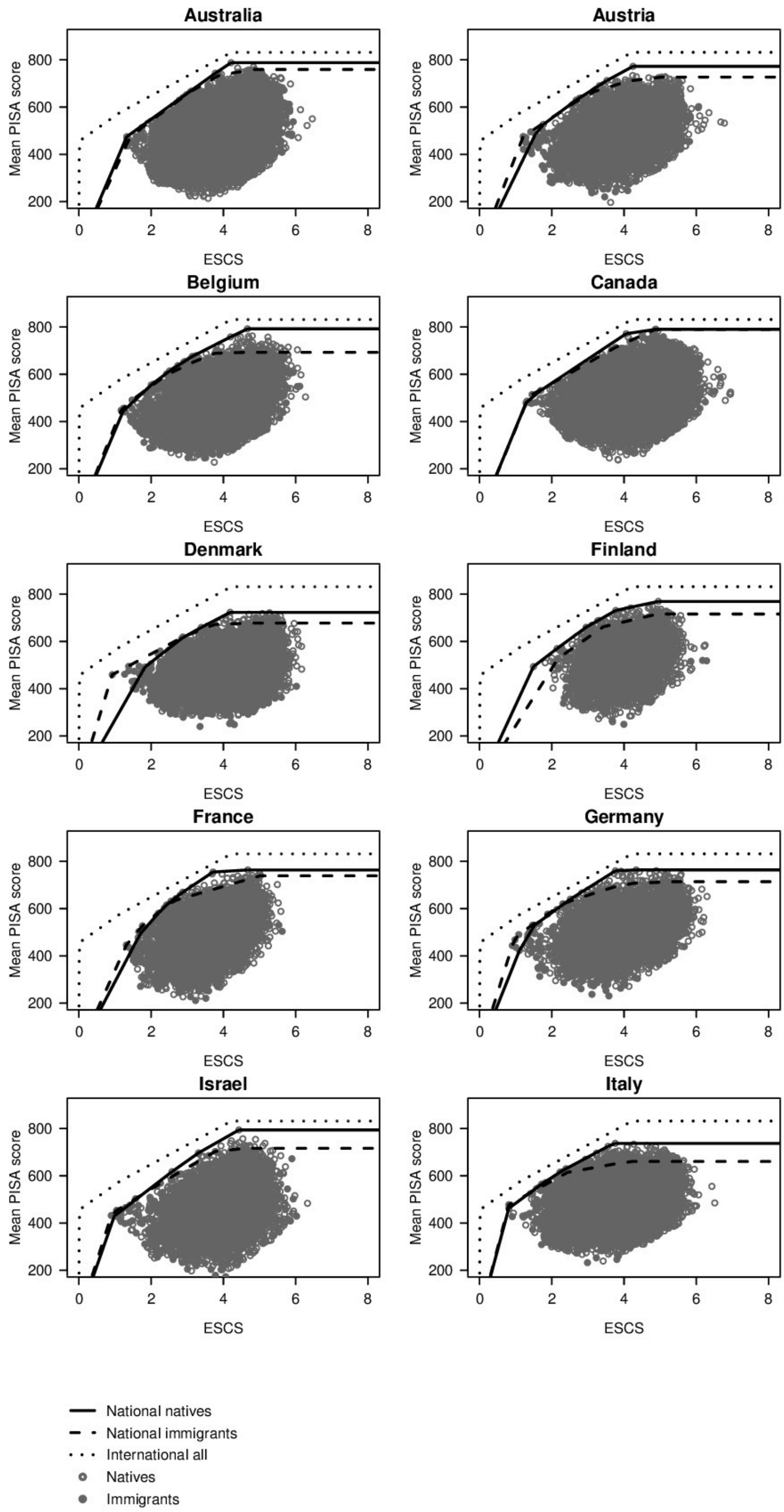


Figure A.5 – Efficiency frontiers



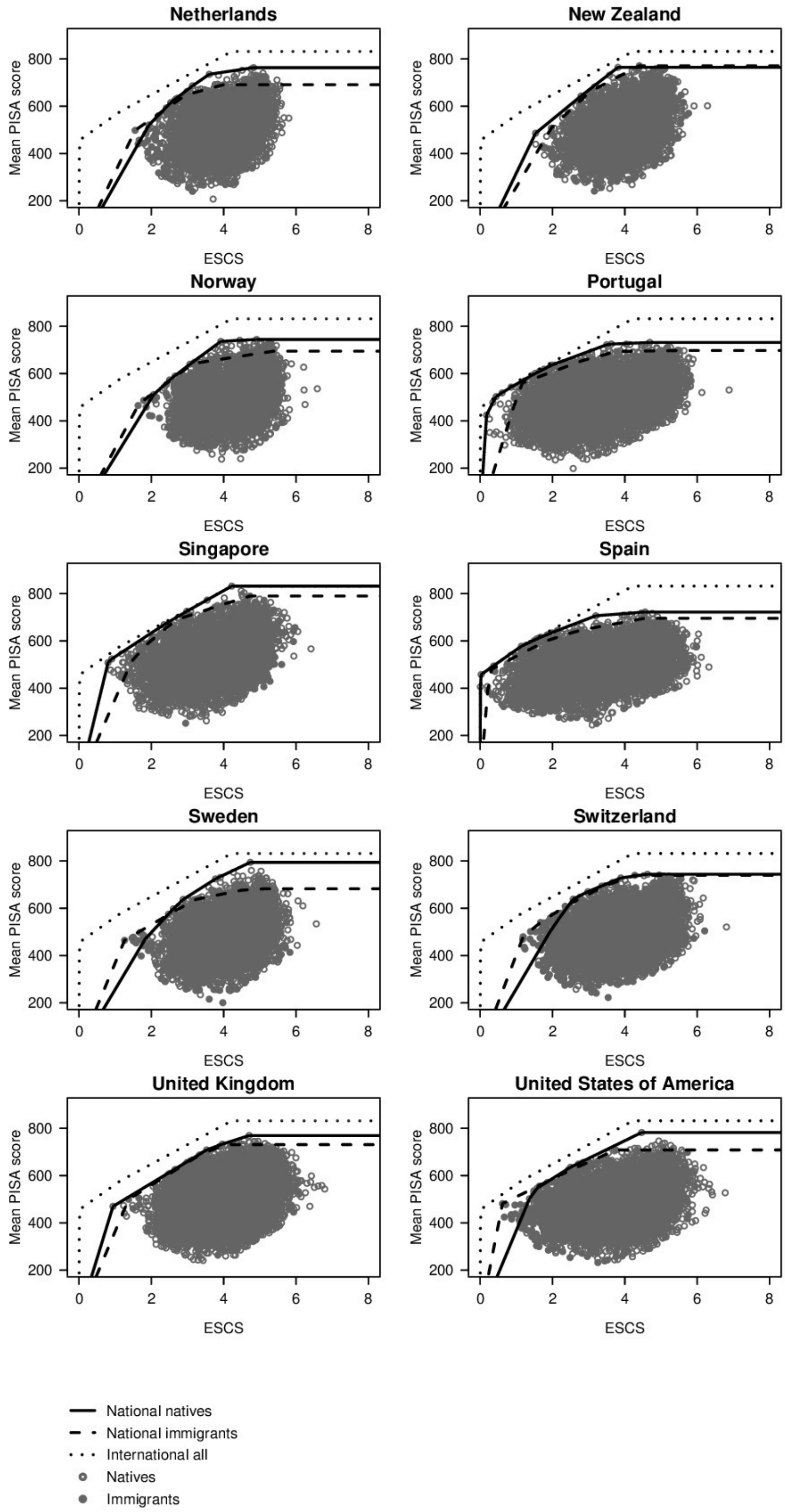

Netherlands
New Zealand
Norway
Portugal
Singapore
Spain
Sweden
Switzerland
United Kingdom
United States of America
Mean PISA score
ESCS
200
400
600
800
0
2
4
6
8
National natives
National immigrants
International all
Natives
Immigrants

Table A.6 – Decomposition, national students and national frontiers, three PISA scores as outputs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$M^{\mathcal{H}_k}(\mathcal{H}_k)$	$M^{\mathcal{I}_k}(\mathcal{I}_k)$	$M^{\mathcal{E}_k}(\mathcal{H}_k)$	$M^{\mathcal{E}_k}(\mathcal{I}_k)$	$M^{\mathcal{E}_k}(\mathcal{E}_k)$	$\left(\frac{M^{\mathcal{I}_k}(\mathcal{I}_k)}{M^{\mathcal{I}_k}(\mathcal{H}_k)} \cdot \frac{M^{\mathcal{H}_k}(\mathcal{I}_k)}{M^{\mathcal{H}_k}(\mathcal{H}_k)}\right)^{\frac{1}{2}}$	$\frac{M^{\mathcal{E}_k}(\mathcal{H}_k)}{M^{\mathcal{E}_k}(\mathcal{I}_k)}$
AU	0.681	0.725	0.680	0.701	0.684	0.970	0.970
AT	0.715	0.704	0.715	0.668	0.706	1.074	1.071
BE	0.718	0.721	0.718	0.672	0.710	1.076	1.068
CA	0.700	0.718	0.698	0.708	0.700	0.986	0.986
DK	0.744	0.718	0.742	0.688	0.730	1.081	1.079
FI	0.744	0.739	0.744	0.676	0.741	1.101	1.101
FR	0.716	0.717	0.716	0.679	0.711	1.059	1.055
DE	0.731	0.725	0.730	0.686	0.723	1.069	1.065
IL	0.670	0.714	0.669	0.675	0.670	0.992	0.991
IT	0.722	0.736	0.722	0.671	0.718	1.080	1.075
NL	0.726	0.734	0.725	0.683	0.721	1.072	1.062
NZ	0.714	0.730	0.711	0.713	0.712	0.999	0.997
NO	0.724	0.753	0.724	0.703	0.721	1.034	1.029
PT	0.717	0.761	0.717	0.713	0.716	1.009	1.005
SG	0.717	0.770	0.716	0.736	0.720	0.977	0.974
ES	0.742	0.751	0.742	0.703	0.738	1.053	1.054
SE	0.710	0.719	0.709	0.657	0.700	1.080	1.078
CH	0.738	0.710	0.735	0.696	0.723	1.049	1.057
GB	0.709	0.720	0.708	0.701	0.707	1.010	1.010
US	0.691	0.734	0.690	0.708	0.694	0.978	0.975
Mean	0.716	0.730	0.716	0.692	0.712	1.037	1.035

Table A.7 – Decomposition, national students and international frontier, three PISA scores as outputs

	(1)	(2)	(3)	(4)	(5)	(6)
	$M^{\mathcal{E}}(\mathcal{H}_k)$	$M^{\mathcal{E}}(\mathcal{I}_k)$	$M^{\mathcal{E}}(\mathcal{E}_k)$	$\frac{M^{\mathcal{E}}(\mathcal{H}_k)}{M^{\mathcal{E}}(\mathcal{E})}$	$\frac{M^{\mathcal{E}}(\mathcal{I}_k)}{M^{\mathcal{E}}(\mathcal{E})}$	$\frac{M^{\mathcal{E}}(\mathcal{E}_k)}{M^{\mathcal{E}}(\mathcal{E})}$
AU	0.635	0.656	0.639	0.968	1.000	0.974
AT	0.655	0.606	0.646	0.999	0.923	0.984
BE	0.668	0.617	0.659	1.018	0.941	1.005
CA	0.654	0.664	0.656	0.997	1.012	1.000
DK	0.650	0.597	0.637	0.991	0.910	0.972
FI	0.677	0.612	0.675	1.033	0.933	1.029
FR	0.667	0.625	0.662	1.017	0.952	1.009
DE	0.684	0.639	0.676	1.042	0.974	1.031
IL	0.620	0.621	0.620	0.946	0.946	0.946
IT	0.662	0.620	0.659	1.010	0.944	1.004
NL	0.663	0.619	0.659	1.011	0.943	1.004
NZ	0.661	0.660	0.661	1.007	1.007	1.007
NO	0.652	0.627	0.649	0.994	0.955	0.989
PT	0.674	0.666	0.673	1.027	1.016	1.026
SG	0.710	0.731	0.714	1.082	1.114	1.088
ES	0.684	0.662	0.682	1.043	1.010	1.039
SE	0.654	0.601	0.645	0.997	0.916	0.983
CH	0.672	0.635	0.660	1.024	0.968	1.007
GB	0.646	0.638	0.645	0.985	0.973	0.984
US	0.637	0.650	0.640	0.970	0.991	0.975
Mean	0.661	0.637	0.658	1.008	0.971	1.003

5

Higher Education Institution Efficiency in Germany and the United Kingdom

Higher Education Institution Efficiency in Germany and the United Kingdom

Gerald Fugger

University Duisburg-Essen, 45141 Essen, Germany.

Abstract

This paper assesses the efficiency of 91 Higher Education Institutions (HEIs) from the United Kingdom (UK) and Germany using non-radial Data Envelopment Analysis (DEA) models. The HEIs' efficiency is calculated relatively to country-specific efficiency frontiers, to an international frontier, and to the frontier of the respective other country using super-efficient non-radial DEA models. Within the countries, 27 out of 46 German HEIs are identified as efficient and 26 out of 45 UK HEIs. In the international comparison, UK HEIs are on average more efficient than their German counterparts. Descriptive results and super-efficient non-radial DEA models indicate country-specific input-output structures.

Keywords: Data Envelopment Analysis, Higher education assessment, Higher education institutions, Super-efficiency
JEL Classification: C14 C52 C61 I22 I23

5.1 Introduction

In the field of higher education there is a long tradition of auditing costs and efficiency of higher education institution (HEIs). Policymakers have strong incentives to focus on minimising costs and HEIs on maximising their outputs. Various benchmarking techniques have been used to evaluate the efficiency of HEIs, mainly Data Envelopment Analysis (DEA) (De Witte et al., 2017). Athanassopoulos et al. (1997) identify over-resourced HEIs and Veiderpass et al. (2016) use a cost minimisation model to compare the input saving efficiency of HEIs. The majority of higher education efficiency assessments concentrate on output maximisation of HEIs. The HEI targets or missions can be grouped into three outputs: research, teaching, and innovation (Frenken et al., 2017).

Performance evaluation of HEIs is mainly conducted within countries, e.g. by Gawellek et al. (2016) for German HEIs and by Chuanyi et al. (2016) for Chinese HEIs. Some studies provide country comparisons like Lehmann et al. (2018) or Veiderpass et al. (2016) for German and Italian HEIs as well as Agasisti et al. (2016) for Dutch and Italian HEIs. Only few studies assess HEIs from several countries, as data are limited (Veiderpass et al., 2016). Data constraints determine that efficiency evaluations are conducted either within a particular HEI at departmental level (Aziz et al., 2013; Göken et al., 2015) or between several HEIs at HEI level (Rhaiem, 2017).

The efficiency assessment of HEIs requires several operational decisions to capture their input-output structures. Some HEIs use more personnel, and others may have more capital resources available. While some researchers may focus on top publications, others may prefer a broader publication strategy (Bornmann et al., 2015). Furthermore, some data can not be classified as inputs or outputs. For instance, third-party funding is perceived as input by some (Fandel, 2007) and as output by others (Agasisti et al., 2016). Most studies use radial efficiency models that allow the exclusion of quantities in the efficiency calculation of the HEIs, which may result in an overestimation of efficiency. In addition, radial DEA models do not account for substitution among the inputs or outputs (Khalili et al., 2010).

While there is sophisticated literature on the efficiency of German HEIs the efficiency results are barely decomposed or compared with similar higher education systems (Gralka et al., 2019; Wohlrabe et al., 2019a). This paper evaluates the efficiency of German and UK HEIs and assesses their efficiency results relative to different efficiency frontiers. The operational challenges are identified and discussed to enable a conscientious efficiency assessment. The German and UK HEIs are compared in several publications such as by (Wolszczak-Derlacz, 2017). UK HEIs are perceived as some of the best performing in Europe and are quite similar in orientation and endowment to the German HEIs (Johnes, 2006; Thanassoulis et al., 2011). However, the UK higher education sector is less federal and has a more performance-based funding structure (Hüther et al., 2018). This can result in country-specific, unique input-output structures (Seeber et al., 2019). In the following, the unique input-output structures are first identified by descriptive results, and then also confirmed by super-efficient Slack-based measurement (SBM) models that benchmark the HEIs of one country against the HEIs of the other.

The introduction is followed by a literature review on the efficiency of higher education institutions, followed by an operational discussion, and in the fourth section, the methodology is outlined. The results of the efficiency analyses and their decomposition are discussed in the fifth section before the conclusion.

5.2 Literature Review

This section provides a review of the international academic literature that has examined ways to evaluate the efficiency of HEIs. Rhaiem (2017) and De Witte et al. (2017) have both provided extensive literature reviews of the studies evaluating HEI efficiency, and these publications, their inputs, outputs, and models are presented in Table B.13.

The inputs in the efficiency assessment reflect all resources used by the assessed HEIs and can be grouped into capital usage, the number of personnel employed, and additional resources. Capital usage is usually approximated by examining general academic expenses and third-party funds (Athanassopoulos et al., 1997). For example, Veiderpass et al. (2016) use total income and Wolszczak-Derlacz (2017) use total revenue to account for capital use. The number of personnel employed is typically split into counts of academic and non-academic personnel (Veiderpass et al., 2016). The tasks of the academic staff are mostly teaching and research, while the non-academic staff are mainly tasked with assisting academics in a supporting role, which can be important for the general operation of the HEIs (Fandel, 2007). Wages are only taken into account if no data on the number of employees are available, since they are weighted floating quantities (Gralka et al., 2019). Besides variables relating to capital and personnel, further information can be included to account for the various inputs HEI are using (Athanassopoulos et al., 1997). For example, Johnes (2006) includes as inputs the number of graduates and undergraduate students and assumes that HEIs should maximise their outputs' given the numbers of students they have to teach. However, most HEIs receive funds related to the number of their students they teach, meaning capital inputs are already partly incorporated into the analyses by the size of the HEI.

The outputs reflect the three missions adopted by most HEIs: research, teaching, and innovation. The number of academic papers that are published, which can be a common indicator for research, can only measure the quantity and not the quality of the research. Weighted publication indices may be preferable, or the number of publications in top ranking journals, which can help to account for the quality of the research (Athanassopoulos et al., 1997; Lehmann et al., 2018). While the numbers of graduates and doctorates are included as teaching outputs in most studies. Veiderpass et al. (2016) believe these to be insufficient indicators compared to the numbers of publications. Gralka et al. (2019) adds that the amount of research grants and the number of publications can be used interchangeably. However, most analyses perceive third-party funding to be used as an input. Data on innovation is rarely available and is often not part of the analyses. If data is available, the value (Lehmann et al., 2018) or the number (Chuanyi et al., 2016) of patents granted or the number of publications that have been co-authored with one or more industrial organisations (Frenken et al., 2017) can be used.

Chuanyi et al. (2016) compare the results of radial and non-radial models and conclude that non-radial models are better suited to assess the efficiency of HEIs. Moreno et al. (2018) prefer non-radial models such as the SBM which can consider slacks and which do not require an equiproportional increase in all considered outputs by allowing for substitution among them. Output-oriented models can assess how effectively HEIs maximise their objectives in terms of inputs. Although political decision makers have incentives to minimise the costs to HEIs, external pressures such as international league tables and growing student choice lead them to prioritise a maximisation of their outputs to showcase their research potential, thereby justifying the use of output-oriented models (Rhaiem, 2017).

Without additional weight restrictions, radial DEA models may assign zero weights to inputs and outputs of certain HEIs. Specialised HEIs may be calculated as efficient because only subsets of the data are included in their efficiency calculation (Cooper et al., 2007). Athanassopoulos et al. (1997) implement additional weight restrictions to prevent zero weights and to allow for a meaningful interpretation of efficiency scores. Additional weight restrictions are not necessary for SBMs as these models have already implemented lower weight restrictions (Moreno et al., 2018). In the context of higher education, Chuanyi et al. (2016) show that the results of radial and non-radial DEA models are overall consistent and that the SBM calculates HEIs on average as being less efficient than radial models.

HEIs can be benchmarked against various efficiency measures. For example, Wolszczak-Derlacz (2017) analyses the efficiency scores of HEIs from several countries relative to an intercontinental efficiency frontier and to country-specific frontiers. The results enable various to be made within and between countries. However, the use of radial models and compromises in data selection in order to take into account as many countries as possible partially limits the usefulness of her results.

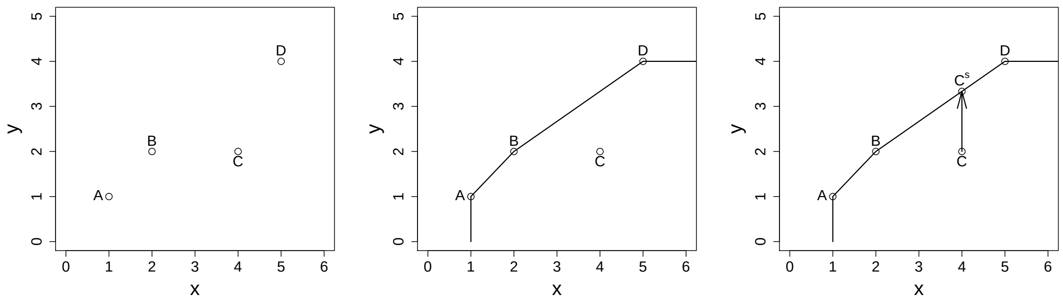
5.3 Operationalisation

Aside from technical efficiency, the efficiency results of HEIs also depend on the level of analysis, the data selected, and the models chosen. This section of the paper discusses the operationalisation of efficiency analyses of HEIs and their implementation in the literature. First, the efficiency assessment in DEA is explained using the radial BCC model introduced by Banker et al. (1984), before moving on to the methodology used in the DEA model.

5.3.1 Efficiency Assessments using DEA

DEA assesses the efficiency of HEIs by comparing their relative productivity. An efficiency score is obtained by benchmarking benchmarking each individual HEI against what is considered to be a group of the most productive institutions. The most productive HEIs form the efficiency frontier enveloping the data (hence the name DEA) and their input-output combinations represent best practice. Figure 5.1 illustrates the idea of DEA for the artificial HEIs A, B, C, and D. In this example, the HEIs consume one input (x) and produce one output (y). Their efficiency scores are calculated using the output-oriented BCC model. The left panel shows the data points. Their output to

Figure 5.1 – Efficiency assessment example



(a) Data set **(b)** Efficiency frontier **(c)** Efficiency assessment

input ratios are $\frac{1}{1} = 1$ for A, $\frac{2}{2} = 1$ for B , $\frac{4}{2} = 0.5$ for C, and $\frac{4}{5} = 0.8$ for D. Assuming constant returns to scale, A and B are efficient and C and D are inefficient. DEA allows for the assumption of variable returns to scale if it is assumed that HEIs do not scale linearly for different input quantities.[1] The middle panel of Figure 5.1 illustrates how the efficient HEIs A, B, and D (assuming variable returns to scale) form the efficiency frontier. The right panel displays how the model evaluates the inefficiency of C. C^s is calculated based on the input of C and shows how much output C could produce if it were efficient (output-oriented model). C^s is a combination of B's (to one third) and D's (two thirds) input-output structures. C^s displays best practice output of C. The efficiency score of C is calculated as the ratio of actual to possible output: $\frac{3.333}{2} = 1.667$. For simplicity, the reciprocal efficiency score is usually used in the calculation of output-oriented models. After the transformation, the efficiency scores are bounded between zero and one. Efficient HEIs have an efficiency score of one. The efficiency scores of C is $\frac{1}{1.667} = 0.6$. The BCC model is formally presented below.

5.3.2 Analysis Level

Higher educational efficiency assessment studies are mostly conducted on the HEI-level or for departments of one specific HEI. Foladi et al. (2019), for example, evaluate the Urmia University in Iran. Their results indicate that the efficiency of HEI departments within the same HEI can differ strongly. The departments have a unique input-output

[1]Constant returns to scale reflect the assumption that outputs will change by the same proportion as inputs are changed (e.g. a 50% increase of all inputs will increase outputs by 50%). Variable returns to scale encompass both increasing (an over-proportional increase) and decreasing returns (an under-proportional increase) to scale (Fadeyi et al., 2019).

structure, with the technical department being the most efficient and the art department the most inefficient in their analysis. Furthermore, even if data for departments are available, many of the HEI's overhead costs can hardly be meaningfully allocated to individual departments (Foladi et al., 2019). The CWTS Leiden Ranking is one of the few data sources providing international data on the HEIs' main fields of science.[2] However, the CWTS Leiden Ranking only contains information on the scientific impact (publications), collaborations, and staff. For an efficiency assessment, further input data are necessary and thus the CWTS Leiden Ranking must be merged with other data sources like the ETER database.[3] Such a data merge may reduces the number of HEIs and limits the level of analysis to the HEI level. Wolszczak-Derlacz (2017) argues that the departmental level should be the preferred level of analysis, but due to data limitations, her international analysis must be conducted at the HEI level.

When entire HEIs are evaluated, the subject structure alone can determine the result. For example, imagine two HEIs that have four departments each. Even if all departments of the first HEI are more efficient than those of the second, the former may be less efficient on the HEI level if all inputs and outputs are aggregated. The efficiency scores of the aggregated HEIs depend on the sizes and thus, the weights of the departments. The various departmental priorities can be partially addressed by selecting appropriate inputs and outputs (Wolszczak-Derlacz, 2017). Moreover, according to Kounetas et al. (2011), the selection of variables in the efficiency evaluation of HEIs is more important than the composition of their departments. In addition, Abramo et al. (2011) conclude that DEA is an appropriate technique for calculating efficiency at the HEI level, since the department structures of HEIs are similar to business units in companies.

The outputs of the different departments are only comparable to a limited extent. Some focus on research excellence, others on innovation, and still others on political and social commitment. Frenken et al. (2017) analyse the CWTS Leiden data and show that HEIs and departments have different focal areas. Similar to the aggregation problem, this leads to substantial distortions if the output measure discriminates against individual departments (e.g. research excellence) and favours others (e.g. innovation). Frenken et al. (2017) propose the inclusion of several outputs to cover as many HEI missions

[2]The CWTS Leiden Ranking, which is published by Leiden University, distinguishes between five main fields of science: biomedical and health sciences, life and earth sciences, mathematics and computer science, physical sciences and engineering, and social sciences and humanities (Waltman et al., 2012).

[3]The European University Register (ETER) contains data on the size of the HEI, the number of students and staff, graduates, and information on research and international activities (Lepori et al., 2015).

as possible. This proposal is in line with the suggestion of studies focusing on the differences of departments within HEIs. Kumar et al. (2017) find for Kurukshetra University and Abdullah et al. (2018) for Malikussaleh University that departments exhibit different foci on quantity and quality of publications. Kao et al. (2008) and Duguleana et al. (2015) propose the inclusion of different performance indicators to evaluate all HEI tasks, namely: research excellence, teaching, and innovation.

5.3.3 Inputs and Outputs

DEA compares the ratio of the aggregated input consumption to output production of similar units. Therefore, all data must be classified into inputs or outputs, and an increase in inputs should increase the outputs. However, some variables, like third-party funding, cannot be clearly categorised. For example, Fandel (2007) argues that third-party funds are an input because they are used to employ personnel for teaching and research. Contrary, Agasisti et al. (2016) use third-party funds as an output to represent the reputation of HEIs through their ability to raise competitive funds. As an alternative measure, Gralka et al. (2019) claim that the amount of external funding and the number of publications are interchangeable research indicators in the evaluation of HEIs. In addition, Agasisti et al. (2016) see information on publications as a more direct measure of the HEIs' research output than third-party funding. Following this argument, more recent studies use the number and quality of publications as outputs and total funding or total expenditure as input (Gralka et al., 2019; Wohlrabe et al., 2019a).

The number of publications represents the research activity of HEIs in the vast majority of recent HEIs' efficiency assessments (Rhaiem, 2017). Concentrating on the number of publications alone favours researchers who publish as many publications as possible. Bornmann et al. (2015) find increasing growth rates of scientific publications across all disciplines.[4]

In most efficiency evaluations of HEIs, the number of graduates represents the teaching output (De Witte et al., 2017). The lack of a further weighting implicitly assumes that graduates are similarly well educated on average between HEIs. MacLeod et al. (2017) show that employment and wages after graduation depend at least in part on the reputation of the completed HEI. The reputation of a HEI, in turn, also depends

[4] The publication growth rate was below 1% before the 19th century, rose to 2 to 3% in the interwar period and finally tripled to 8 to 9% by 2010 (Bornmann et al., 2015).

partly on the quality of its graduates (MacLeod et al., 2017). In Europe, the Bologna reform aimed to harmonise higher education by creating transparent systems with comparable degrees within the participating countries (Vögtle, 2019).[5] Although the successes cannot always be conclusively measured (Scheerlinck et al., 2019), Hahm et al. (2016) and Hahm et al. (2019) find a process of harmonisation among Bachelor's graduates in Germany and Tzanakou et al. (2017) identify the same for the British higher education sector.

5.3.4 Models and Zero Weights

Most efficiency assessments in the context of higher education are conducted using radial DEA models or, to a lesser extent, Stochastic Frontier (SFA) models.[6] Most recently, Chuanyi et al. (2016) use BCC and SFA models to asses the efficiency of HEIs in China and Bayraktar et al. (2013) apply similar models to HEIs in Turkey. SFA is selected to account for unobserved inefficiencies due to managerial inefficiencies, environmental effects, and statistical noise (Bayraktar et al., 2013). The results of the DEA and SFA models are strongly positively correlated. Most studies are conducted at the HEI level and assume that the database is sufficiently robust to neglect assumptions of unobserved inefficiencies or measurement errors and thus use DEA over SFA (Bangi, 2014).

In the following, the radial BCC model (that is presented in the methodology chapter) is referred to as standard or basic DEA model. In the standard DEA models, weights are restricted only to being non-negative and ensure that for each HEI, the ratio of weighted outputs to weighted inputs is less than or equal to one. Furthermore, the weights are calculated so that each HEI is as efficient as possible. HEIs may be assessed on only a subset of inputs and outputs by assigning zero weights to the other variables. Hence, variables with zero weights are not included in the HEI's efficiency assessment. Benneyan et al. (2007) describe the presence of zero weights as "irrational weighting" and propose the inclusion of additional weight restrictions to prevent them. Weight restrictions can increase the discriminatory power of a model to better distinguish

[5]See Feeney et al. (2017) for a comprehensive overview of the Bologna reform, its objectives and recent implementation successes.

[6]In SFA, an additional error term (ϵ) is added to the production functions to model inefficiency. ϵ consists of inefficiency due to inefficient production and of a "random" noise component that the DMU can not control (e.g. weather or measurement errors) (Behr, 2015). The underlying characteristics of the DEA and SFA efficiency frontiers are quite similar (Reinhard et al., 2000).

between efficient and inefficient DMUs (Atici et al., 2015). Only few higher eduction assessments address the problems of fully specialised DMUs and low discriminatory power. Kao et al. (2008) and Athanassopoulos et al. (1997) include additional weight restrictions based on expert opinions to limit the weights towards a reasonable range, to reflect value judgements between the inputs and outputs, and to exclude zero weights. The non-radial SBM includes data-based lower weight restrictions to prevent fully specialised DMUs (Tone, 2001). Chuanyi et al. (2016) compare the SBM and the SFA to standard DEA models. The results of the BCC are rather similar to that of the SBM. The same HEIs are efficient in both models, but the average efficiency is lower in the SBM because it more strongly discriminates between HEIs.

5.4 Methodology

This chapter introduces the DEA models, the decomposition approach, and the notation used to facilitate referencing to specific groups of HEIs.

5.4.1 DEA Models

The output-oriented BCC model for HEI$_o$ (o denotes a specific HEI under consideration) is:

$$\begin{aligned} \underset{\eta}{min} \quad & \eta = \boldsymbol{v}\boldsymbol{x}_o - u_0 \\ \text{s.t.} \quad & \boldsymbol{u}\boldsymbol{y}_o = 1 \\ & \boldsymbol{v}\boldsymbol{X} - \boldsymbol{u}\boldsymbol{Y} - u_0 \geq 0 \\ & \boldsymbol{v} \geq 0, \boldsymbol{u} \geq 0, u_0 \text{ free in sign.} \end{aligned} \tag{30}$$

where $\boldsymbol{x}$ $(m \times 1)$ is the vector of inputs of HEI$_o$, $\boldsymbol{y}$ $(s \times 1)$ the vector of its' outputs, $\boldsymbol{X}$ $(m \times n)$ the input matrix of all reference HEIs, and $\boldsymbol{Y}$ $(s \times n)$ their output matrix.[7] m is the number of inputs, s the number of outputs, and n the number of reference HEIs. $\boldsymbol{v}$ $(1 \times m)$ are the input weights and $\boldsymbol{u}$ $(1 \times s)$ are the output weights. η^* $([1, \infty])$ denotes the solution to the minimisation problem and for convenience η^* is transformed: $\theta^* = \frac{1}{\eta^*}$ $([0, 1])$. A HEI is inefficient, if $\theta^* \neq 1$, otherwise it is efficient. The assumption of variable returns to scale is implemented by the scalar u_0 that is free in sign. The first line in model (34) is the objective or target function and the lines

[7] Bold lower case symbols indicate vectors and bold capitalised symbols matrices.

below are the restrictions.

The linear problem is solved for each HEI and the weights are assigned to maximise its efficiency. The restrictions in model (34) do not prevent zero weights and the BCC model does not account for input excess ($\boldsymbol{s}^-$) and output shortfalls ($\boldsymbol{s}^+$) and outputs can only be radially increased.

The output-oriented SBM excludes zero weights, allows substitution among inputs and outputs, and its efficiency measure is monotone decreasing in each output slack (Tone, 2001). By using a positive scalar variable t the output-oriented SBM with variable returns to scale is given by

$$
\begin{aligned}
\min_{\boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+} \quad & \tau = t \\
\text{s.t.} \quad & 1 = t + \frac{1}{s}\frac{t\boldsymbol{s}^+}{\boldsymbol{y}_o} \\
& t\boldsymbol{x}_o \geq \boldsymbol{X}(t\boldsymbol{\lambda}) + t\boldsymbol{s}^- \\
& t\boldsymbol{y}_o = \boldsymbol{Y}(t\boldsymbol{\lambda}) - t\boldsymbol{s}^+ \\
& \boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq 0, t > 0 \\
& \boldsymbol{e\lambda} = 1.
\end{aligned} \tag{31}
$$

A HEI is efficient if no output shortfalls are present ($\tau^* = 1$). An inefficient HEIs can become efficient by reducing its output slacks. Restriction $\boldsymbol{e\lambda} = 1$ includes the assumption of variable returns to scale in the model, and if it is excluded, constant returns to scale are assumed (Cooper et al., 2007).

The radial model (31) maximises inefficiency based on the maximal relative distance to the efficiency frontier given the HEIs' PPS. If HEI$_o$ is benchmarked against a reference set it is not part of (e.g. the efficiency frontier of another country HEIs'), the HEI can be outside the feasible region. Then model (31) is infeasible. Tone (2002) propose a two step-approach to calculate the efficiency of HEIs for reference sets they are not part of. In the first step, model (31) is calculated for all HEIs. Efficient HEIs have an efficiency score of one and are located on the efficiency frontier. Inefficient HEIs inside the feasible region have an score below one and HEIs for that the model is infeasible are located outside the feasible region. For the latter group, the following model is calculated:

$$
\begin{aligned}
\min_{\boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+} \quad & \delta = t \\
\text{s.t.} \quad & 1 = t - \frac{1}{s}\frac{t\boldsymbol{s}^+}{\boldsymbol{y}_o} \\
& t\boldsymbol{x}_o \geq \boldsymbol{X}(t\boldsymbol{\lambda}) + t\boldsymbol{s}^- \\
& t\boldsymbol{y}_o \geq \boldsymbol{Y}(t\boldsymbol{\lambda}) + t\boldsymbol{s}^+ \\
& \boldsymbol{\lambda}, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq 0, t > 0 \\
& \boldsymbol{e\lambda} = 1.
\end{aligned}
\tag{32}
$$

Compared to model (31), the first and third restrictions are altered. The efficiency frontier is the same but the slacks are calculated from outside the feasible region to the efficiency frontier. δ^*, if it exist, is greater than one. Model (32) under variable returns to scale has no feasible solution if there exist i such that $x_{io} < \min_{j=1,\neq o} \{x_{ij}\} \quad (i = 1 \ldots m \,; j = 1 \ldots n)$. This infeasibility implies that the HEI$_o$ has a unique input-output combination compared to the reference HEIs. The assumption of constant returns to scale may render the model feasible (Cooper et al., 2007). The input-output structure of the reference group and a group of evaluated HEIs can differ in a way that all HEIs are identified as super-efficient.

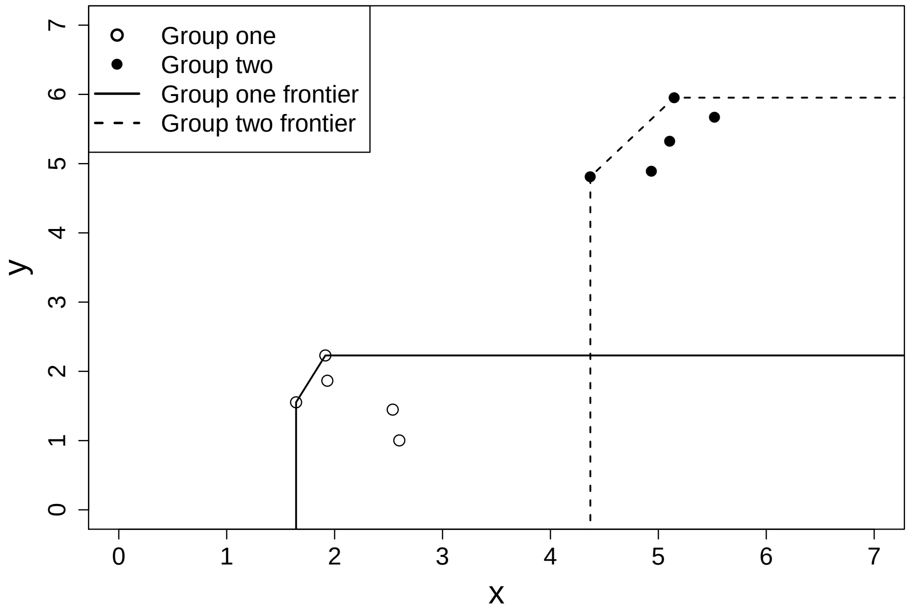


Figure 5.2 – Super-efficiency example

Figure 5.2 provides an example of how all HEIs can be super-efficient when two groups are compared. In this example, the HEIs consume one input (x) and produce one output (y). Within each group, two HEIs are efficient and three are inefficient. In a combined analysis in which all HEIs would be assessed together, the same four universities would be efficient as in the group-specific analyses. Group-specific input-output structures result in all HEIs being super-efficient compared to the efficiency frontier of the other group. Without additional decomposition, such unique input-output structures could remain unnoticed if HEIs are only considered within groups or in a combined analysis.

5.4.2 Decomposition

The efficiency of HEIs can be assessed at country-level relative to different frontiers or a common international frontier. The following notation is used to differentiate between the various compositions:

- $\mathcal{K}$ is the index set of the countries, $k = 1, 2$
- $\mathcal{U}_k$ is the index set of HEIs in country k
- $\mathcal{U}$ is the index set of all HEIs in all countries, $\mathcal{U} = \{\mathcal{U}_1, \mathcal{U}_2\}$

$D^{\mathcal{U}_k}$ are the efficiency scores of the HEIs in country k. $M\left[F^{\mathcal{U}_k}\right]\left(D^{\mathcal{U}_k}\right)$ denotes the arithmetic mean (M) of $D^{\mathcal{U}_k}$ based on their own frontier $F^{\mathcal{U}_k}$. For convenience, the abbreviation $M^{\mathcal{U}_k}(\mathcal{U}_k)$ is used for $M\left[F^{\mathcal{U}_k}\right]\left(D^{\mathcal{U}_k}\right)$ in the following. It is calculated as:

$$M^{\mathcal{U}_k}(\mathcal{U}_k) = \frac{1}{n_k}\left(\sum_{i=1}^{n_k} D_i^{\mathcal{U}_k}(\mathcal{U}_k)\right). \tag{33}$$

n_k is the number of evaluated HEIs in country k. $M^{\mathcal{U}_l}(\mathcal{U}_k)$ is the arithmetic mean of the HEI of country k benchmarked against the efficiency frontier of country l and $M^{\mathcal{U}}(\mathcal{U}_k)$ is the arithmetic mean relative to all HEIs of all countries.

5.4.3 An Illustration

This section demonstrates the efficiency score decomposition of HEIs based on one input and one output. This enables a straightforward interpretation and visual presentation of the efficiency score decomposition. First, an output-oriented BCC model with three inputs and five outputs (which are presented in the following section) is calculated. In the second step, average input ($\bar{\boldsymbol{v}}$) and output ($\bar{\boldsymbol{u}}$) weights are derived from the results and used to calculate a weighted input ($\bar{\boldsymbol{x}} = \bar{\boldsymbol{v}}\boldsymbol{X}$) and a weighted output ($\bar{\boldsymbol{y}} = \bar{\boldsymbol{u}}\boldsymbol{Y}$). $\bar{\boldsymbol{x}}_{GER}$ is the weighted input vector of the German HEIs.

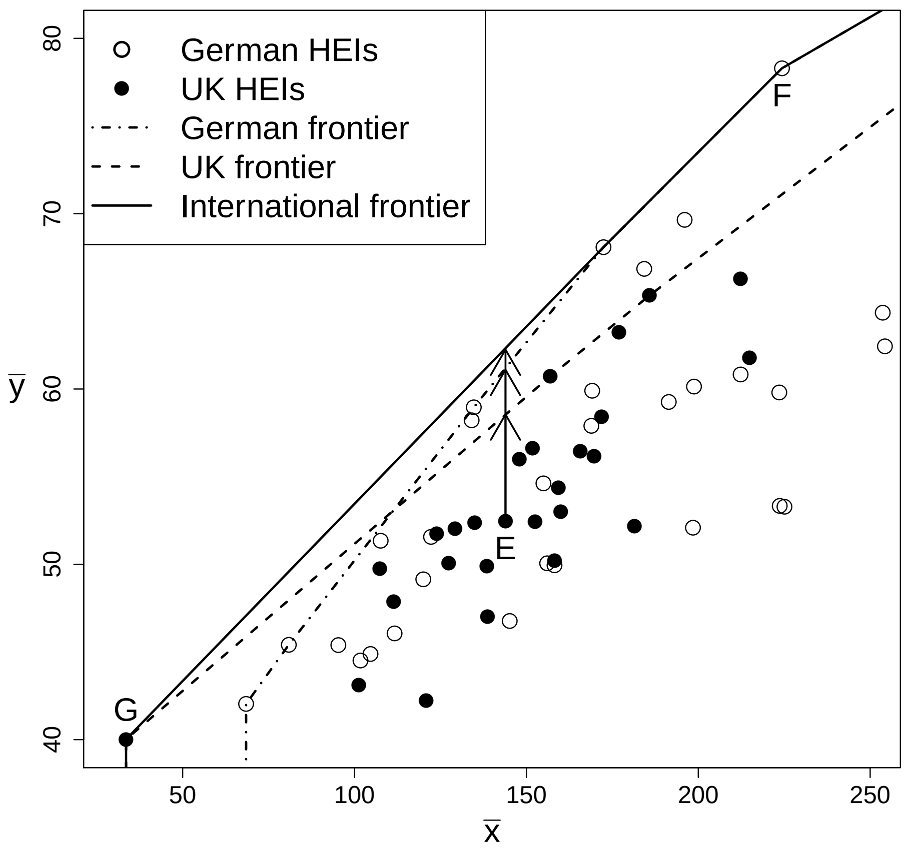


Figure 5.3 – Efficiency decomposition example

Figure 5.3 shows the weighted input (from 30 to 250 units) and weighted output as well as efficiency frontiers under the assumptions of variable returns to scale. The international frontier $F^{\mathcal{U}}$ consists of one UK and two German HEIs for this input range. If the countries are evaluated separately, four German HEIs and three UK HEIs are efficient.

Table 5.1 provides the efficiency frontier for three selected HEIs E, F, and G.[8] Efficiency scores in the columns with index τ are derived from model (31) with index δ from the second-step model (32). The arrows in Figure 5.3 indicate by how much

[8]E is the HEI of Kent, F the Free University of Berlin, and G the York St. John University.

Table 5.1 – Efficiency scores given $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$

	$F_{\tau}^{\mathcal{U}_{UK}}$	$F_{\tau}^{\mathcal{U}_{GER}}$	$F_{\delta}^{\mathcal{U}_{GER}}$	$F_{\tau}^{\mathcal{U}}$	$F_{\tau}^{\mathcal{U}_{GER}}$	$F_{\tau}^{\mathcal{U}_{UK}}$	$F_{\delta}^{\mathcal{U}_{UK}}$
E	0.896	0.857	-	0.842			
F				1.000	1.000	-	1.101
G	1.000	-	1.943	1.000			

E must increase its weighted output to become efficient. E has the least distance to $F^{\mathcal{U}_{UK}}$, followed by $F^{\mathcal{U}_{GER}}$, and is most inefficient (the highest distance) compared to the international frontier ($F^{\mathcal{U}}$). F is efficient relative to the international and the German frontiers. Considering $F^{\mathcal{U}_{UK}}$, F is infeasible in model (31) and super-efficient in model (32). Therefore, the latter model calculates that the efficiency score of F is higher than one. G is efficient compared to $F^{\mathcal{U}_{UK}}$ and $F^{\mathcal{U}}$, and super-efficient when compared to $F^{\mathcal{U}_{GER}}$. The weighted input of G is lower than that of any German HEI ($\bar{x}_G < min\{\bar{\boldsymbol{x}}_{GER}\}$). If G is benchmarked against the German HEIs, model (32) becomes infeasible under variable returns to scale. However, the models remains feasible under constant returns to scale and G is super efficient with an efficiency score of 1.943. Chen (2005) and Cook et al. (2009a) suggest using constant returns to scale if different input-output structures of HEI$_o$ and the reference set make the model otherwise infeasible.

In an international comparison of all HEIs in both countries, HEIs from the UK are on average nearly 6% more efficient than their German counterparts ($M^{\mathcal{U}}(\mathcal{U}_{UK}) = 0.907$ and $M^{\mathcal{U}}(\mathcal{U}_{GER}) = 0.858$).

5.5 The Efficiency of German and UK HEIs

For the following analyses, data from the CWTS Leiden ranking (2014-2017) and ETER (2016) are combined. These data allow a momentary assessment, since education is a cumulative process. HEIs with missing data are excluded. The final data set comprises 45 UK and 46 German HEIs. The specific HEIs and their data are listed in the appendix. The inputs are:

- Number of academic staff (x_1)
- Number of non-academic staff (x_2)
- Total current expenditure minus personnel expenditure (x_3) in PPP (in M €)

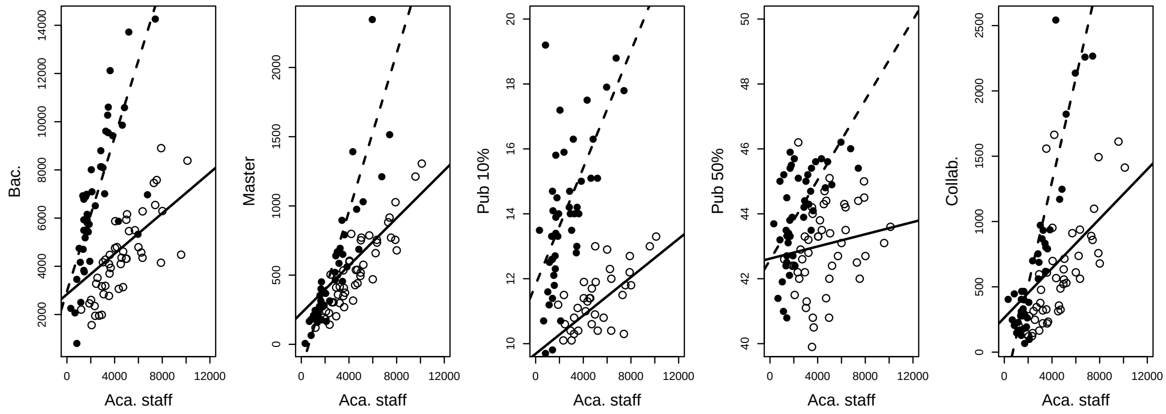
The numbers of academic and non-academic staff account for human resources. By including staff numbers, different remuneration structures do not a priori benefit some HEIs and worsen the results of others. x_3 covers the HEIs' total expenditure including all spendings on physical capital. Personnel costs are excluded as personnel capacities are already accounted for by the other inputs. x_3 is purchasing power and exchange rate adjusted to allow an international comparison. The outputs are:

- Total graduates (Bachelor) (y_1)
- Total graduates (Master) (y_2)
- Number of top 10% publications in the according field (y_3)
- Number of top 50% publications minus the number of top 10% publications (y_4)
- Number of publications that have been co-authored with one or more industrial (manufacturing and services sectors) organisations (y_5).

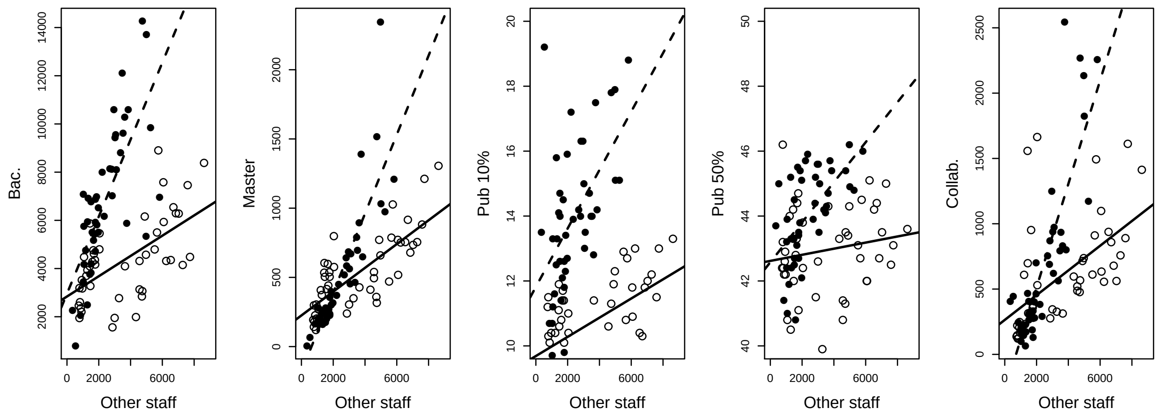
y_1 and y_2 capture the teaching mission of HEIs. The distinction between Bachelor and Master graduates takes into account the different teaching foci and structures of the HEIs. y_3 and y_4 represent different publication strategies. While y_3 accounts for top publications in the according field, y_4 considers a much broader publication strategy. y_5 is a measure of industrial cooperation (innovation mission).

Table 6.1 contains the descriptive results and Table 6.2 provides the correlation coefficients for all HEIs. Table B.3 lists the descriptive results and Table B.4 lists the correlation coefficients for each country. Tables B.2 and B.1 contain all data for each HEI to ensure complete reproducibility. All data are positive, there are no missing values, and all correlation coefficients are positive too. The latter indicates that an increase in inputs will most likely increase outputs. The descriptive results reveal country-specific input-output structures. The German HEIs have 81% more academic and 51% more non-academic personnel but 51% lower average expenditures. They have 36% less bachelor graduates but around 7% more master graduates than the UK HEIs. The German HEIs publish 23 % less frequently top publications, 3% in the broader quantity measurement, and 9% less collaboration publications. The funding system in Germany does not distinguish between teaching and research, qualitative and quantitative research indicators determine funding in the UK. Thus the UK HEIs have strong incentives to publish more frequently (Auranen et al., 2010). Higher core funding of German HEIs increases stability within these institutions and allows them

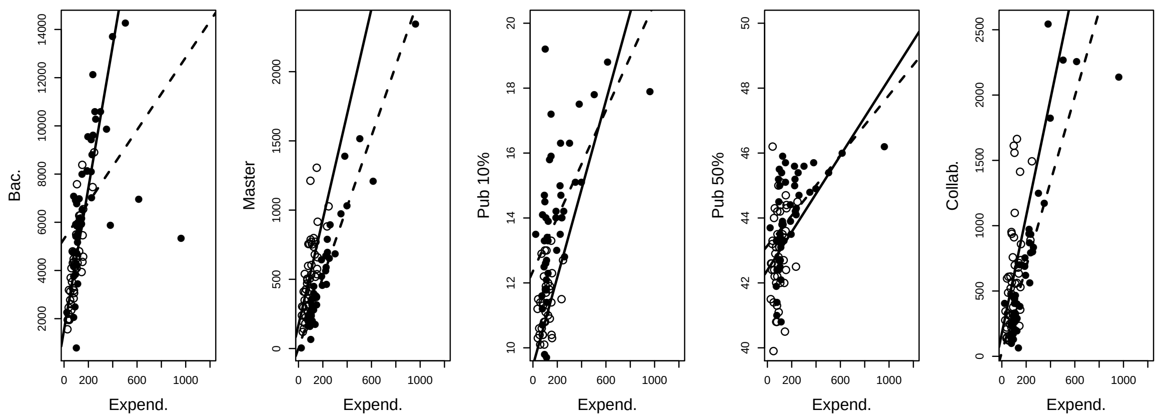
Figure 5.4 – Inputs-outputs relationships for Germany and the UK HEIs (filled dots and dotted line)



(a) Academic staff



(b) Other Staff



(c) Expenditure

Table 5.2 – Descriptive results, all HEIs

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Min	325.00	350.00	23.51	790.00	5.00	7.60	39.90	65.00
Median	3157.50	2007.50	106.10	5041.50	434.00	11.80	43.35	466.00
Mean	3584.34	2896.30	141.31	5547.98	506.22	12.18	43.42	620.70
Max	10112.00	8623.00	962.75	14260.00	2345.00	19.20	46.20	2544.00
Sd	2165.07	2036.63	130.06	2648.49	358.40	2.52	1.47	521.31

Table 5.3 – Correlation coefficients, all HEIs

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Aca. staff	1.000	0.835	0.380	0.331	0.741	0.128	0.171	0.647
Other staff	0.835	1.000	0.348	0.346	0.696	0.182	0.200	0.571
Expen.	0.380	0.348	1.000	0.545	0.803	0.657	0.515	0.748
Bac.	0.331	0.346	0.545	1.000	0.499	0.473	0.383	0.499
Master	0.741	0.696	0.803	0.499	1.000	0.434	0.422	0.876
Pub. 10%	0.128	0.182	0.657	0.473	0.434	1.000	0.729	0.532
Pub. 50%	0.171	0.200	0.515	0.383	0.422	0.729	1.000	0.470
Collab.	0.647	0.571	0.748	0.499	0.876	0.532	0.470	1.000

to employ personnel longer and more consistently (Kloss, 1985). Figure 5.4 depicts the relationship between inputs and outputs in both countries. Overall, UK HEIs produce more outputs with fewer academic and non–academic personnel. In terms of expenditures, the output relationships are quite similar in the two countries. However, the UK HEIs have more financial resources available.

5.5.1 Efficiency Results

Tables B.5 and B.6 contain all efficiency scores and Table 5.4 provides an aggregated overview. The efficiency score reported in column $F^{\mathcal{U}_{GER}}$ (German HEIs), $F^{\mathcal{U}_{UK}}$ (UK), and $F^{\mathcal{U}_{U}}$ (international) are calculated using an output-oriented SBM with variable returns to scale. The $F^{\mathcal{U}_{UK}}$ (German HEIs) and $F^{\mathcal{U}_{GER}}$ (UK HEIs) contain the results when the HEIs are benchmarked against the efficiency frontier of the other country using the super-efficient output-oriented SBM with constant returns to scale. The necessity to use different returns to scale assumptions reduces the explanatory power of results between, but not within, models. Tables B.7 to B.10 provide the slacks of the German and UK HEIs for the country-specific and international frontiers and Table 5.5 an overview of relative slacks.

A similar number of HEIs are efficient in Germany (27) and the UK (26) when they are benchmarked against their country specific frontiers. The two least efficient HEIs in

Table 5.4 – Efficiency results overview

	German HEIs				UK HEIs		
	$F^{\mathcal{U}_{GER}}$	$F^{\mathcal{U}}$	$F^{\mathcal{U}_{UK}}$		$F^{\mathcal{U}_{UK}}$	$F^{\mathcal{U}}$	$F^{\mathcal{U}_{GER}}$
Mean	0.892	0.807	1.090	Mean	0.918	0.907	1.252
Median	1.000	0.817	1.113	Median	1.000	1.000	1.182
Std. Deviation	0.144	0.195	0.139	Std. Deviation	0.118	0.123	0.278
Efficient HEIs	27	19	-	Efficient HEIs	26	24	-

Germany are the Bielefeld University (0.600) and the University of Hannover (0.510). In the UK, Brunel University London (0.642) and the University of Kent (0.607) perform worst. The average efficiency scores are rather similar (0.892 in Germany and 0.918 in the UK). These averages cannot be compared directly, as the HEIs are assessed on the basis of country-specific frontiers. Inefficient German HEIs must increase their efficiency by 12% on average to become efficient. Most inefficient German HEIs have at least one input slack and slacks in most outputs. In regard to their slacks, inefficient German HEIs should employ on average 13.1% less academic staff, 33.8% non-academic staff and reduce their expenditures by 15.5%. Their average relative output slacks of the graduates are 13.8% (bachelor) and 33.4% (master). The average relative slacks for the publication outputs are 12.8% for the quality and 3.6% for the quantity measure. The inefficient HEIs have rather low collaboration outputs that should be increased by 138.1%. The relative slack structure of the inefficient UK HEIs is rather similar compared to the German HEIs. The absolute slacks provide further guidance to increase the HEIs performance. For example, the Leipzig University should reduce its' non-academic staff by around 1847 people (41%) and increase its' Master graduates by 138 students (27.8%), its amount of top publications by 1.861 (19.8%), its broader publications by 2.739 (8.5%) and the number of collaboration by 404.198 (77.6%). The University of Hannover, the Ruhr-University Bochum, and the University of Stuttgart are the only German HEIs that should reduce their academic staff. Relative to their own country-specific frontier, Heriot-Watt University, Swansea University, University of Sheffield, and University of Warwick are the only UK HEIs that should reduce their expenses. Apart from University of Dundee, there are no slacks in bachelor graduates at any UK HEI.

The efficiency scores obtained relative to the international frontier ($F^{\mathcal{U}}$) can be compared directly between the countries. 19 German and 24 UK HEIs are efficient. HEIs of both countries form the efficiency frontier. Compared to the results from the country

Table 5.5 – Relative slacks

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
$F^{\mathcal{U}_{GER}}(U_{GER})$	0.13	0.34	0.15	0.14	0.33	0.13	0.04	1.38
$F^{\mathcal{U}_{U}}\ (U_{GER})$	0.18	0.28	0.14	0.23	0.30	0.24	0.04	2.21
$F^{\mathcal{U}_{UK}}\ (U_{UK})$	0.10	0.13	0.10	0.08	0.23	0.12	0.03	0.98
$F^{\mathcal{U}_{U}}\ (U_{UK})$	0.00	0.13	0.10	0.16	0.26	0.14	0.02	1.00

specific-frontiers, the average efficiency in the UK is only slightly lower ($M^{\mathcal{U}_{UK}}(\mathcal{U}_U) =$ 0.907 to $M^{\mathcal{U}_{UK}}(\mathcal{U}) = 0.918$) but German HEIs are on average more inefficient ($M^{\mathcal{U}_{GER}}(\mathcal{U}_U) = 0.807$ to $M^{\mathcal{U}_{GER}}(\mathcal{U}_{GER}) = 0.892$). TU Dortmund University (-0.432) and Technical University of Berlin (-0.413) are the German HEIs whose efficiency values are decreasing the most. In the UK, the two HEIs with the highest reductions are the University of Aberdeen (-0.110) and Queen Mary University of London (-0.086). In this international analysis, University of Kent (0.601) and University of Lancaster (0.628) perform worst in the UK and University of Kaiserslautern (0.411) and Carl von Ossietzky the University of Oldenburg (0.415) do worst in Germany. Average slacks are lower in the UK and there are no slacks in academic personnel quantities. German HEIs have particularly great potential for improvement in their cooperation with industry.

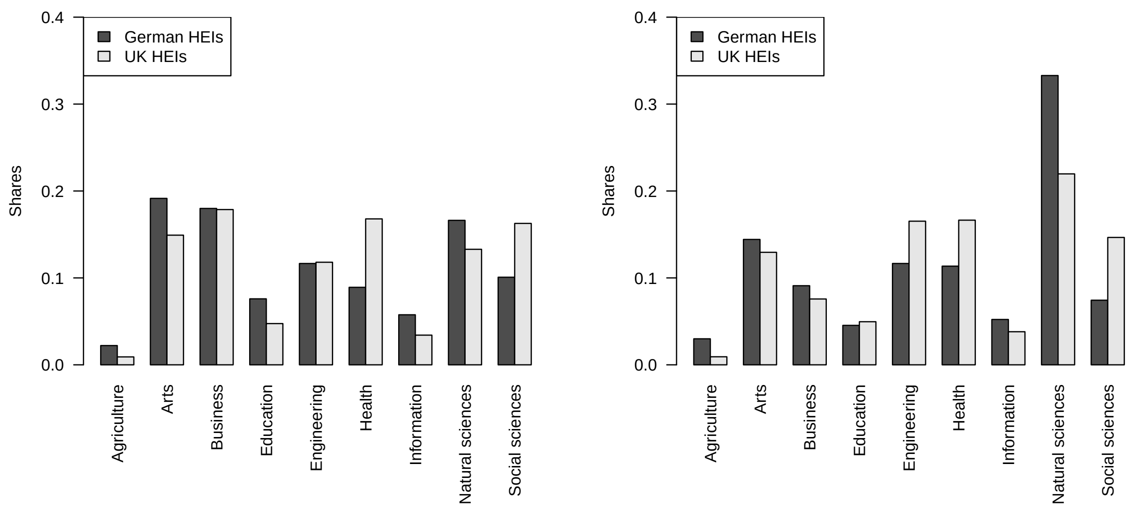
The columns $F^{\mathcal{U}_{UK}}$ for Germany and $F^{\mathcal{U}_{GER}}$ for the UK provide the results of the super-efficient models. The German and UK HEIs have so different country specific input-output structures so that nearly all HEIs are identified as super efficient if the efficiency frontier of the respective other country is used. Bielefeld University (0.487), University of Hannover (0.551), and University of Oxford (0.743) are the only inefficient HEIs. The latter is efficient if the country specific ($F^{\mathcal{U}_{UK}}$) and the international ($F^{\mathcal{U}}$) frontiers are used. The results of the super-efficient models can not be directly compared between the countries as the HEIs are not benchmarked against the same frontiers. The super efficient models reveal similar group specific input-output structures as shown in Figure 5.2.

5.5.2 Result Comparisons and Discussion

This section discusses the department structure of the selected HEIs and the reimbursement structures of higher education sectors in Germany and the UK. . After a discussion of these two aspects, the results of the previous chapter are compared with those of Wohlrabe et al. (2019b), Wolszczak-Derlacz (2017), and the HEIs selected in the Excellence Initiative of German Universities (EIGU).

Different reimbursement structures within HEIs can result in different incentives and thus in unique input-output combinations when we try to examine the differences in HEIs between countries (Blecich, 2020). In the German higher education system, salaries are fixed by legislation. Bonuses are used to reward the fulfilment of agreed administrative tasks, successful teaching, and research (Mellewigt et al., 2017). In the UK, the income of academic staff is more market-oriented and determined by negotiation at the level of each individual HEI. Academic staff, therefore, have strong incentives to increase their market value (Angermuller, 2017). The differences are also illustrated by the descriptive results in Figure 5.4, which show that academic staff in the UK publish more often, which may explains why most HEIs are super-efficient.

Figure 5.5 – Average student shares of the fields of education



(a) Bachelor students **(b)** Master students

Figure 5.5 shows the relative shares of the students that obtain a bachelors degree (left panel) and a masters degree (right panel) for different fields of education. Table B.11 contains the descriptive results. In Germany, the highest average number of bachelor students are art students, while there are relatively higher numbers of master's students in the natural sciences. In the UK, the highest shares are in business for bachelors students and the natural sciences for masters students. Table B.12 provides the student shares for the different educational sectors for the HEIs used in the efficiency analyses and for all HEIs in the respective countries (using data taken from the OECD). In the

45 German HEIs, the shares of students in economics and engineering are lower and the shares of students in the social and natural sciences are higher than in the OECD data (OECD, 2015). For the UK HEIs, the shares are similar. The correlation coefficients between the sample used for the efficiency analyses and the OECD data are 0.7 for German and 0.93 for the UK HEIs.

Table 5.6 – Student shares of the fields of education regression results, country specific SBM efficiency scores as dependent variable

Students	Measure	Agriculture	Arts	Business	Education	Engineering	Health	Information	Natural sciences	Social sciences
German	Coefficient	0.468	0.192	0.080	-0.862	-0.140	0.388	-0.632	0.066	0.939
Bachelor	T value	0.973	0.745	0.313	-1.816	-1.059	1.149	-1.043	0.105	1.866
German	Coefficient	0.297	0.228	-0.436	-0.141	-0.205	0.542	-1.291	0.309	-0.155
Master	T value	1.085	0.822	-0.884	-0.365	-1.475	2.235	-3.129	1.172	-0.249
UK	Coefficient	0.780	0.055	-0.278	0.381	-0.118	0.138	-2.307	-0.157	-0.042
Bachelor	T value	0.677	0.248	-1.252	1.000	-0.702	1.188	-2.358	-0.496	-0.195
UK	Coefficient	0.699	0.076	-0.906	-0.044	-0.122	0.149	-0.612	-0.003	0.050
Master	T value	0.734	0.332	-2.277	-0.135	-1.043	1.316	-0.846	-0.013	0.328

Table 5.6 contains regression coefficients and t values resulting from simple regressions of the country-specific SBM efficiency scores on the respective student shares. The regression results indicate that a HEIs with higher proportion of students in agriculture, art, and health tend to be more efficient. Interestingly, HEIs with a higher proportion of students in engineering and information tend to be more inefficient. The coefficient of the share of engineering master's students in Germany is negative (-0.205) and significant at the ten percent level (t value of -1.475). The negative coefficient is partly due to the fact that the University of Stuttgart and RWTH Aachen University are two of the three HEIs with the largest proportion of master's students in engineering sciences in Germany (56% and 47% respectively) and are inefficient in the country-specific assessment. Frenken et al. (2017) note that HEIs in OECD countries with large engineering departments tend to publish more than HEIs that focus on other fields of education. In the DEA, efficiency scores are calculated based on both input consumption and output production. Thus, departments that produce high numbers of outputs with relatively high inputs may be more inefficient than departments that produce fewer outputs but use relatively fewer inputs. Furthermore, student numbers are only proxies for understanding the structures of HEIs. Data on the number of professorships, doctoral students, and expenditure at the departmental level may better reflect the priorities of HEIs and could provide further insights as they become available. Additional information could give an indication of the weight of the individual departments within each HEI and, if possible, enable efficiency evaluations at a departmental level. The efficiency analyses in the previous section take into account the main performance indicators of HEIs (research, teaching, and innovation) so that the differ-

ences between the priorities of the departments are taken into account as thoroughly as possible.

The aim of EIGU is to promote top HEIs in Germany (Kehm, 2013). In 2012, 11 HEIs were selected for funding by EIGU for the coming years. These HEIs are often referred to as excellence HEIs. They are selected based on their output production, whereby their input consumption is not taken into account (Fischer et al., 2017). In contrast, DEA models consider input consumption and output production simultaneously. Three of the eleven Excellence HEIs are inefficient in the efficiency evaluations in the previous chapter: RWTH Aachen University, Dresden University of Technology and the University of Cologne. These results indicate that these three HEIs achieve their excellence status due to their relatively high capital and personnel inputs.

In a recent publication, Wohlrabe et al. (2019b) assess the efficiency of 70 German HEIs between 2004 and 2015, with expenditure and personnel used as inputs. Their outputs are graduates in the natural sciences, social sciences, and in medicine, as well as the number of top publications. A disaggregation of the personnel data into scientific and non-scientific personnel is not provided and their publication output does not take into account the different publication strategies due to missing data. Wohlrabe et al. (2019b) use an output-oriented BCC model without additional weight restrictions. The model most likely calculates specialised HEIs that are assessed only on subsets of the data, while ignoring input excess and output shortfalls. Their descriptive results show that there are zeros in all outputs and the authors do not report whether they transform their data. Overall, Wohlrabe et al. (2019b) find an average efficiency of 0.87 and state that 21 out of 70 HEIs are efficient. The minimum efficiency is around 0.58 in 2015, while the results being robust over the years. HEIs that are efficient every year include the Technical University of Munich, the Heidelberg University, the Ludwig Maximilian University of Munich, the University of Heidelberg, the University of Flensburg, and the University of Lübeck. A comparison with the results in the previous section shows that the first four HEIs are also efficient in the SBM relative to the German efficiency frontier. The University of Flensburg and the University of Lübeck are excluded in this study due to missing data. Although the selected inputs and outputs and the DEA models used differ, the average efficiency values are remarkably similar (0.892 in SBM in this paper to 0.87) and the results are positively correlated (the correlation coefficient is 0.243). Without a super-efficient decomposition it remains unknown if the inefficiencies are due to different input-output structures or inefficient output maximisation.

Wolszczak-Derlacz (2017) assesses the efficiency of HEIs from Europe and the United States of America using national and international efficiency frontiers. She includes 65 HEIs from Germany and 85 from the UK. The inputs and outputs were chosen in such a way that as many HEIs as possible could be included. Wolszczak-Derlacz (2017) notes that her inputs and outputs reflect the available resources and targets of the HEIs only to a certain extent. For example, graduates are not divided into those with bachelor and master degrees. The selected data directly influences the resulting efficiency scores and limits the comparability of the results with those of the previous section. Overall, Wolszczak-Derlacz (2017) also finds HEIs in the UK on average more efficient than their German counterparts. Wolszczak-Derlacz (2017) uses an output-oriented BCC model, incorporating the same disadvantages discussed above. Her results suggests that HEIs from the United States of America are quite inefficient while HEIs from Poland perform well. Polish HEIs perform well internationally, due to relatively high outputs with relatively low inputs. The next iterations of the ETER and CWTS Leiden Ranking data sets could help validate the results of Wolszczak-Derlacz (2017) with radial and super-efficient radial DEA models for more countries.

5.6 Conclusion and Suggestions for Future Research

This paper discusses the operationalisation of efficiency assessments in the context of higher education, evaluates the efficiency of German and UK HEIs to maximise their outputs (graduates, publications, and collaboration) given their capital and personnel inputs. The HEIs' efficiency is decomposed relatively to country-specific efficiency frontiers, to an international frontier, and to the frontier of the respective other countries using super-efficient SBM models. The latter models are only feasible under the assumption of constant returns to scale, a problem that is still being discussed in recent literature (c.f. Tian et al. (2020)) and should be addressed in further research.

Descriptive results show that UK HEIs use fewer academic and non-academic personnel but can spend more than their German counterparts. German HEIs publish less frequently and have fewer bachelor graduates but more master graduates on average.

An output-oriented SBM with variable returns to scale and country-specific frontiers calculates 27 out of 46 German HEIs as efficient and with an average efficiency score of 0.892. In the UK, 26 HEIs out of 45 are efficient in the country-specific analyses with an average efficiency score of 0.918.

The international comparisons assess all HEIs and identify 19 German HEIs and 24 UK HEIs as efficient. The lower average efficiency in Germany (0.807 compared to 0.907 in the UK) shows the potential for increasing outputs while maintaining the same level of inputs. As one example, most German HEIs have large potentials for improvement in their cooperation with industry. Overall, the findings in this study are in line with those of Wohlrabe et al. (2019b) for Germany and of Wolszczak-Derlacz (2017) for Germany and the UK. However, the latter publications use radial models, do not further decompose the efficiency scores, and their inputs and outputs do not cover the entire spectrum of HEI activities. Three of the eleven Excellence HEIs in Germany are assessed as inefficient due to their relatively high capital and personnel input. DEA can therefore be regarded as more suitable for identifying top HEIs than a comparison that focuses mainly on performance.

Super-radial models calculate nearly all HEIs as super-efficient that indicates country-specific input-output structures. These input-output structures must be taken into consideration if international HEI comparisons are to be conducted and need to be further investigated. This is particularly important when more up to date and comprehensive data becomes available. Data for individual departments could further improve efficiency assessments by allowing inputs and outputs to be allocated to individual departments to reflect their strengths and weaknesses.

B.1 Supplementary Results

Table B.1 – Inputs and outputs of German HEIs

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Bielefeld University	2929	964	81.568	3173	203	10.4	42.2	165
Carl von Ossietzky the University of Oldenburg	2023	816	55.674	2604	181	8.5	42.4	121
Tuebingen University	6342	6529	81.426	4353	755	10.4	44.2	938
Free University of Berlin	4192	2040	124.259	4797	798	11.0	43.8	1664
Friedrich-Schiller-University of Jena	4708	4565	51.343	3135	539	10.6	41.4	487
University of Erlangen Nuernberg	7531	6063	106.098	7579	771	11.8	42.0	1098
Georg August Goettingen University	4999	6243	111.428	4315	752	13.0	45.1	556
Frankfurt University	5042	4893	139.206	6159	771	11.9	43.4	715
University of Hannover	3669	1283	144.335	3935	373	9.1	40.5	233
Heidelberg University	9581	7739	98.009	4482	1212	13.0	43.1	1612
Heinrich Heine University Duesseldorf	4349	4741	82.606	3064	316	11.3	43.5	567
Humboldt University of Berlin	3526	1435	104.044	4094	534	11.1	44.2	1558
Johannes Gutenberg University Mainz	4052	1447	106.259	5872	604	11.9	43.3	698
University of Wuerzburg	6001	5516	70.844	4793	470	12.9	43.1	611
Justus Liebig University Giessen	3018	1776	92.839	4168	504	10.1	41.1	377
Karlsruhe Institute of Technology	4036	1658	79.086	4758	597	11.4	43.3	952
University of Kiel	3575	1217	54.444	3698	408	10.4	44.0	609
Leipzig University	4959	4563	72.352	4314	494	9.4	40.8	521
Ludwig Maximilian University of Munich	10112	8623	149.868	8383	1306	13.3	43.6	1413
Martin Luther University	3529	3274	50.143	2773	349	7.8	39.9	327
Philipps University of Marburg	2455	1473	72.118	3276	504	10.6	42.0	315
Ruhr-University Bochum	5209	2051	152.914	5461	575	10.4	42.4	530
RWTH Aachen University	7293	7593	235.074	7460	882	11.5	42.5	890
Saarland University	3078	4714	79.074	2842	359	8.6	41.3	477
Technical University of Munich	7907	5747	247.359	8903	1027	12.7	44.5	1493
Technical University of Berlin	4582	1758	151.425	4355	537	11.4	44.7	357
Technical University of Darmstadt	3228	1596	91.323	4277	416	10.8	42.8	319
Technische University Dresden	5705	5663	99.685	5497	787	10.8	42.7	911
University of Kaiserslautern	2324	866	59.607	2352	226	10.1	42.2	123
TU Dortmund University	3596	1073	78.861	4476	269	7.6	40.8	220
Ulm University	2923	4340	42.462	1987	411	9.4	43.3	596
University of Hamburg	7407	6696	157.258	6544	916	10.3	44.4	859
University of Regensburg	4517	3648	60.698	4095	428	11.4	44.3	313
University of Bayreuth	1887	742	37.731	2461	193	11.2	42.4	137
University of Bonn	6238	4968	157.041	4573	659	12.3	44.3	737
University of Bremen	2564	934	65.120	3521	304	8.4	42.9	248
University of Cologne	8010	6852	133.838	6293	679	11.8	42.7	679
University of Duisburg-Essen	4982	6085	138.795	5932	518	10.9	42.0	634
University of Freiburg	7890	7281	90.187	4148	755	12.2	45.0	758
Ernst Moritz Arndt University of Greifswald	2073	2861	27.862	1560	239	9.0	41.5	276
University of Hohenheim	1192	910	35.324	2204	120	9.3	42.6	206
University of Konstanz	2375	783	42.971	1946	142	11.5	46.2	146
University of Muenster	6352	7029	124.563	6280	735	12.0	43.5	562
University of Potsdam	3221	785	39.983	3190	298	10.3	43.2	221
University of Rostock	2705	2999	29.996	1951	304	9.3	42.6	345
University of Stuttgart	4718	1693	107.144	4612	441	11.7	44.4	325
Mean	4622	3620	95.962	4362	536	10.8	43.0	606

Table B.2 – Inputs and outputs of UK HEIs

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Brunel University London	1135	1075	78.327	4165	205	11.2	41.0	166
Cardiff University	3465	3065	196.443	9545	455	13.0	43.5	619
City, University of London	2105	1035	79.386	7085	165	10.7	42.4	98
Cranfield University	680	855	79.488	2060	165	10.7	41.4	246
University of Durham	1700	2350	126.120	6160	450	13.9	45.9	291
Heriot-Watt University	840	1030	112.647	3450	180	9.7	43.2	203
Imperial College London	4330	3765	381.494	5870	1390	17.5	45.7	2544
King's College London	4840	2960	300.937	10585	685	16.3	45.6	1248
The University of Lancaster	1930	1530	107.154	4205	280	14.0	42.7	263
London School of Economics and Political Science	1705	1300	136.862	5930	175	15.8	43.3	65
London School of Hygiene and Tropical Medicine	840	540	101.396	790	65	19.2	45.0	445
Loughborough University	1535	1675	111.671	5175	275	12.1	43.1	278
Newcastle University	2910	2820	189.671	8125	640	14.0	44.4	688
Queen Mary University of London	2385	1975	148.949	6515	315	15.9	45.1	700
The Queen's University of Belfast	1665	1865	124.668	6985	400	12.3	43.8	401
Swansea University	1425	1590	113.760	5495	200	11.4	40.8	291
The University of Edinburgh	4650	5265	349.679	9855	975	15.1	44.8	1171
The University of Manchester	5195	5005	398.897	13710	1030	15.1	44.9	1822
The University of Sheffield	3270	3535	238.219	9610	695	14.0	44.3	933
The University of Warwick	2855	3370	230.201	8805	465	14.7	44.2	563
University College London	7415	4745	504.456	14260	1515	17.8	45.4	2267
The University of Aberdeen	1460	1440	79.887	3760	240	14.1	43.3	467
The University of Bath	1370	1785	106.103	4710	235	12.6	43.5	276
University of Birmingham	3625	3475	236.856	12120	790	14.0	44.1	792
The University of Bristol	3175	2845	227.018	7020	585	16.3	45.2	869
The University of Cambridge	5965	4975	962.752	5330	2345	17.9	46.2	2136
The University of Dundee	1425	1515	94.508	3830	175	14.7	44.5	311
University of East Anglia	1800	1730	98.152	5425	270	14.5	45.5	186
The University of Exeter	2055	2215	148.869	8000	370	17.2	45.7	382
The University of Glasgow	3840	3025	223.065	9420	560	15.0	45.6	937
The University of Hull	1035	1185	74.413	4750	210	11.6	41.9	146
The University of Kent	1425	1785	96.181	5925	225	9.8	42.7	128
The University of Leeds	3485	3860	254.056	10595	650	14.2	45.4	801
The University of Leicester	1645	1975	111.779	5510	315	12.7	42.1	465
The University of Liverpool	2835	2700	190.915	8140	520	14.2	43.9	753
University of Nottingham	3420	3640	260.476	10275	895	12.8	44.7	833
The University of Oxford	6770	5830	614.688	6960	1210	18.8	46.0	2257
University of Plymouth	1345	1340	94.892	6920	160	13.3	42.4	171
The University of Reading	1700	1885	118.492	5810	300	13.4	45.4	280
The University of Southampton	3030	3110	224.320	8100	670	13.5	45.0	972
The University of St Andrews	1165	1285	91.220	2495	245	12.5	45.2	227
The University of Strathclyde	1580	1785	106.744	6885	355	11.8	43.4	333
The University of Surrey	1425	1515	104.329	6775	245	12.6	42.5	405
The University of Sussex	1865	1065	125.175	5745	280	13.3	43.9	193
York St John University	325	350	23.508	2255	5	13.5	43.7	404
Mean	2548	2393	195.085	6781	502	14.0	44.1	667

Table B.3 – Descriptive results for each country

German HEIs								
	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Min	1192.00	742.00	27.86	1560.00	120.00	7.60	39.90	121.00
Median	4270.50	3136.50	86.40	4295.50	504.00	10.85	43.00	543.00
Mean	4621.83	3620.11	95.96	4361.85	536.11	10.76	42.96	606.50
Max	10112.00	8623.00	247.36	8903.00	1306.00	13.30	46.20	1664.00
Sd	2108.38	2442.26	48.95	1726.56	272.62	1.41	1.33	416.93
UK HEIs								
	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Min	325.00	350.00	23.51	790.00	5.00	9.70	40.80	65.00
Median	1865.00	1885.00	125.17	6515.00	315.00	14.00	44.20	405.00
Mean	2547.56	2392.56	195.08	6780.78	501.78	13.97	44.05	667.24
Max	7415.00	5830.00	962.75	14260.00	2345.00	19.20	46.20	2544.00
Sd	1630.07	1342.13	167.29	2936.34	444.32	2.27	1.42	628.87

Table B.4 – Correlation coefficients for each country

German HEIs								
	x_1	x_2	x_3	y_1	y_2	y_3	y_4	y_5
x_1	1.000	0.856	0.649	0.794	0.898	0.657	0.281	0.691
x_2	0.856	1.000	0.473	0.595	0.770	0.512	0.172	0.555
x_3	0.649	0.473	1.000	0.836	0.684	0.473	0.216	0.527
y_1	0.794	0.595	0.836	1.000	0.801	0.567	0.185	0.628
y_2	0.898	0.770	0.684	0.801	1.000	0.650	0.300	0.827
y_3	0.657	0.512	0.473	0.567	0.650	1.000	0.634	0.536
y_4	0.281	0.172	0.216	0.185	0.300	0.634	1.000	0.313
y_5	0.691	0.555	0.527	0.628	0.827	0.536	0.313	1.000
UK HEIs								
	x_1	x_2	x_3	y_1	y_2	y_3	y_4	y_5
x_1	1.000	0.930	0.875	0.725	0.883	0.624	0.589	0.903
x_2	0.930	1.000	0.839	0.731	0.862	0.533	0.584	0.846
x_3	0.875	0.839	1.000	0.429	0.960	0.609	0.549	0.869
y_1	0.725	0.731	0.429	1.000	0.511	0.242	0.349	0.503
y_2	0.883	0.862	0.960	0.511	1.000	0.563	0.563	0.901
y_3	0.624	0.533	0.609	0.242	0.563	1.000	0.731	0.669
y_4	0.589	0.584	0.549	0.349	0.563	0.731	1.000	0.580
y_5	0.903	0.846	0.869	0.503	0.901	0.669	0.580	1.000

Table B.5 – Efficiency results of German HEIs

HEIs	$F^{\mathcal{U}_{GER}}$	$F^{\mathcal{U}}$	$F^{\mathcal{U}_{UK}}$
Frankfurt University	1.000	0.804	1.085
TU Dortmund University	1.000	0.568	1.036
University of Bremen	1.000	0.638	1.092
University of Rostock	1.000	1.000	1.214
Georg August Goettingen University	1.000	1.000	1.113
University of Bayreuth	1.000	0.640	1.146
Johannes Gutenberg University Mainz	1.000	1.000	1.114
University of Kiel	1.000	1.000	1.178
Philipps University of Marburg	1.000	1.000	1.126
University of Erlangen Nuernberg	1.000	1.000	1.173
Humboldt University of Berlin	1.000	1.000	1.196
University of Potsdam	1.000	1.000	1.177
Tuebingen University	1.000	1.000	1.185
Ludwig Maximilian University of Munich	1.000	1.000	1.170
Free University of Berlin	1.000	1.000	1.183
Heidelberg University	1.000	1.000	1.239
Karlsruhe Institute of Technology	1.000	1.000	1.179
Friedrich-Schiller-University of Jena	1.000	1.000	1.186
Ulm University	1.000	1.000	1.211
University of Freiburg	1.000	1.000	1.155
Technical University of Munich	1.000	0.923	1.040
University of Konstanz	1.000	1.000	1.088
Technical University of Berlin	1.000	0.587	1.011
University of Regensburg	1.000	1.000	1.142
Ernst Moritz Arndt University of Greifswald	1.000	1.000	1.213
University of Hohenheim	1.000	0.665	1.113
University of Wuerzburg	1.000	0.907	1.151
Technische University Dresden	0.922	0.922	1.161
University of Hamburg	0.870	0.829	1.093
Justus Liebig University Giessen	0.847	0.661	1.084
RWTH Aachen University	0.825	0.737	1.008
Leipzig University	0.791	0.760	1.138
University of Bonn	0.768	0.725	1.033
University of Muenster	0.751	0.741	1.095
University of Stuttgart	0.750	0.601	1.040
Martin Luther University	0.746	0.653	1.143
University of Duisburg-Essen	0.742	0.689	1.015
Ruhr-University Bochum	0.730	0.656	1.015
Carl von Ossietzky the University of Oldenburg	0.729	0.416	1.047
University of Cologne	0.712	0.710	1.074
Saarland University	0.708	0.663	1.070
Heinrich Heine University Duesseldorf	0.681	0.664	1.045
Technical University of Darmstadt	0.678	0.610	1.059
University of Kaiserslautern	0.653	0.411	1.065
Bielefeld University	0.600	0.448	0.487
University of Hannover	0.510	0.479	0.551
Mean	0.892	0.807	1.090

Table B.6 – Efficiency results of UK HEIs

HEIs	$F^{\mathcal{U}_{UK}}$	$F^{\mathcal{U}}$	$F^{\mathcal{U}_{GER}}$
King's College London	1.000	1.000	1.064
University College London	1.000	1.000	1.090
The University of Oxford	1.000	1.000	0.743
Cranfield University	1.000	1.000	1.908
London School of Economics and Political Science	1.000	1.000	1.146
Queen Mary University of London	1.000	0.914	1.142
The University of St Andrews	1.000	1.000	1.282
The University of Glasgow	1.000	1.000	1.081
Imperial College London	1.000	1.000	1.227
York St John University	1.000	1.000	2.689
The University of Aberdeen	1.000	0.890	1.321
The University of Manchester	1.000	1.000	1.131
University of Birmingham	1.000	1.000	1.166
The University of Leeds	1.000	1.000	1.127
The University of Surrey	1.000	1.000	1.385
University of Plymouth	1.000	1.000	1.229
The University of Strathclyde	1.000	1.000	1.302
University of Durham	1.000	1.000	1.300
London School of Hygiene and Tropical Medicine	1.000	1.000	1.519
The University of Exeter	1.000	1.000	1.266
City, University of London	1.000	1.000	1.121
The University of Sussex	1.000	1.000	1.180
The Queen's University of Belfast	1.000	1.000	1.327
University of East Anglia	1.000	1.000	1.163
The University of Cambridge	1.000	1.000	1.116
University of Nottingham	1.000	1.000	1.199
Newcastle University	0.954	0.916	1.170
The University of Sheffield	0.939	0.939	1.172
The University of Southampton	0.929	0.923	1.182
The University of Liverpool	0.917	0.902	1.148
The University of Leicester	0.890	0.881	1.292
The University of Reading	0.885	0.859	1.219
The University of Bristol	0.882	0.855	1.115
Cardiff University	0.861	0.802	1.085
The University of Edinburgh	0.855	0.855	1.103
The University of Warwick	0.823	0.823	1.115
The University of Dundee	0.767	0.740	1.180
Heriot-Watt University	0.741	0.741	1.618
Swansea University	0.732	0.732	1.200
The University of Hull	0.728	0.728	1.438
Loughborough University	0.728	0.728	1.224
The University of Bath	0.726	0.726	1.268
The University of Lancaster	0.699	0.628	1.082
Brunel University London	0.642	0.634	1.302
The University of Kent	0.607	0.601	1.183
Mean	0.918	0.907	1.252

Table B.7 – Slacks of German HEIs relative to $F^{\mathcal{U}_{GER}}$

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Bielefeld University			23.375	74.436	134.410	0.393	1.065	425.462
Carl von Ossietzky the University of Oldenburg			11.522	8.056	42.853	2.622	0.184	157.898
Tuebingen University								
Free University of Berlin								
Friedrich-Schiller-University of Jena								
University of Erlangen Nuernberg								
Georg August Goettingen University								
Frankfurt University								
University of Hannover	198.843		55.241		107.765	1.836	3.439	986.029
Heidelberg University								
Heinrich Heine University Duesseldorf		2676.829		1463.735	302.191	0.086		509.461
Humboldt University of Berlin								
Johannes Gutenberg University Mainz								
University of Wuerzburg								
Justus Liebig University Giessen		333.086	8.067			0.699	1.812	298.491
Karlsruhe Institute of Technology								
University of Kiel								
Leipzig University		1847.407			137.543	1.861	2.739	404.198
Ludwig Maximilian University of Munich								
Martin Luther University				77.208	114.140	2.302	3.090	317.475
Philipps University of Marburg								
Ruhr-University Bochum	834.997		27.653		164.450	1.046	1.256	759.619
RWTH Aachen University		2458.678	8.061	764.377	107.152	0.919	1.884	631.262
Saarland University		2562.539		454.163	91.223	1.823	2.462	656.813
Technical University of Munich								
Technical University of Berlin								
Technical University of Darmstadt		261.396	0.219		57.829	0.194	0.800	700.870
Technische University Dresden		2025.475				0.929	0.509	297.465
University of Kaiserslautern			10.228	474.405	37.795	0.978	0.602	267.776
TU Dortmund University								
Ulm University								
University of Hamburg		1066.864		447.613	21.238	2.277		370.650
University of Regensburg								
University of Bayreuth								
University of Bonn				1733.254	239.823			566.857
University of Bremen								
University of Cologne		607.495		315.735	439.846	0.670	0.899	847.824
University of Duisburg-Essen		3330.622			265.771	0.666	1.672	714.849
University of Freiburg								
Ernst Moritz Arndt University of Greifswald								
University of Hohenheim								
University of Konstanz								
University of Muenster		2657.725			191.086		0.043	783.123
University of Potsdam								
University of Rostock								
University of Stuttgart	849.523		4.476	115.518	64.722			486.340
Mean	470.841	1802.556	14.884	538.954	148.226	1.135	1.497	509.123

Table B.8 – Slacks of German HEIs relative to $F^{\mathcal{U}_{ALL}}$

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Bielefeld University	792.559		12.485	122.688	101.360	1.742	1.783	892.047
Carl von Ossietzky the University of Oldenburg	387.328			388.188	38.953	4.050	1.494	744.628
Tuebingen University								
Free University of Berlin								
Friedrich-Schiller-University of Jena								
University of Erlangen Nuernberg								
Georg August Goettingen University								
Frankfurt University		1623.559			92.952		0.470	778.307
University of Hannover	597.911		51.438		82.214	2.302	3.587	1135.522
Heidelberg University								
Heinrich Heine University Duesseldorf		1714.581		777.413	324.501	1.104	0.074	651.925
Humboldt University of Berlin								
Johannes Gutenberg University Mainz								
University of Wuerzburg	1357.471	1682.234			47.662			252.677
Justus Liebig University Giessen		162.721			36.181	1.726	2.582	851.600
Karlsruhe Institute of Technology								
University of Kiel								
Leipzig University		1090.446			133.963	3.051	2.412	483.256
Ludwig Maximilian University of Munich								
Martin Luther University		383.257		272.481	77.319	5.499	3.597	504.338
Philipps University of Marburg								
Ruhr-University Bochum	998.764				178.277	1.319	1.727	1139.202
RWTH Aachen University		1717.664			376.291	2.906	1.850	943.539
Saarland University		2836.595		848.635	143.535	3.652	2.390	648.704
Technical University of Munich	663.503				159.237	2.261		122.280
Technical University of Berlin	1020.170		6.446		82.661	0.876		1171.619
Technical University of Darmstadt					142.334	0.866	0.835	878.345
Technische University Dresden		2025.475				0.929	0.509	297.465
University of Kaiserslautern	534.937			729.762	19.454	2.332	1.719	798.926
TU Dortmund University	1203.353		0.322		41.296	3.952	2.795	672.253
Ulm University								
University of Hamburg	288.338	462.366		1621.944	67.501	4.496		233.458
University of Regensburg								
University of Bayreuth	30.592			466.895		0.838	1.375	345.097
University of Bonn				2397.540	279.339	1.454		612.883
University of Bremen	332.025					3.784	0.868	587.375
University of Cologne		606.566			445.773	0.778	0.885	880.212
University of Duisburg-Essen		2993.037			352.158	0.685	1.805	929.970
University of Freiburg								
Ernst Moritz Arndt University of Greifswald								
University of Hohenheim				402.826	18.164	3.962	1.059	357.003
University of Konstanz								
University of Muenster		2417.172			153.381			865.108
University of Potsdam								
University of Rostock								
University of Stuttgart	1220.019				79.375	0.170		1015.498
Mean	725.152	1408.262	17.673	802.837	138.955	2.280	1.691	696.046

Table B.9 – Slacks of UK HEIs relative to $F^{\mathcal{U}_{UK}}$

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Brunel University London	36.068	10.345			4.978	2.581	2.867	409.598
Cardiff University	548.743	281.774			190.007	1.078	0.597	180.796
City, University of London								
Cranfield University								
University of Durham								
Heriot-Watt University		162.710	33.536			3.703	0.619	273.474
Imperial College London								
King's College London								
The University of Lancaster	538.294	225.834			42.824	0.165	1.354	513.824
London School of Economics and Political Science								
London School of Hygiene and Tropical Medicine								
Loughborough University		206.009			56.787	1.827	0.797	415.595
Newcastle University	259.700	199.557				0.115		159.081
Queen Mary University of London								
The Queen's University of Belfast								
Swansea University		157.396	3.080		95.951	1.925	2.335	326.939
The University of Edinburgh	1.600	1424.078			209.672	1.121	0.232	651.153
The University of Manchester								
The University of Sheffield		398.647	0.773		80.199	1.686	0.614	71.818
The University of Warwick		616.319	20.742		203.275	0.063		355.177
University College London								
The University of Aberdeen								
The University of Bath		455.848			65.477	1.164	0.123	418.861
University of Birmingham								
The University of Bristol	326.245	157.465			157.968	0.049		346.415
The University of Cambridge								
The University of Dundee	219.526	264.490		300.238	84.231			298.280
University of East Anglia								
The University of Exeter								
The University of Glasgow								
The University of Hull		35.346				1.160	1.755	251.986
The University of Kent		380.050			27.822	3.445	0.667	350.720
The University of Leeds								
The University of Leicester	52.538	448.877			19.121	1.209	1.850	194.658
The University of Liverpool	140.365	160.191			110.693	0.146	0.304	169.022
University of Nottingham								
The University of Oxford								
University of Plymouth								
The University of Reading					44.357	1.676		105.767
The University of Southampton	72.085	293.168			60.634	2.411		110.217
The University of St Andrews								
The University of Strathclyde								
The University of Surrey								
The University of Sussex								
York St John University								
Mean	219.516	326.561	11.626	300.238	90.875	1.418	1.086	294.915

Table B.10 – Slacks of UK HEIs relative to $F^{\mathcal{U}_{ALL}}$

	Aca. staff	Other staff	Expen.	Bac.	Master	Pub. 10%	Pub. 50%	Collab.
Brunel University London		3.109			10.773	2.535	2.863	421.340
Cardiff University		171.693			278.173	0.380	0.531	359.440
City, University of London								
Cranfield University								
University of Durham								
Heriot-Watt University		162.710	33.536			3.703	0.619	273.474
Imperial College London								
King's College London								
The University of Lancaster		105.123			118.718		1.577	657.602
London School of Economics and Political Science								
London School of Hygiene and Tropical Medicine								
Loughborough University		206.009			56.787	1.827	0.797	415.595
Newcastle University		236.186			14.317	0.456		276.474
Queen Mary University of London					119.956		0.111	59.806
The Queen's University of Belfast								
Swansea University		157.396	3.080		95.951	1.925	2.335	326.939
The University of Edinburgh		1424.345			209.736	1.119	0.232	651.473
The University of Manchester								
The University of Sheffield		398.647	0.773		80.199	1.686	0.614	71.818
The University of Warwick		616.319	20.742		203.275	0.063		355.177
University College London								
The University of Aberdeen		336.310			21.197		0.893	236.450
The University of Bath		455.848			65.477	1.164	0.123	418.861
University of Birmingham								
The University of Bristol		83.056			202.925		0.163	430.466
The University of Cambridge								
The University of Dundee		253.836		604.960	87.511			341.029
University of East Anglia								
The University of Exeter								
The University of Glasgow								
The University of Hull		35.346				1.160	1.755	251.986
The University of Kent		381.066			26.796	3.324	0.537	364.950
The University of Leeds								
The University of Leicester		438.337			27.562	1.142	1.844	211.762
The University of Liverpool		131.233			132.579		0.306	212.739
University of Nottingham								
The University of Oxford								
University of Plymouth								
The University of Reading					55.480	1.515		146.081
The University of Southampton		278.325			71.898	2.335		132.741
The University of St Andrews								
The University of Strathclyde								
The University of Surrey								
The University of Sussex								
York St John University								
Mean	NaN	309.205	11.626	604.960	98.911	1.622	0.956	300.737

Table B.11 – Descriptive results of the fields of education

Students	Measure	Agriculture	Arts	Business	Education	Engineering	Health	Information	Natural sciences	Social sciences
	Min	0.000	0.000	755.000	15.000	0.000	0.000	0.000	646.000	192.000
	Median	0.000	4836.000	3939.000	1568.000	394.000	2591.000	1298.000	4267.500	2834.000
German	Mean	466.860	5021.222	4216.356	2089.956	3650.953	2315.978	1655.644	4432.174	2741.114
Bachelor	Max	3298.000	13725.000	9973.000	10081.000	21968.000	6612.000	5260.000	9015.000	6980.000
	Sd.	857.151	3622.636	2080.834	1876.795	5830.674	1782.983	1314.274	2020.589	1622.169
	Numbers	20075	225955	189736	94048	156991	104219	74504	203880	120609
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	57.000	0.000
	Median	0.000	216.000	124.500	41.000	26.500	147.000	77.000	569.000	110.000
German	Mean	57.087	316.196	179.522	87.068	250.886	219.761	87.622	639.196	162.956
Master	Max	543.000	1182.000	611.000	465.000	2179.000	685.000	319.000	1605.000	733.000
	Sd.	119.137	305.209	154.075	113.621	447.819	208.694	69.711	385.048	160.620
	Numbers	2626	14545	8258	3831	11039	10109	3943	29403	7333
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Median	0.000	2405.000	3140.000	740.000	1960.000	2790.000	590.000	2445.000	2990.000
UK	Mean	191.556	2883.556	3262.333	940.556	2134.778	3049.222	637.778	2574.889	3098.333
Bachelor	Max	1765.000	6630.000	8025.000	4565.000	6140.000	10680.000	1615.000	5970.000	5700.000
	Sd.	371.690	1743.600	1581.556	980.462	1597.217	2263.965	355.538	1322.996	1530.984
	Numbers	8620	129760	146805	42325	96065	137215	28700	115870	139425
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Median	0.000	195.000	100.000	60.000	240.000	200.000	50.000	350.000	200.000
UK	Mean	17.556	232.889	104.000	86.778	300.667	312.667	68.556	448.778	233.000
Master	Max	150.000	1100.000	245.000	720.000	1205.000	1440.000	280.000	1685.000	770.000
	Sd.	36.550	227.128	50.549	122.489	283.947	322.852	62.710	417.186	187.917
	Numbers	790	10480	4680	3905	13530	14070	3085	20195	10485

Table B.12 – Sample HEIs and OECD student shares of the fields of education, source: OECD (2015)

	Agriculture	Arts	Business	Education	Engineering	Health	Information	Natural sciences	Social sciences
German sample HEIs	0.018	0.188	0.155	0.076	0.131	0.089	0.061	0.182	0.100
German OECD	0.015	0.139	0.227	0.072	0.210	0.073	0.063	0.104	0.076
UK sample HEIs	0.010	0.151	0.164	0.050	0.118	0.163	0.034	0.147	0.162
UK OECD	0.011	0.158	0.183	0.071	0.093	0.145	0.041	0.149	0.110

Table B.13 – Literature review

Authors	Inputs	Outputs	Method	Sample
Athanas-sopoulos et al., 1997	Students, academic staff, research income, expenditures	Graduates, research rating	CCR-model, weight restrictions	45 UK HEIs, 1992
Avkiran, 2001	Academic staff, non-academic staff	Students, research quantum, graduate employment rate, fee-paying enrolments	BCC-model	36 Australian HEIs ,1995
Warning, 2004	Expenditure on personnel, other expenditure	Publications, graduates	CCR-model	73 German HEIs, 1998
Johnes, 2006	Students, staff, administration expenditure, library expenditure, total depreciations and interests	Graduates, research grants	BCC-model	109 UK HEIs, 2000/2001
Fandel, 2007	Students, personnel, outside funding	Graduates, doctorates	BCC-model	15 German HEIs, 1997
Thanassoulis et al., 2011	Operating costs	Graduates, research grants, other income	BCC-model	358 UK HEIs, 2000-2003
Agasisti et al., 2012	Students, academic staff, expenditures	Graduates, research grants, contracts	BBC-model	122 German and Italian HEIs, 2001-2007

Continued on next page

Authors	Inputs	Outputs	Method	Sample
Aziz et al., 2013	Academic staff, non-academic staff, operating expenses	graduates, research grants, publications	BCC-model	20 Malaysian HEI departments, 2011
Bakaya et al., 2014	Budget, staff	Graduates, third-party funds	BCC-, CCR-models	33 German HEIs, 2011
Nazarko et al., 2014	Budget	Students, scholarly achievements, alumni employment rate	CCR-model	19 HEIs
Göken et al., 2015	Area, academic staff, non-academic staff	Publications, graduates	BCC-model	26 Turkish departments, 2012
Mikuová, 2015	Expenditure, academic staff	Graduates, students	BCC	26 Czech HEIs, 2013
Agasisti et al., 2016	Students, staff, expenditures	Graduates, research grants	SFA-model	71 Dutch and Italian HEIs, 2005-2009
Chuanyi et al., 2016	Expenditure, equipment	Graduates, publications, patents	BCC, SBM, SFA	48 Chinese HEIs
Gawellek et al., 2016	Professors, funding	Students, third-party funding	BCC-model	164 German HEIs,2001-2011
Veiderpass et al., 2016	Academic staff, non-academic staff, non-personnel expenditures, income	Graduates	BCC-model	944 HEIs in 17 European countries, 2008

Continued on next page

Authors	Inputs	Outputs	Method	Sample
Visbal-Cadavid et al., 2017	Academic staff, non-academic staff, financial and physical resources	Students, faculty mobility, published articles, journals	Augmented BCC-model	32 Colombian HEIs, 2011/2012
Wolszczak-Derlacz, 2017	Academic staff, non-academic staff, total revenue, students	Publications, graduates	BCC-model	500 HEIs in 11 countries, 2000-2010
Lehmann et al., 2018	Funding	Graduates, publications, patents	BCC-model	133 German and Italian HEIs, 2006-2011
Moreno et al., 2018	Students, teachers, expenditures	Graduates, publications, citations,	CCR-model, SBM	47 Spanish HEIs, 2008-2015
Gralka et al., 2019	Expenditures, wages	Graduates, grants, publications	BCC-model, SFA	72 German HEIs, 2004-2013
Wohlrabe et al., 2019a	Expenditures, staff	Graduates, publications	CCR-, BCC-model, SFA	14 German HEIs, 2010-2012

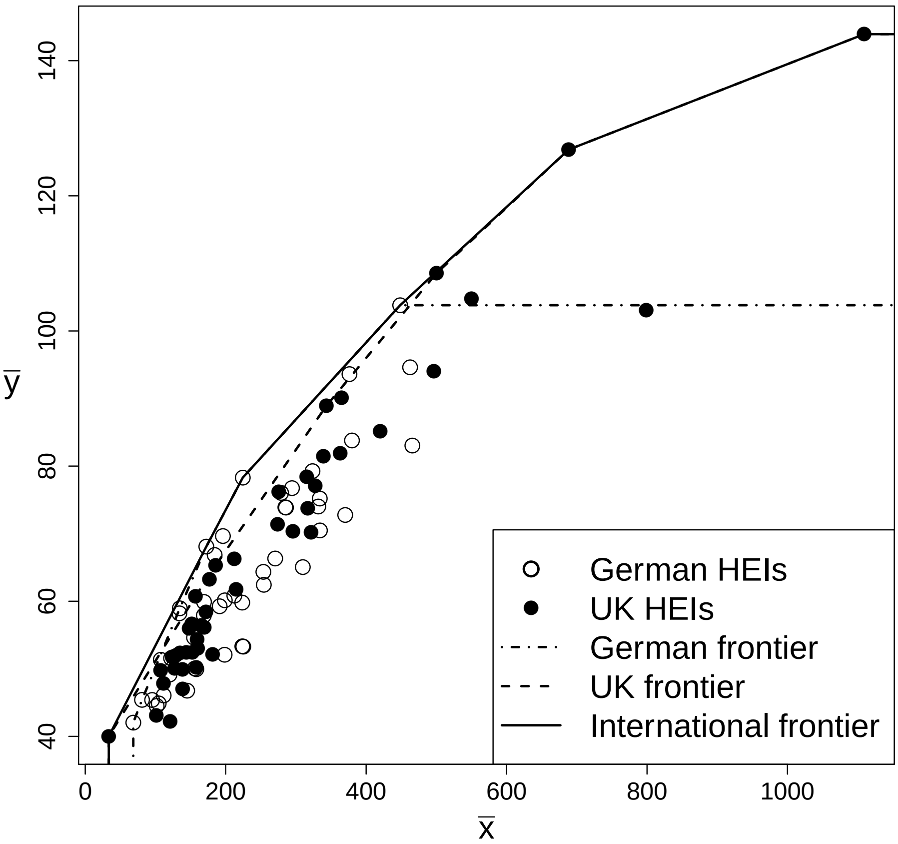


Figure B.1 – Efficiency decomposition example

6

An Encompassing Assessment of OECD Countries

An encompassing assessment of OECD countries using weight restricted DEA models

Prof. Andreas Behr, Gerald Fugger

University Duisburg-Essen, 45141 Essen, Germany.

Abstract

This paper assesses the capacity of 33 OECD countries to provide their citizens with a long and fulfilling life, given their economic, environmental, and health endowments. Such comparisons are usually made with composite indices that do not distinguish between inputs and outputs and give equal weight to all variables. Therefore, these composite indices do not consider the individual characteristics of countries. We assess the countries' efficiency using Data Envelopment Analysis (DEA) models to allow for an assessment which considers the countries' specific strengths and weaknesses. Additional weight restrictions ensure that each country is evaluated based on all variables and provide a better distinction between efficient and inefficient countries. Moreover, robustness checks reveal that the countries' efficiency ranks are quite robust to alternative weight restrictions.

Keywords: Data Envelopment Analysis, Human development, Well-being, Assurance regions

JEL Classification: C14 C52 C61 I31

6.1 Introduction

Offering its citizens the highest well-being and a life worth living is the primary goal of all societies and thus should also be prioritised by their political decision makers (Patrizii et al., 2017). However, long-used measures of economic prosperity do not adequately describe the standard of living of a nation's citizens. Recent international surveys like the Gallup World Poll better reflect the socio-economic conditions of a country and have gained increasing attention in recent literature (Peiró-Palomino et al., 2018). We focus on countries' ability to provide their citizens a as long, healthy and rewarding life as possible. Countries have different starting levels of available resources and ensure the quality of life of their citizens to varying degrees. An efficiency analysis can take into account both the countries' endowments (inputs) and the quality of life achieved (outputs) simultaneously. We therefore assess the countries' abilities with radial Data Envelopment Analysis (DEA) models.

DEA models reveal inefficiencies and allow the assessment of individual countries strengths and weaknesses (Xu et al., 2017). Moreover, these models do not a priori assume specific functional relations between the inputs and outputs and have been widely used to assess efficiency at country-level (Mizobuchi, 2017). Countries' macroeconomic performance has been studies by Lovell et al. (1995), followed by health-related studies like those of Afonso et al. (2006) and environmental studies e.g. from Rashidi et al. (2015), and finally for gauging the countries socio-economic performance, see for example Mariano et al. (2015).

In standard DEA models weights of the inputs and outputs are determined to maximise the efficiency of each country. The weights are restricted only to being non-negative and ensure that for each country, the ratio of weighted outputs to weighted inputs is less than or equal to one. This allows countries to be assessed as specialised on specific inputs and outputs as well as to assign zero weights to the other variables. Hence, variables having zero weights are not included in the country's efficiency assessment. Benneyan et al. (2007) describe the presence of zero weights as "irrational weighting" and propose the inclusion of additional weight restrictions to prevent them. Furthermore, weight restrictions can increase the discriminatory power of a model to better distinguish between efficient and inefficient DMUs (Atici et al., 2015). Only few human development assessments address the problems of fully specialised DMUs and low discriminatory power. For instance, Benneyan et al. (2007) and Peiró-Palomino et al. (2018) include additional weight restrictions to exclude zero weights and limit the weights towards a reasonable range. A standard DEA without additional restrictions provides zero weights at least for one input or one output in each country. Therefore, we include additional weight restrictions in a second step. The additional weight restrictions reduce the average efficiencies, but rank correlation coefficients indicate a rather robust ranking of the efficiency scores compared to alternative restrictions.

In addition to the usual input and output measures, we include environmental and health related inputs and subjective well-being as an output into the efficiency assessment of 33 countries.[1] The efficiency analysis not only provides an overall assessment of the efficiency but also identifies for each country relevant reference sets of countries for a benchmark.[2]

6.2 Literature Overview

An early first efficiency study at country level focused on macroeconomic performance (Lovell et al., 1995). Most subsequent studies used measures such as the GDP and price stability as outputs and reduced the countries' inputs to a simple one. See Table C.3 in the appendix for an overview of relevant studies. This input represented the policy makers, the so-called helmsmen, who provide the macroeconomic services. Follow-up publications replaced the helmsmen with quantifiable variables such as government spending (Mohamad et al., 2011). On the basis of these initial studies, the research questions were later extended to include health and environmental related aspects to cover a broader perspective of societies (Afonso et al., 2006; Rashidi et al., 2015). Recent analyses examine countries' effectiveness in maximising human development, taking into account a variety of socio-economic quantities. One of the most important outputs thereby is the subjective well-being of citizens (Mariano et al., 2015).

In the following literature review, the most relevant studies from each literature strand as well as the most important composite indices and important DEA applications relevant for our analysis are explained in more detail. Tables C.3 to C.6 in the appendix provide a comprehensive overview of the literature strands.

[1] In the following we use the terms well-being and happiness interchangeably. Both variables are subjective, relative, and based on a broad literature foundation. High happiness is a fundamental part of a fulfilled life (Gilligan et al., 2017). Suh et al. (2018) provide an overview of well-being predictors and relevant literature.

[2] Table C.1 provides the country names and their abbreviations.

6.2.1 Economic Performance

The first macroeconomic efficiency studies evaluate either the countries' ability to maximise economic performance (the only output) given their financial and human capital (inputs), or their relative ability to maximise the quantities of the OECD's magic diamond (Färe et al., 1994; Brockett et al., 1999).[3] The latter approach assumes a helmsmen who provides the four macroeconomic services. In DEA models, the helmsmen is included as an input that equals unity for each country. The use of a helmsman simplifies the model but ignores the resources consumed by the economy and reduces the models' complexity (Pavone et al., 2015). More recent studies consider the assumption of the helmsman to be unrealistic since the countries have different material and social endowments (for example, Mohamad (2007) and Mohamad et al. (2011)).

6.2.2 Health

In most healthcare systems efficiency studies, the inputs represent measures of physical capital, medical technology indicators, and human capital. Retzlaff-Roberts et al. (2004) proxy the latter by the number of doctors. Their other health-related inputs include health expenditure, the number of hospital beds, and the amount of high-tech diagnostic equipment. These inputs are assumed to be within the control of the healthcare system and measure the access to healthcare services and the invested capital. Retzlaff-Roberts et al. (2004) include the infant mortality and the life expectancy at birth as health related outputs. The authors admit that their outputs (especially the infant mortality) may capture factors beyond the control of the health system such as poverty. In the analysis of Retzlaff-Roberts et al. (2004), countries with both relatively high (e.g. Sweden and Japan) and relatively low (e.g. Mexico and Turkey) health outcomes found to be efficient. Inefficient countries are either utilizing their resources inefficiently (like Switzerland) or should have higher health outcomes given their inputs (like Hungary).

Asandului et al. (2014) calculate the efficiency of European public healthcare systems. Their outputs are the life expectancy at birth, health adjusted life expectancy, and infant mortality rate. The health adjusted life expectancy represents the quality of

[3]The OECD magic diamond states that the following four macroeconomic quantities are important targets that the decision makers in each country should maximise, but contradict each other in their fulfilment: GDP per capita, price stability rate, employment rate, and the trade balance. (Lovell et al., 1995; Moesen et al., 1998)

life and is assumed to be one of the main objectives of policy makers. Any increase decrease the costs for the healthcare systems, increase employee productivity, and includes further obvious social benefits. Asandului et al. (2014) employ the health adjusted life expectancy and life expectancy at birth in two separate DEAs because the two outputs are strongly positive correlated. The inputs of the healthcare systems are approximated by the number of doctors, number of hospital beds, and public health expenditures. Six out of 30 countries are efficient. Austria and the Czech Republic are the least efficient countries.

Behr et al. (2017) criticise the use of intermediate outputs like the number of surgeries as outputs in many health-related studies. They propose variables such as years gained or health-adjusted life expectancy as preferred outputs in a DEA, as these more accurately capture the impact of healthcare systems.

6.2.3 Environmental Studies

Environmental studies, or short eco-efficiency studies, focus on the relationship between environmental costs and the gains of social or economic activities (Rashidi et al., 2015). Political and social shareholders have to balance the trade-off between the economic output and the environmental factors to achieve a sustainable development (Zhou et al., 2018). Environmental issues and eco-efficiency analysis have gained increasing attention in recent years (Alsahlawi, 2013). Energy efficiency studies and environmental efficiency are closely related, and both aspects are simultaneously considered in most eco-efficiency analyses e.g. Sueyoshi et al. (2017).

Suzuki et al. (2016) assess the national energy-environment-economic efficiency strategies of 27 countries. Therefore, they include energy consumption and the population as inputs and GDP and CO_2 emissions as outputs. The relationship between their outputs is complementary because an increase in GDP most likely increases the CO_2 emissions, too. However, the authors argue that this relationship has been reduced with the introduction of new energy technologies that produce less CO_2 emissions for the same amount of energy. Overall the European countries are the most efficient countries and Asian countries are mostly inefficient.

Guo et al. (2017) calculate the eco-efficiency of OECD countries and China. They include the land area, population, and energy use as inputs. The energy use and the population proxy the environmental costs of the economic activities and policy decisions. The authors select the GDP and the CO_2 emissions as outputs to consider the trade-offs between CO_2 emissions and energy consumption. These trade-offs represent the tension between environmental protection and economic growth. Nine of the 27 countries are eco-efficient in each DEA, calculated separately for each year from 2000 to 2010. Overall China is the most eco-inefficient country.

6.2.4 Well-being Studies

Several efficiency studies focus on well-being and its implication on the country-level. Mariano et al. (2015) provide an overview of 57 efficiency studies that assess human development using DEA. They state that DEA is a well-suited tool to assess the efficiency of countries to provide an as high standard of human development as possible given their economic, social, and environmental endowments. Patrizii et al. (2017) argue that multidimensional well-being measures better capture countries' socio-economic conditions than simple measures like the GDP. In their efficiency analysis, they include various desirable (employment rate, educational attainment, life satisfaction,...) and undesirable (air pollution, homicide rate,...) outputs. The hours worked per person and per capita consumption are the inputs. They find that relatively poor countries focus more on low inputs and relatively rich countries weight environmental and social outputs higher (Patrizii et al., 2017).

Ülengin et al. (2011) assess the efficiency of countries to maximise human development (the primary objective of every society) given their social-economic endowments. Three dimensions (education, health and income) represent human development and are the outputs in a DEA. Three indices (basic requirements, efficiency enhancers, and innovation and sophistication factors) are the inputs. The three inputs are based on 177 variables of the Global Competitiveness Report published by the World Economic Forum. However, the interactions between the 177 variables that make up the three inputs remain unclear and the variables are partly highly correlated. In addition, the individual variables are equally weighted to obtain the final three inputs (Schwab et al., 2016). In the efficiency assessment of Ülengin et al. (2011), any inefficiencies due to the underlying variables cannot be identified because only the aggregated indices are used as inputs. Norway, Italy, Argentina, and the United States are efficient and Kenya and Nigeria are the least efficient countries.

Mizobuchi (2017) finds health and income related factors to have the greatest positive impact and community-based factors the least impact on well-being. He includes ten socio-economic indices and the Gini coefficient as input in a DEA. The average well-being of nations is the sole output. The ten socio-economic indices are based on 24 underlying socio-economic quantities. Six of the underlying variables are negatively correlated with subjective well-being which is against the more is better assumption of DEA. The correlation coefficients between the ten indices (the DEA inputs) and the output are not reported.

Guardiola et al. (2014) use fixed weights derived form a DEA to calculate a well-being composite index. In the DEA, they include the self-reported life satisfaction as output and include ten normalised socio-economic variables as inputs. Wealth and community related variables are weighted higher by the DEA than quantities from the areas of leisure, nutrition, and love. A regression for well-being provides negative coefficients of wealth and trust related variables (Dyson et al., 2001). It remains unclear whether the negative signs result from the interaction of the explanatory variables or contradict the assumption of the DEA that an increase in inputs must increase outputs. Bivariate correlation coefficients could improve the understanding of the relationship between inputs and well-being.

6.2.5 The Human Development Index and the Better Life Index

The Human Development Index (HDI) measures human development and is one of the most popular composite indices (Ray, 2008). It is the geometric mean of normalised indices for three dimensions (health, education, and income). The first dimension is based on life expectancy at birth. The second dimension is calculated as an average of expected school years and average school years. The third dimension, the GNI, represents the standard of living, (United Nations Development Programme, 2018). The HDI calculation can be found in the appendix.

The Better Life Index (BLI) uses the same procedure to aggregate 24 variables into ten dimensions. Bad outputs like air pollution are subtracted from one so that higher values are better. In the final step, all indices are averaged, using equal weights, to obtain the BLI for each country (Patrizii et al., 2017).

One of the main criticisms of the HDI is the lack of areas of societal interest for a reliable measurement of human development (Ray, 2008). In addition, the equal weighting implies a certain appreciation, is not objective and static. Each country is assumed to equally weight the included dimensions, neglecting individual preferences and policy targets (Greco et al., 2019). Pinar et al. (2017) find that the health dimension dominates the others in the HDI. Countries can more easily increase their HDI ranking by focusing on the health-related dimension than on the others. Paruolo et al. (2013) state that equal weights are an oversimplification and propose the use of more objective and flexible approaches to calculate human development indices. Despotis (2005) shows that the country rankings in the HDI depend on the weighting and Greco et al. (2019) propose the use of data-driven weight selection techniques such as DEA to avoid any subjective weight selection.

The HDI and the BLI are inflexible, and the results cannot be further decomposed to identify inefficiency sources. Overall, DEA models are regarded as superior for calculating appropriate weights, identifying inefficiencies, and assessing the strengths and weaknesses of individual countries (Cherchye et al., 2008).

6.2.6 Additional Weight Restrictions in DEA

Standard DEA models without additional weight restrictions are perceived as insufficient, as zero weights are not prevented and the models may lack discriminatory power (Khalili et al., 2010). Without additional weight restrictions, the linear program may assign zero weights to inputs and outputs of certain DMUs. Thus, certain variables may be excluded entirely in the DMU efficiency assessment. DEA models that lack discriminatory power may overstate efficiency and inefficient DMUs may be incorrectly calculated as efficient (Atici et al., 2015).

Moesen et al. (1998) and Cherchye (2001) calculate the macroeconomic efficiency of 20 countries using four macroeconomic quantities as outputs. Both studies interpret the weights as proxies for the true policy priorities (all variables are normalised). The weights are limited to equal or to be greater than 10% of the sum of all weights to prevent zero weights and unreasonable weights distributions. Benneyan et al. (2007) measure the efficiency of the health sector of 39 countries. 62% of the countries' weights of infant mortality (one of the outputs) equal zero and are thus ignored in the efficiency assessment. The authors describe the zero weights as "irrational weighting". They reduce the number of zero weights by including ARI restrictions based on their subjective preferences (Benneyan et al., 2007). In the efficiency analysis of Peiró-Palomino et al. (2018), the weights are restricted to equal at least a small non-Archimedean number.

As a result, all variables are taken into account in their human development assessment of 35 countries.

Weight restrictions enable the decision maker (DM) to better distinguish between efficient and inefficient DMUs (Carboni et al., 2015). Morais et al. (2011) calculate the efficiency of 206 European cities to maximise several socio-economic outputs. Without additional weight restrictions, the DEA lacks discriminatory power and all but 14 cities are efficient. Additional weight restrictions ensure that each weighted output must be at least 1.5% of the sum of all weighted outputs. These restrictions reveal how the respective dimension contributes to the efficiency score and do not depend on the measurement units.

Additional weight restrictions limit the available weights and thus alter the DEA results. It is, therefore, necessary to elaborate on the effects of the restrictions. Most efficiency studies which include additional weight restrictions do not provide comparisons between the subjective restrictions chosen by the DM and alternative thresholds. In addition, results with and without the additional restrictions are not sufficiently elaborated (Decancq et al., 2013).

6.3 Operationalisation and Variable Selection

The primary goal of any society is to increase the happiness of its citizens as much as possible (Ülengin et al., 2011). Subjective well-being is defined as the perception of citizens that their lives are desirable and proceeding well (Diener et al., 2015). Our first output is happiness and like Brulé et al. (2017), we operationalise it as the national average of subjective well-being. The subjective well-being values are reported by the Gallup World Poll (GWP).[4] Our second output, healthy life expectancy, takes into account how long citizens can enjoy their respective happiness. We have opted for healthy life expectancy in order to take the life span into account. The healthy life expectancy corrects the life expectancy at birth with the years lost due to poor health (World Health Organization, 2018). The correction for years lost relates our output closer to the ideal-typical healthcare output of the number of additional quality-adjusted life years gained by the healthcare system (Behr et al., 2017). In addition, our output healthy life expectancy encompasses the benefits of intermediate health-related

[4]It is the national average to the question: "All things considered, how satisfied are you with your life as a whole these days? Use a 0 to 10 scale, where 0 is dissatisfied and 10 is satisfied." (Gallup, 2017).

inputs such as the number of operations and is one of the most important factors for the operationalisation of human development (Prados de la Escosura, 2015).

Because we focus on the countries' abilities to provide their citizens a long and fulfilled life, we include quantities as inputs which have been regarded as outputs in other country efficiency studies. For instance, Lovell et al. (1995) use the GDP per capita as an output to calculate countries' economic performance. All empirical evidence suggests that happier people tend to live in countries with higher GDP. That links our first economic input to our output well-being (Welsch et al., 2016). Human development, however, depends not only on countries' are but on the benefits that citizens derive from this wealth (Veenhoven, 2015). Therefore, we perceive income as an input and not as an ultimate political target for societies. Oishi et al. (2015) find that increasing income inequality decreases overall happiness as the perception of wealth is relative. Hence, we include both the absolute wealth of nations (GDP) and inequality within societies (Gini) in our study (Mizobuchi, 2017).

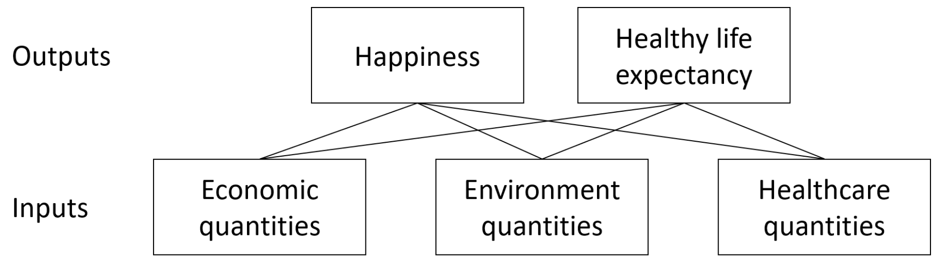


Figure 6.1 – Our operationalisation

Figure 6.1 shows how we operationalise the endowment of countries. In our framework, medical inputs are beneficial if they increase healthy life expectancy or happiness. Our medical-related inputs are the number of physicians and nurses and proxy access to healthcare and the resources employed in the health sector (Afonso et al., 2006). According to Cantor et al. (2017), these two inputs are considered in 73% and 51% of the relevant healthcare studies, respectively. The application areas of nurses and doctors vary between countries, and both groups are vital for functional health sectors. Therefore, the DEA is better suited to compare the countries efficiency than an approach that uses equal weights for all countries (Behr et al., 2017). Even after surgeries medical personal is still necessary to ensure the healing process (Buerhaus et al., 2016).

Environmental quantities should be considered in human development assessments for two reasons. Firstly, environmental factors are important for well-being and a healthy life (Diener et al., 2015). Secondly, environmental concerns have become increasingly important for societies in recent years (Alsahlawi, 2013). Environmental aspects have not yet been sufficiently addressed in human development efficiency studies. So far, eco-efficiency has been a largely independent domain in efficiency analysis (Sueyoshi et al., 2017). For instance, Suzuki et al. (2016) include energy consumption and population as inputs and CO_2 emissions and the GDP per capita as outputs in a DEA to provide policy-makers with an assessment of national energy-environmental efficiency. The environmental variables we include in the DEA are in line with most environmental efficiency studies, which focus on the relationship between environmental costs and the benefits of social or economic activities (Rashidi et al., 2015; Zhou et al., 2018). However, we do not include CO_2 emissions because they mainly represent environmental pollution that we consider undesirable (Moesen et al., 1998). As inputs we consider energy use and the population density to adequately capture the environmental costs of the human and economic developments. A higher population density of a country is most likely related to higher consumption of resources and a higher environmental impact (Guo et al., 2017). Energy consumption is a measure of human and technological development, as most infrastructure and technical devices require some form of energy (Suzuki et al., 2016).

Tables C.3 to C.6 in the Appendix provide an overview of the typical economic, environmental, health, and well-being performance literature. We regard the variables considered in the DEA as the best available proxies for human development inputs and outputs. The data are taken from the OECD, the WHO, the World Happiness Report, and the World Bank. Our outputs capture the countries abilities to provide their citizens with a long and fulfilled life:

- long: healthy life expectancy (y_1)
- happy: happiness (y_2).

y_2 is the national average subjective well-being ranging from zero to ten (highest well-being) and is published by the Gallup World Poll (Welsch et al., 2016). Our inputs are:

- economic: GNI per capita (x_1) and 1- the Gini coefficient (x_2)
- health: the number of physicians (x_3) and the number of nurses (x_4)
- ecological: energy use (x_5) and population density (x_6).

The GNI per capita in 1,000 \$ (x_1) represents the standard of living in purchasing power

(United Nations Development Programme, 2018). We include x_1 to proxy the different financial endowments of the countries. Our second economic input (x_2) represents the wealth equality within a society. Following Reig-Martínez (2013) and Carboni et al. (2015), x_2 is calculated as 1 minus the Gini coefficient based on disposable income, post taxes and, transfers.

Our health inputs consist of the number of physicians (x_3) and nurses and midwives (x_4) per 1,000 inhabitants. Both measures proxy access to healthcare and the resources employed in the health sector (Afonso et al., 2006).

The selection of our ecological inputs, the energy use (tons of oil equivalent per capita, x_5) and population density (people per sq. km of land area, x_6), represent the environmental costs of the human and economic developments (Suzuki et al., 2016; Guo et al., 2017).

The exact data sources are provided in Table C.7 in the appendix. All data are from 2015. Following Adang et al. (2007) and Behr et al. (2017) we use data from 2014 to replace missing values in 2015. Figure 6.2 depicts normalised inputs and outputs for each country. The normalisation allows straightforward comparability because the lowest value of each variable is zero and the highest value is one. Norway, followed by Switzerland, Denmark, and Finland have the highest average happiness scores. Japan, Spain, and Switzerland have the highest healthy life expectancy. Switzerland is the only country that enables its citizens a relatively long and happy life. Japan and Portugal demonstrate why it is important to take both outputs into account and to allow some specialisation in the efficiency evaluation of countries. Japanese have the longest health adjusted life expectancy but report relatively little happiness. Portuguese state the lowest subjective well-being but have a relatively long life expectancy. An equal weighting would not allow a trade-off between both outputs whereas DEA maximises their efficiency by selecting the most favourable weights. Russians have the lowest healthy life expectancy and report a below-average happiness. Countries with the lowest inputs are Mexico (GNI, Gini, and energy consumption), Turkey (lowest number of doctors and nurses), and Australia with the lowest population per square kilometre. The latter indicates a relatively low environmental impact of society. Norway (GNI), Slovenia (Gini), Greece (number of doctors), Switzerland (number of nurses), Canada (energy consumption) and the Netherlands (population density) have the highest inputs. Table 6.1 provides descriptive results and Table 6.2 the Pearson correlation coefficients of all variables. The positive correlation between inputs and outputs is a prerequisite

Table 6.1 – Descriptive results

	Outputs		Economics		Health		Environment	
	Happy	Life.exp	GNI	Gini	Doctors	Nurses	Energy	Population
Mean	6.524	70.182	37.158	0.681	3.453	9.336	3.724	131.235
Median	6.515	71.051	38.116	0.682	3.430	8.683	3.175	102.446
Min	5.081	64.083	17.074	0.541	1.749	2.617	1.537	3.100
Max	7.603	74.825	66.584	0.750	6.255	18.230	7.631	502.818

Table 6.2 – Correlation coefficients

	Outputs		Economics		Health		Environment	
	Happy	Life.exp	GNI	Gini	Doctors	Nurses	Energy	Population
Happy	1.00	0.52	0.75	0.31	0.05	0.72	0.58	0.09
Life Exp.	0.52	1.00	0.59	0.21	0.11	0.45	0.28	0.39
GNI	0.75	0.59	1.00	0.39	0.10	0.81	0.63	0.19
Gini	0.31	0.21	0.39	1.00	0.24	0.56	0.27	0.11
Doctors	0.05	0.11	0.10	0.24	1.00	0.10	-0.09	-0.09
Nurses	0.72	0.45	0.81	0.56	0.10	1.00	0.50	0.04
Energy	0.58	0.28	0.63	0.27	-0.09	0.50	1.00	-0.17
Pop-density	0.09	0.39	0.19	0.11	-0.09	0.04	-0.17	1.00

of the DEA. All variables are strictly positive.

6.4 Methodology

We use the output-oriented BCC-model, introduced by Banker et al. (1984). The output-orientation implies that countries maximise their output, given their input levels. For country$_o$ (o denotes a specific country under consideration) the model is defined as:

$$
\begin{aligned}
\underset{\eta}{min} \quad & \eta = \sum_i v_i x_{io} - u_0 \\
\text{subject to } & \sum_r u_r y_{ro} = 1 \\
& \sum_i v_i x_{ij} - \sum_r u_r y_{rj} - u_0 \geq 0 \quad (j = 1, \ldots, n) \\
& v_i \geq 0 \quad (i = 1, \ldots, m) \\
& u_r \geq 0 \quad (r = 1, \ldots, s) \\
& u_0 \text{ free in sign.}
\end{aligned}
\tag{34}
$$

y_r is output r and its weight is given by u_r $(r = 1, \ldots, s)$. Input i is given by x_i and is weighted by v_i $(i = 1, \ldots, m)$. s is the number of outputs, m the number of inputs, and n the number of countries. η^* ($[1, \infty]$) denotes the solution to the minimisation problem. For convenience, we define $\theta = \frac{1}{\eta}$ ($[0, 1]$). A country is efficient, only if $\eta^* = \theta^* = 1$, otherwise it is inefficient. Efficient countries are part of the reference set, span the efficiency frontier, and serve as benchmarks for the inefficient countries. The scalar u_0 is free in sign and implements the assumption of variable returns to scale (Fadeyi et al., 2019).

v_i and u_r can be interpreted as shadow prices, virtual prices, variable multipliers or weights (Iribarren et al., 2013). The linear problem is solved for each country and the prices are assigned to the inputs and outputs so that the efficiency of country o is maximised. The prices obtained from the linear model for country o are donated by v_{io} and u_{ro}. The restrictions in model (34) do not prevent zero weights. Zero weights can be avoided by implementing additional weight restrictions which also increase the discriminatory power of the model. Khalili et al. (2010) provide an overview of different weight restriction types. For example, assurance regions of type I (ARI) additively or relatively link the input or output weights to each other.

6.5 Results

Table 6.3 provides the efficiency scores, the weights, and the number of zero weights calculated with the BCC-model from equation (34). 16 out of the 33 countries are efficient and the mean efficiency is 0.990. The Slovak Republic (with an efficiency score of 0.945) and Lithuania (0.959) are the most inefficient countries. The program calculates zero weights at least for one input or one output in each country. The weights of healthy life expectancy are zero in 14 countries and for happiness in 9 countries. Both outputs are only included in the efficiency assessment of a few countries such as Belgium and Germany. Due to a higher number of inputs, a larger number of zero weights is calculated for four of the inputs. The weights of the equality input are set to zero for 31 countries to maximise their efficiency. Only one input and one output are included in the efficiency calculation of Greece, Italy, Norway, and Portugal. For example, Greece is assessed as efficient because only healthy life expectancy and the number of nurses are taken into account. Thus, the Greek efficiency score does not reflect how satisfied its citizens are with their lives and what economic or environmental resources it possesses. Due to the high number of zero weights and the lack of discriminatory power, we include additional weight restrictions in the next step.

Table 6.3 – Efficiency scores, weights, and number of zero weights of the BCC model

	Efficiency	Happy	Life.exp	GNI	Gini	Doctors	Nurses	Energy	Population	Zeros
Australia	1.000	0.137	0.000	0.000	0.000	0.000	0.050	0.061	0.013	4
Austria	0.990	0.078	0.006	0.000	0.000	0.000	0.013	0.005	0.000	3
Belgium	0.980	0.027	0.011	0.000	0.000	0.011	0.001	0.001	0.000	3
Canada	1.000	0.135	0.000	0.000	0.000	0.393	0.000	0.000	0.001	5
Denmark	1.000	0.133	0.000	0.000	0.000	0.000	0.000	0.071	0.001	5
Estonia	0.980	0.000	0.015	0.006	0.000	0.000	0.012	0.001	0.003	3
Finland	1.000	0.134	0.000	0.000	0.000	0.009	0.000	0.000	0.000	4
France	0.983	0.000	0.014	0.000	0.000	0.017	0.000	0.004	0.000	4
Germany	0.978	0.019	0.012	0.000	0.000	0.000	0.000	0.000	0.000	4
Greece	1.000	0.000	0.014	0.000	0.000	0.000	0.020	0.000	0.000	6
Hungary	0.947	0.000	0.015	0.005	0.000	0.001	0.000	0.000	0.000	5
Ireland	1.000	0.043	0.010	0.000	0.000	0.019	0.002	0.023	0.000	2
Israel	1.000	0.141	0.000	0.000	0.000	0.000	0.019	0.007	0.000	4
Italy	0.994	0.000	0.014	0.000	0.000	0.000	0.000	0.075	0.000	6
Japan	1.000	0.000	0.013	0.000	0.000	0.055	0.000	0.008	0.000	4
Latvia	1.000	0.000	0.015	0.000	0.000	0.000	0.046	0.000	0.005	5
Lithuania	0.959	0.000	0.015	0.000	0.000	0.000	0.000	0.060	0.002	5
Luxembourg	0.988	0.023	0.012	0.000	0.000	0.006	0.001	0.000	0.000	4
Mexico	1.000	0.160	0.000	0.000	0.000	0.000	0.122	0.000	0.011	5
Netherlands	0.992	0.137	0.000	0.000	0.000	0.000	0.005	0.010	0.000	5
New Zealand	1.000	0.135	0.000	0.000	0.000	0.045	0.036	0.065	0.010	3
Norway	1.000	0.132	0.000	0.000	0.000	0.000	0.000	0.000	0.002	6
Poland	0.960	0.067	0.009	0.000	0.000	0.090	0.000	0.014	0.000	4
Portugal	0.992	0.000	0.014	0.000	0.000	0.000	0.000	0.077	0.000	6
Russia	1.000	0.000	0.016	0.012	0.000	0.102	0.000	0.038	0.013	3
Slovak Republic	0.945	0.023	0.013	0.005	0.000	0.000	0.000	0.000	0.000	4
Slovenia	0.991	0.000	0.014	0.003	0.000	0.020	0.000	0.000	0.000	4
Spain	1.000	0.000	0.014	0.000	0.000	0.001	0.016	0.000	0.001	4
Sweden	0.992	0.038	0.010	0.000	0.000	0.000	0.002	0.000	0.000	3
Switzerland	1.000	0.048	0.009	0.000	0.129	0.000	0.000	0.031	0.000	3
Turkey	1.000	0.000	0.015	0.000	0.000	0.342	0.021	0.000	0.003	4
United Kingdom	0.990	0.051	0.009	0.000	0.000	0.048	0.001	0.018	0.000	3
United States	1.000	0.000	0.014	0.000	0.417	0.025	0.000	0.000	0.000	4
Zeros	-	14	9	26	31	17	17	14	9	-

Following Morais et al. (2011), our AR1 restrict the weighted outputs and inputs:

$$\frac{u_r y_{ro}}{\sum_{r=1}^{s} u_r y_{ro}} \geq \alpha_{out}, \qquad \frac{v_i x_{io}}{\sum_{i=1}^{m} v_i x_{io}} \geq \alpha_{in} \tag{35}$$

α_{out} and α_{in} are the lower thresholds for the weighted outputs and inputs, respectively. If the outputs and inputs are equally weighted, α_{out} and α_{in} equal the reciprocal values of the respective numbers of variables that is $\alpha_{out} = \frac{1}{2} = 50\%$ and $\alpha_{in} = \frac{1}{8} = 12.5\%$. These lower boundaries would prevent countries from specialising and would be a simple composite index with equal weights. In the standard BCC DEA model in equation (34), α_{out} and α_{in} are zero and countries can fully specialise so that only one output and one input are included in their assessment. Between these two extremes, any thresholds can be imposed. For example, averaging the extreme values leads to $\alpha_{out} = 25\%$ and $\alpha_{in} = 6.25\%$. However, the boundaries of the lower weights must be carefully chosen in order not to render the linear program unsolvable and to minimise their impact on the efficiency scores (Khalili et al., 2010). We consider $\alpha_{out} = 25\%$ and $\alpha_{in} = 6.25\%$ to be too strict and therefore use half of them in the following analysis: $\alpha_{out} = 12.5\%$ and $\alpha_{in} = 3.125\%$. We donate the restricted model as BCC AR1 and its efficiency scores, and the shares of weighted inputs and outputs in percent are presented in Table 6.4. Table C.2 in the appendix contains the input and output weights of the BCC AR1 to allow a direct comparison with Table 6.3.

Table 6.4 – Efficiency scores, shares of weighted inputs and outputs in percent, and number of zero weights of the BCC AR1 ($\alpha_{out} = 12.5\%$ and $\alpha_{in} = 3.125\%$)

	Efficiency	Happy	Life.exp	GNI	Gini	Doctors	Nurses	Energy	Population	Zeros
Australia	1.000	87.500	12.500	3.125	3.125	3.125	3.125	65.649	21.851	0
Austria	0.986	52.738	47.262	3.125	3.125	3.125	67.841	11.375	11.409	0
Belgium	0.980	16.528	83.472	17.344	3.125	56.904	16.377	3.125	3.125	0
Canada	1.000	87.500	12.500	3.125	3.125	3.125	3.125	3.125	84.375	0
Denmark	1.000	87.500	12.500	3.125	3.125	3.125	46.045	36.648	7.932	0
Estonia	0.960	12.500	87.500	42.095	3.125	3.125	21.591	3.125	26.939	0
Finland	0.999	87.500	12.500	33.117	3.125	54.383	3.125	3.125	3.125	0
France	0.978	12.500	87.500	3.125	3.125	40.157	20.839	22.925	9.829	0
Germany	0.977	13.389	86.611	80.037	3.125	3.125	3.125	3.125	7.463	0
Greece	0.998	12.500	87.500	3.125	3.125	3.125	83.838	3.125	3.662	0
Hungary	0.930	12.500	87.500	84.375	3.125	3.125	3.125	3.125	3.125	0
Ireland	1.000	22.148	77.852	3.125	3.125	34.013	3.125	53.487	3.125	0
Israel	1.000	87.500	12.500	3.125	3.125	3.125	52.603	3.125	34.897	0
Italy	0.979	12.500	87.500	3.125	3.125	3.125	3.125	84.375	3.125	0
Japan	1.000	12.500	87.500	3.125	3.125	56.413	3.125	12.799	21.413	0
Latvia	1.000	12.500	87.500	3.125	3.125	3.125	35.869	15.596	39.160	0
Lithuania	0.939	12.500	87.500	3.125	3.125	3.125	3.125	46.755	40.745	0
Luxembourg	0.988	15.294	84.706	3.125	3.125	55.974	31.526	3.125	3.125	0
Mexico	1.000	12.500	87.500	3.125	3.125	3.125	3.125	36.846	50.654	0
Netherlands	0.988	40.735	59.265	3.125	3.125	3.125	83.453	4.047	3.125	0
New Zealand	1.000	87.500	12.500	3.125	3.125	3.125	3.125	66.521	20.979	0
Norway	1.000	87.500	12.500	3.125	3.125	3.125	3.125	3.125	84.375	0
Poland	0.959	12.500	87.500	26.056	3.125	58.370	3.125	3.125	6.199	0
Portugal	0.960	12.500	87.500	84.375	3.125	3.125	3.125	3.125	3.125	0
Russia	1.000	12.500	87.500	80.021	3.125	3.125	3.125	3.125	7.479	0
Slovak Republic	0.943	16.690	83.310	84.146	3.125	3.125	3.125	3.125	3.354	0
Slovenia	0.971	12.500	87.500	34.977	3.125	36.069	3.125	19.578	3.125	0
Spain	1.000	12.500	87.500	3.125	3.125	3.125	70.576	3.125	16.924	0
Sweden	0.991	29.346	70.654	3.125	3.125	3.125	84.375	3.125	3.125	0
Switzerland	1.000	36.905	63.095	3.125	31.271	3.125	3.125	37.658	21.697	0
Turkey	1.000	12.500	87.500	3.125	3.125	84.375	3.125	3.125	3.125	0
United Kingdom	0.989	28.164	71.836	3.125	3.125	59.729	3.185	27.711	3.125	0
United States	0.987	12.500	87.500	3.125	66.416	21.084	3.125	3.125	3.125	0
Zeros	-	0	0	0	0	0	0	0	0	-

The average efficiency is only slightly lower in the BCC AR1 (0.985) compared to the BCC-model without additional restrictions (0.990). Finland and Greece are no longer efficient because the model can no longer maximise their efficiency by excluding certain inputs or outputs in their efficiency calculation. Countries with rather equal weights in the BCC are only marginally influenced by the imposed AR1. For example, the Netherlands is inefficient in the BCC-model and its efficiency score remains nearly unchanged by the additional restrictions. Countries whose efficiency scores heavily depend on specialisation obtain lower scores in the BCC AR1. Portugal (0.032) and Slovenia (0.020) are the countries with the highest reductions in efficiency scores. Although the additional restrictions lower the average efficiency and increase the discriminatory power of the model, the efficiency scores of the BCC-model and the BCC AR1 model are highly correlated ($\rho = 0.921$).

The BCC AR1 characterises 14 countries as efficient. These countries are Australia, Canada, Denmark, Ireland, Israel, Japan, Latvia, Mexico, New Zealand, Norway, Russia, Spain, Switzerland, and Turkey. Figure 6.2(b) shows that Russia has the lowest GNI and the lowest well-being values. Therefore, Russia would be disadvantaged in an analysis with equal weights where countries are not assessed individually. In our analysis, Russia is efficient because the DEA assigns relatively high weight to the GNI and the highest allowed weight to the healthy life expectancy. Interestingly, Switzerland and the United States are the only countries with high Gini weights. Japanese have the highest healthy life expectancy and therefore, Japan is efficient although its inputs are not especially low. The opposite is true for Mexico. Mexico is efficient although its outputs are relatively low because Mexico uses its rather low inputs (GNI, Gini, and the energy consumption) relatively efficient. Hungary has an efficient score of 0.930, has the greatest potential for efficiency improvement, and potentially could increase its outputs by 7.5% while keeping its inputs constant.

6.5.1 Robustness Checks

In the following, we calculate results for different restriction thresholds to validate the robustness of our country ranking. For comparison, we choose lower boundaries where all variables are equally weighted ($\alpha_{out} = 50\%$ and $\alpha_{in} = 12.5\%$), half of the thresholds ($\alpha_{out} = 25\%$ and $\alpha_{in} = 6.25\%$), half of the boundaries we use in the BCC AR1 ($\alpha_{out} = 6.25\%$ and $\alpha_{in} = 1.56\%$), and the BCC model which has no additional lower boundaries ($\alpha_{out} = 0$ and $\alpha_{in} = 0$). In the left panel of Figure 6.3, the countries are ordered differently according to their efficiency scores for each threshold. The model with equal weights provides the lowest efficiency scores and thus forms the bottom line. Decreasing levels of α increase the overall efficiency by increasing the allowed weight flexibility of the linear program. In Figure 6.3(b) the countries are sorted throughout different thresholds based on the ranking obtained with the boundaries of the BCC AR1. On both panels of Figure 6.3, the thresholds are equally altered. In the right panel, the country rankings are unchanged. Overall, the rankings are quite similar for increasing values of α.

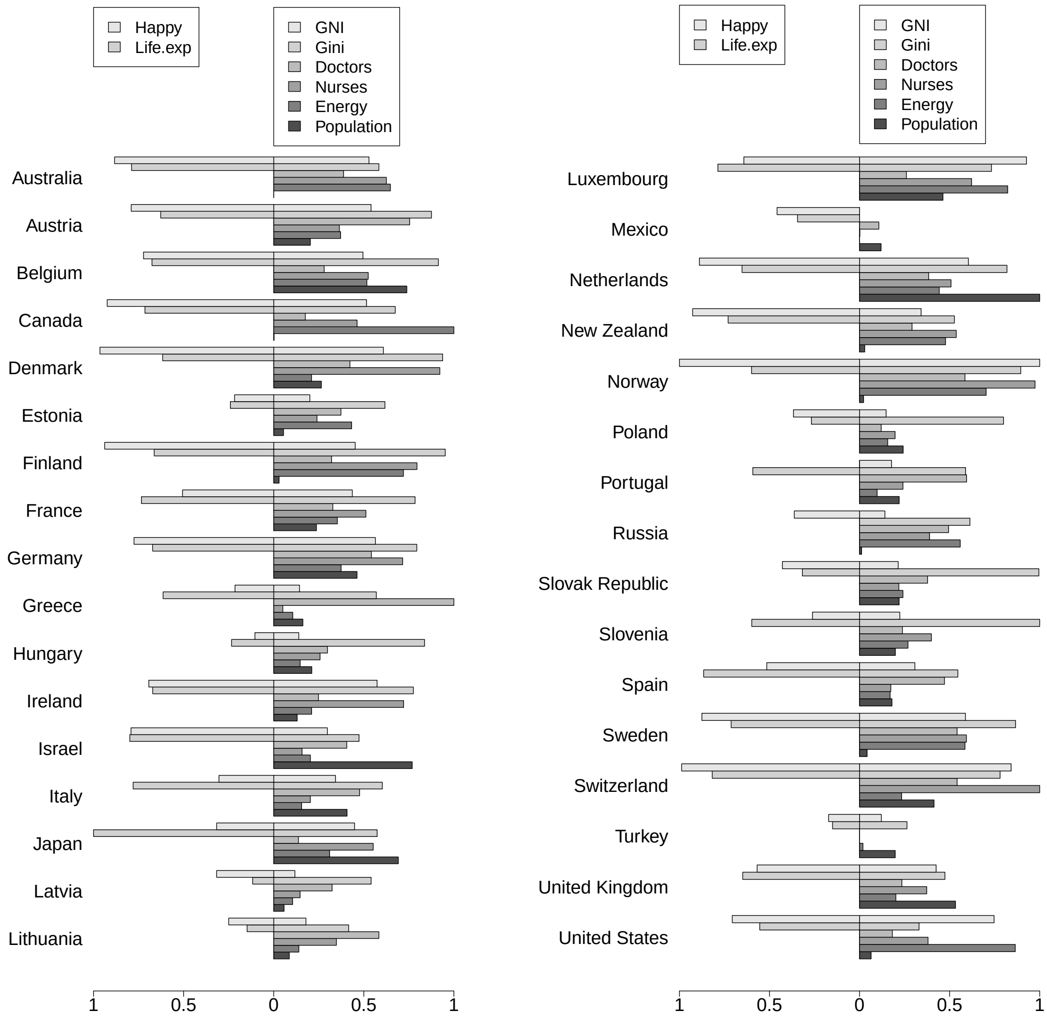
Table 6.5 contains the efficiency scores, the average efficiency, and the Pearson correlation coefficients with the BCC AR1 results for the different thresholds. The results in the second column are obtained with equal output and equal input weights. Ten countries are efficient, and Hungary (0.857) is the most inefficient country. If the lower boundaries are halved (third column of the table), three more countries are efficient (Denmark, Ireland, and Spain). The efficiencies of Japan (0.096) and Portugal (0.071) increase mostly compared to the model without weight flexibility. In the BCC AR1, the lower boundaries are again halved (fourth column). The resulting efficiency scores are highly correlated among different models, and the largest deviations result from the model with rigid weights (second column). The last column includes the results of the BCC model without additional weight restrictions. Zero weights are present in all variables and in all countries, and the efficiency of all countries is maximised in the standard model.

If the results of the BCC are compared with those of the model with fixed weights, the efficiency values of Portugal (0.134) and Slovenia (0.108) benefit the most. This is because Portugal has the lowest happiness score and Slovenia's only relatively low input is the Gini. If equal weights are used, Japan (0.902) is one of the most inefficient countries. Japan is almost efficient if the linear program can select the weights more flexibly (see the second column) and efficient for lower values of α. The opposite is true for the Netherlands, as its outputs and inputs are all relatively high compared to the other countries. Thus, the efficiency values of the Netherlands are only slightly increased if the thresholds are lowered. The BCC efficiency score of the Netherlands is only 0.011 higher than it is for equal weights. This is the lowest observed increase of all countries that are inefficient in all assessments.

6.6 Conclusion

Our analysis focuses on the ability of countries to make the lives of their citizens as long, healthy, and happy as possible, given their endowments. Our two outputs, self-reported well-being and healthy life expectancy, cover the most important aspects of human development. We include economic, environmental, and health-related inputs in the DEA. DEA models allow to account for the countries' individual strength and weaknesses. However, a DEA without additional restrictions often lacks discriminatory power, and frequently zero weights are observed for all inputs and outputs.

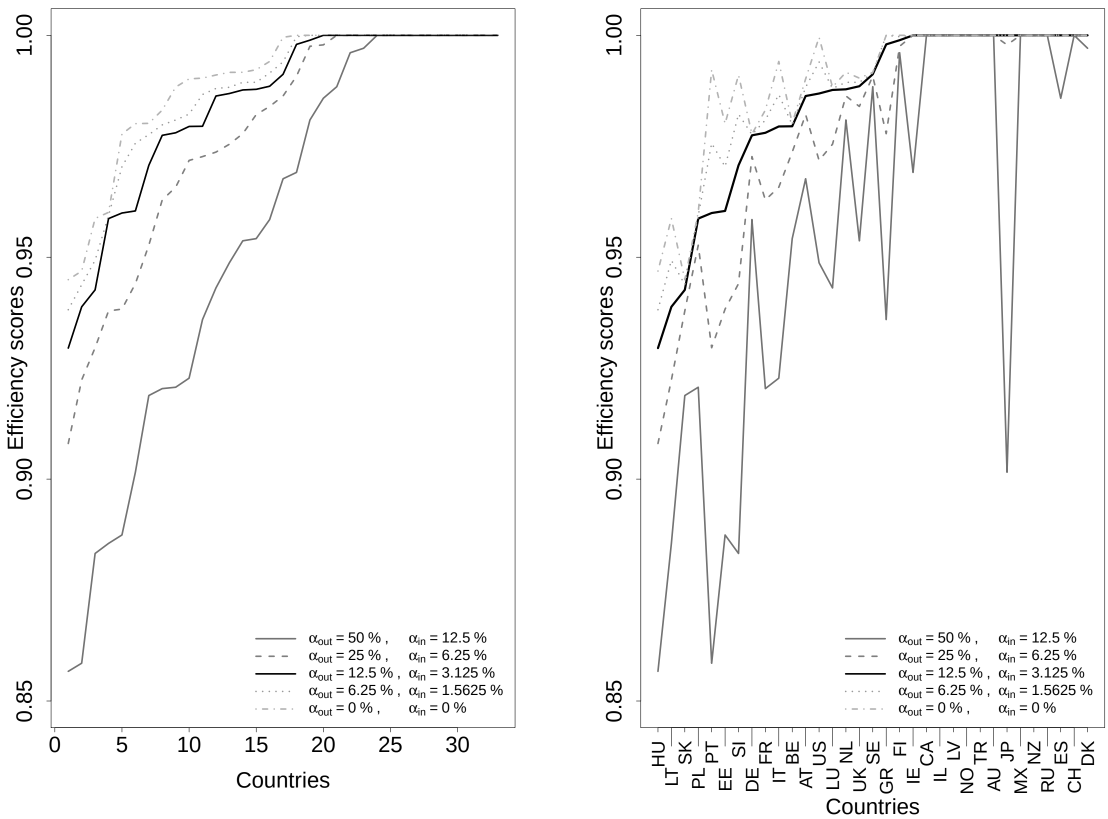
In our application using the BCC model, zero weights are present for for each country and each variable (e.g. the Gini is only considered in the efficiency evaluation in two countries). In the evaluation of Greece, Italy, Norway, and Portugal only one input and one output are included in order to maximise their efficiency. Therefore, we restrict the weighted outputs and the weighted inputs so that they are greater than or equal to a certain threshold value. Compared to the DEA without additional restrictions, our selected weight boundaries render two more countries inefficient and slightly reduce the average efficiency from 0.990 to 0.985. We identify Australia, Canada, Denmark, Ireland, Israel, Japan, Latvia, Mexico, New Zealand, Norway, Russia, Spain, Switzerland, and Turkey as efficient. We find that high efficiency scores are obtained through rather heterogeneous combinations. E.g. Japan, Switzerland and Norway, perform well due to their high outputs. In contrast, Mexico and Russia are efficient because they attain their low outputs with relatively low inputs. These different approaches to reach efficiency demonstrate how DEA accounts for the individual characteristics of the countries. Overall, DEA may be regarded better suited to calculate the efficiency of countries in maximising human development than a composite index with equal weights. Furthermore, our analysis shows that additional constraints can improve the discriminatory power of the efficiency analysis and ensure that all countries are assessed on the basis of all variables. Although any weight restriction potentially changes the obtained results dramatically, we find the efficiency results in our analysis and the rankings of the countries to remain relatively stable.



(a) Normalised outputs (left) and inputs (right)

(b) Normalised outputs (left) and inputs (right)

Figure 6.2 – Normalised data



(a) Sorted from lowest to highest scores for each threshold; sorting of countries can be different for each threshold

(b) Sorted according to the country ranking from $\alpha_{out} = 12.5\%$ and $\alpha_{in} = 3.125\%$ (thick line); constant sorting of countries

Figure 6.3 – Efficiency scores sorted according to country rankings for different thresholds

Table 6.5 – Efficiency scores for different thresholds

α_{in} α_{out}	50% 12.5%	25% 6.25%	12.5% 3.125%	6.25% 1.563%	0% 0%
Australia	1.000	1.000	1.000	1.000	1.000
Austria	0.968	0.982	0.986	0.988	0.990
Belgium	0.954	0.974	0.980	0.980	0.980
Canada	1.000	1.000	1.000	1.000	1.000
Denmark	0.997	1.000	1.000	1.000	1.000
Estonia	0.887	0.938	0.960	0.970	0.980
Finland	0.996	0.998	0.999	1.000	1.000
France	0.920	0.963	0.978	0.981	0.983
Germany	0.958	0.973	0.977	0.978	0.978
Greece	0.936	0.978	0.998	1.000	1.000
Hungary	0.857	0.908	0.930	0.938	0.947
Ireland	0.969	1.000	1.000	1.000	1.000
Israel	1.000	1.000	1.000	1.000	1.000
Italy	0.923	0.966	0.979	0.987	0.994
Japan	0.902	0.998	1.000	1.000	1.000
Latvia	1.000	1.000	1.000	1.000	1.000
Lithuania	0.886	0.922	0.939	0.949	0.959
Luxembourg	0.943	0.975	0.988	0.988	0.988
Mexico	1.000	1.000	1.000	1.000	1.000
Netherlands	0.981	0.986	0.988	0.989	0.992
New Zealand	1.000	1.000	1.000	1.000	1.000
Norway	1.000	1.000	1.000	1.000	1.000
Poland	0.921	0.953	0.959	0.959	0.960
Portugal	0.859	0.930	0.960	0.976	0.992
Russia	1.000	1.000	1.000	1.000	1.000
Slovak Republic	0.919	0.938	0.943	0.944	0.945
Slovenia	0.883	0.944	0.971	0.982	0.991
Spain	0.986	1.000	1.000	1.000	1.000
Sweden	0.988	0.991	0.991	0.992	0.992
Switzerland	1.000	1.000	1.000	1.000	1.000
Turkey	1.000	1.000	1.000	1.000	1.000
United Kingdom	0.954	0.984	0.989	0.989	0.990
United States	0.949	0.972	0.987	0.994	1.000
Mean	0.956	0.978	0.985	0.987	0.990
Correlation	0.823	0.967	1.000	0.984	0.921

C.1 Supplementary Results

Table C.1 – Country names and abbreviations

Country	Abbreviation	Country	Abbreviation
Australia	AU	Luxembourg	LU
Austria	AT	Mexico	MX
Belgium	BE	Netherlands	NL
Canada	CA	New Zealand	NZ
Denmark	DK	Norway	NO
Estonia	EE	Poland	PL
Finland	FI	Portugal	PT
France	FR	Russia	RU
Germany	DE	Slovak Republic	SK
Greece	GR	Slovenia	SI
Hungary	HU	Spain	ES
Ireland	IE	Sweden	SE
Israel	IL	Switzerland	CH
Italy	IT	Turkey	TR
Japan	JP	United Kingdom	UK
Latvia	LV	United States	US
Lithuania	LT		

Table C.3 – Overview of macroeconomic performance literature

Authors	Inputs	Outputs	Method	Sample
Färe et al., 1994	Capital stock and employment rate	GDP	Malmquist indices and distance functions	17 OECD countries, 1979 - 1988
Brockett et al., 1999	Capital stock and employment rate	GDP	BCC-model and ranking indices	17 OECD countries, 1979 - 1988
Lovell et al., 1995	One	OECD magic diamond	Additive DEA, BCC-model, and weight restrictions	19 OECD countries, 1970-1990
Moesen et al., 1998	One	OECD magic diamond	Linear program with prior defined weights categories	19 OECD countries, 1987-1996
Cherchye, 2001	One	OECD magic diamond	Additive DEA with lower weights bounds	20 OECD countries, 1992-1996

Table C.3 Continued: Overview of macroeconomic performance literature

Authors	Inputs	Outputs	Method	Sample
Pavone et al., 2015	One	OECD magic diamond, CO_2, and the HDI	BCC-model with bad outputs	60 OECD countries, 2008-2011
Mohamad, 2007	Government spending	OECD magic diamond	BCC-model	22 OECD countries, in 1996, 2000, and 2003
Mohamad et al., 2011	Government spending	OECD magic diamond	BCC-model	54 countries, 2003-2007
Staníková et al., 2012	Expenditure on research and development, employment rate, gross fixed capital formation, number of students	GDP, labour productivity	BCC-model	EU 27 member states, 2000-2010
Hsu et al., 2008	Government efficiency index, business efficiency index, infrastructure advancements index	Economic performance index	BCC-model	60 countries, 2004
Chattopadhyay et al., 2015	-	OECD magic diamond split in six dimensions	Linear program	48 countries, 2000-2012

SBM: slack based measurement, CCR-model: DEA model with constant returns to scale introduced by Charnes et al. (1978)

Table C.4 – Overview of health-related efficiency literature

Authors	Inputs	Outputs	Method	Sample
Puig-Junoy, 1998	Physicians, non physician personnel, hospital beds, tobacco and alcohol consumption	Variation of life expectancy at birth	BCC-model, weights restrictions	All OECD countries (if available), 1960s, 1970s, and 1980s.

Table C.4 Continued: Overview of health-related efficiency literature

Authors	Inputs	Outputs	Method	Sample
Retzlaff-Roberts et al., 2004	Practising physicians, in-patient beds, magnetic resonance imagers (MRIs), health expenditure	Infant mortality and life expectancy	BCC-model	27 OECD countries, 2000
Afonso et al., 2005	Doctors, nurses, in-patient beds	Infant mortality and life expectancy	BCC- and FHD-models, input- and output-oriented	24 OECD countries, 2000
Bhat, 2005	Practising physicians, nurses, inpatient beds, pharmaceuticals	Population aged 0-19years; population aged 20-64 years; population aged 65 years and older	CCR-model	24 OECD countries, 2002/2003
Spinks et al., 2005	School expectancy years, unemployment rates, and total health expenditure	Life expectancy	DEA based Malmquist indexes	28 OECD countries, 1995 and 2000
Afonso et al., 2006	Doctors, nurses, in-patient beds, high-tech diagnostic medical equipment	Infant mortality, life expectancy, years of life not lost	PCA, BCC-model	24 OECD countries, 2000
Adang et al., 2007	Health expenditure, physicians, and tobacco use	Life expectancy at birth, infant mortality	CCR-model, Malmquist index	15 OECD countries, 1995-2002
Benneyan et al., 2007	Health expenditure, number of doctors and nurses, hospital beds, immunisation rate, median age	Healthy life expectancy, adult mortality rate, infant mortality, morbidity surrogate measure, an equity index, and the incidence rate of medical misadventure	CCR-model, BCC-model, weights restrictions	39 countries

Table C.4 Continued: Overview of health-related efficiency literature

Authors	Inputs	Outputs	Method	Sample
Asandului et al., 2014	Doctors, hospital beds, and public health expenditures	life expectancy at birth, health adjusted life expectancy, and infant mortality rate	CCR-model, BCC-model	30 European countries in 2010
Behr et al., 2017	Multiple inputs in several partial analysis	Multiple outputs in several partial analysis	BCC-model, weight restrictions	34 OECD countries, 2012

SBM: slack based measurement, CCR-model: DEA model with constant returns to scale introduced by Charnes et al. (1978)

Table C.5 – Overview of environmental efficiency literature

Authors	Inputs	Outputs	Method	Sample
Arcelus et al., 2005	Capital stock, labour force	GDP, CO_2 emissions	CCR-models with and without bad outputs	14 countries, 1970-1991
Chien et al., 2007	Capital stock, labour force, energy consumption	GDP	CCR-model	45 countries, 2001-2002
Hu et al., 2007	Capital stock, labour force, energy consumption	GDP	CCR-model, slack corrected	17 countries, 1991-2000
Gomes et al., 2008	CO_2 emissions	GDP, labour force, energy consumption	CCR-model, input-oriented	64 countries, 2001
Alsahlawi, 2013	Capital stock, labour force, energy consumption	GDP	CCR-model	6 countries, 2001-2008
Camarero et al., 2013	CO_2 emissions, NO_X emissions, and SO_X emissions	GDP	Additive slack based model	22 countries, 1980 and 2008

Table C.5 Continued: Overview of environmental efficiency literature

Authors	Inputs	Outputs	Method	Sample
Simsek, 2014	Capital stock, labour force, several different energy consumptions	GDP, CO_2 emissions	SBM with bad outputs	23 countries, 1995-2009
Vlontzos et al., 2014	Capital stock, labour force, energy consumption	GDP, CO_2 emissions, gross nutrient balance	Non-radial DEA	25 countries, 2001-2008
Rashidi et al., 2015	Labour force, energy consumption, precipitation average	GDP, CO_2 emissions	Non-radial additive model	19 countries, 2012
Suzuki et al., 2016	Energy consumption, population	GDP, CO_2 emissions	Super-efficient models	27countries, 2003-2012
Tsai et al., 2016	Labour force, energy consumption, government Expenditures	GDP, CO_2 emissions	SBMs	73 countries, 2006-2010
Guo et al., 2017	Land area, population, energy use	GDP, CO_2 emissions	SBMs	27 countries, 2000-2010

SBM: slack based measurement, CCR-model: DEA model with constant returns to scale introduced by Charnes et al. (1978)

Table C.6 – Overview of human development efficiency literature

Authors	Inputs	Outputs	Method	Sample
Cravioto et al., 2011	Electricity consumption, CO_2 emissions	GDP, HDI	CCR-model	40 countries, 2007
Morais et al., 2011	-	GDP, HDI	Modified CCR-model	206 European cities, 2007
Ülengin et al., 2011	Basic requirements, efficiency enhancers, innovation and sophistication indices	Life expectancy at birth, schools gross enrolment ratios, and GDP	Super efficiency CCR-model	45 countries, 2006/2007

Table C.6 Continued: Overview of human development efficiency literature

Authors	Inputs	Outputs	Method	Sample
Reig-Martínez, 2013	-	GDP, Life expectancy, gross enrolment rates, government effectiveness, environmental effectiveness, Gini coefficient, gender gap index	Modified CCR-model	42 countries, mainly 2008
Morais et al., 2013	-	15 socio, economic, and environmental outputs	Modified CCR-model	246 cities, aggregated from 2003, 2006, and 2009
Chansarn, 2014	Electricity consumption, CO_2 emissions, energy use	GDP, life expectancy, years of schooling, expected years of schooling	SBM	115 countries, 2008
Debnath et al., 2014	GDP, three political indicators	Average well-being, well-being dispersion	CCR-model, BCC-model	113 countries
Guardiola et al., 2014	10 socio-economic quantities	Subjective well-being	CCR-model, common weights	177 people, 2008
Mizobuchi, 2014	Produced, natural capital, and intangible capital	BLI variables	CCR-model	34 countries, 2011
Carboni et al., 2015	8 economic, inequality, ecological inputs	4 socio-economic outputs	BCC-model, Malmquist indices	20 Italian regions, 2005-2012
Cordero et al., 2017a	Individual income, health, and education status	Subjective well-being	Conditional efficiency models	31854 individuals from 26 countries, 2005/2006
Patrizii et al., 2017	Hours worked, consumption capital	BLI variables	SBMs	35 countries 2012/2013
Mizobuchi, 2017	10 BLI indices	SWB	Modified CCR-model	36 countries, 2014

Table C.6 Continued: Overview of human development efficiency literature

Authors	Inputs	Outputs	Method	Sample
Peiró-Palomino et al., 2018	-	BLI variables	Modified CCR-model	38 countries, 2013-2016
DiMaria et al., 2019	GDP, employment, capital stock	Life satisfaction	BCC-model, Malmquist indices	20 countries, 2004-2010

SBM: slack based measurement, CCR-model: DEA model with constant returns to scale introduced by Charnes et al. (1978)

The Human Development Index calculation

The HDI is the geometric mean of normalised indices for three dimensions. The normalised indices (I) are calculated for each variable (j) and each country (k) (United Nations Development Programme, 2018):

$$\mathrm{I}_{j,k} = \frac{x_{j,k} - min(x_j)}{max(x_j) - min(x_j)} \quad j = \text{Health, Education, Income} \tag{36}$$

where $min(x_j)$ and $max(x_j)$ are the minimum and maximum values of the respective variable over all countries (Ray, 2008). Equation (36) is applied separately to the expected school years and the average school years. The education dimension results from the arithmetic mean of the two variables. Finally, the HDI is calculated as the geometric mean of the three dimensions: $\mathrm{HDI}_k = (\mathrm{I}_{\text{Health},k} \cdot \mathrm{I}_{\text{Education},k} \cdot \mathrm{I}_{\text{Income},k})^{\frac{1}{3}}$ (United Nations Development Programme, 2018).

Table C.2 – Efficiency scores, input and output weights, and number of zero weights of the BCC AR1 ($\alpha_{out} = 12.5\%$ and $\alpha_{in} = 3.125\%$)

	Efficiency	Happy	Life.exp	GNI	Gini	Doctors	Nurses	Energy	Population	Zeros
Australia	1.0000	0.1197	0.0017	0.0000	0.0017	0.0003	0.0001	0.0044	0.0026	0
Austria	0.9863	0.0745	0.0067	0.0001	0.0067	0.0009	0.0127	0.0047	0.0002	0
Belgium	0.9795	0.0239	0.0117	0.0001	0.0015	0.0064	0.0005	0.0002	0.0000	0
Canada	1.0000	0.1180	0.0017	0.0000	0.0004	0.0001	0.0000	0.0000	0.0019	0
Denmark	1.0000	0.1164	0.0018	0.0002	0.0102	0.0021	0.0065	0.0314	0.0001	0
Estonia	0.9604	0.0222	0.0131	0.0070	0.0210	0.0041	0.0153	0.0034	0.0040	0
Finland	0.9989	0.1175	0.0018	0.0003	0.0016	0.0063	0.0001	0.0002	0.0001	0
France	0.9780	0.0197	0.0122	0.0000	0.0018	0.0049	0.0008	0.0025	0.0000	0
Germany	0.9775	0.0190	0.0121	0.0005	0.0011	0.0002	0.0001	0.0002	0.0000	0
Greece	0.9980	0.0222	0.0124	0.0001	0.0048	0.0005	0.0248	0.0014	0.0000	0
Hungary	0.9295	0.0234	0.0131	0.0046	0.0057	0.0013	0.0006	0.0017	0.0000	0
Ireland	1.0000	0.0324	0.0109	0.0001	0.0069	0.0183	0.0003	0.0293	0.0001	0
Israel	1.0000	0.1236	0.0017	0.0002	0.0093	0.0017	0.0198	0.0021	0.0002	0
Italy	0.9795	0.0214	0.0121	0.0002	0.0082	0.0014	0.0009	0.0596	0.0000	0
Japan	1.0000	0.0213	0.0117	0.0002	0.0108	0.0547	0.0006	0.0086	0.0001	0
Latvia	1.0000	0.0213	0.0134	0.0004	0.0132	0.0027	0.0201	0.0197	0.0034	0
Lithuania	0.9388	0.0219	0.0133	0.0004	0.0171	0.0024	0.0013	0.0672	0.0030	0
Luxembourg	0.9877	0.0228	0.0117	0.0000	0.0014	0.0058	0.0008	0.0001	0.0000	0
Mexico	1.0000	0.0200	0.0129	0.0018	0.0578	0.0140	0.0118	0.2397	0.0081	0
Netherlands	0.9878	0.0556	0.0083	0.0000	0.0012	0.0002	0.0022	0.0003	0.0000	0
New Zealand	1.0000	0.1180	0.0017	0.0009	0.0480	0.0102	0.0028	0.1497	0.0120	0
Norway	1.0000	0.1151	0.0018	0.0000	0.0013	0.0002	0.0001	0.0002	0.0018	0
Poland	0.9587	0.0208	0.0131	0.0018	0.0076	0.0439	0.0009	0.0022	0.0001	0
Portugal	0.9600	0.0246	0.0124	0.0043	0.0063	0.0009	0.0007	0.0019	0.0000	0
Russia	1.0000	0.0208	0.0137	0.0113	0.0159	0.0027	0.0012	0.0021	0.0029	0
Slovak Republic	0.9426	0.0271	0.0123	0.0043	0.0059	0.0013	0.0007	0.0015	0.0000	0
Slovenia	0.9707	0.0218	0.0124	0.0019	0.0062	0.0191	0.0005	0.0092	0.0000	0
Spain	1.0000	0.0196	0.0119	0.0000	0.0005	0.0001	0.0014	0.0001	0.0000	0
Sweden	0.9912	0.0403	0.0098	0.0000	0.0009	0.0001	0.0014	0.0001	0.0000	0
Switzerland	1.0000	0.0487	0.0087	0.0001	0.1034	0.0017	0.0004	0.0296	0.0002	0
Turkey	1.0000	0.0227	0.0133	0.0014	0.0524	0.4824	0.0119	0.0189	0.0003	0
United Kingdom	0.9885	0.0432	0.0101	0.0001	0.0088	0.0385	0.0007	0.0181	0.0000	0
United States	0.9869	0.0182	0.0125	0.0002	0.3960	0.0299	0.0013	0.0017	0.0003	0
Zeros	-	0	0	0	0	0	0	0	0	-

Table C.7 – Data sources

Variable	Data	Source
Happiness	Subjective well-being	World Happiness Report 2016
Healthy life expectancy	Healthy life expectancy (HALE) at birth (years) (Mortality and global health estimates)	WHO
GNI	GNI per capita, PPP (constant 2011 international $)	World Bank
Gini	Gini (disposable income, post taxes and transfers)	OECD
Number of nurses	Nurses and midwives (per 1,000 people)	World Bank
Number of doctors	Physicians (per 1,000 people)	World Bank
Energy use	Energy use (kg of oil equivalent per capita)	World Bank
Population density	Population density (people per sq. km of land area)	World Bank

7

Summary and General Conclusion

The first part of this thesis is more technically focused, and the second part, which draws on uses the insights of the first, consists of three empirical studies. These studies are efficiency assessments of secondary and tertiary educational institutions as well as a comprehensive country evaluation.

In addition to the technical aspects of efficiency assessment with DEA, I also focus on operationalisation and methodology decisions in this thesis. Operationalisation decisions can directly influence the results of efficiency assessments. The decision maker must identify and select relevant inputs and outputs before a DEA model can be calculated. With an increasing number of variables relative to the number of DMUs, the efficiency results are less discriminatory and the DEA model may calculate unreasonably high weights or zero weights (Moreno et al., 2018). Ideal-typical inputs and outputs should reflect all available resources and all produced results for the production processes but ideal-typical inputs and outputs are rarely available. Behr et al. (2017), for example, assessed the efficiency of health care systems and noted that an ideal-typical output would be the number of additional quality-adjusted life years provided by the health care systems. However, the authors had to use proxy variables due to the unavailability of the ideal-typical output. In many other efficiency studies, variable selection is only superficially discussed, or variable transformations are missing. For instance, Lehmann et al. (2018), who evaluated the efficiency of HEIs in Germany and Italy, justified their choice of variables mainly with reference to earlier literature without a more detailed analysis of their own variable selection process.

Data transformation is another important aspect of operationalisation, since most efficiency models can only be calculated if all data are positive. Furthermore, an important prerequisite for DEA is that an increase in inputs should result in an increase in outputs. Therefore, the original data frequently have to be transformed, but these changes are often not mentioned in the respective efficiency studies. As an example, Veiderpass et al. (2016), who assessed the efficiency of HEIs, and Aparicio et al. (2017b), who measured the efficiency of schools, provided incomplete descriptive results of their data or no description at all. Agasisti et al. (2018) and Wohlrabe et al. (2019b), who calculated the efficiency of schools and HEIs respectively, reported descriptive results that indicate zeros or negative data that would render their DEA models infeasible. Their respective data transformations are not provided. Without this information, the results of all cited studies are of limited comprehensibility, and are difficult to reproduce by other researchers.

Once the variables have been selected, the appropriate DEA model must be calculated, and all results should be reported. When evaluating the results, zero weights or inappropriate weight distributions should be considered. A multitude of efficiency studies, such as Johnes (2006) and Kounetas et al. (2011), have focused only on the efficiency scores and do not report the weights. Yang et al. (2019) noted that efficiency models without weight restrictions may overestimate efficiency. It is therefore ambiguous whether the efficiency models estimate the production processes intended by the decision maker, or whether they unintentionally calculate a high number of specialised DMUs if the weights are not assessed. The foregoing considerations show that operationalisation is an important part of efficiency evaluations and, thus, an essential part of the empirical studies in this thesis. As a basis for the empirical studies, the first two studies address the implementation of additional weight restrictions in radial DEA models and examine differences between radial and non-radial DEA models.

The first study highlighted the necessity for additional restrictions in radial DEA models, compared absolute and relative restrictions, and demonstrated their implementation. Without additional weight restrictions, DEA models may overestimate efficiency and incorrectly calculate DMUs as specialised. Such specialised DMUs are only assessed on subsets of the data due to zero weights. Furthermore, the marginal rates of substitution and transformation cannot be defined if zero weights are calculated. In my first paper, I used five artificial DMUs to demonstrate how additional weight restrictions limit the available weighting space and impact the efficiency scores. Absolute weight restrictions (AWR) constrain weights to vary within set boundaries, but AWRs do not necessarily yield meaningful marginal transformation and substitution rates. Therefore, AWRs are rarely used in the literature (Olesen et al., 1996). Assurance regions (AR) are more flexible than AWRs as they relatively link inputs and outputs and thus directly reflect marginal rates of substitution or transformation (Atici et al., 2015).

Additional weight restrictions are either data-driven or based on expert opinions. Relative market prices belong to the former category and offer marginal substitution or conversion rates, but are they not available for most efficiency applications (Joro et al., 2004). Alternatively, data-driven restrictions can be derived from additional calculations such as regressions, correlations, or previous DEAs. Expert statements can be an alternative source of additional restrictions. Opinions can be collected by pair-wise comparison techniques or questionnaires. The judgements of experts may differ and must be aggregated (Castelo Gouveia et al., 2016). Once the restrictions have been identified and implemented, the results should be validated for alternative weight restriction thresholds. In my first paper, I demonstrated that restriction thresholds directly influence the efficiency results, and specialised DMUs tend to become more inefficient with stricter restrictions than DMUs with more balanced input-output combinations. My results also show that additional weight restrictions may enable the identification of specialised DMUs and prevent zero weights, thereby allowing a more realistic representation of production processes. This has the added benefit of increasing discrimination between efficient and inefficient DMUs. Additional weight restrictions were already used in early applications, (e.g., by Thompson et al. (1986)), and since then, this approach has been continuously improved. Additional weight restrictions directly influence the efficiency results and limit the available production possibility set. The latter contradicts the intention of the DEA to assess DMUs strictly according to their inputs and outputs. But even with these disadvantages, additional weight restrictions can improve efficiency assessments to such an extent that they are worth considering.

My second paper focused on non-radial DEA models, mainly the SBM-Min and the SBM-Max, and compared them with radial models. The latter ignore input excess and output shortfalls and do not allow substitution among inputs and outputs. Radial models tend to overestimate efficiency. Efficiency measures of the two non-radial models decrease monotonically with each input and output slack.

A simulation, in which the efficiencies of 1000 DMUs are calculated 1000 times with radial and non-radial models, enables a large-scale model comparison. The main results confirm that these models tend to calculate similar efficiency scores overall, which are positively correlated but interpret inefficiency differently. The SBM-Max provides an upper efficiency score bound while the SBM-Min provides a lower efficiency score bound, and the results of the radial models are located in between. The average computing time for the SBM-Max is about 70 times longer than for the other models because several linear programs have to be calculated. The assumptions of non-radial DEA models differ substantially from radial models, and the models calculate inefficiencies differently. Non-radial DEA models allow substitution between inputs and outputs, and thus can provide additional information on the production processes of the DMUs even if no input excess and output shortfalls are present. So far, the SBM-Max has only been used for small sample applications, such as the work of Johnes et al. (2017), who measured the efficiency of 118 HEIs. The second paper showed how the SBM-Max maximises the calculated efficiency and illustrated that the calculation takes considerably longer. The calculation effort could be reduced by minimising the number of DMUs to be included in some of the linear calculations within the SBM-Max. The implementation of such a preselecting algorithm remains a task for further research.

In the second part of this thesis, the previous results were implemented in three empirical applications. In the third paper, my co-author and I applied a radial DEA model with variable scale returns to assess secondary education systems. This approach allowed us to present a straightforward interpretation of our results and a further decomposition of the efficiency scores. The fourth study evaluated the efficiency of HEIs, for which non-radial DEA models were selected based on the results of my second study. The last study was a comprehensive performance assessment of OECD countries using additional weight restrictions, as proposed in my first paper. Operationalisation decisions are circumstantiated and compared with those in the relevant literature.

In the third paper, my co-author and I used DEA to assess the performance of secondary education systems to maximise the academic performance of students. The students' ESCS values are our input, and their average PISA scores are our output. The efficiency of each student is calculated in relation to national and international efficiency frontiers, and the results of native and immigrant students are compared. Comparisons within countries show that in Denmark, Finland, and Italy, natives perform better than immigrants on average. In Austria, Belgium, Germany, the Netherlands, Spain, and Sweden, natives also perform better, but the differences are not that large. In contrast, immigrants in Australia, Canada, Israel, Singapore, and the United States of America perform better on average than their native peers. Compared to the international frontier, which consists of all students, students perform best on average in Spain, Finland, and Denmark, and worst in Israel, Australia, and the United States of America.

Our main results are obtained by comparing the average PISA scores of immigrants with the average efficiency scores for each country. Students' average efficiency scores indicate the performance of education systems in maximising students' PISA scores given their socio-economic endowments. Compared only on their absolute PISA scores, immigrants in Spain perform relatively worse, but on average they perform very well in terms of efficiency. In both analyses, immigrants in Singapore perform best. Our study of secondary educational systems using PISA scores shows that performance gaps between natives and immigrants persist in most countries, even if we account for the students' socio-economic backgrounds. Borgna et al. (2014) found that the socio-economic endowments of students are one of their most important performance determinants, and Hwang et al. (2018) noted that the ESCS covers the most important socio-economic factors. Our results indicate that countries with relatively selective immigrant policies perform well not only in absolute PISA scores, but are also quite efficient given their ESCS input levels.

The good performance of countries with relatively selective immigrant policies implies that the selection process not only influences the level of ESCS students in these countries but also that educational systems enable immigrants to use their resources efficiently. Furthermore, we found that Spanish immigrants tend to have low levels of ESCS, but the Spanish educational system maximises their PISA scores relatively well. Our results may enable other educational systems to improve by comparing themselves with the Spanish educational system and adopting its successful approach to the integration of immigrants.

In the field of tertiary education, data from the CWTS Leiden Ranking and the ETER datasets was used to assess the efficiency of 46 German and 45 UK HEIs. The outputs were selected to cover the different departmental priorities and to reflect the objectives of HEIs in research, teaching, and innovation. The inputs were the personnel employed and expenditures. Based on the insights of my second paper, I used non-radial DEA models and calculated the efficiency relative to country-specific and an international efficiency frontier. In addition, I applied super-efficient non-radial DEA models to compare the input-output structures of HEIs between countries.

One of the main findings of the higher education assessment is that UK HEIs, when measured by a common international efficiency frontier, are on average more efficient than their German counterparts. The better performance of UK HEIs is in line with the findings of Wolszczak-Derlacz (2017), who conducted an international efficiency analysis. Further results show that the average efficiency within the two countries is quite similar when using country-specific frontiers. The super-efficient models indicate country-specific input-output structures by identifying almost all HEIs as super-efficient compared to the efficiency frontier of the other country. Such unique inputoutput structures could remain unnoticed if HEIs are measured only within countries or compared to a common frontier. The different input-output structures can result from different departmental structures within the educational sectors or the different reimbursement structures of HEIs in the UK and Germany.

The impact of incentive systems in higher education is the subject of ongoing discussions and, like the departmental structure of HEIs, should be included in future efficiency research as soon as more recent data and findings are available (Blecich, 2020). In the Excellence Initiative of German Universities (EIGU), eleven top-performing HEIs in Germany were selected and supported with increased funding (Fischer et al., 2017). My results identified RWTH Aachen University, Dresden University of Technology, and the University of Cologne as inefficient, even though they were selected as HEIs of excellence under EIGU. This finding indicates that these three HEIs achieved their excellence status due to their relatively high capital and personnel input, which demonstrates that DEA offers an alternative approach to purely output-focused comparisons.

In the fifth paper, my co-author and I assessed the performance of OECD countries to provide their citizens with long and fulfilling lives, given their economic, environmental, and health endowments. Such country comparisons are commonly made with composite indices that ignore the countries' endowments and use equal weights for all variables (Mizobuchi, 2017). Of the 33 countries studied, Australia, Canada, Denmark, Ireland, Israel, Japan, Latvia, Mexico, New Zealand, Norway, Russia, Spain, Switzerland, and Turkey are efficient. Our efficiency assessment accounts for country-specific strengths and weaknesses. Mexico and Russia, for example, are efficient, even though they provide their citizens with relatively low outputs. However, these low outputs are achieved with relatively low inputs. In contrast, Japan, Switzerland, and Norway, which have relatively high inputs, provide relatively high outputs.

We used a radial DEA model with additional weight restrictions as proposed by my first paper. The additional weight restrictions are based on a previous DEA model, which prevents zero weights and increases the discrimination between efficient and inefficient countries. Our lower weight thresholds were selected in such a way that their influence on the efficiency scores is as minimal as possible, and the linear program remains feasible. The additional weight restrictions ensure that each country is assessed based on all variables. Additional robustness tests for alternative thresholds indicated that the country ranking based on the efficiency scores is quite stable.

In contrast to the efficiency assessment, composite indices, which are commonly used for country comparisons such as the HDI, use static weights that neglect individual preferences and policy objectives (Ray, 2008). Furthermore, our approach accounts for the country's resources and outputs simultaneously and follows the suggestion of Greco et al. (2019) to apply data-driven weights. Future research questions might focus on other measures such as human development to calculate alternative country or sector comparisons. Additional or alternative – preferably ideal-typical – inputs and outputs could also provide further insights.

Over the last 40 years, DEA models have been continuously improved, extended, and applied in a large number of studies to evaluate the efficiency of DMUs in a wide field of applications (Emrouznejad et al., 2018). This thesis complements the ever-growing DEA literature by discussing the technical aspects of some of the most important model enhancements and examining their empirical applications. Apart from the technical aspects, the studies also emphasised the operationalisation of efficiency evaluations in the context of secondary and tertiary education and a comprehensive country analysis. In the field of secondary education efficiency evaluation, we found large differences in educational system efficiency, when controlling for negative selection effects caused by immigration regimes. We identified the Spanish educational system as one of the best and found that countries with relatively selective immigration policies perform well in absolute PISA scores and also achieve high efficiency scores. With regard to higher education efficiency assessments, the results showed that UK HEIs on average perform better than their German counterparts. Non-radial super-efficient models provided further details and indicated country-specific input-output structures that should be considered in further international evaluations. Our comprehensive assessment of OECD countries showed quite heterogeneous input-output combinations. We therefore included additional weight restrictions, which increases the discriminatory power of the analysis and ensures that all countries are assessed on the basis of all variables. Our analysis ultimately identified 14 out of 33 OECD countries as efficient.

References

Ababneh, Emad, Ahmad Al-Tweissi and Khattab Abulibdeh (2016). „TIMSS and PISA impact the case of Jordan“. In: *Research Papers in Education* 31.5, pp. 542–555.

Abdullah, Dahlan, Muhammad Zarlis, Darmawan Napitupulu, Hartono Hartono, S Sriadhi, Cut Ita Erliana, Rohman Dijaya, Yulian Findawati, Heri Nurdiyanto, Robbi Rahim et al. (2018). „A Slack-Based Measures within Group Common Benchmarking using DEA for Improving the Efficiency Performance of Departments in Universitas Malikussaleh“. In: *MATEC Web of Conferences*. Vol. 197. EDP Sciences, p. 16005.

Abramo, Giovanni, Tindaro Cicero and Ciriaco Andrea DAngelo (2011). „A field-standardized application of DEA to national-scale research assessment of universities“. In: *Journal of Informetrics* 5.4, pp. 618–628.

Adang, Eddy M. M. and George F. Borm (2007). „Is there an association between economic performance and public satisfaction in health care?“ In: *The European Journal of Health Economics* 8.3, pp. 279–285.

Aduand, Derick Taylor and Elisha Kwaku Denkyirah (2017). „Education and economic growth: a co-integration approach“. In: *International Journal of Education Economics and Development* 8.4, pp. 228–249.

Afonso, António and Miguel St Aubyn (2005). „Non-parametric approaches to education and health efficiency in OECD countries“. In: *Journal of Applied Economics* 8.2, pp. 227–246.

Afonso, António and Miguel St Aubyn (2006). „Relative efficiency of health provision: A DEA approach with non-discretionary inputs“. In: *ISEG-UTL Economics Working Paper* 33.

Agasisti, Tommaso and Carla Haelermans (2016). „Comparing Efficiency of Public Universities among European Countries: Different Incentives Lead to Different Performances“. In: *Higher Education Quarterly* 70.1, pp. 81–104.

Agasisti, Tommaso and Carsten Pohl (2012). „Comparing German and Italian Public Universities: Convergence or Divergence in the Higher Education Landscape?“ In: *Managerial and Decision Economics* 33.2, pp. 71–85.

Agasisti, Tommaso and Pablo Zoido (2018). „Comparing the Efficiency of Schools Through International Benchmarking: Results From an Empirical Analysis of OECD PISA 2012 Data“. In: *Educational Researcher* 47.6, pp. 352–362.

Ali, Agha Iqbal and Lawrence M. Seiford (1990). „Translation Invariance in Data Envelopment Analysis“. In: *Operations Research Letters* 9, pp. 403–405.

Allen, Robert, Antreas Athanassopoulos, Robert G Dyson and Emmanuel Thanassoulis (1997). „Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions“. In: *Annals of Operations Research* 73.0, pp. 13–34.

Alsahlawi, Mohammed A (2013). „Measuring energy efficiency in GCC countries using data envelopment analysis“. In: *The Journal of Business Inquiry* 12.1, pp. 15–30.

Amado, Carla Alexandra da Encarnação Filipe and Sérgio Pereira dos Santos (2009). „Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres“. In: *Health Policy* 91.1, pp. 43–56.

Ammermueller, Andreas (2007). „Poor Background or Low Returns? Why Immigrant Students in Germany Perform so Poorly in the Programme for International Student Assessment“. In: *Education Economics* 15.2, pp. 215–230.

Andreu, Laura, José Luis Sarto and Luis Vicente (2014). „Efficiency of the strategic style of pension funds: an application of the variants of the slacks-based measure in DEA“. In: *Journal of the Operational Research Society* 65.12, pp. 1886–1895.

Angermuller, Johannes (2017). „Academic careers and the valuation of academics. A discursive perspective on status categories and academic salaries in France as compared to the US, Germany and Great Britain“. In: *Higher Education* 73.6, pp. 963–980.

Angulo-Meza, Lidia and Marcos Pereira Estellita Lins (2002). „Review of methods for increasing discrimination in data envelopment analysis“. In: *Annals of Operations Research* 116.1-4, pp. 225–242.

Aparicio, Juan, Jose M. Cordero, Martin Gonzalez and Jose J. Lopez-Espin (2017a). „Using non-radial DEA to assess school efficiency in a cross-country perspective: An empirical analysis of OECD countries“. In: *Omega*.

Aparicio, Juan, Eva Crespo-Cebada, Francisco Pedraja-Chaparro and Daniel Santín (2017b). „Comparing school ownership performance using a pseudo-panel database: A Malmquist-type index approach“. In: *European Journal of Operational Research* 256.2, pp. 533–542.

Arcelus, Francisco J. and Pablo Arocena (2005). „Productivity differences across OECD countries in the presence of environmental constraints“. In: *Journal of the Operational Research Society* 56.12, pp. 1352–1362.

Arikan, Serkan, Fons JR Van de Vijver and Kutlay Yagmur (2017). „PISA mathematics and reading performance differences of mainstream European and Turkish immigrant students“. In: *Educational Assessment, Evaluation and Accountability* 29.3, pp. 229–246.

Asandului, Laura, Monica Roman and Puiu Fatulescu (2014). „The efficiency of healthcare systems in Europe: A data envelopment analysis approach“. In: *Procedia Economics and Finance* 10, pp. 261–268.

Athanassopoulos, Antreas D and Estelle Shale (1997). „Assessing the comparative efficiency of higher education institutions in the UK by the means of data envelopment analysis“. In: *Education Economics* 5.2, pp. 117–134.

Atici, Kazim and Victor Podinovski (2015). „Using data envelopment analysis for the assessment of technical efficiency of units with different specialisations: An application to agriculture“. In: *Omega* 54, pp. 72–83.

Auranen, Otto and Mika Nieminen (2010). „University research funding and publication performanceAn international comparison“. In: *Research Policy* 39.6, pp. 822–834.

Avkiran, Necmi K (2001). „Investigating technical and scale efficiencies of Australian universities through data envelopment analysis“. In: *Socio-economic Planning Sciences* 35.1, pp. 57–80.

Avkiran, Necmi Kemal, Kaoru Tone and Miki Tsutsui (2008). „Bridging radial and non-radial measures of efficiency in DEA“. In: *Annals of Operations Research* 164.1, pp. 127–138.

Aziz, Nur Azlina Abd, Roziah Mohd Janor and Rasidah Mahadi (2013). „Comparative departmental efficiency analysis within a university: a DEA approach“. In: *Procedia-Social and Behavioral Sciences* 90, pp. 540–548.

Bangi, Yustin (2014). „Efficiency assessment of tanzanian private universities: Data envelopment analysis (DEA)“. In: *International Journal of Education and Research* 2.5, pp. 455–472.

Banker, Rajiv, Abraham Charnes and William W. Cooper (1984). „Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis“. In: *Management Science* 30.9, pp. 1078–1092.

Barros, Carlos Pestana, Shunsuke Managi and Roman Matousek (2012). „The technical efficiency of the Japanese banks: non-radial directional performance measurement with undesirable output“. In: *Omega* 40.1, pp. 1–8.

Bakaya, Sait and Matthias Klumpp (2014). „International data envelopment analysis in higher education: how do institutional factors influence university efficiency“. In: *Journal of Business and Economics* 5.11, pp. 2085–2090.

Basso, Antonella, Francesco Casarin and Stefania Funari (2018). „How well is the museum performing? A joint use of DEA and BSC to measure the performance of museums“. In: *Omega* 81, pp. 67–84.

Bayraktar, Erkan, Ekrem Tatoglu and Selim Zaim (2013). „Measuring the relative efficiency of quality management practices in Turkish public and private universities". In: *Journal of the Operational Research Society* 64.12, pp. 1810–1830.

Becker, Sascha (2012). *Brain drain and brain gain: The global competition to attract high-skilled migrants*. Oxford University Press.

Behr, Andreas (2015). *Production and Efficiency Analysis with R*. Springer.

Behr, Andreas and Katja Theune (2017). „Health System Efficiency: A Fragmented Picture Based on OECD Data". In: *PharmacoEconomics - Open* 1.3, pp. 203–221.

Beine, Michel, Anna Boucher, Brian Burgoon, Mary Crock, Justin Gest, Michael Hiscox, Patrick McGovern, Hillel Rapoport, Joep Schaper and Eiko Thielemann (2016). „Comparing immigration policies: An overview from the IMPALA database". In: *International Migration Review* 50.4, pp. 827–863.

Benneyan, James, Mehmet Erkan Ceyhan and Aysun Sunnetci (2007). „Data envelopment analysis of national healthcare systems and their relative efficiencies". In: *The 37th International Conference on Computers and Industrial Engineering*, pp. 251–261.

Bertoli, Simone, Vianney Dequiedt and Yves Zenou (2016). „Can selective immigration policies reduce migrants' quality?" In: *Journal of Development Economics* 119, pp. 100–109.

Bhat, Vasanthakumar N. (2005). „Institutional arrangements and efficiency of health care delivery systems". In: *The European Journal of Health Economics* 6.3, pp. 215–222.

Bjerre, Liv, Marc Helbling, Friederike Römer and Malisa Zobel (2015). „Conceptualizing and measuring immigration policies: A comparative perspective". In: *International Migration Review* 49.3, pp. 555–600.

Blecich, Andrea Arbula (2020). „Factors affecting relative efficiency of higher education institutions of economic orientation". In: *Management: Journal of Contemporary Management Issues* 25.1, pp. 45–67.

Bonjour, Saskia and Sébastien Chauvin (2018). „Social class, migration policy and migrant strategies: an introduction". In: *International Migration* 56.4, pp. 5–18.

Borgna, Camilla and Dalit Contini (2014). „Migrant achievement penalties in Western Europe: Do educational systems matter?" In: *European Sociological Review* 30.5, pp. 670–683.

Bornmann, Lutz and Rüdiger Mutz (2015). „Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references". In:

Journal of the Association for Information Science and Technology 66.11, pp. 2215–2222.

Bremberger, Christoph, Francisca Bremberger, Mikulas Luptacik and Stephan Schmitt (2015). „Regulatory impact of environmental standards on the eco-efficiency of firms“. In: *Journal of the Operational Research Society* 66.3, pp. 421–433.

Brockett, Patrick L, Boaz Golany and Shan Li (1999). „Analysis of intertemporal efficiency trends using rank statistics with an application evaluating the macro economic performance of OECD nations“. In: *Journal of Productivity Analysis* 11.2, pp. 169–182.

Brulé, Gaël and Ruut Veenhoven (2017). „The ‘10 Excess’ Phenomenon in Responses to Survey Questions on Happiness“. In: *Social Indicators Research* 131.2, pp. 853–870.

Buerhaus, Peter I, David I Auerbach and Douglas O Staiger (2016). „Recent changes in the number of nurses graduating from undergraduate and graduate programs“. In: *Nursing Economics* 34.1, p. 46.

Camanho, A.S. and R.G. Dyson (2005). „Cost efficiency measurement with price uncertainty: a DEA application to bank branch assessments“. In: *European Journal of Operational Research* 161.2. Financial Modelling, pp. 432–446.

Camarero, Mariam, Juana Castillo, Andrés J. Picazo-Tadeo and Cecilio Tamarit (2013). „Eco-Efficiency and Convergence in OECD Countries“. In: *Environmental and Resource Economics* 55.1, pp. 87–106.

Camarota, Steven A and Karen Zeigler (2016). „Immigrants in the United States: A profile of the foreign-born using 2014 and 2015 Census Bureau data“. In: *Center for Immigration Studies.*

Camehl, Georg F, Pia S Schober and C Katharina Spiess (2018). „Information asymmetries between parents and educators in German childcare institutions“. In: *Education Economics* 26.6, pp. 624–646.

Cantor, Victor John M. and Kim Leng Poh (2017). „Integrated Analysis of Healthcare Efficiency: A Systematic Review“. In: *Journal of Medical Systems* 42.1, p. 8.

Carboni, Oliviero A. and Paolo Russu (2015). „Assessing Regional Wellbeing in Italy: An Application of Malmquist–DEA and Self-organizing Map Neural Clustering“. In: *Social Indicators Research* 122.3, pp. 677–700.

Castelo Gouveia, Maria do, Luis C. Dias, Carlos Henggeler Antunes, Maria Augusta Mota, Eduardo M. Duarte and Eduardo M. Tenreiro (2016). „An application of value-based DEA to identify the best practices in primary health care“. In: *OR Spectrum* 38.3, pp. 743–767.

Chang, Young-Tae, Hyo-soo Park, Jae-beom Jeong and Jin-woo Lee (2014). „Evaluating economic and environmental efficiency of global airlines: A SBM-DEA approach“. In: *Transportation Research Part D: Transport and Environment* 27, pp. 46–50.

Chansarn, Supachet (2014). „The Evaluation of the Sustainable Human Development: A Cross-country Analysis Employing Slack-based DEA“. In: *Procedia Environmental Sciences* 20. The 4th International Conference on Sustainable Future for Human Security SUSTAIN 2013, pp. 3–11.

Charnes, Abraham, William W. Cooper, Boaz Golany, Larry Seiford and J Stutz (1985). „Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions“. In: *Journal of Econometrics* 30.1-2, pp. 91–107.

Charnes, Abraham, William W. Cooper, Zhimin Huang and Dee Sun (1990). „Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks“. In: *Journal of Econometrics* 46.1-2, pp. 73–91.

Charnes, Abraham, William W. Cooper and Edwardo Rhodes (1978). „Measuring the efficiency of decision making units“. In: *European Journal of Operational Research* 2.6, pp. 429–444.

Chattopadhyay, Somnath and Suchismita Bose (2015). „Global macroeconomic performance: a comparative study based on composite scores“. In: *Journal of Reviews on Global Economics* 4, pp. 51–68.

Chen, Yao (2005). „Measuring super-efficiency in DEA in the presence of infeasibility“. In: *European Journal of Operational Research* 161.2. Financial Modelling, pp. 545–551.

Cherchye, Laurens (2001). „Using data envelopment analysis to assess macroeconomic policy performance“. In: *Applied Economics* 33, pp. 407–416.

Cherchye, Laurens, Willem Moesen, Nicky Rogge, Tom Van Puyenbroeck, M Saisana, A Saltelli, R Liska and S Tarantola (2008). „Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index“. In: *Journal of the Operational Research Society* 59.2, pp. 239–251.

Chien, Taichen and Jin-Li Hu (2007). „Renewable energy and macroeconomic efficiency of OECD and non-OECD economies“. In: *Energy Policy* 35.7, pp. 3606–3615.

Chuanyi, Wang, Lv Xiaohong and Zhao Shikui (2016). „The Relative Efficiencies of Research Universities of Science and Technology in China: Based on the Data Envelopment Analysis and Stochastic Frontier Analysis.“ In: *Eurasia Journal of Mathematics, Science & Technology Education* 12.10.

Coelli, Tim (1998). „A multi-stage methodology for the solution of orientated DEA models“. In: *Operations Research Letters* 23.3-5, pp. 143–149.

Cook, Dennis R (1979). „Influential observations in linear regression“. In: *Journal of the American Statistical Association* 74.365, pp. 169–174.

Cook, Wade D., Juan Du and Joe Zhu (2017). „Units invariant DEA when weight restrictions are present: ecological performance of US electricity industry“. In: *Annals of Operations Research* 255.1, pp. 323–346.

Cook, Wade D., Liang Liang, Y Zha and Joe Zhu (2009a). „A modified super-efficiency DEA model for infeasibility“. In: *Journal of the Operational Research Society* 60.2, pp. 276–281.

Cook, Wade, Alex Kazakov, Yaakov Roll and Lawrence Seiford (1991). „A data envelopment approach to measuring efficiency: Case analysis of highway maintenance patrols“. In: *The Journal of Socio-Economics* 20.1, pp. 83–103.

Cook, Wade, Moshe Kress and Lawrence Seiford (1992). „Prioritization models for frontier decision making units in DEA“. In: *European Journal of Operational Research* 59.2, pp. 319–323.

Cook, Wade and Larry Seiford (2009b). „Data envelopment analysis (DEA) Thirty years on“. In: *European Journal of Operational Research* 192.1, pp. 1–17.

Cooper, William W., José Ruiz and Inmaculada Sirvent (2009). „Selecting non-zero weights to evaluate effectiveness of basketball players with DEA“. In: *European Journal of Operational Research* 195.2, pp. 563–574.

Cooper, William W., Lawrence M. Seiford and Joe Zhu (2011). *Handbook on Data Envelopment Analysis*. Second Edition. Springer.

Cooper, William W., Lawrence Seiford and Kaoru Tone (2007). *Data Envelopment Analysis - A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Springer.

Cordero, José Manuel, Javier Salinas-Jiménez and M Mar Salinas-Jiménez (2017a). „Exploring factors affecting the level of happiness across countries: A conditional robust nonparametric frontier analysis“. In: *European Journal of Operational Research* 256.2, pp. 663–672.

Cordero, José Manuel, Daniel Santín and Rosa Simancas (2017b). „Assessing European primary school performance through a conditional nonparametric model“. In: *Journal of the Operational Research Society* 68.4, pp. 364–376.

Cravioto, Jordi, Eiji Yamasue, Hideki Okumura and Keiichi N Ishihara (2011). „Performance analysis between well-being, energy and environmental indicators using data envelopment analysis“. In: *Zero-carbon Energy Kyoto 2010*. Springer, pp. 49–55.

Damiani, Valeria (2016). „Large-scale assessments and educational policies in Italy“. In: *Research Papers in Education* 31.5, pp. 529–541.

De Castro, Vinicius Ferreira and Enzo Morosini Frazzon (2017). „Benchmarking of best practices: an overview of the academic literature“. In: *Benchmarking: An International Journal*.

De Witte, Kristof and Laura Lopez-Torres (2017). „Efficiency in education: a review of literature and a way forward“. In: *Journal of the Operational Research Society* 68.4, pp. 339–363.

Debnath, Roma Mitra and Ravi Shankar (2014). „Does Good Governance Enhance Happiness: A Cross Nation Study“. In: *Social Indicators Research* 116.1, pp. 235–253.

Decancq, Koen and María Ana Lugo (2013). „Weights in Multidimensional Indices of Wellbeing: An Overview“. In: *Econometric Reviews* 32.1, pp. 7–34.

Despotis, D K (2005). „A reassessment of the human development index via data envelopment analysis“. In: *Journal of the Operational Research Society* 56.8, pp. 969–980.

Diener, Ed, Shigehiro Oishi and Richard E Lucas (2015). „National accounts of subjective well-being.“ In: *American Psychologist* 70.3, p. 234.

DiMaria, Charles Henri, Chiara Peroni and Francesco Sarracino (2019). „Happiness Matters: Productivity Gains from Subjective Well-Being“. In: *Journal of Happiness Studies*.

Dimitrov, Stanko and Warren Sutton (2010). „Promoting symmetric weight selection in data envelopment analysis: A penalty function approach“. In: *European Journal of Operational Research* 200.1, pp. 281–288.

Doyle, John and Rodney Green (1994). „Efficiency and Cross-efficiency in DEA: Derivations, Meanings and Uses“. In: *Journal of the Operational Research Society* 45.5, pp. 567–578.

Dronkers, Jaap, Mark Levels and Manon de Heus (2014). „Migrant pupils scientific performance: the influence of educational system features of origin and destination countries“. In: *Large-scale Assessments in Education* 2.1, p. 3.

Duguleana, Liliana and Constantin Duguleana (2015). „Data envelopment analysis for the efficiency of academic departments“. In: *Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V* 8.2, p. 453.

Dustmann, Christian, Tommaso Frattini and Gianandrea Lanzara (2012). „Educational achievement of second-generation immigrants: an international comparison“. In: *Economic Policy* 27.69, pp. 143–185.

Dyson, Robert G, Rachel Allen, Ana S Camanho, Victor V Podinovski, Claudia S Sarrico and Estelle A Shale (2001). „Pitfalls and protocols in DEA“. In: *European Journal of Operational Research* 132.2. Data Envelopment Analysis, pp. 245–259.

Dyson, Robert G and Emmanuel Thanassoulis (1988). „Reducing weight flexibility in data envelopment analysis“. In: *Journal of the Operational Research Society* 39.6, pp. 563–576.

Emrouznejad, Ali and Guo-liang Yang (2018). „A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016“. In: *Socio-Economic Planning Sciences* 61, pp. 4–8.

Entorf, Horst and Nicoleta Minoiu (2005). „What a difference immigration policy makes: A comparison of PISA scores in Europe and traditional countries of immigration“. In: *German Economic Review* 6.3, pp. 355–376.

Facchini, Giovanni and Elisabetta Lodigiani (2014). „Attracting skilled immigrants: An overview of recent policy developments in advanced countries“. In: *National Institute Economic Review* 229.1, R3–R21.

Fadeyi, Oluwaseun, Petra Maresova, Ruzena Stemberkova, Micheal Afolayan and Funminiyi Adeoye (2019). „Perspectives of University-Industry Technology Transfer in African Emerging Economies: Evaluating the Nigerian Scenario via a Data Envelopment Approach“. In: *Social Sciences* 8.10, p. 286.

Falck, Oliver, Constantin Mang and Ludger Woessmann (2018). „Virtually No Effect? Different Uses of Classroom Computers and their Effect on Student Achievement“. In: *Oxford Bulletin of Economics and Statistics* 80.1, pp. 1–38.

Fandel, Günter (2007). „On the performance of universities in North Rhine-Westphalia, Germany: Governments redistribution of funds judged using DEA efficiency measures“. In: *European Journal of Operational Research* 176.1, pp. 521–533.

Färe, Rolf and Shawna Grosskopf (2004). „Modeling undesirable factors in efficiency evaluation: comment“. In: *European Journal of Operational Research* 157.1, pp. 242–245.

Färe, Rolf, Shawna Grosskopf, Mary Norris and Zhongyang Zhang (1994). „Productivity growth, technical progress, and efficiency change in industrialized countries“. In: *The American Economic Review*, pp. 66–83.

Farrell, Michael James (1957). „The measurement of productive efficiency“. In: *Journal of the Royal Statistical Society: Series A (General)* 120.3, pp. 253–281.

Feeney, Sharon and John Hogan (2017). „A path dependence approach to understanding educational policy harmonisation: the qualifications framework in the European higher education area“. In: *Higher Education Policy* 30.3, pp. 279–298.

Fischer, Mira and Patrick Kampkötter (2017). „Effects of German universities excellence initiative on ability sorting of students and perceptions of educational quality“. In: *Journal of Institutional and Theoretical Economics* 173.4, p. 662.

Foladi, Samira, Maghsud Solimanpur and Mustafa Jahangoshai Rezaee (2019). „Inverse Dynamic Data Envelopment Analysis for Evaluating Faculties of University with Quasi-Fixed Inputs“. In: *Social Indicators Research*, pp. 1–25.

Førsund, Finn R. (2013). „Weight restrictions in DEA: misplaced emphasis?“ In: *Journal of Productivity Analysis* 40.3, pp. 271–283.

Frenken, Koen, Gaston J. Heimeriks and Jarno Hoekman (2017). „What drives university research performance? An analysis using the CWTS Leiden Ranking data“. In: *Journal of Informetrics* 11.3, pp. 859–872.

Gallup, INC (2017). *Worldwide research methodology and codebook.*

Gawellek, Bastian and Marco Sunder (2016). *The German excellence initiative and efficiency change among universities, 2001-2011.* Tech. rep. Working Paper, Universität Leipzig, Wirtschaftswissenschaftliche Fakultät.

Gilligan, Megan, Jill J Suitor, Sangbo Nam, Brianna Routh, Marissa Rurka and Gulcin Con (2017). „Family networks and psychological well-being in midlife“. In: *Social Sciences* 6.3, p. 94.

Göken, Ylmaz, Onur Doan and Bilge Özkarabacak (2015). „A data envelopment analysis application for measuring efficiency of university departments“. In: *Procedia Economics and Finance* 19, pp. 226–237.

Golany, Boaz and Yaakov Roll (1989). „An application procedure for DEA“. In: *Omega* 17.3, pp. 237–250.

Gomes, EG and Marcos Pereira Estellita Lins (2008). „Modelling undesirable outputs with zero sum gains data envelopment analysis models“. In: *Journal of the Operational Research Society* 59.5, pp. 616–623.

Gralka, Sabine, Klaus Wohlrabe and Lutz Bornmann (2019). „How to measure research efficiency in higher education? Research grants vs. publication output“. In: *Journal of Higher Education Policy and Management* 41.3, pp. 322–341.

Greco, Salvatore, Alessio Ishizaka, Menelaos Tasiou and Gianpiero Torrisi (2019). „On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness“. In: *Social Indicators Research* 141.1, pp. 61–94.

Guardiola, Jorge and Andrés J. Picazo-Tadeo (2014). „Building Weighted-Domain Composite Indices of Life Satisfaction with Data Envelopment Analysis". In: *Social Indicators Research* 117.1, pp. 257–274.

Guo, Xiaoying, Ching-Cheng Lu, Jen-Hui Lee and Yung-Ho Chiu (2017). „Applying the dynamic DEA model to evaluate the energy efficiency of OECD countries and China". In: *Energy* 134, pp. 392–399.

Hadi-Vencheh, Abdollah, Josef Jablonsky and Abbasi Esmaeilzadeh (2015). „The slack-based measure model based on supporting hyperplanes of production possibility set". In: *Expert Systems with Applications* 42.19, pp. 6522–6529.

Hahm, Sabrina and Jochen Kluve (2016). „Effects of the Bologna Reform on educational outcomes: Micro evidence from Germany". In: *IZA Discussion Paper*.

Hahm, Sabrina and Jochen Kluve (2019). „Better with Bologna? Tertiary education reform and student outcomes". In: *Education Economics* 27.4, pp. 425–449.

Harris, Richard, Louise Courtney, Zain Ul-Abadin and Katharine Burn (2019). „Student access to the curriculum in an age of performativity and accountability: an examination of policy enactment". In: *Research Papers in Education*, pp. 1–21.

Ho, Esther Sui Chu (2016). „The use of large-scale assessment (PISA): insights for policy and practice in the case of Hong Kong". In: *Research Papers in Education* 31.5, pp. 516–528.

Hochschild, Jennifer L and Porsha Cropper (2010). „Immigration regimes and schooling regimes: Which countries promote successful immigrant incorporation?" In: *School Field* 8.1, pp. 21–61.

Hsu, Maxwell, Xueming Luo and Gary H Chao (2008). „The fog of OECD and non-OECD country efficiency: a data envelopment analysis approach". In: *The Journal of Developing Areas*, pp. 81–93.

Hu, Jin-Li and Chih-Hung Kao (2007). „Efficient energy-saving targets for APEC economies". In: *Energy policy* 35.1, pp. 373–382.

Hüther, Otto and Georg Krücken (2018). *Higher Education in Germany–Recent Developments in an International Perspective*. Vol. 49. Springer.

Hwang, Jihyun, Kyong Mi Choi, Yejun Bae and Dong Hoon Shin (2018). „Do Teachers Instructional Practices Moderate Equity in Mathematical and Scientific Literacy?: an Investigation of the PISA 2012 and 2015". In: *International Journal of Science and Mathematics Education* 16.1, pp. 25–45.

Iribarren, Diego and Ian Vázquez-Rowe (2013). „Is labor a suitable input in LCA+DEA studies? Insights on the combined use of economic, environmental and social parameters". In: *Social Sciences* 2.3, pp. 114–130.

Isphording, Ingo E., Marc Piopiunik and Nuria Rodriguez-Planas (2016). „Speaking in numbers: The effect of reading performance on math performance among immigrants“. In: *Economics Letters* 139, pp. 52–56.

Jain, Vipul, Amit Kumar, Sameer Kumar and Charu Chandra (2015). „Weight restrictions in Data Envelopment Analysis: A comprehensive Genetic Algorithm based approach for incorporating value judgments“. In: *Expert Systems with Applications* 42.3, pp. 1503–1512.

Jerrim, John (2015). „Why do East Asian children perform so well in PISA? An investigation of Western-born children of East Asian descent“. In: *Oxford Review of Education* 41.3, pp. 310–333.

Johnes, Jill (2006). „Data envelopment analysis and its application to the measurement of efficiency in higher education“. In: *Economics of Education Review* 25.3, pp. 273–288.

Johnes, Jill, Maria Portela and Emmanuel Thanassoulis (2017). „Efficiency in education“. In: *Journal of the Operational Research Society* 68.4, pp. 331–338.

Joro, Tarja and Esa-Jussi Viitala (2004). „Weight-restricted DEA in action: from expert opinions to mathematical models“. In: *Journal of the Operational Research Society* 55.8, pp. 814–821.

Kao, Chiang and Hsi-Tai Hung (2008). „Efficiency analysis of university departments: An empirical study“. In: *Omega* 36.4. Special Issue on Logistics: New Perspectives and Challenges, pp. 653–664.

Karatheodoros, Anastasios (2017). „The contribution of secondary education on regional economic growth in Greece, over the period 1995-2012s“. In: *International Journal of Education Economics and Development* 8.1, pp. 46–64.

Kehm, Barbara M. (2013). „To Be or Not to Be? The Impacts of the Excellence Initiative on the German System of Higher Education“. In: *Institutionalization of World-Class University in Global Competition*. Ed. by Jung Cheol Shin and Barbara M. Kehm. Dordrecht: Springer Netherlands, pp. 81–97.

Khalili, Masoud, Ana Camanho, Maria Silva Portela and Mohammadreza Alirezaee (2010). „The measurement of relative efficiency using data envelopment analysis with assurance regions that link inputs and outputs“. In: *European Journal of Operational Research* 203.3, pp. 761–770.

Kirjavainen, Tanja (2015). „Immigrant students and the efficiency of basic education“. In: *National Audit Office of Finland*.

Kloss, Gunther (1985). „The academic restructuring of British and German universities and greater efficiency: A comparative perspective“. In: *Oxford Review of Education* 11.3, pp. 271–282.

Kong, Wei-Hsin and Tsu-Tan Fu (2012). „Assessing the performance of business colleges in Taiwan using data envelopment analysis and student based value-added performance indicators“. In: *Omega* 40.5, pp. 541–549.

Kounetas, Kostas, Athanasios Anastasiou, Panagiotis Mitropoulos and Ioannis Mitropoulos (2011). „Departmental efficiency differences within a Greek university: An application of a DEA and Tobit analysis“. In: *International Transactions in Operational Research* 18.5, pp. 545–559.

Kumar, Ashok, Jivesh Bansal, BM Gupta and Manoj K Joshi (2017). „Scientometric Assessment of Kurukshetra University Research Output during 2006-15“. In: *International Journal of Information Dissemination and Technology* 7.1, p. 26.

Kunz, Johannes S. (2016). „Analyzing Educational Achievement Differences between Second-Generation Immigrants: Comparing Germany and German-Speaking Switzerland“. In: *German Economic Review* 17.1, pp. 61–91.

Kuosmanen, Timo and Victor Podinovski (2009). „Weak disposability in nonparametric production analysis: reply to Färe and Grosskopf“. In: *American Journal of Agricultural Economics* 91.2, pp. 539–545.

Lehmann, Erik E., Michele Meoli, Stefano Paleari and Sarah A. E. Stockinger (2018). „Approaching effects of the economic crisis on university efficiency: a comparative study of Germany and Italy“. In: *Eurasian Business Review* 8.1, pp. 37–54.

Lepori, Benedetto, Andrea Bonaccorsi, Alessandro Daraio, Cinzia Daraio, Hebe Gunnes et al. (2015). „ETER project“. In: *Handbook for Data Collection. Brussels.*

Levels, Mark, Jaap Dronkers and Gerbert Kraaykamp (2008). „Immigrant Children's Educational Achievement in Western Countries: Origin, Destination, and Community Effects on Mathematical Performance“. In: *American Sociological Review* 73.5, pp. 835–853.

Liu, John S., Louis Y.Y. Lu, Wen-Min Lu and Bruce J.Y. Lin (2013). „A survey of DEA applications“. In: *Omega* 41.5, pp. 893–902.

Lovell, CA Knox, Jesus T Pastor and Judi A Turner (1995). „Measuring macroeconomic performance in the OECD: A comparison of European and non-European countries“. In: *European Journal of Operational Research* 87.3, pp. 507–518.

MacLeod, W Bentley, Evan Riehl, Juan E Saavedra and Miguel Urquiola (2017). „The big sort: College reputation and labor market outcomes“. In: *American Economic Journal: Applied Economics* 9.3, pp. 223–61.

Mariano, Enzo Barberio, Vinicius Amorim Sobreiro and Daisy Aparecida do Nascimento Rebelatto (2015). „Human development and data envelopment analysis: A structured literature review“. In: *Omega* 54, pp. 33–49.

Mecit, Emine Demet and Ihsan Alp (2013). „A new proposed model of restricted data envelopment analysis by correlation coefficients“. In: *Applied Mathematical Modelling* 37.5, pp. 3407–3425.

Mellewigt, Thomas and Erik Schäfer (2017). „Professorale Wurstschnappaktionen“. In: *Managementforschung* 27.1, pp. 137–146.

Mikuová, Pavla (2015). „An application of DEA methodology in efficiency measurement of the Czech public universities“. In: *Procedia Economics and Finance* 25, pp. 569–578.

Mizobuchi, Hideyuki (2014). „Measuring World Better Life Frontier: A Composite Indicator for OECD Better Life Index“. In: *Social Indicators Research* 118.3, pp. 987–1007.

Mizobuchi, Hideyuki (2017). „Measuring Socio-economic Factors and Sensitivity of Happiness“. In: *Journal of Happiness Studies* 18.2, pp. 463–504.

Moesen, Willem and Laurens Cherchye (1998). „The macroeconomic performance of nations measurement and perception“. In: *CES-Dischussion Paper Series DPS 98.22*, pp. 1–29.

Mohamad, Nordin (2007). „A linear programming formulation of macroeconomic performance: the case of Asia Pacific“. In: *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics* 23, pp. 29–40.

Mohamad, Nordin Haji and Fatimah Binti Said (2011). „Comparing macroeconomic performance of OIC member countries“. In: *International Journal of Economics and Management Sciences* 1.3, pp. 90–104.

Morais, Paulo and Ana S. Camanho (2011). „Evaluation of performance of European cities with the aim to promote quality of life improvements“. In: *Omega* 39.4, pp. 398–409.

Morais, Paulo, Vera L. Miguéis and Ana S. Camanho (2013). „Quality of Life Experienced by Human Capital: An Assessment of European Cities“. In: *Social Indicators Research* 110.1, pp. 187–206.

Moreno, Justo de Jorge, Adriana González Robles, Antonio Martinez, Roberto Minero Calvo and Andrada Georgiana Miron (2018). „Assessing efficiency in the Spanish public universities through comparative non-radial and radial data envelopment analysis“. In: *Tertiary Education and Management*, pp. 1–16.

Moshtaghi, Hojatollah Rajabi, Gholam Reza Faramarzi and Reza Farzipoor Saen (2018). „Developing new methods for determining weights of components in network data envelopment analysis“. In: *International Journal of Operational Research* 32.2, pp. 223–250.

Nazarko, Joanicjusz and Jonas Saparauskas (2014). „Application of DEA method in efficiency evaluation of public higher education institutions“. In: *Technological and Economic Development of Economy* 20.1, pp. 25–44.

OECD (2015). *Education at a Glance 2015*, p. 564.

OECD (2016). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. Paris: OECD Publishing.

OECD (2017). *PISA 2015 Technical Report*. Tech. rep. OECD Publishing.

Oishi, Shigehiro and Selin Kesebir (2015). „Income Inequality Explains Why Economic Growth Does Not Always Translate to an Increase in Happiness“. In: *Psychological Science* 26.10. PMID: 26338882, pp. 1630–1638.

Olesen, Ole B. and Niels C. Petersen (1996). „Indicators of Ill-Conditioned Data Sets and Model Misspecification in Data Envelopment Analysis: An Extended Facet Approach“. In: *Management Science* 42.2, pp. 205–219.

Olesen, Ole B. and Niels C. Petersen (2002). „The Use of Data Envelopment Analysis with Probabilistic Assurance Regions for Measuring Hospital Efficiency“. In: *Journal of Productivity Analysis* 17.1, pp. 83–109.

Overton, Kathleen, Seong-Jong Joo and Philipp A. Stoeberl (2016). „Benchmarking public school performance by unionized status“. In: *Benchmarking: An International Journal* 23.7, pp. 1626–1642.

Pannu, Husanbir Singh, U. Dinesh Kumar and Jamal A. Farooquie (2010). „Efficiency and productivity analysis of Indian pharmaceutical industry using data envelopment analysis“. In: *International Journal of Operational Research* 10.1, pp. 121–136.

Parr, Alyssa K. and Verena S. Bonitz (2015). „Role of Family Background, Student Behaviors, and School-Related Beliefs in Predicting High School Dropout“. In: *The Journal of Educational Research* 108.6, pp. 504–514.

Paruolo, Paolo, Michaela Saisana and Andrea Saltelli (2013). „Ratings and rankings: voodoo or science?“ In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176.3, pp. 609–634.

Patrizii, Vincenzo, Anna Pettini and Giuliano Resce (2017). „The Cost of Well-Being“. In: *Social Indicators Research* 133.3, pp. 985–1010.

Pavone, Antonio and Paola Pianura (2015). „A Comparison of Macroeconomic Performance in a Context of Confliting Policy Objectives“. In: *Proceedings of the 11th*

International Conference of ASECU Openness, Innovation, Efficiency and Democratization as Preconditions for Economic Development, pp. 164–165.

Peiró-Palomino, Jesús and Andrés J. Picazo-Tadeo (2018). „OECD: One or Many? Ranking Countries with a Composite Well-Being Indicator“. In: *Social Indicators Research* 139.3, pp. 847–869.

Pinar, Mehmet, Thanasis Stengos and Nikolas Topaloglou (2017). „Testing for the implicit weights of the dimensions of the Human Development Index using stochastic dominance“. In: *Economics Letters* 161, pp. 38–42.

Podinovski, Victor V. (1999). „Side effects of absolute weight bounds in DEA models“. In: *European Journal of Operational Research* 115.3, pp. 583–595.

Podinovski, Victor V. (2004). „Production trade-offs and weight restrictions in data envelopment analysis“. In: *Journal of the Operational Research Society* 55.12, pp. 1311–1322.

Prados de la Escosura, Leandro (2015). „World Human Development: 18702007“. In: *Review of Income and Wealth* 61.2, pp. 220–247.

Puig-Junoy, Jaume (1998). „Measuring health production performance in the OECD“. In: *Applied Economics Letters* 5.4, pp. 255–259.

Ramón, Nuria, José L. Ruiz and Inmaculada Sirvent (2010). „A multiplier bound approach to assess relative efficiency in DEA without slacks“. In: *European Journal of Operational Research* 203.1, pp. 261–269.

Rangvid, Beatrice Schindler (2007). „Sources of immigrants' underachievement: Results from PISACopenhagen“. In: *Education Economics* 15.3, pp. 293–326.

Rashidi, Kamran and Reza Farzipoor Saen (2015). „Measuring eco-efficiency based on green indicators and potentials in energy saving and undesirable output abatement“. In: *Energy Economics* 50, pp. 18–26.

Ray, Amal Kanti (2008). „Measurement of social development: an international comparison“. In: *Social Indicators Research* 86.1, pp. 1–46.

Ray, Subhash C., Lawrence M. Seiford and Joe Zhu (1998). „Market entity behavior of Chinese state-owned enterprises“. In: *Omega* 26.2, pp. 263–278.

Reig-Martínez, Ernest (2013). „Social and Economic Wellbeing in Europe and the Mediterranean Basin: Building an Enlarged Human Development Indicator“. In: *Social Indicators Research* 111.2, pp. 527–547.

Reinhard, Sti, CA Knox Lovell and Geert J Thijssen (2000). „Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA“. In: *European Journal of Operational Research* 121.2, pp. 287–303.

Reparaz, Charo and María Angeles Sotés-Elizalde (2019). „Parental involvement in schools in Spain and Germany: Evidence from PISA 2015“. In: *International Journal of Educational Research* 93, pp. 33–52.

Retzlaff-Roberts, Donna, Cyril F Chang and Rose M Rubin (2004). „Technical efficiency in the use of health care resources: a comparison of OECD countries“. In: *Health policy* 69.1, pp. 55–72.

Rhaiem, Mehdi (2017). „Measurement and determinants of academic research efficiency: a systematic review of the evidence“. In: *Scientometrics* 110.2, pp. 581–615.

Rindermann, Heiner and James Thompson (2016). „The cognitive competences of immigrant and native students across the world: An analysis of gaps, possible causes and impact“. In: *Journal of Biosocial Science* 48.1, pp. 66–93.

Rogiers, Amelie, Hilde Van Keer and Emmelien Merchie (2020). „The profile of the skilled reader: An investigation into the role of reading enjoyment and student characteristics“. In: *International Journal of Educational Research* 99, p. 101512.

Ruiz, José L., José V. Segura and Inmaculada Sirvent (2015). „Benchmarking and target setting with expert preferences: An application to the evaluation of educational performance of Spanish universities“. In: *European Journal of Operational Research* 242.2, pp. 594–605.

Saaty, Thomas (1980). *The Analytic Hierarchy Process, Planning, Piority Setting, Resource Allocation*. McGraw-Hill.

Saen, Reza Farzipoor (2010). „Performance measurement of power plants in the existence of weight restrictions via slacks-based model“. In: *Benchmarking: An International Journal* 17.5, pp. 677–691.

Sarrico, Claudia S. and Robert G. Dyson (2004). „Restricting virtual weights in data envelopment analysis“. In: *European Journal of Operational Research* 159.1, pp. 17–34.

Schaffnit, Claire, Dan Rosen and Joseph C. Paradi (1997). „Best practice analysis of bank branches: An application of DEA in a large Canadian bank“. In: *European Journal of Operational Research* 98.2, pp. 269–289.

Scheerlinck, Eef, Klaas De Brucker and Pamela Kerkhove (2019). „The Pan-European Bologna Reform: An Analysis Based on Employer Perceptions in a Belgian Context“. In: *European Education* 51.4, pp. 309–328.

Schneebaum, Alyssa, Bernhard Rumplmaier and Wilfried Altzinger (2016). „Gender and migration background in intergenerational educational mobility“. In: *Education Economics* 24.3, pp. 239–260.

Schneeweis, Nicole (2011). „Educational institutions and the integration of migrants". In: *Journal of Population Economics* 24.4, pp. 1281–1308.

Schnepf, Sylke Viola (2007). „Immigrants educational disadvantage: an examination across ten countries and three surveys". In: *Journal of Population Economics* 20.3, pp. 527–545.

Schwab, Klaus and Xavier Sala-i-Mart'in (2016). „The global competitiveness report 2013–2014: Full data edition". In: World Economic Forum.

Seeber, Marco, Vitaliano Barberio, Jeroen Huisman and Jelle Mampaey (2019). „Factors affecting the content of universities mission statements: an analysis of the United Kingdom higher education system". In: *Studies in Higher Education* 44.2, pp. 230–244.

Silverman, Bernard W. (1986). *Density estimation for statistics and data analysis*. eng. London [u.a.]: Chapman and Hall.

Simsek, Nevzat (2014). „Energy efficiency with undesirable output at the economy-wide level: cross country comparison in OECD sample". In: *American Journal of Energy Research* 2.1, pp. 9–17.

Spinks, Jean and Bruce Hollingsworth (2005). „Health production and the socioeconomic determinants of health in OECD countries: the use of efficiency models". In: *Monash University. Centre for Health Economics, Working Paper* 151.

Staníková, Michaela and Karel Skokan (2012). „Evaluation of the EU member states efficiency by data envelopment analysis method". In: *International Journal of Mathematical Models and Methods in Applied Sciences* 6.2, pp. 349–357.

Sueyoshi, Toshiyuki and Mika Goto (2012a). „DEA radial and non-radial models for unified efficiency under natural and managerial disposability: Theoretical extension by strong complementary slackness conditions". In: *Energy Economics* 34.3, pp. 700–713.

Sueyoshi, Toshiyuki and Mika Goto (2012b). „DEA radial measurement for environmental assessment and planning: Desirable procedures to evaluate fossil fuel power plants". In: *Energy Policy* 41, pp. 422–432.

Sueyoshi, Toshiyuki and Kazuyuki Sekitani (2009). „An occurrence of multiple projections in DEA-based measurement of technical efficiency: Theoretical comparison among DEA models from desirable properties". In: *European Journal of Operational Research* 196.2, pp. 764–794.

Sueyoshi, Toshiyuki, Yan Yuan and Mika Goto (2017). „A literature study for DEA applied to energy and environment". In: *Energy Economics* 62, pp. 104–124.

Suh, Eunbook and Soyeon Choi (2018). „Predictors of subjective well-being across cultures“. In: *Handbook of Well-being. Salt Lake City: DEF.*

Sutherland, Douglas, Robert Price and Frédéric Gonand (2009). „Improving public spending efficiency in primary and secondary education“. In: *OECD Journal: Economic Studies* 2009.1, pp. 1–30.

Suzuki, Soushi and Peter Nijkamp (2016). „An evaluation of energy-environment-economic efficiency for EU, APEC and ASEAN countries: Design of a Target-Oriented DFM model with fixed factors in Data Envelopment Analysis“. In: *Energy Policy* 88, pp. 100–112.

Suzuki, Soushi and Peter Nijkamp (2017). „Overview of DEA and Its Improvements“. In: *Regional Performance Measurement and Improvement: New Developments and Applications of Data Envelopment Analysis.* Singapore: Springer Singapore. Chap. Regional Performance Measurement and Improvement, pp. 13–18.

Teltemann, Janna and Reinhard Schunck (2016). „Education systems, school segregation, and second-generation immigrants educational success: Evidence from a country-fixed effects approach using three waves of PISA“. In: *International Journal of Comparative Sociology* 57.6, pp. 401–424.

Thanassoulis, Emmanuel, Mika Kortelainen, Geraint Johnes and Jill Johnes (2011). „Costs and efficiency of higher education institutions in England: a DEA analysis“. In: *Journal of the operational research society* 62.7, pp. 1282–1297.

Thanassoulis, Emmanuel, Maria Conceicao Portela and Rachel Allen (2004). „Handbook on Data Envelopment Analysis“. In: ed. by William W. Cooper, Lawrence M. Seiford and Joe Zhu. Kluwer Acamedic Publishers. Chap. Incorporating Value Judgments in DEA, pp. 99–138.

Thanassoulis, Emmanuel, Kristof De Witte, Jill Johnes, Geraint Johnes, Giannis Karagiannis and Conceição S. Portela (2016). „Applications of Data Envelopment Analysis in Education“. In: *Data Envelopment Analysis: A Handbook of Empirical Studies and Applications.* Ed. by Joe Zhu. Boston, MA: Springer US, pp. 367–438.

Theodoridis, Alexandros M. and Athanasios Ragkos (2015). „A Restricted Data Envelopment Analysis Application to Dairy Farming“. In: *Data Envelopment Analysis Journal* 1.2, pp. 171–193.

Thompson, Russell, Larry Langemeier, Chih-Tah Lee, Euntaik Lee and Robert Thrall (1990). „The role of multiplier bounds in efficiency analysis with application to Kansas farming“. In: *Journal of Econometrics* 46.1, pp. 93–108.

Thompson, Russell, FD Singleton, Robert Thrall and Barton Smith (1986). „Comparative site evaluations for locating a high-energy physics lab in Texas“. In: *Interfaces* 16.6, pp. 35–49.

Tian, Na, Shusong Tang, Ada Che and Peng Wu (2020). „Measuring regional transport sustainability using super-efficiency SBM-DEA with weighting preference“. In: *Journal of Cleaner Production* 242, p. 118474.

Tobin, Mollie, Dita Nugroho and Petra Lietz (2016). „Large-scale assessments of students learning and education policy: synthesising evidence across world regions“. In: *Research Papers in Education* 31.5, pp. 578–594.

Tone, Kaoru (2001). „A Slacks-Based Measure of Efficiency in Data Envelopment Analysis“. In: *European Journal of Operational Research* 130.3, pp. 498–509.

Tone, Kaoru (2002). „A slacks-based measure of super-efficiency in data envelopment analysi“. In: *European Journal of Operational Research* 143.1, pp. 32–41.

Tone, Kaoru (2015). „SBM variations revisited“. In: *GRIPS Discussion Papers* 15.

Tone, Kaoru (2017). „On the Consistency of Slacks-based Measure-Max Model and Super-Slacks-based Measure Model“. In: *Universal Journal of Management* 5.3, pp. 160–165.

Tong, Goh Chong (2001). „National day speech“. In: *National Archives of Singapore*.

Tsai, Wen-Hsien, Hsiu-Li Lee, Chih-Hao Yang and Chung-Chen Huang (2016). „Input-output analysis for sustainability by using DEA method: A comparison study between European and Asian countries“. In: *Sustainability* 8.12, p. 1230.

Tzanakou, Charikleia and Heike Behle (2017). „The intra-European transferability of graduates skills gained in the UK“. In: *Journal of Ethnic and Migration Studies* 43.8, pp. 1379–1397.

Ülengin, Füsun, Özgür Kabak, Sule Önsel, Emel Aktas and Barnett R. Parker (2011). „The competitiveness of nations and implications for human development“. In: *Socio-Economic Planning Sciences* 45.1, pp. 16–27.

United Nations Development Programme (2018). *Human development indices and indicators: 2018 Statistical update*.

Veenhoven, Ruut (2015). „Social conditions for human happiness: A review of research“. In: *International Journal of Psychology* 50.5, pp. 379–391.

Veiderpass, Ann and Maureen McKelvey (2016). „Evaluating the performance of higher education institutions in Europe: a nonparametric efficiency analysis of 944 institutions“. In: *Applied Economics* 48.16, pp. 1504–1514.

Visbal-Cadavid, Delimiro, Mónica Martínez-Gómez and Francisco Guijarro (2017). „Assessing the efficiency of public universities through DEA. A case study“. In: *Sustainability* 9.8, p. 1416.

Vlontzos, George, Spyros Niavis and Basil Manos (2014). „A DEA approach for estimating the agricultural energy and environmental efficiency of EU countries“. In: *Renewable and Sustainable Energy Reviews* 40, pp. 91–96.

Vögtle, Eva Maria (2019). „20 years of Bologna - a story of success, a story of failure“. In: *Innovation: The European Journal of Social Science Research* 32.4, pp. 406–428.

Waltman, Ludo, Clara Calero-Medina, Joost Kosten, Ed CM Noyons, Robert JW Tijssen, Nees Jan van Eck, Thed N van Leeuwen, Anthony FJ van Raan, Martijn S Visser and Paul Wouters (2012). „The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation“. In: *Journal of the American Society for Information Science and Technology* 63.12, pp. 2419–2432.

Warning, Susanne (2004). „Performance differences in German higher education Empirical analysis of strategic groups“. In: *Review of Industrial Organization* 24.4, pp. 393–408.

Welsch, Heinz and Jan Kühling (2016). „Macroeconomic performance and institutional change: evidence from subjective well-being data“. In: *Journal of Applied Economics* 19.2, pp. 193–217.

Woessmann, Ludger (2016). „The Importance of School Systems: Evidence from International Differences in Student Achievement“. In: *Journal of Economic Perspectives* 30.3, pp. 3–32.

Wohlrabe, Klaus, Lutz Bornmann, Sabine Gralka and Felix de Moya Anegon (2019a). „Wie effizient forschen Universitäten in Deutschland, deren Zukunftskonzepte im Rahmen der Exzellenzinitiative ausgezeichnet wurden? Ein empirischer Vergleich von Input-und Output-Daten“. In: *Zeitschrift Für Evaluation* 2.

Wohlrabe, Klaus, Sabine Gralka and Lutz Bornmann (2019b). „Zur Effizienz deutscher Universitäten und deren Entwicklung zwischen 2004 und 2015“. In: *ifo Schnelldienst* 72.21, pp. 15–21.

Wolszczak-Derlacz, Joanna (2017). „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“. In: *Research Policy* 46.9, pp. 1595–1605.

Wong, Y-HB and JE Beasley (1990). „Restricting weight flexibility in data envelopment analysis“. In: *Journal of the Operational Research Society* 41.9, pp. 829–835.

World Health Organization (2018). „WHO methods and data sources for life tables 1990-2016“. In: *Global Health Estimates Technical Paper WHO/HIS/IER/GHE*.

Xu, Huan and Fangtao Liu (2017). „Measuring the efficiency of education and technology via DEA approach: Implications on national development“. In: *Social Sciences* 6.4, p. 136.

Yang, Guo-Liang and Ali Emrouznejad (2019). „Modelling efficient and anti-efficient frontiers in DEA without explicit inputs“. In: *International Journal of Operational Research* 35.4, pp. 505–528.

Yeasmin, Nafisa and Satu Uusiautti (2018). „Finland and Singapore, Two Different Top Countries of PISA and the Challenge of Providing Equal Opportunities to Immigrant Students.“ In: *Journal for Critical Education Policy Studies (JCEPS)* 16.1.

Zhou, Haibo, Yi Yang, Yao Chen and Joe Zhu (2018). „Data envelopment analysis application in sustainability: The origins, development and future directions“. In: *European Journal of Operational Research* 264.1, pp. 1–16.

Zhou, Yan, Xinpeng Xing, Kuangnan Fang, Dapeng Liang and Chunlin Xu (2013). „Environmental efficiency analysis of power industry in China based on an entropy SBM model“. In: *Energy Policy* 57, pp. 68–75.

Zhu, Yuanchang, Yongjun Li and Liang Liang (2018). „A variation of two-stage SBM with leader-follower structure: an application to Chinese commercial banks“. In: *Journal of the Operational Research Society* 69.6, pp. 840–848.

Attachments

Überblick über die Einzelbeiträge der vorliegenden Dissertation:

- Gerald Fugger: “Absolute and Relative Weight Restrictions in DEA - An Comparison”
  Diese Studie wurde zur Veröffentlichung bei einer Zeitschrift eingereicht.

- Gerald Fugger: “Comparing the Slack-Based Maximum Measure of Efficiency: A Simulation Application”
  Diese Studie wurde zur Veröffentlichung bei einer Zeitschrift eingereicht.

- Andreas Behr, Gerald Fugger: “PISA Performance of Natives and Immigrants: Selection versus Efficiency”
  Diese Studie wurde in *Open Education Studies* angenommen und publiziert.

- Gerald Fugger: “Higher Education Institution Efficiency in Germany and the United Kingdom”
  Diese Studie wurde zur Veröffentlichung bei einer Zeitschrift eingereicht.

- Andreas Behr, Gerald Fugger: “An Encompassing Assessment of OECD Countries Using Weight Restricted DEA Models”
  Diese Studie wurde zur Veröffentlichung bei einer Zeitschrift eingereicht.

Gerald Fugger
Herwarthstr. 50
45138 Essen

Eidesstattliche Erklärung zu meiner Dissertation mit dem Titel:
"Data Envelopment Analysis: Methodological Aspects and Empirical Applications"

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und keinen anderen als die angegebenen Hilfsmittel verwendet habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet.

Die Gelegenheit zum vorliegenden Promotionsverfahren ist mir nicht kommerziell vermittelt worden. Insbesondere habe ich keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt. Mir ist bekannt, dass Unwahrheiten hinsichtlich der vorstehenden Erklärung die Zulassung zur Promotion ausschließen bzw. später zum Verfahrensabbruch oder zur Rücknahme des Titels führen können.

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig ohne unzulässige Hilfe Dritter verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und alle wörtlich oder inhaltlich übernommenen Stellen unter der Angabe der Quelle als solche gekennzeichnet habe. Die Grundsätze für die Sicherung guter wissenschaftlicher Praxis an der Universität Duisburg-Essen sind beachtet worden. Ich habe die Arbeit keiner anderen Stelle zu Prüfungszwecken vorgelegt.

Ich versichere außerdem, dass ich die vorliegende Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und dass ich in keinem laufenden oder früheren Promotionsverfahren zum Erwerb desselben Grades Dr.rer.pol. endgültig gescheitert bin.

Ort, Datum	Gerald Fugger

Gerald Fugger
Herwarthstr. 50
45138 Essen

Erklärung zur Koautorenschaft

Kapitel 4 der vorliegenden Dissertation – "PISA Performance of Natives and Immigrants: Selection versus Efficiency" – ist eine gemeinsame Forschungsarbeit mit Prof. Behr.
Ich erkläre hiermit, dass ich an der Entstehung dieser Arbeit maßgeblich beteiligt war. Prof. Behr hat den methodischen Ansatz konzipiert und in R auf Basis simulierter Daten implementiert. Mein Beitrag umfasst: Einleitung, Stand der Forschung, Literaturübersicht und -diskussion, Beschreibung der DEA Modelle, Variablen- und Datensatzbeschreibung, Einordnung der Ergebnisse in den Literaturstand und das Fazit. Zudem habe ich die Daten aufbereitet und die deskriptiven Ergebnisse erstellt. Die inhaltliche Interpretation der Ergebnisse und deren Diskussion wurde gemeinsam vorgenommen.

Ort, Datum	Gerald Fugger

Hiermit bestätige ich die obigen Angaben.

Ort, Datum	Prof. Dr. Behr

Gerald Fugger
Herwarthstr. 50
45138 Essen

Erklärung zur Koautorenschaft

Kapitel 6 der vorliegenden Dissertation – "An Encompassing Assessment of OECD Countries Using Weight Restricted DEA Models" – ist eine gemeinsame Forschungsarbeit mit Prof. Behr.
Ich erkläre hiermit, dass ich an der Entstehung dieser Arbeit maßgeblich beteiligt war. Prof. Behr hat den methodischen Ansatz konzipiert und in R auf Basis simulierter Daten implementiert. Mein Beitrag umfasst: Einleitung, Stand der Forschung, Literaturübersicht und -diskussion, Beschreibung der DEA Modelle, Beschreibung der Auswirkung von zusätzlichen Gewichtsbeschränkungen und die Motivation für deren Aufnahme, Variablen- und Datensatzbeschreibung, Einordnung der Ergebnisse in den Literaturstand und das Fazit. Zudem habe ich die Daten aufbereitet und die deskriptiven Ergebnisse erstellt. Die inhaltliche Interpretation der Ergebnisse und deren Diskussion wurde gemeinsam vorgenommen.

Ort, Datum	Gerald Fugger

Hiermit bestätige ich die obigen Angaben.

Ort, Datum	Prof. Dr. Behr