Bootstrapping in Network Data Envelopment Analysis

Maria Michali
Doctor of Philosophy

ASTON UNIVERSITY
August 2022

Aston University
Bootstrapping in Network Data Envelopment Analysis

Maria Michali
Doctor of Philosophy
2022

Abstract

Data Envelopment Analysis (DEA) is a linear-programming method used to measure the relative efficiency of firms. The objective of this thesis is (i) to study the efficiency of the railway transport process in Europe considering its inner structure and the impact of railway noise on humans and (ii) to study the performance of bootstrapping approaches in obtaining DEA efficiency estimates when the production process has a network structure and the relation between the different stages is considered. First, the railway transport process is divided into two stages, related to assets and service provision, respectively. The negative impact of railways on people is measured as the number of people that are exposed to high levels of railway noise. The number of rail wagons in each country that is retrofitted with more silent braking technology is used as a proxy to measure the effort to reduce railway noise pollution. Data is extracted from Eurostat (2016), ERA 006REC1072 Impact Assessment (2018), and EEA (2020) and the additive efficiency decomposition approach is used. Based on the results, asset-efficient countries are usually service-efficient, but the inverse does not hold. Sensitivity analysis revealed that efficiency rankings are robust to alterations in the decomposition weight restrictions. Subsampling bootstrap was chosen as the most appropriate as it does not require any restrictive assumptions. The performance of subsampling is examined through Monte Carlo simulations for various sample and subsample sizes for general two-stage series structures. Results indicate great sensitivity both to the sample and subsample size, as well as to the data generating process-higher than in one-stage structures. A practical approach is suggested to overcome some result inconsistencies that are due to the peculiarities of the additive decomposition algorithm. The method is applied to obtain confidence interval estimates for the overall and stage efficiency scores of European railways.

Keywords: Data Envelopment Analysis, Network Efficiency Decomposition, Subsampling Bootstrap, Monte Carlo, European Railways

Acknowledgements

This work could not have been achieved without the help of my supervisors. First, I would like to express my sincere gratitude to my main supervisor Prof. Ali Emrouznejad. His support during all these years was invaluable, and with his experience and advice, he significantly contributed to my scientific and personal development. I am also deeply grateful to Dr Akram Dehnokhalaji for her guidance and constructive collaboration. Our discussions were always a source of encouragement and motivation. I would also like to thank Prof. Ben Clegg for the collaboration and his multi-faceted support.

Next, I would like to thank the members of my examination committee Prof. Emmanuel Thanassoulis and Prof. Sebastián Lozano for accepting the invitation to take part in the *viva voce* and for the very constructive feedback and discussion; it was a great honour for me. My thanks to Dr Leonidas Anastasakis for agreeing to act as the Chair of the *viva voce*.

I would like to extend my sincere thanks to Prof. Emmanuel Thanassoulis and Dr Dimitris Giraleas for taking the time to read/attend my annual reports, for posing constructive questions and for providing insightful comments and suggestions.

I would like to thank Aston University for offering me the studentship that allowed me to conduct this research. I would also like to thank Farat Ara, and Jennifer Knight for their willingness to offer their help whenever I needed it. My thanks to Dr Shahin Ashkiani and Dr Ozren Despić for the collaboration and their understanding regarding the balance of my teaching and research duties.

I would also like to thank my *alma mater*: the Department of Mathematics of the Aristotle University of Thessaloniki. Especially, I would like to thank Prof. Georgios Tsaklidis for his continuous guidance and for enlightening me the first glance of research and research ethics. I would also like to thank Associate Prof. Alexandra Papadopoulou for bringing to me this PhD opportunity and express my deep gratitude for her multilateral and steady support and guidance.

A special thanks to my fellow PhD students, whom I shared the working space with and countless unforgettable moments inside and outside the university; their support was priceless.

Last but not least, I would like to express my deep thanks and gratitude to my family and friends. With or without distance, their encouragement and support during all these years was the most valuable company in this long journey.

Contents

List of Figures

List of Tables

1

Introduction

Introduction Measuring the performance of firms and organisations has become an integrated part of the management process in the industry. Performance measurement was defined by (Moullin 2002, pg. 188) as the process of "evaluating how well organisations are managed and the value they deliver for customers and other stakeholders". In other words, it helps managers to assess whether the decisions made are effective and how far an organisation or firm is from achieving its goals and objectives.

Among the different performance measures, this thesis focuses on the measurement of efficiency. Data Envelopment Analysis (DEA) is a data-driven approach that was introduced by Charnes et al. (1978) to measure the relative efficiency of organisations that consume inputs to produce outputs. In DEA there is no functional assumption about the relation between the inputs and outputs. This allows DEA to be applied to the efficiency measurement of entities that consume multiple inputs to produce multiple outputs. In such multidimensional cases, the estimation of a production function that captures well the input-output relationship would probably be a complicated task. In DEA, the efficient frontier is shaped by the best performing entities in the sample, and inefficient units use the efficient ones as benchmarks to improve their activity. The efficiency score of each entity is measured as the proportional decrease that a unit needs to apply in its input consumption without decreasing the output production or the proportional expansion that needs to achieve in its output production without increasing the input consumption until it reaches the best practice frontier.

Apart from the input-output activities of the observed Decision Making Units (DMUs), the shape of the efficient frontier also depends on some economic assumptions that are made about the production process of DMUs. Charnes et al. (1978) suggested the first DEA model under the assumption that if a DMU increases its inputs this would lead to an equivalent increase in the output production, i.e. it is assumed that DMUs operate under constant returns to scale (CRS). However, in real market conditions, it is very common for DMUs to experience economies or diseconomies of scale. For example, buying materials in larger quantities usually guarantees a cost-per-unit reduction. That means that firms operating at a larger scale are able to buy inputs at larger amounts and therefore can increase their production at a reduced cost. On the other hand, in large-scale operating companies, there might be some productivity loss due to difficulties in managing a large workforce. To also capture economies or diseconomies of scale, Banker et al. (1984) introduced a DEA model under the variable returns to scale (VRS) assumption. Later studies also extended the main DEA models by making different assumptions about the production frontier and/or using different metrics to measure the DMUs' inefficiency. The main concepts of DEA as well as its main developments are outlined by Thanassoulis (2001) and Cooper et al. (2011), whereas Ray (2004) discusses the economic rationale of the DEA models, based on the neoclassical model of production.

1.1 Motivation and research framework

Due to its non-parametric nature, the DEA methodology has been implemented to measure the efficiency of DMUs in a wide range of fields and industries, such as banking, education, hotel industry, health care, agriculture, supply chain, etc. A collection of DEA applications in different sectors can be found in Zhu (2016). Within the globally increasing concern regarding energy consumption and climate change, during the last two years, transportation, energy and the environment are among the most common areas in which DEA models are implemented (Emrouznejad, Yang, Khoveyni & Michali 2022, pg. 346). In this Thesis, the environmental efficiency of railway transport in Europe is studied, in terms of how noise pollution affects humans and the measures that countries take to eliminate this issue, considering the inner structure of the railway transport process. To take into account sampling noise, and get more realistic efficiency score estimations, a statistical framework is established for the cases when the production process of DMUs has a two-stage series structure, and the performance of bootstrapping in such structures is examined.

The transportation sector, which is an important sector for achieving economic growth, is responsible for a significant percentage of the total energy consumption. According to the European Environmental Agency (EEA), in 2007 the energy consumption from the transportation industry in the EEA-33[1] was 38% higher than in 1990. In 2012, the share of the transportation sector in energy consumption in the EU-28 was 32% (Ntovantzi et al. 2015). This small decrease that was observed in the energy consumption from this sector was only temporary and was due to the economic crisis (EEA 2019*a*). Furthermore, during the same period, greenhouse gas emissions increased by 36% in this sector, whereas other sectors decreased their emissions by 15%, and the transportation sector is responsible for about 25% of the greenhouse gas emissions in the EU-28 (Ntovantzi et al. 2015).

Despite the aforementioned environmental issues of the transportation sector that arise from energy consumption and greenhouse gas emissions, the railway industry seems to be the most environmentally friendly among the different transportation industries. Railways are responsible for a very low percentage of energy consumption-only 0.3% of the total transport energy, and are the least carbon-intensive mode of transport - they are responsible only for 0.5% of the transportation greenhouse gas emissions in 2017 in the EEA-33 (EEA 2019*b*). At the same time, railways are also more efficient than other means of transport in terms of safety, traffic congestion or land consumption. Therefore, it is crucial for the environment, the economy and people to further investigate the inefficiencies in the railway transport process, so its operations get improved and further expanded.

The main environmental impact of railways is noise pollution. Apart from other negative effects of railways on the natural habitat, such as collisions with animals, or soil and water pollution caused by the herbicides used to maintain the rail lines, crossing trains cause wheel and rail vibrations, and at high speeds, they also produce aerodynamic noise. The older braking technology that is currently still being used in freight wagons (LL-type composite brake blocks) also plays a major role in the sound levels produced (Pyrgidis 2016, 428-429). This distorts both the wildlife and humans living close to the railway lines. In this thesis, only the impact of railway noise on humans is considered. Especially in Europe, railway transport is the second largest source of noise pollution after road transport, and about 22 million people every year are estimated to be exposed to high levels of railway noise that can be harmful to their health (EEA Report No 22/2019 2020).

Country-level measurements that are now reported by the EEA every five years re-

[1]EEA-33 includes EU-28 plus Switzerland, Iceland, Norway, Liechtenstein and Turkey

veal that railway noise levels inside and outside urban areas are above the safety levels defined by the World Health Organisation (WHO) (EEA 2020). Exposure to high noise levels for a long period can cause health problems such as chronic annoyance, stress or insomnia or increase the risk of heart attacks (EEA Technical Report No 11/2010 2010). Recognising the severe environmental impact of railway noise pollution, the European Commission, in 2006, set noise emission limits for the new rail wagons. Members of the Union Internationale des Chemins de Fer (UIC) (International Union of Railways) and the Community of European Railway and Infrastructure Companies (CER) have agreed on noise reduction plans and goals to eliminate railway noise-pollution by 2030 and 2050 (UIC & CER 2012).

In order for the countries to be able to comply with these goals and stay within the noise-emission limits set by the European Commission, a necessary measure that needs to be taken is to retrofit their wagon fleet with more silent braking technology. Therefore, to measure the environmental efficiency of railway transport in a country both its negative impact on people as well as measures to mitigate that impact need to be considered. In this study, the negative impact of railways on people is measured as the number of people in each country that are exposed to high levels of railway noise. The number of rail wagons in each country that is retrofitted with more silent braking technology is used as a proxy to measure the effort that each country makes to reduce railway noise pollution.

Conventional DEA only considers the initial inputs that enter the production and the final outputs produced and ignores any inner structure of the production process. Nevertheless, in many applications, it is more useful to study the efficiency of the production process capturing the different stages and any intermediate products that are involved in the production of the final outputs. The railway transport process for example, similarly to other production processes, involves different stages; To give the final outputs, which are passenger and freight carrying services, in such a capital-intensive industry, an initial stage of building a good network of rail lines and acquiring the necessary number of wagons at a not only affordable but at an also competitive cost is the first and critical step for achieving efficiency. Therefore both stages of the railway transport process need to be considered in order to get a better insight into the strengths and weaknesses of its operation and set goals for improvement.

The first DEA studies that considered the different stages of a production process were in mid-1980's (Färe & Primont 1984; Charnes et al. 1986). Färe & Grosskopf (2000) officially introduced the term Network DEA to describe the DEA models that are developed to deal with efficiency measurement when more than one sub-processes are considered. The different sub-processes can be arranged in series, in parallel, or series-parallel structures and intermediate products from one stage may become inputs to another stage. There is a high volume of literature on Network DEA, concerning the calculation of the efficiency scores of the sub-processes. This study focuses on two-stage series structures.

Various algorithms have been developed to obtain the efficiency score of the overall and the stage processes. In some approaches, the efficiency of each stage is calculated independently and the linkage between the stages is not taken into account (Seiford & Zhu 1999; Zhu 2000). In the network approaches (Färe 1991; Färe & Grosskopf 1996), the inter-connection between the stages is considered in the calculation of the overall process efficiency, but no measure of the stage efficiencies is provided.

The most broadly used approaches in the DEA literature are the efficiency decomposition approaches. Two main efficiency decomposition approaches are used in the NDEA literature; the multiplicative and the additive one. Kao & Hwang (2008) suggested decomposing the overall efficiency as the product of the two-stage efficiencies. They linked the two stages assuming that the aggregated outputs of the first stage are introduced unchanged in the second stage. However, the multiplicative decomposition approach can only be applied under the CRS assumption, as under the VRS, the resulting models cannot be linearised. Another limitation of the conventional multiplicative approach is that it can be generalised to multi-stage series structures only in the cases when there are no stage-specific inputs and outputs. For general network structures, alternative multiplicative approaches have been developed, such as converting the original model to a parametric linear one (see for example Zha & Liang (2010)). Chen et al. (2009) decomposed the overall efficiency of a DMU as the weighted average of the stage efficiencies. They defined the decomposition weights endogenously, so they can reflect the relative contribution of each stage to the overall process. The additive decomposition approach has the advantage that the resulting models can be linearised both under the CRS and the VRS assumptions. Furthermore, although under the CRS assumption, the additive decomposition approach favours one stage against the other by assigning a higher value to the decomposition weight of a stage, under the VRS assumption there is not such an inherent bias in the optimisation process.

Conventional DEA is a deterministic approach that does not consider any sort of noise, such as sampling noise, measurement and specification errors. DEA models have been extended in different directions to account for such deficiencies. Regarding sampling noise, in conventional DEA, the efficient frontier is defined empirically based on the best-performing DMUs included in the observed sample. However, alterations in the observed sample (inclusion or omission of DMUs) may affect the shape of the efficient frontier and the efficiency scores of DMUs. For example, when measuring the efficiency of different supermarkets in a city, there is a chance that not all the supermarkets of the city are included in the sample, or a new supermarket is going to open soon, or we may need to consider efficiency measurement at a country-level.

In practice, the observed set of DMUs is just a random sample drawn from an underlying population, with an unknown efficient frontier. DEA can only provide an estimate of the true efficiency score. Kneip et al. (1998) derived the rate of convergence of the DEA estimator, i.e. the rate at which the estimation error decreases as the size of the observed sample increases, under the VRS assumption. For the single input-single output case, Gijbels et al. (1999) derived the analytical form of the distribution of the efficiency differences between the sample frontier and the true unknown frontier. However, in higher dimensions, the closed expression of this distribution cannot be obtained, and the only way to approximate it is through bootstrapping techniques.

The idea in bootstrapping is that the observed sample which is drawn from an unknown population through an unknown data generating process (DGP) mimics the population that it comes from. Therefore if a bootstrap sample is drawn from the original sample through a known DGP that is a consistent estimator of the unknown DGP, then it will be like drawing a sample from the population itself. By repeating this process several times, many bootstrap samples are generated and can be used to make statistical inference.

The DGP, i.e. the bootstrapping approach, that is used to generate the bootstrap sample must mimic well the true unknown DGP. Naïve bootstrapping which consists of drawing with replacement a bootstrap sample of equal size as the original sample provides inconsistent estimates of the true efficiency scores. This happens because with naïve bootstrapping, asymptotically, the probability that the bootstrap frontier coincides with the sample frontier is positive, whereas the probability that the sample frontier coincides with the true unknown frontier is zero. Different bootstrap approaches based on smoothing or subsampling, and under different assumptions have been proven to provide consistent estimates of the true efficiency scores.

The consistency of the DEA estimator obtained through different bootstrapping techniques based on smoothing or subsampling, and under different assumptions has been broadly studied. The homogeneous smooth bootstrap approach (Simar & Wilson 1998) requires the restrictive assumption that the distribution of inefficiencies is the same for all DMUs. This assumption implies that all DMUs have a positive probability of reaching the unknown efficient frontier. As it was discussed by Olesen & Petersen (2016) this is not always the case. In the heterogeneous smooth bootstrap introduced by Simar & Wilson (2000), the homogeneity assumption is relaxed, however, in practice, it is very difficult to implement as it requires a large amount of data to be available to estimate a DMU-specific inefficiency distribution. On the other hand, subsampling bootstrap does not require any homogeneity assumption and is computationally easier than the heterogeneous smooth bootstrap.

Numerous applications implement bootstrapping to get confidence interval estimates for the efficiency scores of DMUs. However, in these studies, the transformation of inputs into outputs is considered to occur at a single stage and any inner structure of the production process is ignored. There are very limited attempts to make statistical inference in the cases when the production process has a network structure, and there is no study that provides confidence interval estimates for the efficiency scores of the overall and the stage processes considering the inter-connection between the stages.

Therefore, in this thesis, we extend the application of the bootstrapping methodology in the cases when the production process involves two stages arranged in series. Subsampling bootstrap is being used as the most appropriate one, since it allows for heterogeneity in the inefficiency distribution of DMUs, without requiring the computational burden of the heterogeneous smooth bootstrap. The efficiency estimates for the overall and stage efficiency scores are obtained using the additive decomposition methodology since this can be applied under the VRS assumption and is computationally easier than the efficiency composition methodologies. A practical approach is suggested to overcome some peculiar results obtained from the additive decomposition algorithm.

First, Monte Carlo simulations are run to assess the performance of subsampling bootstrap in different two-stage series structures. Then, the methodology is applied to construct confidence interval estimates about the efficiency scores of European railways. Confidence interval estimates are particularly insightful in this case; the additive decomposition approach yields many efficient DMUs in the second stage, but the lower bounds of the confidence interval estimates show that in reality this is not the case, and provide the possible range of the true inefficiency of railway services in those countries.

1.2 Research aims and objectives

Despite the great number of studies that assess the efficiency of railways, there is a limited number of studies that consider the inner structure of the railway transport process. Additionally, even though noise pollution seems to be a major aftereffect of railway operations under the current train technology, no study incorporates this issue in the efficiency measurement.

Furthermore, although bootstrapping techniques have been widely applied in one-stage production structures to make estimations about the true efficiency scores of DMUs, and consider sampling noise, there are very limited attempts to make statistical inference in the cases when the production process has a network structure, and there is no study that provides confidence interval estimates for the efficiency scores of the overall and the stage processes considering the inter-connection between the stages.

Therefore, the aim of this research is on the one hand to assess the efficiency of European railways considering its inner structure and the major environmental problem of railway noise pollution, and on the other hand, to set a statistical framework for Network DEA to consider for sampling noise, while taking into account the relations between the different stages.

Therefore, the objectives of this thesis are formulated as follows:

- Unveil the sources of inefficiency in the railway transport process in the European countries
- Investigate the impact of railway noise pollution on humans
- Provide a deep discussion on the decomposition weights in the additive decomposition approach
- Assess the performance of subsampling bootstrap in two-stage series structures through Monte Carlo simulations
- Investigate the impact of the sample and subsample size in the coverage probabilities
- Provide suggestions on how to deal with the peculiarities of the additive decomposition algorithm in the bootstrap context

1.3 Structure of the thesis

The structure of this thesis is given below.

In Chapter 2, the main concepts and assumptions in DEA are discussed. The main DEA models, i.e. the CCR, the BCC, the SBM and the additive model are formulated, and the underlying assumptions behind each model are discussed and compared.

Chapter 3, is a review of the most influential approaches to efficiency measurement when the inner structure of the production process is considered. More specifically, this chapter focuses on the cases when the production process has a two-stage series structure. The available efficiency measurement approaches for such structures are categorised into five different general categories. In the independent approach, the efficiency of each stage is calculated independently using the standard DEA models and interdependencies between the stages are not considered. Therefore, efficiency decomposition and efficiency composition approaches, as well as slacks-based measure approaches, and the network approaches were developed to jointly assess the efficiency of the overall and the stage processes. The strengths and weaknesses of each approach are discussed.

In Chapter 4, the additive decomposition approach is used to assess the efficiency of European railways. First, the sources, as well as the impact of the railway noise pollution, are discussed, and noise emission limits and goals set by the European Commission are given. The railway transport process is divided into two stages, related to assets and service provision, respectively. The negative impact of railways on people is measured as the number of people in each country that are exposed to high levels of railway noise. The number of rail wagons in each country that is retrofitted with more silent braking technology is used as a proxy to measure the effort that each country makes to reduce railway noise pollution. Data is extracted from Eurostat (2016), ERA 006REC1072 Impact Assessment (2018), and EEA (2020). Sensitivity analysis is performed in order to choose the decomposition weight restrictions that will prevent one of the two stages to be assigned zero contribution to the overall process, and at the same time, will not significantly affect the efficiency scores. Next, based on the findings, policy suggestions for the countries are given, and finally, some limitations and future research directions are provided. This Chapter has been published in the *Transportation Research Part D: Transport and Environment* (Michali et al. 2021).

In Chapter 5, the bootstrap approaches used in the DEA context are discussed. First, the main statistical concepts and definitions that will be used are given, and then the bootstrap idea is explained. In practice, bootstrapping is the only way to account for sampling noise, and obtain estimations about the true efficiency scores of DMUs. A necessary condition to get good approximations of the true efficiency scores is that the bootstrap is a consistent estimator of the true unknown data generating process. Naïve bootstrap, homogeneous and non-homogeneous smooth bootstrap and subsampling bootstrap within the DEA context are discussed, and consistency issues and computational difficulties for each approach are explained in detail.

In Chapter 6, the statistical framework is extended to Network DEA. The frontier model is defined for a two-stage general series structure. Among the different bootstrap approaches, subsampling bootstrap was chosen as the most appropriate since -as it is discussed in Chapter 5 - it does not require any restrictive assumptions and is computationally easier compared to the heterogeneous smooth bootstrap. The general two-stage series structure is studied in the cases of a five and a seven-dimensional structure. The performance of subsampling is examined through Monte Carlo simulations, in which the true probabilities that the true efficiency scores lie in the estimated confidence intervals were calculated, for various sample and subsample sizes. The use of the algorithm suggested by Simar & Wilson (2010) for the choice of the optimal subsample size is then examined, and some practical considerations of its applicability are provided. Then, a two-step approach is suggested to overcome some result inconsistencies that are due to the peculiarities of the additive decomposition algorithm. Finally, the suggested methodology is applied to obtain confidence interval estimates for the overall and stage efficiency scores of the European railways. This Chapter, is based on an article published in the European Journal of Operational Research (Michali et al. 2022).

Finally, in Chapter 7, conclusions and limitations of this research, as well as future research directions are provided.

2

Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric technique based on linear programming (LP), which is used to measure the comparative efficiency of a homogeneous set of decision making units (DMUs) and provide targets for achieving efficiency improvement. Such a set of DMUs can be the branches of a bank, or a supermarket, hotels, airports, utility companies, etc., which consume multiple resources to produce multiple outputs.

In DEA, the performance of each unit is evaluated by measuring its distance from an empirically constructed efficient frontier; among the observed set of DMUs, those with the best performance, i.e. those with the maximum attainable output level produced by a given input level, are used to define the efficient frontier. The less well-performing DMUs will be enveloped by the best practice frontier and use it as a benchmark against which they will compare their performance. In contrast to the econometric approaches where a production function needs to be estimated, the great advantage of DEA is that no functional form is needed to describe the transformation of inputs into outputs since the efficient production frontier is empirically constructed. This allows the inclusion of multiple inputs/outputs in the production model without the risk of increasing the estimation error.

The origins of DEA lie on the seminal works of Farrell (1957), Koopmans (1951) and Debreu (1951). Farrell (1957) approximated the production function as a convex, piecewise linear frontier obtained from the observed input-output data, solving a LP. Based on the activity analysis of Koopmans (1951) and Debreu (1951)'s "coefficient of resource utilisation", Farrell also defined a relative, radial input-based index to assess the technical efficiency of a firm, measuring the proportional reduction of inputs needed, given the production of a single output, to reach the best practice frontier.

The concept of technical efficiency was firstly introduced by Koopmans (1951) who defined the feasible activity of a firm as efficient if an increase in the production of any of its outputs (decrease in the consumption of any of its inputs) is only possible only by decreasing the production of another output (increasing the consumption of another input). Koopmans's definition of technical efficiency is similar to the Pareto optimality and is therefore known as Pareto-Koopmans efficiency. Farrell's technical efficiency differs from Koopmans's as the first one is a radial measure of technical efficiency and therefore, some input excesses and/or output shortfalls may remain after a unit becomes efficient. Farrell (1957) noted the difficulty in generalising the defined index to the multiple input-multiple output setting.

About two decades later, Charnes et al. (1978) introduced the first DEA model in the form of a fractional program, by generalising Farrell's measure to the multiple input-output setting; efficiency is measured as the ratio of the weighted sum of inputs to the weighted sum of outputs, with the weights being assigned by the optimisation process. Banker et al. (1984) extended Charnes et al. (1978) work to incorporate different assumptions about the production technology under which units operate. Since then, a considerable number of studies have focused both on the theoretical development of DEA and its applications. According to Emrouznejad, Yang, Khoveyni & Michali (2022), the rapid growth of the number of DEA publications started in the mid-2000s and during the last decade, more than 1000 DEA-related articles are published each year. A full list of publications on DEA since 1978 is provided by Emrouznejad & Yang (2018).

The aim of this chapter is to discuss the main concepts needed to set the DEA framework. In Section 2.1, the production model is defined and its properties are given. In Section 2.2 efficiency measurement in DEA is discussed and a simple example is provided. Section 2.3 reviews the four basic DEA models. In Section 2.4, the curse of dimensionality in DEA is briefly discussed, and finally, in Section 2.5 the main conclusions of the Chapter are provided.

2.1 The production model

In DEA, a decision making unit (DMU) is a unit that transforms multiple inputs into multiple outputs (see Figure 2.1). The main idea behind DEA is to assess the efficiency of a DMU in transforming its inputs into outputs compared to other DMUs going through the same transformation process. The term DMU was introduced by Charnes et al. (1978) to emphasise, on the one hand, the control that a unit has over its consumption and production quantities and, on the other hand, that the developed methodology "is centred on decision making by not-for-profit entities rather than the more customary 'firms' and 'industries'" (Charnes et al. 1978, pg. 429).

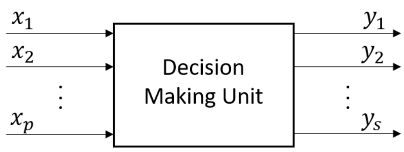


Figure 2.1: A DMU representation

Suppose there is a set of N homogeneous DMUs, each consuming P inputs to produce S outputs. The set needs to be homogeneous in the sense that all the DMUs perform the same tasks, i.e. they consume the same type of inputs and produce similar outputs, but in varying quantities. Let $x_j = (x_{1j}, ..., x_{Pj}) \in \mathbb{R}^P_+$ and $y_j = (y_{1j}, ..., y_{Sj}) \in \mathbb{R}^S_+$ denote the non-negative input and output vectors of DMU_j, $j = 1, 2, ..., N$. It is supposed that all data are non-negative, but each vector has at least one positive element. Therefore, each DMU has at least one positive input and output. The pair $(x, y) \in \mathbb{R}^{P+S}_+$ is called an activity vector. An input-output correspondence is a feasible activity if the specific output can be produced from that input quantity. The set that includes all the feasible input-output activities is called the production possibility set (PPS) and is defined as

$$T = \{(x, y) \in \mathbb{R}^{P+S}_+ | x \text{ can produce } y\}.$$

The boundary of the PPS comprises the efficient frontier and is defined empirically upon the observed set of DMUs and some assumptions that are made about the returns to scale (RTS) features of the production technology under which the DMUs operate. The RTS refers to how a proportionate increase in the inputs of a unit affects the amounts of the outputs produced. If such an increase in the inputs of a unit results in the amount of outputs being increased by the same proportion, the production exhibits constant returns to scale (CRS). When the change in the outputs is not proportionate to the change in the inputs, the DMU operates under the variable returns to scale (VRS). If the outputs are increased by a higher proportion, the units operate under increasing returns to scale (IRS). Such economies of scale occur when the efficiency of a unit increases as the unit increases the scale size of its operations. The production process exhibits decreasing returns to scale (DRS) when the proportional increase in outputs is less than the input increase, i.e. a unit becomes less efficient as it moves to a larger scale.

More specifically, the PPS has the following properties (see Banker, 1984):

(i) (Inclusion of observations) $(x_j, y_j) \in T \; \forall \; \mathrm{DMU}_j, \; j = 1, ..., N$.

(ii) (Strong disposability) If $(x, y) \in T$ and $x' \geq x, \; y' \leq y,$ where $x, x' \in \mathbb{R}^m, y, y' \in \mathbb{R}^r$, then $(x', y') \in T$.

(iii) (Convexity) If $(x_j, y_j) \in T, j = 1, ..., N$, then, for any vector of non-negative scalars λ_j such that $\sum_{j=1}^{N} \lambda_j = 1 \Rightarrow (\sum_{j=1}^{N} \lambda_j x_j, \sum_{j=1}^{N} \lambda_j y_j) \in T$.

(iv) (Minimum Extrapolation) If T_1 satisfies (i)-(iv) then $T \subseteq T_1$.

Under the CRS assumption, the PPS also possesses the following property:

(v) If $(x, y) \in T \Rightarrow (tx, ty) \in T$ for any positive scalar t.

Based on the observed input-output data and the properties (i)-(v) assumed, the PPS under all the different production technologies is formulated as follows:

$$T = \big\{(x, y) \in \mathbb{R}_+^{P+S} | \sum_{j=1}^{N} \lambda_j x_j \leq x, \sum_{j=1}^{N} \lambda_j y_j \geq y, \lambda \in \Lambda\big\}, \tag{2.1}$$

where Λ is defined depending on the RTS assumptions; under the CRS assumption, $\Lambda = \mathbb{R}_+^N$, when the VRS is assumed, $\Lambda = \{\mathbb{R}_+^N | \sum_{j=1}^{N} \lambda_j = 1\}$, $\Lambda = \{\mathbb{R}_+^N | \sum_{j=1}^{N} \lambda_j \leq 1\}$, under the IRS assumption, and $\Lambda = \{\mathbb{R}_+^N | \sum_{j=1}^{N} \lambda_j \geq 1\}$ under the DRS assumption.

The most common assumptions made in DEA are the CRS and the VRS. Figure 2.2 illustrates the PPSs under the CRS and VRS assumptions, respectively, based on a set of eight DMUs which consume a single input and produce a single output (Table 2.1).

Table 2.1: One input and one output case

DMU	A	B	C	D	E	F	G	H
Input	2	7	6	5	3	8	6	4
Output	1	3	7	5	4	5	4	3

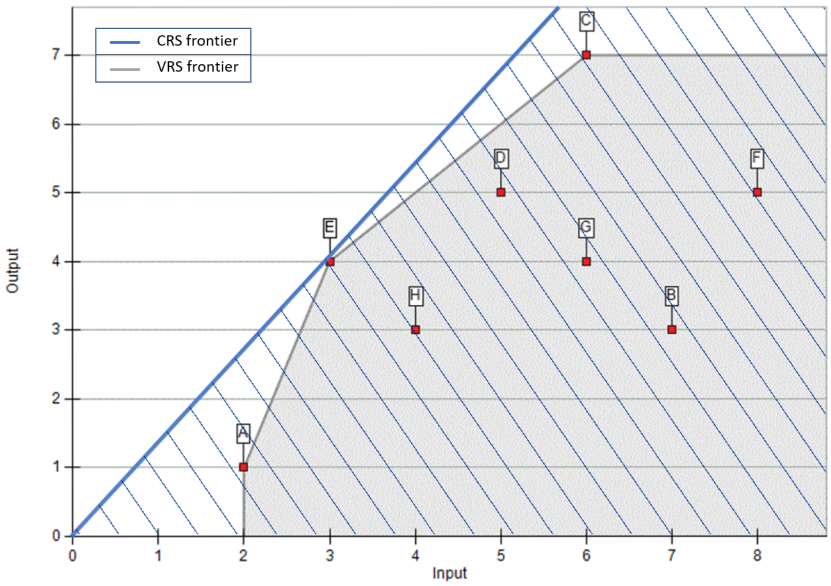


Figure 2.2: PPSs under the CRS and the VRS assumptions

In Figure 2.2, the lined area represents the PPS under the CRS assumption, whereas the grey shaded area represents the PPS when the VRS is assumed. The boundary of the PPS under the CRS assumption is estimated by the blue ray starting from the origin and passing through DMU E and is the best practice frontier under the CRS assumption. Under the VRS assumption, the best practice frontier is assumed to be given by the grey, piecewise linear curve defined by DMUs A, E and C. In each case, the inefficient DMUs are enveloped by the best practice frontier. Although DMUs A and C are on the efficient frontier under the VRS assumption, they are inefficient compared to the CRS frontier. Therefore, even from this trivial example it can be seen that in DEA, the efficiency of DMUs is a relative measure that depends on the RTS assumptions and the activity of the DMUs included in the observed sample. Furthermore, the VRS frontier is ”closer” to inefficient units and, therefore, suggests a more favourable efficiency measurement.

2.2 Efficiency measurement in DEA

Efficiency measurement in DEA is relative rather than absolute. Comparing the amount of outputs produced by a DMU, given the input quantities, with the benchmark output amounts, can give a measure of technical efficiency for a DMU. The objective of each DMU is twofold: a DMU aims to increase its output production to a maximum possible level, given that it continues to consume the same amount of input resources, or, to minimise its input consumption while producing the same output quantities. Therefore, efficiency measurement can be input or output oriented. Then, input technical efficiency is measured as the maximum feasible radial contraction of a DMUs' inputs (i.e. the input mix is maintained) given that the production level of outputs is not reduced. Output technical efficiency can be defined similarly, as the maximum feasible radial expansion of outputs without increasing its input levels.

Technical efficiency can be defined in terms of the PPS and the input requirement and output correspondence sets that fully characterise the PPS. Consider a set of N DMUs, and let $X(y)$ be the input requirement set which consists of all the input vectors $x \in \mathbb{R}^P_+$ that can produce $y \in \mathbb{R}^S_+$, and similarly, let $Y(x)$ be the output correspondence set containing all the output vectors $y \in \mathbb{R}^S_+$ which can be produced by the input vector $x \in \mathbb{R}^P_+$. Then,

$$X(y) = \{x \in \mathbb{R}^P_+ | (x, y) \in T\}, \tag{2.2}$$

$$Y(x) = \{y \in \mathbb{R}^S_+ | (x, y) \in T\}. \tag{2.3}$$

The radial (or Farrell's) input and output technical efficiency measures can now be defined as follows:

Definition 2.2.1. *The Farrell's input and output technical efficiencies, respectively, are defined as*

$$\theta^* = inf\{\theta | \theta x \in X(y)\}, \tag{2.4}$$

$$\phi^* = sup\{\theta | \theta y \in Y(x)\}. \tag{2.5}$$

Consider the same set of eight DMUs given in the previous section, where each DMU consumes a single input and produces a single output and let us consider the CRS case first. Under the CRS assumption, the maximum radial input contraction (output expansion) given the current output (input) levels is represented by the ray starting from the origin and passing through the point E, i.e. the CCR-efficient boundary of the PPS. Therefore, in the CRS case, all the inefficient DMUs will use DMU E as a benchmark to improve their efficiency. The technical efficiency (TE) of a DMU$_{j0}$ belonging to the sample, can be calculated relatively to the best performing DMU as the following ratio of ratios, which is always bounded on both sides:

$$0 \leq \frac{\text{Output of DMU}_{j0} \text{ per input of DMU}_{j0}}{\text{Output of the efficient DMU per input of the efficient DMU}} \leq 1. \tag{2.6}$$

Figure 2.3 illustrates the input and output technical efficiencies of the inefficient DMU H, under the CRS and the VRS assumption, respectively.

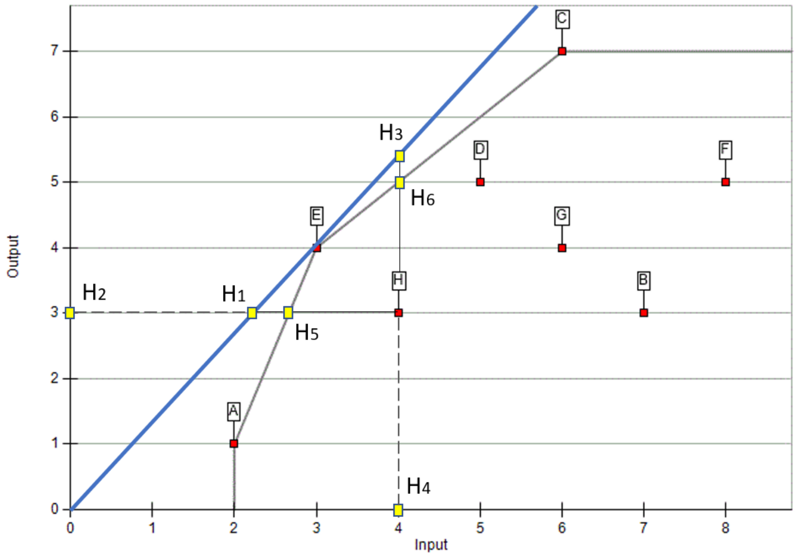


Figure 2.3: CRS and VRS technical efficiency of DMU H

The inefficient DMU H for example, could become technically efficient by minimising its input level to operate at point H_1, or maximise its output production to operate at point H_3. The output technical efficiency is given as the proportion that the DMU's observed output is of the maximum output it could produce, given it consumes the same input quantities, i.e. $1/\phi_H^{CRS*} = TE_{H,O}^{CRS} = \frac{H_4H}{H_4H_3} = 0.56$. The input technical efficiency of DMU H can be defined in a similar way as $\theta_H^{CRS*} = TE_{H,I}^{CRS} = \frac{H_2H_1}{H_2H} = 0.56$, where H_2H_1 is the minimum input level that DMU H could consume to produce the same output level, and H_2H is the actual input level of DMU H. It can be seen that under the CRS assumption the input and output orientation yield the same technical efficiency score.

Under the VRS assumption, the locus of the points with the maximum output production per unit of input is approximated by a piecewise linear curve. In this case, an inefficient DMU can have more than one efficient DMUs as benchmarks, depending on which line segment of the locus approximation its radial projection lies. Therefore, in order to define the technical efficiency of a DMU using the ratio 2.6, we need to define a "virtual", efficient DMU, by taking the linear interpolation between the DMU's benchmarks, i.e. the virtual DMU lies on the linear segment defined by the benchmark DMUs.

In this example, under the VRS assumption, the efficient operation of DMU H is given by the points H_5 or H_6, which are the projections of point H to the VRS frontier, in the input and output orientation, respectively. Therefore, in the output orientation, DMU H uses DMUs E and C as benchmarks, and its VRS output technical efficiency is given as $1/\phi_H^{VRS*} = TE_{H,O}^{VRS} = \frac{H_4H}{H_4H_6} = 0.6$. In the input orientation, DMUs A and E become the benchmarks and DMU H's input technical efficiency is $\theta_H^{VRS*} = TE_{H,I}^{VRS} = \frac{H_2H_5}{H_2H} = 0.67$. Unlike the CRS case, under the VRS assumption the input and output technical efficiency scores can be different. The CRS and VRS technical efficiencies are called global and pure technical efficiencies, respectively. Since the CRS frontier always envelops the VRS frontier, we conclude to the following corollary:

Corollary 2.2.1. *Given a set S_N of N DMUs, for a DMU$_{j_0} \in S_N$, it always holds that $TE_{j_0}^{CRS} \leq TE_{j_0}^{VRS}$, in both orientations.*

Therefore, the scale size of a DMU's operation has an impact on its performance. The divergence between the CRS and the VRS technical efficiency of a DMU is defined as its scale efficiency. Banker et al. (1984) showed that the CRS technical efficiency can be decomposed to the product of VRS technical efficiency and scale efficiency, i.e. for a DMU$_{j_0}$

$$SE_{j_0} = \frac{TE_{j_0}^{CRS}}{TE_{j_0}^{VRS}}, \tag{2.7}$$

where $TE_{j_0}^{VRS}$ is measured either in the input or in the output orientation.

An efficient DMU operating at IRS or DRS can maximise its average productivity if it changes its scale size to CRS, as the local CRS is the most productive scale size (MPSS) for a DMU to operate. The MPSS is defined as follows:

Definition 2.2.2. *(Banker & Thrall 1992) A DMUj_0 operates at the MPSS iff $TE_{j_0}^{CRS} = 1$ and $SE_{j_0} = 1$.*

Therefore, a CRS-efficient DMU is operating at the MPSS.

Table 2.2 depicts the output/input ratio, the CRS and VRS technical efficiencies and the input and output scale efficiencies for the eight DMUs included in the foregoing single input and output example.

Table 2.2: Efficiency measurement in the single input- single output case

DMU	A	B	C	D	E	F	G	H
Input	2	7	6	5	3	8	6	4
Output	1	3	7	5	4	5	4	3
Output/Input	0.5	0.429	1.167	1	1.333	0.625	0.667	0.75
TE^{CRS}	0.375	0.321	0.875	0.75	1	0.469	0.5	0.563
TE_I^{VRS}	1	0.381	1	0.8	1	0.5	0.5	0.667
TE_O^{VRS}	1	0.429	1	0.833	1	0.714	0.571	0.6
SE_I	0.375	0.842	0.875	0.938	1	0.938	1	0.844
SE_O	0.375	0.748	0.875	0.900	1	0.657	0.876	0.938

Farrell's technical efficiency discussed above, requires the radial projection of DMUs on the efficient frontier. However, a DMU radially projected on the efficient frontier may still be able to further improve its activity by further reducing (increasing) its inputs (outputs) without affecting the output (input) levels, because of the existence of input excesses (output shortfalls). Therefore, in DEA Farrell's technical efficiency is often related to weak efficiency. Pareto-Koopmans technical efficiency considers a DMU as strongly efficient if and only if it is not possible to improve any of its inputs or outputs without worsening at least one of its other inputs and/or outputs. Therefore, weak and strong efficiency can be defined as follows:

Definition 2.2.3. *Let $(x, y) \in P$ be the activity vector of DMU_j, $j = 1, 2, ..., N$. DMU_j is strongly (weakly) efficient if and only if there is no other feasible activity $(x', y') \neq (x, y)$, such that $x' \leq x$ $(x' < x)$ and $y' \geq y$ $(y' > y)$.*

2.3 The basic DEA models

In this section, the four basic DEA models, namely the CCR, the BCC, the additive and the SBM model, will be outlined. The first two models use radial projections of DMUs to the efficient boundary, in the sense that they require proportionate reduction (increase) of inputs (outputs), whereas the additive and the SBM models allow for non-radial projections, i.e. changes to a DMU's resources can be non-proportional.

2.3.1 The CCR model

The CCR model was introduced by Charnes, Cooper & Rhodes (1978) to measure the relative, radial efficiency of DMUs under the CRS assumption. In this section, models will be explained based on the input-orientation, and then they will be briefly extended to the output orientation.

Consider a set of N DMUs, each consuming P inputs $x_j = (x_{1j}, ..., x_{Pj}) \in \mathbb{R}^P_+$ and producing S outputs $y_j = (y_{1j}, ..., y_{Sj}) \in \mathbb{R}^S_+$. For each DMU$_{j_0}$ belonging to the sample, a virtual DMU is constructed by the inputs and outputs of DMU$_{j_0}$, with (virtual input)$_{j_0}$ = $\nu_1 x_{10} + ... + \nu_P x_{P0}$ and (virtual output)$_{j_0} = w_1 y_{10} + ... + w_S y_{S0}$. The weights ν_p and w_s are unknown and will be determined so they maximise the ratio of the virtual output to the virtual input. Charnes et al. (1978) suggested the following model for obtaining the optimal weights and the efficiency score of each DMU:

$$
\begin{aligned}
\max\ \theta_{j_0} &= \frac{\sum_{s=1}^{S} w_s y_{sj_0}}{\sum_{p=1}^{P} \nu_p x_{pj_0}} \\
\text{s.t.}\ & \frac{\sum_{s=1}^{S} w_s y_{sj}}{\sum_{p=1}^{P} \nu_p x_{pj}} \leq 1, \quad j = 1, ..., N, \\
& \nu_p, w_s > 0, p = 1, ..., P, s = 1, ..., S.
\end{aligned}
\tag{2.8}
$$

Model (2.8) needs to be solved N times, one for each DMU in the sample, and the constraints ensure that the optimal objective value $\theta^*_{j_0}$ of θ_{j_0} for a DMU under evaluation does not exceed 1.

The CCR model can be converted from a fractional to a linear form by applying the Charnes-Cooper (C-C) transformation (Charnes & Cooper 1962). Let a scalar $t \in \mathbb{R}_+$ such that $t \sum_{p=1}^{P} \nu_p x_{pj_0} = 1$. Multiplying the nominators and denominators in model (2.8) by t does not affect the values of the ratios. Therefore, setting $u_s = t w_s$, $s = 1, ..., S$ and $v_p = t\nu_p$, $p = 1, ..., P$, model (2.8) is transformed into the following equivalent linear programme:

$$
\begin{aligned}
\max\ \theta_{j_0} &= \sum_{s=1}^{S} u_s y_{sj_0} \\
\text{s.t.}\ & \sum_{p=1}^{P} v_p x_{pj_0} = 1 \\
& \sum_{s=1}^{S} u_s y_{sj} - \sum_{p=1}^{P} v_p x_{pj} \leq 0, \quad j = 1, ..., N, \\
& u_s, v_p \geq 0, p = 1, ..., P, s = 1, ..., S.
\end{aligned}
\tag{2.9}
$$

Let $(\theta^*_{j_0}, v^*_p, u^*_s)$ be the optimal solution to model 2.9. We can then define CCR-efficiency as follows:

Definition 2.3.1. *A DMU_{j_0} is CCR-efficient if and only if the optimal solution $(\theta^*_{j_0}, v^*_p, u^*_s)$ satisfies $\theta^*_{j_0} = 1$, and there exists at least one optimal (v^*_p, u^*_s), with $v^*_p, u^*_s > 0$. It is weakly efficient if $\theta^*_{j_0} = 1$, and at least one element of (v^*_p, u^*_s) is zero. Otherwise, it is CCR-inefficient.*

For a CCR-inefficient DMU_{j_0}, it holds that $\theta^*_{j_0} < 1$. The reference set R_{j_0} for the inefficient DMU_{j_0} consists of all $\text{DMU}_j, j \in \{1, ..., N\} \backslash \{j_0\}$ such that at the optimal weights (v^*_p, u^*_s), at least one one of the inequality constraints in model (2.9) is binding, i.e.

$$R_{j_0} = \{j : \sum_{s=1}^{S} u^*_s y_{sj} - \sum_{p=1}^{P} v^*_p x_{pj} = 0\}. \tag{2.10}$$

Model 2.9 is the dual or multiplier model. The variable-constraint correspondence between model 2.3.1 and the primal model is provided in Table 2.3 below.

Table 2.3: Primal-dual correspondences in the input orientation, under the CRS assumption

duals \ primal	λ_1	λ_2	...	λ_N	θ_{j_0}		primal RHS/ dual obj.
v_1	$-x_{11}$	$-x_{12}$	...	$-x_{1N}$	x_{1j_0}	$\geq$	0
v_2	$-x_{21}$	$-x_{22}$	...	$-x_{2N}$	x_{2j_0}	$\geq$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
v_P	$-x_{P1}$	$-x_{P2}$	...	$-x_{PN}$	x_{Pj_0}	$\geq$	0
u_1	y_{11}	y_{12}	...	y_{1N}	y_{1j_0}	$\geq$	y_{1j_0}
u_2	y_{21}	y_{22}	...	y_{2N}	y_{2j_0}	$\geq$	y_{2j_0}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_S	y_{S1}	y_{S2}	...	y_{SN}	y_{Sj_0}	$\geq$	y_{Sj_0}
	$\leq$	$\leq$	$\vdots$	$\leq$	$=$		
prim. obj./ dual RHS	0	0	...	0	1		

The primal or envelopment model is given below:

$$\begin{aligned} &\min\ \theta_{j_0} \\ &\text{s.t.}\ \ \theta_{j_0} x_{pj_0} - \sum_{j=1}^{N} \lambda_j x_{pj} \geq 0 \quad p = 1, ..., P, \\ &\qquad \sum_{j=1}^{N} \lambda_j y_{sj} - y_{sj_0} \geq 0, \quad s = 1, ..., S, \\ &\qquad \lambda_j \geq 0, j = 1, ..., N. \end{aligned} \tag{2.11}$$

It can be seen that the constraints of the envelopment model are related to the PPS T, defined in relation 2.1. The constraints of the envelopment model require the activity $(\theta_{j_0} x_{pj_0}, y_{sj_0})$ to belong to P, while the objective of the model requires minimising θ_{j_0}, so that all the elements of the input vector are proportionally reduced as much as possible, while the same production of outputs is secured, and the improved activity still belongs to T.

The above models provide radial measures of a DMU's efficiency (Farrell efficiency), and therefore, an efficient DMU may still have some excess in its inputs or a shortfall in its outputs. Let $\mathfrak{s}^- \in \mathbb{R}^p$ and $\mathfrak{s}^+ \in \mathbb{R}^s$ denote the slack vectors, representing the input excesses and the output shortfalls, respectively. These can be obtained by solving the following model:

$$\begin{aligned} &\min\ \theta_{j_0} - \varepsilon\Big(\sum_{p=1}^{P} \mathfrak{s}_p^- + \sum_{s=1}^{S} \mathfrak{s}_s^+\Big) \\ &\text{s.t.}\ \ \theta_{j_0} x_{pj_0} - \sum_{j=1}^{N} \lambda_j x_{pj} - \mathfrak{s}_p^- = 0 \quad p = 1, ..., P, \\ &\qquad y_{sj_0} - \sum_{j=1}^{N} \lambda_j y_{sj} + \mathfrak{s}_s^+ = 0, \quad s = 1, ..., S, \\ &\qquad \lambda_j, \mathfrak{s}_p^-, \mathfrak{s}_s^+ \geq 0, j = 1, ..., N, p = 1, ..., P, s = 1, ..., S, \end{aligned} \tag{2.12}$$

where $\varepsilon > 0$ is a non-Archimedean infinitesimal.

Model 2.12 can be solved in two phases:

Phase I

Model 2.11 is solved to obtain the optimal objective value $\theta_{j_0}^*$.

Phase II

Using the optimal objective value θ^*_{j0} obtained in Phase I, the following linear model is solved to obtain the slack values:

$$\begin{aligned}
\max \ & \sum_{p=1}^{P} \mathfrak{s}_p^- + \sum_{s=1}^{S} \mathfrak{s}_s^+ \\
\text{s.t. } & \theta^*_{j0} x_{pj0} - \sum_{j=1}^{N} \lambda_j x_{pj} - \mathfrak{s}_p^- = 0 \quad p = 1, ..., P, \\
& y_{sj0} - \sum_{j=1}^{N} \lambda_j y_{sj} + \mathfrak{s}_s^+ = 0, \quad s = 1, ..., S, \\
& \lambda_j, \mathfrak{s}_p^-, \mathfrak{s}_s^+ \geq 0, j = 1, ..., N, p = 1, ..., P, s = 1, ..., S,
\end{aligned} \tag{2.13}$$

so that input excesses and output shortfalls are maximised, while the optimal objective value θ^*_{j0} is maintained.

Let $(\theta^*_{j0}, \lambda^*, \mathfrak{s}_p^{-*}, \mathfrak{s}_s^{+*})$ be the optimal solution of the two-phase model 2.12, i.e. of models 2.11 and 2.13.

Definition 2.3.2. *A DMU_{j0} is CCR-strongly efficient if and only if the optimal solution $(\theta^*_{j0}, \lambda^*, \mathfrak{s}_p^{-*}, \mathfrak{s}_s^{+*})$ satisfies the conditions*

(i) $\theta^*_{j0} = 1$,

(ii) $\mathfrak{s}_p^{-*} = 0$ *and* $\mathfrak{s}_s^{+*} = 0$.

It is CCR-weakly efficient if the optimal solution satisfies condition ((i)), but not condition ((ii)). Otherwise, it is CCR-inefficient.

Definitions 2.3.1 and 2.3.2 both refer to strong and weak CCR efficiency, based on the primal (multiplier) and the dual (envelopment) models, respectively. Due to the condition of complementary slackness, these two definitions are equivalent.

The following Theorem gives the condition of complementary slackness.

Theorem 2.3.1. *(Complementary Slackness) If a dual variable is greater than zero ($v^*_p, u^*_s > 0$) then the corresponding primal constraint is binding (zero slack). If the primal constraint is not binding (positive slack) then the corresponding dual variable is zero ($v^*_p = 0$ and/or $u^*_s = 0$).*

Let $(\hat{x}_{pj0}, \hat{y}_{sj0})$ be the improved activity of DMU_{j0}, based on the optimal solution $(\theta^*_{j0}, \lambda^*, \mathfrak{s}_p^{-*}, \mathfrak{s}_s^{+*})$. Then, the efficient input and output targets (projections) of DMU_{j0} are given as

$$\hat{x}_{pj_0} = \theta^* x_{pj_0} - \mathfrak{s}_p^{-*} = \sum_{j=1}^{N} \lambda_j^* x_{pj}, \tag{2.14}$$

$$\hat{y}_{sj_0} = y_{sj_0} + \mathfrak{s}_p^{+*} = \sum_{j=1}^{N} \lambda_j^* y_{sj}, \tag{2.15}$$

for $p = 1, ..., P$ and $s = 1, ..., S$.

Models 2.9 and 2.11 are the input-oriented CCR multiplier and envelopment models, respectively. Depending on the nature of the problem and the decision maker's preferences, an output expansion given the current input levels may be desirable to achieve instead. The output-oriented multiplier and envelopment CCR models, respectively, are given in Table 2.4 below:

Table 2.4: The output-oriented CCR models

Multiplier form	Envelopment form
$$\begin{aligned} \min\ \phi_{j_0} = \frac{1}{\theta_{j_0}} &= \sum_{s=1}^{S} v_p x_{pj_0} \\ \text{s.t.}\ \sum_{p=1}^{P} u_s y_{sj_0} &= 1 \qquad (2.16) \\ \sum_{s=1}^{S} u_s y_{sj} - \sum_{p=1}^{P} v_p x_{pj} &\leq 0, \quad \forall j, \\ u_s, v_p \geq 0, p = 1, ..., P, s &= 1, ..., S. \end{aligned}$$	$$\begin{aligned} &\max\ \phi_{j_0} \\ &\text{s.t.}\ x_{pj_0} - \sum_{j=1}^{N} \mu_j x_{pj} \geq 0 \quad \forall p, \quad (2.17) \\ &\phi_{j_0} y_{sj_0} - \sum_{j=1}^{N} \mu_j y_{sj} \leq 0, \quad \forall s, \\ &\mu_j \geq 0, j = 1, ..., N. \end{aligned}$$

The technical output efficiency score of DMU_{j_0} is given by the optimal value $\phi_{j_0}^*$ which is the reciprocal of the input-oriented efficiency $\theta_{j_0}^*$. Therefore, in the output orientation, the higher the value of $\phi_{j_0}^*$, the less efficient the DMU_{j_0} is.

Let $t^- \in \mathbb{R}^P$ and $t^+ \in \mathbb{R}^S$ denote the input excesses and the output shortfalls in the output orientation. Then these are defined similarly to the input-oriented case as

$$x_{pj_0} - \sum_{j=1}^{N} \mu_j x_{pj} - t_p^- = 0, \tag{2.18}$$

$$\phi_{j_0} y_{sj_0} - \sum_{j=1}^{N} \mu_j y_{sj} + t_s^+ = 0, \tag{2.19}$$

for $p = 1, ..., P$ and $s = 1, ..., S$, and can be obtained by solving a two-phase model as in the input orientation. Let the optimal solution of the output-oriented two-phase envelopment model for DMU_{j_0} be $(\phi^*_{j_0}, \mu^*, t^{-*}_s, t^{+*}_p)$. Then the optimal solutions of the input and output oriented envelopment models are related as given below:

$$\theta^*_{j_0} = 1/\phi^*_{j_0}, \tag{2.20}$$

$$\lambda^*_j = \mu^*_j/\phi^*_{j_0}, j = 1, ..., N \tag{2.21}$$

$$\mathfrak{s}^{-*}_p = t^{-*}_p/\phi^*_{j_0}, p = 1, ..., P \tag{2.22}$$

$$\mathfrak{s}^{+*}_s = t^{+*}_s/\phi^*_{j_0}, s = 1, ..., S. \tag{2.23}$$

Example:

The input and output oriented models discussed above, are going to be illustrated with the following example. Consider the set of eight DMUs given in the previous section, but this time, a two input-one output case is examined. The data set is given in Table 2.5 below.

Table 2.5: Two input and one output case

DMU	A	B	C	D	E	F	G	H
Input 1	2	7	6	5	3	8	6	4
Input 2	1	2	5	2	4	6	4	6
Output	1	3	7	5	4	5	4	3
Input 1/Output	2	2.33	0.86	1	0.75	1.6	1.5	1.33
Input 2/Output	1	0.67	0.71	0.4	1	1.2	1	2

In the three-dimensional case, it is possible to get the graphical representation of the PPS under the CRS assumption, by using the ratios of inputs per unit of output. Figure 2.4 illustrates the PPS.

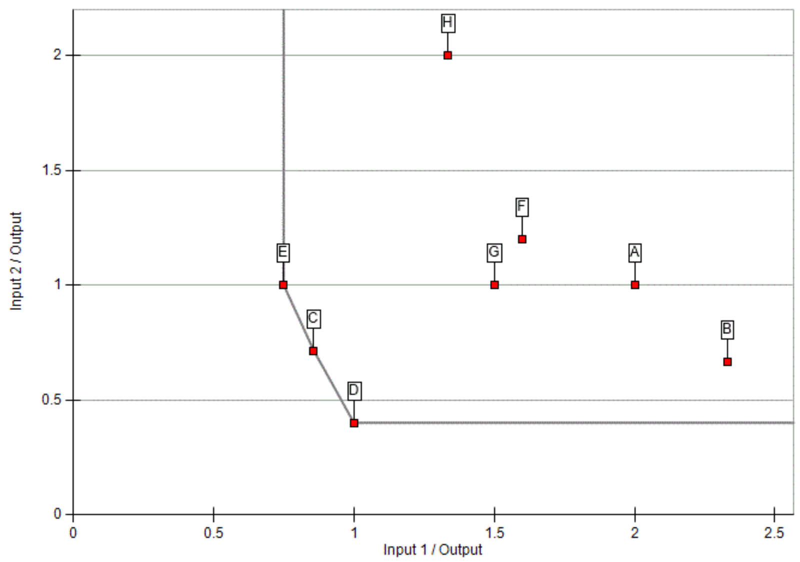


Figure 2.4: PPS for the two input-single output case, under the CRS assumption

Table 2.6 illustrates the results for the input-oriented case, obtained both from the multiplier and the envelopment models; for each DMU, its efficiency score θ^*, the optimal weights (v_1^*, v_2^*, u_1^*) from the multiplier model, the reference set, the optimal λ^*'s, the input excesses $\mathfrak{s}_1^{-*}, \mathfrak{s}_2^{-*}$ and output shortfalls $\mathfrak{s}_1^{+*}$ from the envelopment model are given[1]. Only the non-zero λ_j^*, $j = 1, ..., N$ for a DMU$_{j_0}$ are given; those λ_j^*'s that are not mentioned, are implied to be equal to zero, i.e. the corresponding DMU$_j$, $j = 1, ..., N$ does not belong to the reference set of DMU$_{j_0}$.

Table 2.6: Results for the two input-single output case, under the CRS

DMU	θ^*	v_1^*	v_2^*	u_1^*	Reference Set	λ^*	$\mathfrak{s}_1^{-*}$	$\mathfrak{s}_2^{-*}$	$\mathfrak{s}_1^{+*}$
A	0.48	0.41	0.19	0.48	C, D	λ_C^*=0.04, λ_D^*=0.15	0	0	0
B	0.60	0	0.5	0.2	D	λ_D^*=0.6	1.2	0	0
C	1	0.12	0.05	0.14	C	λ_C^*=1	0	0	0
D	1	0.17	0.08	0.2	D	λ_D^*=1	0	0	0
E	1	0.22	0.08	0.25	E	λ_E^*=1	0	0	0
F	0.55	0.09	0.04	0.11	C,D	λ_C^*=0.59, λ_D^*=0.17	0	0	0
G	0.60	0.13	0.06	0.15	C,D	λ_C^*=0.37, λ_D^*=0.28	0	0	0
H	0.56	0.25	0	0.19	E	λ_E^*=0.75	0	0.37	0

There are three efficient DMUs, C,D and E. In order to find the efficiency score and the efficient targets of the inefficient DMUs, the latter are projected radially on the efficient frontier towards the origin. The projections for each DMU can be obtained as in relations 2.14 and 2.15. Therefore, for the inefficient DMU G, for example, the efficient targets are

[1]Only one set of any alternative optimal weights per DMU is listed in Table 2.6

given as

$$\hat{x}_{1G} = \theta_G^* x_{1G} - \mathfrak{s}_1^{-*} = 0.60 \cdot 6 - 0 = 3.6$$
$$\hat{x}_{2G} = \theta_G^* x_{2G} - \mathfrak{s}_2^{-*} = 0.60 \cdot 4 - 0 = 2.4$$
$$\hat{y}_{1G} = y_{1G} + \mathfrak{s}_1^{+*} = 4 + 0 = 4$$

Efficient targets can also be calculated using the optimal values of λ^*, as

$$\hat{x}_{1G} = \lambda_C^* x_{1C} + \lambda_D^* x_{1D} = 0.37 \cdot 6 + 0.28 \cdot 5 = 3.62$$
$$\hat{x}_{2G} = \lambda_C^* x_{2C} + \lambda_D^* x_{2D} = 0.37 \cdot 5 + 0.28 \cdot 2 = 2.41$$
$$\hat{y}_{1G} = \lambda_C^* y_{1C} + \lambda_D^* y_{1D} = 0.37 \cdot 7 + 0.28 \cdot 5 = 3.99.$$

The projections of all the other inefficient DMUs can be obtained in a similar way.

2.3.2 The BCC model

Banker, Charnes & Cooper (1984) introduced the BCC model, as an extension of the CCR model to consider for the VRS cases. The only difference between the CCR and the BCC models is the extra convexity constraint $\sum_{j=1}^{N} \lambda_j = 1, \lambda_j \geq 0$ added in the latter, in the envelopment form (see in Section 2.1 how the constraint $\lambda \in \Lambda$ in relation 2.1 varies depending on the different RTS assumptions that are made). This convexity constraint corresponds to the free in sign variable u_{j_0} in the primal model. In Table 2.7, the input and output oriented BBC models are given, in the multiplier and envelopment form respectively.

Table 2.7: The BCC models

Multiplier form	Envelopment form
Input orientation	
$\begin{aligned} \max\ \theta_{j_0} &= \sum_{s=1}^{S} u_s y_{sj_0} - u^{j_0} \\ \text{s.t.}\ & \sum_{p=1}^{P} v_p x_{pj_0} = 1 \qquad (2.24) \\ & \sum_{s=1}^{S} u_s y_{sj} - \sum_{p=1}^{P} v_p x_{pj} - u^{j_0} \leq 0, \quad \forall j \\ & u_s, v_p \geq 0, s = 1, ..., S, p = 1, ..., P, \\ & u^{j_0} \text{ free in sign.} \end{aligned}$	$\begin{aligned} & \min\ \theta_{j_0} \\ \text{s.t.}\ & \theta_0 x_{pj_0} - \sum_{j=1}^{N} \lambda_j x_{pj} \geq 0, \quad \forall p \\ & \sum_{j=1}^{N} \lambda_j y_{sj} - y_{sj_0} \geq 0, \quad \forall s, \quad (2.25) \\ & \sum_{j=1}^{N} \lambda_j = 1, \quad \forall j \\ & \lambda_j \geq 0, \mu_j \geq 0, j = 1, ..., N. \end{aligned}$
Output orientation	
$\begin{aligned} \min\ \phi_{j_0} &= \sum_{s=1}^{S} v_p x_{pj_0} - u^{j0} \\ \text{s.t.}\ & \sum_{p=1}^{P} u_s y_{sj_0} = 1 \qquad (2.26) \\ & \sum_{s=1}^{S} u_s y_{sj} - \sum_{p=1}^{P} v_p x_{pj} + u^{j_0} \leq 0, \quad \forall j, \\ & u_s, v_p \geq 0, p = 1, ..., P, s = 1, ..., S. \\ & u^{j_0} \text{ free in sign.} \end{aligned}$	$\begin{aligned} & \max\ \phi_{j_0} \\ \text{s.t.}\ & x_{pj_0} - \sum_{j=1}^{N} \mu_j x_{pj} \geq 0 \quad \forall p, \\ & \phi_{j_0} y_{sj_0} - \sum_{j=1}^{N} \mu_j y_{sj} \leq 0, \quad \forall s, \qquad (2.27) \\ & \sum_{j=1}^{N} \lambda_j = 1, \quad \forall j, \\ & \lambda_j \geq 0, \mu_j \geq 0, j = 1, ..., N. \end{aligned}$

The variable-constraint correspondences between the primal and the dual BCC models, in the input orientation, are presented in Table 2.8.

Unlike the CCR model where the efficiency score of a DMU remains unchanged in both orientations, the input and output oriented BCC models may yield different efficiency scores. Similarly to the CCR model, in the BCC model the input excesses $\mathfrak{s}_p^-$ and output shortfalls $\mathfrak{s}_s^+$ can be obtained by solving a two-phase model. A BCC-efficient DMU$_{j_0}$ is defined analogously to a CCR-efficient DMU$_{j_0}$, as given in Definitions 2.3.1 and 2.3.2 based on the multiplier and the envelopment models, respectively. The definition of the reference set for inefficient DMUs, as well as the calculation of the efficient projections in the BCC model are also similar to the CCR model and are therefore omitted. However, it should be noted that, as distinct to the CCR model, the input and the output reference

Table 2.8: Primal-dual correspondences between models (2.24) and (2.25)

duals \ primal	λ_1	λ_2	...	λ_N	θ_{j_0}		primal RHS/ dual obj.
v_1	$-x_{11}$	$-x_{12}$	...	$-x_{1N}$	x_{1j_0}	$\geq$	0
v_2	$-x_{21}$	$-x_{22}$	...	$-x_{2N}$	x_{2j_0}	$\geq$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
v_P	$-x_{P1}$	$-x_{P2}$	...	$-x_{PN}$	x_{Pj_0}	$\geq$	0
u_1	y_{11}	y_{12}	...	y_{1N}	y_{1j_0}	$\geq$	y_{1j_0}
u_2	y_{21}	y_{22}	...	y_{2N}	y_{2j_0}	$\geq$	y_{2j_0}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_S	y_{S1}	y_{S2}	...	y_{SN}	y_{Sj_0}	$\geq$	y_{Sj_0}
u^{j_0}	1	1	...	1	1	$=$	1
	$\leq$	$\leq$	$\vdots$	$\leq$	$=$		
prim. obj./ dual RHS	0	0	...	0	1		

set for a DMU can differ. Also, Tone (1996) proved that the reference set of a point cannot include increasing and decreasing returns to scale DMUs at the same time.

Let u^{j_0*} be the optimal value of the free variable u^{j_0} for DMU$_{j_0}$, obtained by solving the BCC multiplier model. Banker (1984) and Banker & Thrall (1992) examined how the sign of the optimal value of this free variable can be used to determine the RTS prevailing at a DMU$_{j_0}$, and established this relationship as follows:

Proposition 2.3.1. *Let a DMU$_{j_0}$ with activity vector (x_{pj_0}, y_{sj_0}) be on the efficient frontier. Then, as for the returns to scale at the point (x_{pj_0}, y_{sj_0}), it holds that:*

(i) Increasing returns to scale prevails at (x_{pj_0}, y_{sj_0}) if and only if $u^{j_0*} < 0$ for all optimal solutions.

(ii) Decreasing returns to scale prevails at (x_{pj_0}, y_{sj_0}) if and only if $u^{j_0*} > 0$ for all optimal solutions.

(iii) Constant returns to scale prevails at (x_{pj_0}, y_{sj_0}) if and only if $u^{j_0*} = 0$ in any optimal solution.

Optimal solutions from the CCR envelopment and the BCC multiplier models are related as for the returns to scale characterisations as follows:

Proposition 2.3.2. *(Banker & Thrall 1992)*

(i) $u^{j_0*} > 0$ in all optimal solutions to the BCC model if and only if $\sum_{j=1}^{N} \lambda_j^* > 1$ for all the optimal solutions of the corresponding CCR model.

(ii) $u^{j_0*} < 0$ *in all optimal solutions to the BCC model if and only if* $\sum_{j=1}^{N} \lambda_j^* < 1$ *for all the optimal solutions of the corresponding CCR model.*

(iii) $u^{j_0*} = 0$ *in all optimal solutions to the BCC model if and only if* $\sum_{j=1}^{N} \lambda_j^* = 1$ *for some optimal solutions of the corresponding CCR model.*

Both CCR and BCC models are radial and either input or output oriented. In the next two subsections, two more basic DEA models, the addtive and the SBM models, are going to be discussed. These models combine both orientations and deal directly with input and output slacks.

2.3.3 The additive model

The additive model was introduced by Charnes et al. (1985) and as distinct form the CCR and the BCC models, it is non-oriented. In the additive model, an inefficient DMU can be projected to the efficient frontier in any direction within the quadrant created by the point and its input-output oriented projections to the frontier, and therefore, improvements to inputs and outputs can be non-proportional.

Additive models can be formulated under the CRS or the VRS assumption, in a similar way to the radial DEA models, by omitting or including the free in sign variable u^{j_0} in the multiplier form and the convexity constraint $\sum_{j=1}^{N} \lambda_j = 1$ in the envelopment form. The additive models in the multiplier and the envelopment form, respectively, are given in Table 2.9 that follows.

Table 2.9: The Additive models

Multiplier form	Envelopment form
CRS	
$$\begin{aligned} \min\ & \sum_{p=1}^{P} v_p x_{pj_0} - \sum_{s=1}^{S} u_s y_{sj_0} \\ \text{s.t.}\ & \sum_{s=1}^{S} u_s y_{sj} - \sum_{p=1}^{P} v_p x_{pj} \leq 0, \quad \forall j \\ & u_s, v_p > 0, s = 1, ..., S, p = 1, ..., P, \end{aligned} \tag{2.28}$$	$$\begin{aligned} \max\ & \sum_{p=1}^{P} \mathfrak{s}_p^- + \sum_{s=1}^{S} \mathfrak{s}_s^+ \\ \text{s.t.}\ & \sum_{j=1}^{N} \lambda_j x_{pj} + \mathfrak{s}_p^- = x_{pj_0}, \quad \forall p \\ & \sum_{j=1}^{N} \lambda_j y_{sj} - \mathfrak{s}_s^+ = y_{sj_0}, \quad \forall s \\ & \lambda_j \geq 0, \mu_j \geq 0, j = 1, ..., N. \end{aligned} \tag{2.29}$$
VRS	
$$\begin{aligned} \min\ & \sum_{p=1}^{P} v_p x_{pj_0} - \sum_{s=1}^{S} u_s y_{sj_0} + u^{j_0} \\ \text{s.t.}\ & \sum_{s=1}^{S} u_s y_{sj} - \sum_{p=1}^{P} v_p x_{pj} - u^{j_0} \leq 0, \quad \forall j \\ & u_s, v_p > 0, s = 1, ..., S, p = 1, ..., P, \\ & u^{j_0} \text{ free in sign.} \end{aligned} \tag{2.30}$$	$$\begin{aligned} \max\ & \sum_{p=1}^{P} \mathfrak{s}_p^- + \sum_{s=1}^{S} \mathfrak{s}_s^+ \\ \text{s.t.}\ & \sum_{j=1}^{N} \lambda_j x_{pj} + \mathfrak{s}_p^- = x_{pj_0}, \quad \forall p \\ & \sum_{j=1}^{N} \lambda_j y_{sj} - \mathfrak{s}_s^+ = y_{sj_0}, \quad \forall s \\ & \sum_{j=1}^{N} \lambda_j = 1, \quad \forall j \\ & \lambda_j \geq 0, \mu_j \geq 0, j = 1, ..., N. \end{aligned} \tag{2.31}$$

From the objective function of the envelopment additive model, it can be seen that the l_1-distance for the slack vectors is being used; therefore, an inefficient DMU is projected to the most distant point of the frontier by maximising the sum of input and output slacks. Let $(\mathfrak{s}_p^{-*}, \mathfrak{s}_s^{+*})$ be the optimal values of input excesses and output shortfalls for a DMU_{j_0}.

Definition 2.3.3. *A DMU_{j_0} is Additive-efficient if and only if $\mathfrak{s}_p^{-*} = 0$ and $\mathfrak{s}_s^{+*} = 0$.*

Furthermore, when the CRS is assumed, a DMU_{j_0} is additive-efficient if and only if it is CCR-efficient. Similarly, under the VRS assumption, DMU_{j_0} is additive-efficient if and only if it is BCC-efficient.

Although the additive model can discriminate among efficient and inefficient DMUs, there is no scalar θ to measure the actual efficiency score of DMUs.

2.3.4 The Slacks-Based Measure (SBM)

Tone(1997; 2001) introduced the Slacks-Based Measure (SBM) as an enhancement of the additive model. SBM preserves the non-oriented nature of the additive model dealing directly with slacks, while it introduces a scalar to measure the efficiency score of DMUs.

Consider the following fractional programme, under the CRS assumption:

$$\begin{aligned} \min\ \rho_{j_0} &= \frac{1-\frac{1}{P}\sum_{p=1}^{P}\mathfrak{s}_p^-/x_{pj_0}}{1+\frac{1}{S}\sum_{s=1}^{S}\mathfrak{s}_s^+/y_{sj_0}} \\ \text{s.t. } x_{pj_0} &= \sum_{j=1}^{N}\lambda_j x_{pj} + \mathfrak{s}_p^-, \forall p \\ y_{sj_0} &= \sum_{j=1}^{N}\lambda_j y_{sj} - \mathfrak{s}_s^+, \forall s \\ &\lambda_j \geq 0, \forall j, \mathfrak{s}_p^-, \mathfrak{s}_s^+ \geq 0, \forall p, \forall s. \end{aligned} \tag{2.32}$$

The aforementioned formulation of the SBM model refers to the CRS assumption. The SBM model under the VRS assumption could be obtained by introducing the additional constraint $\sum_{j=1}^{N}\lambda_j = 1$. In the remainder of this section, the SBM model under the CRS assumption will be considered.

The scalar ρ_{j_0} in model 2.32 is equivalent to the following product:

$$\rho_{j_0} = \Big(\frac{1}{P}\sum_{p=1}^{P}\frac{x_{pj_0}-\mathfrak{s}_p^-}{x_{pj_0}}\Big)\Big(\frac{1}{S}\sum_{s=1}^{S}\frac{y_{sj_0}+\mathfrak{s}_s^+}{y_{sj_0}}\Big)^{-1}. \tag{2.33}$$

The ratio $\frac{x_{pj_0}-\mathfrak{s}_p^-}{x_{pj_0}}$ in the first component of relation 2.33, represents the relative reduction rate of each input p, $p = 1, ...P$, and the first component is the mean reduction rate of inputs. Similarly, the ratio $\frac{y_{sj_0}+\mathfrak{s}_s^+}{y_{sj_0}}$ represents the relative expansion rate of each output s, $s = 1, .., S$, and $\frac{1}{S}\sum_{s=1}^{S}\frac{y_{sj_0}+\mathfrak{s}_s^+}{y_{sj_0}}$ measures the mean expansion rate of outputs. Therefore, the first component of the SBM scalar, measures the input efficiency and the second component measures the output efficiency of a DMU_{j_0}.

Model 2.32 can be transformed into the non-linear programme below by multiplying both the nominator and the denominator of the objective function by a positive scalar t, such that $t = 1 - \frac{1}{S}\sum_{s=1}^{S} \mathfrak{s}_s^+/y_{sj_0}$. Then, it follows that

$$\begin{aligned} \min\ \tau_{j_0} &= t - \frac{1}{P}\sum_{p=1}^{P} t\mathfrak{s}_p^-/x_{pj_0} \\ \text{s.t. } t &+ \frac{1}{S}\sum_{s=1}^{S} t\mathfrak{s}_s^+/y_{sj_0} = 1 \\ x_{pj_0} &= \sum_{j=1}^{N} \lambda_j x_{pj} + \mathfrak{s}_p^-, \forall p \\ y_{sj_0} &= \sum_{j=1}^{N} \lambda_j y_{sj} - \mathfrak{s}_s^+, \forall s \\ \lambda_j &\geq 0, \forall j, \mathfrak{s}_p^-, \mathfrak{s}_s^+ \geq 0, \forall p, \forall s, t > 0. \end{aligned} \tag{2.34}$$

Let $\sigma^- = t\mathfrak{s}_p^-$, $\sigma^+ = t\mathfrak{s}_s^+$, and $\mu_j = t\lambda_j$. Then, model 2.34 can be transformed into the following linear model:

$$\begin{aligned} \min\ \tau_{j_0} &= t - \frac{1}{P}\sum_{p=1}^{P} \sigma_p^-/x_{pj_0} \\ \text{s.t. } t &+ \frac{1}{S}\sum_{s=1}^{S} \sigma_s^+/y_{sj_0} = 1 \\ tx_{pj_0} &= \sum_{j=1}^{N} \mu_j x_{pj} + \sigma_p^-, \forall p \\ ty_{sj_0} &= \sum_{j=1}^{N} \mu_j y_{sj} - \sigma_s^+, \forall s \\ \mu_j &\geq 0, \forall j, \sigma_p^-, \sigma_s^+ \geq 0, \forall p, \forall s, t > 0. \end{aligned} \tag{2.35}$$

Let $(\tau_{j_0}^*, t^*, \mu^*, \sigma_p^{-*}, \sigma_p^{+*})$ be the optimal solution to model 2.35. Then, the SBM-efficiency of a DMU_{j_0} can be defined as follows:

Definition 2.3.4. *A DMU_{j_0} is SBM-efficient if and only if $\tau^* = 1$.*

The efficiency condition $\tau_{j_0}^* = 1$ is equivalent to $\sigma_p^{-*} = 0$ and $\sigma_s^{+*} = 0$.

Tone (1997) noted that since the SBM model considering the input/output slacks deals with all short of inefficiencies, the SBM efficiency score of a DMU_{j_0} cannot be greater than its CCR efficiency score. The following theorem states the relationship between the SBM and CCR efficiency

Theorem 2.3.2. *A DMU_{j_0} is CCR-efficient if and only if it is SBM-efficient.*

Two of the modifications of the SBM model existing in the DEA literature are the input and the output-oriented SBM models. These are obtained by introducing a small positive number $\epsilon << 1$ in the objective function of model 2.32, as it is presented in Table 2.10.

Table 2.10: Objective function of the SBM-input and SBM-output oriented model

Input orientation	Output orientation
$$\rho_{j_0}^{in} = \frac{1 - \frac{1}{P}\sum_{p=1}^{P} \mathfrak{s}_p^- / x_{pj_0}}{1 + \frac{\epsilon}{S}\sum_{s=1}^{S} \mathfrak{s}_s^+ / y_{sj_0}} \quad (2.36)$$	$$\rho_{j_0}^{out} = \frac{1 - \frac{\epsilon}{P}\sum_{p=1}^{P} \mathfrak{s}_p^- / x_{pj_0}}{1 + \frac{1}{S}\sum_{s=1}^{S} \mathfrak{s}_s^+ / y_{sj_0}} \quad (2.37)$$

By introducing the small number ϵ in the denominator of relation 2.36, more importance is given in identifying the input surplus, whereas the small number ϵ in the nominator of relation 2.37 aims to give greater importance to the output slacks.

2.4 The curse of dimensionality in DEA

As it was discussed in the previous sections, the efficiency scores calculated with the DEA methodology are relative and not absolute measures of the performance of a DMU. Therefore, the efficiency score assigned to a DMU depends on the performance of the other DMUs included in the sample, and what are the input/output bundles those DMUs are using. As a result, as the number of inputs and outputs increases, the discrimination power of the DEA models diminishes and a number of inefficient DMUs may be falsely rated as efficient, or an inefficient DMU may falsely appear as less inefficient. This issue, known as “the curse of dimensionality” is a common issue of non-parametric approaches.

There are several rules of thumb in the DEA literature for deciding about the acceptable number of inputs and outputs depending on the sample size. Homburg (2001) suggested that $N \geq 2(p+q)$, whereas Raab & Lichty (2002) used the stricter rule that $N \geq 3(p+q)$, where N is the number of DMUs, p is the number of inputs, q the number of outputs. Dyson et al. (2001) suggested that $N \geq 2 \times p \times q$. The empirical rule which is the most commonly used in the DEA literature is that introduced by Cooper et al. (2007), who suggested that $N \geq max\{p \times q, 3(p+q)\}$.

Therefore, in DEA, the selection of inputs and outputs is a task of great importance, as the selected input/output variables should reflect the production process, but at the same time their numbers should not exceed certain vague thresholds. There are various approaches in the DEA literature which are used to increase the discrimination power of DEA models. Such methods include, for example, the aggregation of inputs or outputs, the use of panel data to increase the size of the data set, or the use of weight restrictions to limit the importance of a input or output variable in the production model. Podinovski & Thanassoulis (2007) provide a detailed discussion of the main methods that can be used to address the curse of dimensionality in DEA. Daraio & Simar (2007) suggest the use of a scaled principal component analysis, i.e. a weighted aggregation of inputs and outputs. Since in DEA applications inputs and outputs are usually highly correlated, this approach would reduce the dimensions, without resulting in loosing too much information. As a different approach, Charles et al. (2019) suggest the use of the pure DEA model that considers either only inputs or only outputs to deal with the curse of dimensionality issue.

2.5 Conclusion

DEA has been a technique very broadly used in the efficiency evaluation of DMUs, as its non-parametric nature does not require the estimation of a production function allowing for the consideration of multiple inputs and outputs in the production model. Many theoretical studies have focused on improving the existing DEA models, and there is a thriving number of DEA applications.

In this Chapter, the production model was defined as a process that transforms multiple inputs into multiple outputs and the properties of the PPS based on different returns to scale assumptions were discussed. Among the four basic DEA models, in CCR and BCC models the concept of Farrell's radial efficiency is being used and efficiency of DMUs is measured as their radial distance to the efficient boundary of the PPS without considering the existence of any slacks. Conversely, Pareto-Koopmans technical efficiency (or strong efficiency) requires an efficient DMU not to be able to improve any of its inputs or outputs without worsening the others. Therefore, non-radial models were developed to deal directly with slacks. The additive model allows for non-proportional improvements to inputs and outputs but does not include any scalar to measure their efficiency scores, whereas, the SBM model still allows for non-proportional changes, while providing a scalar for the efficiency ranking of DMUs. Finally, in this Chapter the important issue of "the curse of dimensionality" in DEA was discussed in brief and some empirical rules for the choice of the appropriate sample size were given.

In all the models discussed in this Chapter, inputs are considered to be directly transformed into outputs, and any intermediate stages that may be involved in the production process are not taken into account. Therefore, the production process is considered as a "black box", where inputs enter and outputs exit the process, and no inner information is provided. Network DEA is a branch of the DEA literature that studies the efficiency assessment of DMUs whose operation involves different stages before the final products are generated. In the following chapter, the main developments in Network DEA are discussed.

3

Network Data Envelopment Analysis

In conventional DEA, the production of outputs is considered to occur in one stage and any inner structure of the production process is not taken into account. However, the operational process of a DMU may involve intermediate sub-processes that interact with each other; each sub-process can have different external inputs and outputs and some outputs from one sub-process may need to be used as inputs to the next sub-process. As an example, consider the supply chain process that involves both suppliers and manufacturers. If it is assumed that these two sub-divisions work cooperatively, then the objective is to maximise the profits of the whole supply chain process. Standard DEA approaches would evaluate the efficiency of the whole supply chain as a system and any existing inefficiencies either in the supplier's or manufacturer's stage would not be investigated. Ignoring the inner structure of DMUs may result in misleading conclusions about the efficiency level of a DMU; a DMU can be overall efficient, while all of its sub-processes are inefficient (see for example Kao & Hwang (2008)). Also, a DMU may have lower stage efficiency scores compared to another DMU, but still have higher overall efficiency (Kao & Hwang 2010). What is more, assessing the efficiency of each sub-process can offer a better insight into what is the main source of inefficiency for the whole production process and help decision-makers in improving a DMUs' performance.

Network DEA (NDEA) models were developed as an extension of standard DEA to consider the inner structure of DMUs in the efficiency evaluation process. The first studies that considered the division of the DMUs' operating process into different stages were from Färe & Primont (1984) and Charnes et al. (1986). Färe & Primont (1984) assessed the efficiency of utility firms each operating multiple electricity generation plants by treating each plant as a different sub-DMU. Charnes et al. (1986) divided the army recruitment into two sub-processes and evaluated the efficiency of each sub-process separately. Later, Chilingerian & Sherman (1990) divided the medical care services into managerial and physician-controlled stages. Färe & Whittaker (1995) modelled the dairy production process as a two-stage process composed of the crop and milk production sub-stages and Färe et al. (1997) used the land as an input allocated to different crops. Wang et al. (1997) divided the operation of the banking industry into the capital collection and investment processes. Despite the network concept being already used in the DEA literature for more than one decade, Färe & Grosskopf (2000) were the first to use the term "Network DEA" to describe the different DEA approaches developed to handle multi-stage processes.

Nowadays, there is a high volume of DEA studies that consider various structures of the production process and different approaches have been developed to calculate the efficiency scores of each stage and the whole network system. Castelli et al. (2010) provide a classification of the main contributions in the NDEA literature and the relevant model formulations for shared-flow, multilevel and network models, and Cook & Zhu (2014) discuss the existing modelling techniques for two-stage network structures. Kao (2014) provides a classification of NDEA studies based on the network structure they use. Koronakos (2017) discusses the most influential approaches in NDEA and provides a classification of the multi-stage DEA applications.

This Chapter aims to review the main approaches that have been developed in the NDEA literature to measure the efficiency of two-stage production structures. The rest of this Chapter is structured as follows. In Section 3.1, the PPS in a network structure is defined and the different types of series structures are presented. In Section 3.2, the main approaches to measuring the efficiency of DMUs in two-stage structures are discussed, and Section 3.3 is the conclusion of the chapter.

3.1 The production model

The structure of the production process can vary among the different operational sectors and may involve two or more stages. The different stages can be arranged in series, in parallel or in a way that combines series and parallel structures. In this Section, the main network structures that are going to be discussed in this Chapter are presented, and the mathematical notation is introduced. The main focus will be on the two-stage series structures.

Consider the K-stage production process depicted in Figure 3.1 and a set of N DMUs. For each DMU$_j$, $j = 1, .., N$ let $x_j^k = \left(x_{1j}^k, ..., x_{P_k j}^k\right) \in \mathbb{R}_+^{P_k}$ denote the exogenous inputs to stage k, let $z_j^{(a,k)} = \left(z_{1j}^{(a,k)}, ..., z_{Q_{(k-1)}j}^{(a,k)}\right) \in \mathbb{R}_+^{Q_{(k-1)}}$ denote the intermediate products of stage a that are introduced to stage k, and let $z_j^{(k,I)} = \sum_{a=1}^K z_j^{(a,k)}$ denote the sum of intermediate products produced from the different stages that become inputs to stage k. Also, let $z_j^{(k,b)} = \left(z_{1j}^{(k,b)}, ..., z_{Q_k j}^{(k,b)}\right) \in \mathbb{R}_+^{Q_k}$ be the intermediate products of stage k that are used as inputs to another stage b, and let $z_j^{(k,O)} = \sum_{b=1}^K z_j^{(k,b)}$ be the sum of intermediate products that exit stage k and are used as inputs to the next stages. Finally, let $y_j^k = \left(y_{1j}^k, ..., y_{S_k j}^k\right) \in \mathbb{R}_+^{S_k}$ be the final outputs that exit the system at stage k. The sum of all stage-specific inputs $\sum_{k=1}^K x_j^k = x_j$ will be the overall process input, and $\sum_{k=1}^K y_j^k = y_j$ will be the overall process output.

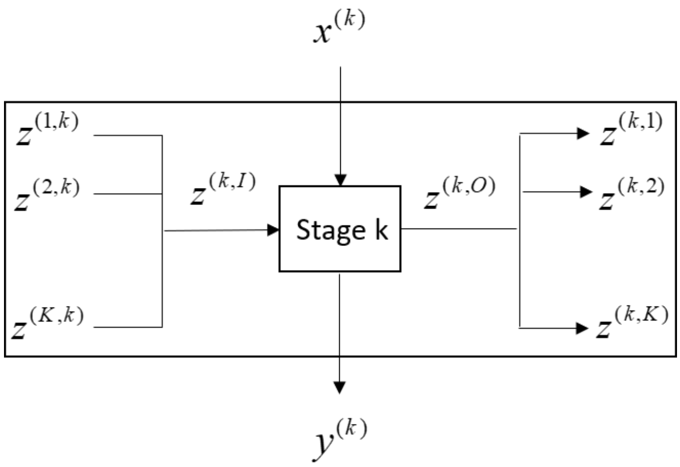


Figure 3.1: General network structure

Then, the PPS of the overall process can be defined as:

$$T = \left\{ (x, z, y) \middle| \sum_{j=1}^N \lambda_j^k x_{pj}^k \leq x_p, p = 1, ..., P_k, \sum_{j=1}^N \lambda_j^k y_{sj}^k \geq y_s, s = 1, ..., S_k, \right.$$

$$\left. \sum_{j=1}^N \lambda_j^k z_{qj}^{(k,I)} \leq z_q, q = 1, ..., Q_{(k-1)}, \sum_{j=1}^N \lambda_j^k z_{qj}^{(k,O)} \geq z_q, q = 1, ..., Q_k, \lambda \in \Lambda, k = 1, ..., K \right\},$$

where $\Lambda = \mathbb{R}_+^N$ under the CRS, $\Lambda = \{\mathbb{R}_+^N | \sum_{j=1}^N \lambda_j^k = 1\}$ under the VRS, when IRS prevails $\Lambda = \{\mathbb{R}_+^N | \sum_{j=1}^N \lambda_j^k \leq 1\}$, and $\Lambda = \{\mathbb{R}_+^N | \sum_{j=1}^N \lambda_j^k \geq 1\}$ under the DRS assumption. The properties of the PPS are defined accordingly to those of the one-stage structures (see Chapter 2, pg. 22).

One type of network structure that is extensively studied in the DEA literature is the series structure. The simplest form of a series structure consists of two stages. There are four types of two-stage series network structures, which are depicted in Figure 3.2.

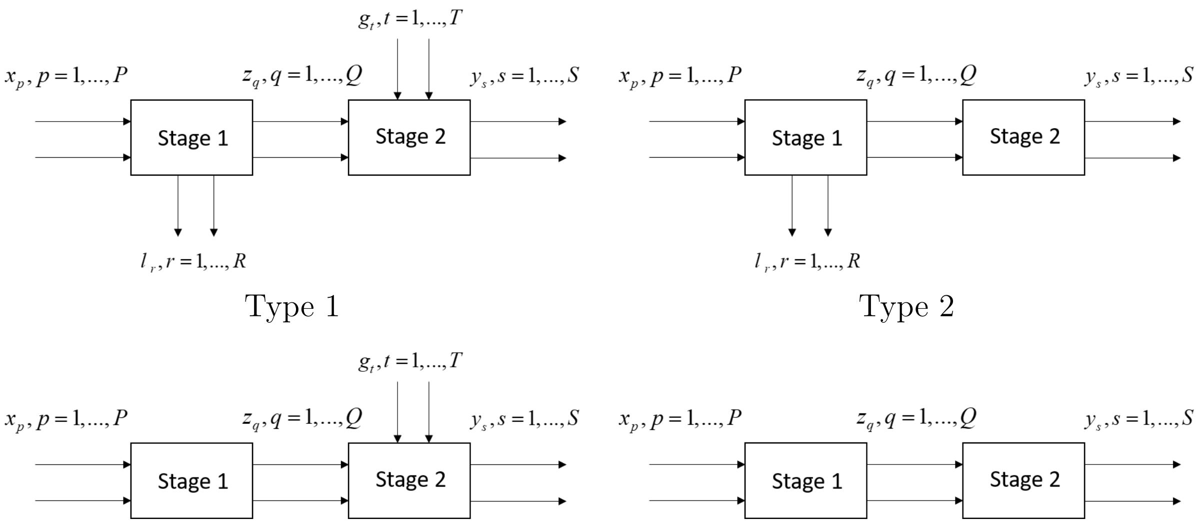


Figure 3.2: The four types of two-stage series network structures

Consider the general two-stage series structure with exogenous inputs and outputs (Type 1). In this case, a Decision Making Unit (DMU) in a first stage consumes P inputs $x_p = (x_1, ..., x_P) \in \mathbb{R}^P_+$ to produce R final first stage outputs $l_r = (l_1, ..., l_R) \in \mathbb{R}^R_+$ and Q intermediate outputs $z_q = (z_1, ..., z_Q) \in \mathbb{R}^Q_+$. In the second stage, intermediate products obtained from the first stage, and external second stage inputs $g_t = (g_1, ..., g_T) \in \mathbb{R}^T_+$ are consumed to produce S final outputs $y_s = (y_1, ..., y_S) \in \mathbb{R}^S_+$. In Type 2 structures, final first stage outputs l_r are produced, but no exogenous inputs enter the second stage. In Type 3 structures, no outputs leave the system in the first stage, but exogenous inputs g_t are introduced in the second stage. Finally, in Type 4 structures, there are no exogenous inputs and outputs; first stage inputs (x_p) are used to produce intermediate outputs (z_q) that are introduced to the second stage to produce the final outputs (y_s) that leave the system.

The aforementioned structures can be generalised to the general series structure depicted in Figure 3.3, where multiple stages arranged in series are involved in the production process. In this case, each stage may have stage specific-exogenous inputs, final and intermediate outputs.

This Chapter will focus on the most influential efficiency measurement approaches in two-stage structures. Not all of those approaches can be directly extended to general series structures or other multistage structures. The strengths and weaknesses of each approach will be discussed in more detail in the following Sections. A list of the notations that are being used in the rest of the Chapter can be found in the Appendix.

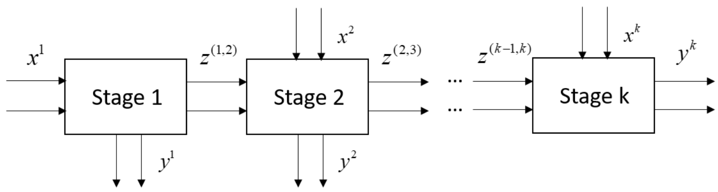


Figure 3.3: General multi-stage series structure

3.2 Efficiency measurement in two-stage processes

When the production process has a network structure, the efficiency score of each stage as well as of the overall process is of interest. Consider a two-stage process, and let θ_j^k be the efficiency of the k-th stage, $k = 1, 2$, and θ_j^0 be the overall process efficiency of DMU$_j$, for $j = 1, 2, .., N$.

Definition 3.2.1. *DMU_{j_0} is considered to be overall efficient if and only if $\theta_{j_0}^{0*} = 1$ and $\theta_{j_0}^{k*} = 1$, $k = 1, 2$. It is stage-k efficient if $\theta_{j_0}^{k*} = 1$, $k = 1, 2$.*

Several approaches have been developed to evaluate θ_j^k, $k = 1, 2$ and θ_j^0. The simplest method of efficiency assessment when the production process has a network structure is, for each DMU, to evaluate the efficiency of each stage independently from the others by using a standard DEA approach. In these approaches, although the existence of the different stages is considered, the interdependencies between the different stages are not taken into account.

In the main volume of NDEA literature, the interdependencies among the different stages are taken into account by jointly assessing the efficiency of the overall process and the different stages. The different approaches that have been developed to consider the dependencies among the different stages can be divided into four main categories: (i) the efficiency decomposition approaches, (ii) the efficiency composition approaches, (iii) the slacks-based measure approaches, (iv) and the network approaches. In the efficiency decomposition approaches, the overall efficiency of the system is calculated first and the stage efficiencies are obtained by the multiplicative or the additive decomposition of the overall efficiency score. In efficiency composition approaches, stages efficiencies are calculated first by considering the stage interdependencies. The different models for the stage efficiencies are combined in a multiple objective linear programme (MOLP) and the overall efficiency is obtained by aggregating the multiple objective functions. The slacks-based measure approaches also decompose the overall efficiency into stage efficiencies, but they allow for non-proportional changes in the inputs and outputs using the SBM model. Finally, in network approaches, the overall efficiency of the process is evaluated considering the different stages that are involved in the process and their interdependencies, but do not provide any information about the stage efficiency scores. In the sections that follow, the aforementioned approaches will be discussed in more detail.

3.2.1 The standard DEA approach

In this approach, the efficiency of each stage is calculated using the standard DEA models. Charnes et al. (1986), Chilingerian & Sherman (1990) and Wang et al. (1997) who were among the first DEA studies to consider the inner structure of DMUs, applied the independent approach in two-stage structures. Similarly, Seiford & Zhu (1999) and Zhu (2000) evaluated the efficiency of the top 55 US commercial banks and of the Fortune 500 companies, respectively, both dividing the production process into the marketability and profitability stages. Sexton & Lewis (2003) assessed the efficiency of 30 baseball teams of the USA Major League Baseball, by separating front-office and on-field operations. They calculated the efficiency of the second stage by using the optimal amount of inputs given from the first stage. The independent approach is also applied to more complex structures; Lewis & Sexton (2004) further divided the operation process of baseball teams by splitting the front-office operations into two processes and the on-field operations into three. Adler et al. (2013) evaluated the efficiency of 43 airports in Europe, where the operation process was divided into two stages, the efficiency of each one of which was evaluated independently, and the first stage was further split into passenger and cargo transportation and the second into aeronautical and non-aeronautical activities.

Consider a two-stage production process with no exogenous inputs or outputs (Type 4 structure in Figure 3.2). Under the CRS assumption, the first and second stage efficiency scores for DMU_{j_0} in the input orientation, can be calculated independently by solving the following mathematical models, respectively:

$$\begin{aligned} \max\ \theta^1_{j_0} &= \frac{\sum_{q=1}^{Q} \gamma^A_q z_{qj_0}}{\sum_{p=1}^{P} v_p x_{pj_0}} \\ \text{s.t.}\ \ \theta^1_j &\leq 1, \quad \forall j \\ & v_p, \gamma^A_q > 0, \forall p, \forall q, \end{aligned} \tag{3.1a}$$

$$\begin{aligned} \max\ \theta^2_{j_0} &= \frac{\sum_{s=1}^{S} \eta_s y_{sj_0}}{\sum_{q=1}^{Q} \gamma^B_q z_{qj_0}} \\ \text{s.t.}\ \ \theta^2_j &\leq 1, \quad \forall j \\ & \gamma^B_q, \eta_s > 0, \forall q, \forall s \end{aligned} \tag{3.1b}$$

whereas the overall efficiency of the system is obtained as follows:

$$\begin{aligned} \theta_{j_0}^0 = & \max \frac{\sum_{s=1}^{S} \eta_s y_{sj_0}}{\sum_{p=1}^{P} v_p x_{pj_0}} \\ & \text{s.t. } \theta_j^0 \leq 1, \quad \forall j \\ & \quad v_p, \gamma_q, \eta_s > 0, \forall p, \forall q, \forall s. \end{aligned} \tag{3.2}$$

In models (3.1a) and (3.1b) stage efficiency scores are calculated independently. Although this makes the independent approach computationally easy, the conflicting role of intermediate products is not considered; for a DMU$_{j_0}$ that reduces the second stage inputs to become efficient, the first stage outputs are also reduced affecting the first stage's efficiency score. Furthermore, the overall efficiency of the system is calculated based only on the initial inputs x_{j_0} that enter the system and the final outputs y_{j_0} that exit the system, and there is no consideration in relating the overall efficiency to the stage efficiencies.

3.2.2 Efficiency decomposition approaches

The efficiency decomposition approaches take into account the relationship between the different stages of the production process, and provide evaluations of the overall as well as of the stage efficiency scores. There are two main decomposition approaches in the DEA literature, the multiplicative and the additive, introduced by Kao & Hwang (2008) and Chen et al. (2009), respectively. In both approaches, the overall efficiency is obtained first and then, the stage efficiencies are calculated while maintaining the optimal overall efficiency level. In the multiplicative approach, the overall efficiency is decomposed as the product of stage efficiencies, while in the additive approach, overall efficiency is defined as the weighted sum of the stage efficiencies. In the following subsections, these two decomposition approaches and some of their extensions are discussed.

3.2.2.1 The multiplicative decomposition approach

Kao & Hwang (2008) introduced the multiplicative decomposition approach for a two-stage structure with no exogenous inputs and outputs (Type 4 structure in Figure 3.2). The first stage, the second stage, and the overall process efficiency score of a DMU$_j$, $j = 1, ..., N$, with a Type 4 structure, in the input orientation, under the CRS assumption, are defined as

$$\theta_j^1 = \frac{\sum_{q=1}^{Q} \gamma_q^A z_{qj}}{\sum_{p=1}^{P} v_p x_{pj}}, \theta_j^2 = \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q^B z_{qj}}, \text{ and } \theta_j^0 = \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{p=1}^{P} v_p x_{pj}}, \tag{3.3}$$

respectively. For a DMU$_{j_0}$, $\theta_{j_0}^0, \theta_{j_0}^1$ and $\theta_{j_0}^2$ can be calculated independently using models (3.2), (3.1a) and (3.1b), as discussed in Section 3.2.1.

In order to take into consideration the series relationship between the two stages, Kao & Hwang (2008) assumed that the multipliers related to the intermediate products z_{qj}, are the same in both stages, i.e. $\gamma_q^A = \gamma_q^B = \gamma_q$. That means that the optimal aggregated outputs from the first stage become inputs to the second stage.

The overall efficiency θ_j^0, $j = 1, 2, ..., N$ can then be decomposed to the product of stage efficiencies, i.e.

$$\theta_j^0 = \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{p=1}^{P} v_p x_{pj}} = \frac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj}} \cdot \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} = \theta_j^1 \theta_j^2. \tag{3.4}$$

The two-stage structure is taken into account in the calculation of the overall efficiency score, by incorporating the ratio constraints of the two stages (models (3.1a) and (3.1b), where $\gamma_q^A = \gamma_q^B = \gamma_q$) into model 3.2. Therefore, in the multiplicative decomposition approach, the overall efficiency of DMU$_{j_0}$ is given by the following model:

$$\begin{aligned} \max\ \theta_{j_0}^0 &= \frac{\sum_{s=1}^{S} \eta_s y_{sj_0}}{\sum_{p=1}^{P} v_p x_{pj_0}} \\ \text{s.t.}\ & \frac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj}} \leq 1, j = 1, ..., N \\ & \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \leq 1, j = 1, ..., N \\ & v_p, \gamma_q, \eta_s, \forall p, \forall q, \forall s. \end{aligned} \tag{3.5}$$

The constraint $\frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{p=1}^{P} v_p x_{pj}} \leq 1, j = 1, ..., N$ is omitted, as it is implied by the two inequality constraints included in model (3.5).

Applying the Charnes-Cooper transformation, the fractional model (3.5) can be transformed into a linear one:

$$\begin{aligned} \theta_{j0}^{0*} = \ & \max \sum_{s=1}^{S} \eta_s' y_{sj_0} \\ & \text{s.t.} \sum_{p=1}^{P} \upsilon_p' x_{pj_0} = 1 \\ & \sum_{q=1}^{Q} \gamma_q' z_{qj} - \sum_{p=1}^{P} \upsilon_p' x_{pj_0} \leq 0, j = 1, ..., N \\ & \sum_{s=1}^{S} \eta_s' y_{sj} - \sum_{q=1}^{Q} \gamma_q' z_{qj} \leq 0, j = 1, ..., N \\ & \upsilon_p', \gamma_q', \eta_s' > 0, \forall p, \forall q, \forall s. \end{aligned} \tag{3.6}$$

Let $(\theta_{j0}^{0*}, \upsilon_p'^*, \gamma_q'^*, \eta_s'^*)$ be the optimal solution to model (3.6). The stage efficiency scores for DMU_{j_0} can be obtained as $\theta^{1*} = \frac{\sum_{q=1}^{Q} \gamma_q'^* z_{qj}}{\sum_{p=1}^{P} \upsilon_p'^* x_{pj}}$, and $\theta^{2*} = \frac{\sum_{s=1}^{S} \eta_s'^* y_{sj}}{\sum_{q=1}^{Q} \gamma_q'^* z_{qj}}$.

However, it is possible that model (3.6) has multiple optimal solutions, and in this case, the decomposition $\theta_{j0}^{0*} = \theta_{j0}^{1*} \theta_{j0}^{2*}$ is not unique. This happens because the term $\gamma_q' z_{qj}$ does not appear in the objective function or the normalisation constraint. Therefore, its value can vary while continuing to satisfy the constraints and without affecting the value of the objective function. Kao & Hwang (2008) noted that in such cases the comparison of the stage efficiency scores among the DMUs lacks a common basis.

The solution that Kao & Hwang (2008) suggested to overcome the non-uniqueness of the overall efficiency decomposition, was to obtain the overall efficiency score first and then, give pre-emptive priority to one of the two stages. The efficiency score of the priority stage is maximised while the overall efficiency level is maintained at its optimum level. Then, the efficiency of the other stage is obtained using the decomposition equation.

Let p denote the priority stage, and θ_j^{kp*}, $k = 1, 2$ denote the optimal efficiency score of the priority stage. First, suppose that stage one is the priority stage. After solving model (3.6) to get the overall efficiency score θ_{j0}^{0*}, the efficiency score of the first stage is obtained by solving the following linearised model, which ensures that $\theta_{j0}^{0} = \theta_{j0}^{0*}$:

$$
\begin{aligned}
\theta_{j_0}^{1p*} = \ & \max \sum_{q=1}^{Q} \gamma'_q z_{qj_0} \\
& \text{s.t.} \ \sum_{p=1}^{P} \upsilon'_p x_{pj_0} = 1 \\
& \sum_{s=1}^{S} \eta'_s y_{sj} = \theta_{j_0}^{0*} \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj_0} \leq 0, j = 1, ..., N \\
& \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, j = 1, ..., N \\
& \upsilon'_p, \gamma'_q, \eta'_s > 0, \forall p, \forall q, \forall s.
\end{aligned}
\tag{3.7}
$$

Then, the efficiency score of the second stage is derived from the decomposition equation as $\theta_{j_0}^{2*} = \theta_{j_0}^{0*}/\theta_{j_0}^{1p*}$. Similarly, if stage two is considered as the priority stage, then its efficiency score is obtained first, as follows

$$
\begin{aligned}
\theta_{j_0}^{2p*} = \ & \max \sum_{s=1}^{S} \eta'_s y_{sj_0} \\
& \text{s.t.} \ \sum_{q=1}^{Q} \gamma'_q z_{qj} = 1 \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} \leq 0, j = 1, ..., N \\
& \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, j = 1, ..., N \\
& \upsilon'_p, \gamma'_q, \eta'_s > 0, \forall p, \forall q, \forall s,
\end{aligned}
\tag{3.8}
$$

and then, form the decomposition equation, $\theta_{j_0}^{1*} = \theta_{j_0}^{0*}/\theta_{j_0}^{2p*}$. If the decomposition is unique, the choice of the priority stage does not affect the efficiency scores, i.e. $\theta_{j_0}^{0*} = \theta_{j_0}^{1p*}\theta_{j_0}^{2*} = \theta_{j_0}^{1*}\theta_{j_0}^{2p*}$.

As noted by Chen et al. (2009), a shortcoming of the multiplicative approach is that it can not be applied under the VRS assumption. The product of the first and the second stage efficiency scores of a DMU_j, $j = 1, ..., N$, under the VRS assumption would be

$$\theta_j^1 \theta_j^2 = \frac{\sum_{q=1}^{Q} \gamma_q z_{qj} + u^A}{\sum_{p=1}^{P} v_p x_{pj}} \cdot \frac{\sum_{s=1}^{S} \eta_s y_{sj_0} + u^B}{\sum_{q=1}^{Q} \gamma_q z_{qj}}. \tag{3.9}$$

However, the above quantity cannot be linearised. Therefore, under the VRS assumption, alternative approaches should be used to obtain the overall and stage efficiency scores. As an alternative, Kao & Hwang (2011) used an input-oriented VRS model in the first stage and an output-oriented model in the second stage and calculated the overall efficiency as the product of the overall technical and scale efficiencies, which were obtained as the product of the stage technical and scale efficiencies, respectively.

Liang et al. (2008) studied the two stage processes using the game theory concepts of cooperative and non-cooperative game. They showed that when there is only one intermediate output from the first stage to the second stage, and no exogenous inputs enter the second stage, the non-cooperative game approach coincides with the cooperative approach of Kao & Hwang (2008). In the non-cooperative approach, one of the two stages is considered as the leader, and the other stage as the follower. Let the exponent L denote the leader stage, and the exponent F denote the follower stage, and let the first stage be the leader and the second stage the follower. Then, the efficiency score of the leader stage $\theta_{j_0}^{1L}$ of a DMU_{j_0} is calculated first using the standard DEA approach. The linear model for obtaining the efficiency score of the first, leader stage, under the CRS assumption, is given below:

$$\begin{aligned} \theta_{j_0}^{1L*} = \ & \max \sum_{q=1}^{Q} \gamma_q' z_{qj_0} \\ & \text{s.t.} \sum_{p=1}^{P} v_p' x_{pj_0} = 1 \\ & \sum_{q=1}^{Q} \gamma_q' z_{qj} - \sum_{p=1}^{P} v_p' x_{pj} \leq 0, j = 1, ..., N \\ & v_p', \gamma_q' > 0, \quad \forall p, \forall q. \end{aligned} \tag{3.10}$$

The efficiency score of the follower $\theta_{j_0}^{2F*}$ is obtained next, while keeping the efficiency of the leader at its optimum level, i.e. $\theta_{j0}^{1L} = \theta_{j0}^{1L*}$, as follows:

$$\begin{aligned}
\theta_{j0}^{2F*} = \ & \max \frac{\sum_{s=1}^{S} \eta_s y_{sj_0}}{W \sum_{q=1}^{Q} \gamma'_q z_{qj_0}} \\
& \text{s.t.} \ \sum_{p=1}^{P} \upsilon'_p x_{pj_0} = 1 \\
& \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} W \gamma'_q z_{qj}} \leq 1, j = 1, ..., N \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj_0} = \theta_{j0}^{1L*} \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} \leq 0, j = 1, ..., N \\
& \upsilon'_p, \gamma'_q, W, \eta_s > 0, \quad \forall p, \forall q, \forall s.
\end{aligned} \tag{3.11}$$

Let $\eta' = \eta_s / W$. Then, model 3.11 is transformed to the following linear model:

$$\begin{aligned}
\theta_{j0}^{2F*} = \ & \max \left(\sum_{s=1}^{S} \eta'_s y_{sj_0} \right) / \theta_{j0}^{1L*} \\
& \text{s.t.} \ \sum_{p=1}^{P} \upsilon'_p x_{pj_0} = 1 \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj_0} = \theta_{j0}^{1L*} \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} \leq 0, j = 1, ..., N \\
& \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, j = 1, ..., N \\
& \upsilon'_p, \gamma'_q, \eta'_s > 0, \quad \forall p, \forall q, \forall s.
\end{aligned} \tag{3.12}$$

Similarly, if stage two is considered as the leader and stage one as the follower, the efficiency score of DMU_{j_0} for the second stage is calculated first, and then, the efficiency score of the first stage is maximised while $\theta_{j0}^{2L} = \theta_{j0}^{2L*}$. An advantage of the non-cooperative approach is that in contrast to the cooperative approach, it always yields a unique overall efficiency decomposition.

In contrast to one stage DEA, in NDEA, due to the existence of intermediate measures, adjusting the inputs or outputs of a DMU by its input or output efficiency score, respectively, will not necessarily project the DMU on the frontier. Chen et al. (2013) demonstrated that in NDEA the duality between the envelopment and the multiplier model does not hold and suggested using the first one to get the frontier projections and the second one for calculating the overall and stage efficiency scores. Chen et al. (2010) suggested the following envelopment model as the equivalent to model (3.6), under the CRS assumption:

$$
\begin{aligned}
&\min\ \theta_{j_0} \\
&\text{s.t.}\ \sum_{j=1}^{N} \lambda_j x_{pj} \leq \theta_{j_0} x_{pj_0}, \quad p = 1, ..., P \\
&\qquad \sum_{j=1}^{N} \mu_{j_0} y_{sj} \geq y_{sj_0}, \quad s = 1, ..., S \\
&\qquad \sum_{j=1}^{N} (\lambda_j - \mu_j) z_{qj} \geq 0, \quad q = 1, ..., Q \\
&\qquad \lambda_j, \mu_j \geq 0.
\end{aligned}
\tag{3.13}
$$

However, as noted by Chen et al. (2010), this model provides an overall efficiency score, but fails to provide correct frontier projections for the DMUs. Therefore, they suggested a new model to obtain the frontier projections introducing a new set of variables $\tilde{z}_{qj_0}, q = 1, ..., Q$ to define the intermediate product levels, and dividing the intermediate product constraints in model (3.13) into two new set of constraints where the intermediate products are treated as inputs and as outputs, respectively. Then, the model takes the following form:

$$
\begin{aligned}
\min\ & \tilde{\theta}_{j_0} \\
\text{s.t.}\ & \sum_{j=1}^{N} \lambda_j x_{pj} \leq \tilde{\theta}_{j_0} x_{pj_0}, \quad p = 1, ..., P \\
& \sum_{j=1}^{N} \mu_{j_0} y_{sj} \geq y_{sj_0}, \quad s = 1, ..., S \\
& \sum_{j=1}^{N} \lambda_j z_{qj} \geq \tilde{z}_{qj_0}, \quad q = 1, ..., Q \\
& \sum_{j=1}^{N} \mu_j z_{qj} \leq \tilde{z}_{qj_0}, \quad q = 1, ..., Q \\
& \tilde{z}_{qj_0} \geq 0, \quad q = 1, ..., Q \\
& \lambda_j, \mu_j \geq 0, \quad j = 1, ..., N.
\end{aligned}
\tag{3.14}
$$

Model (3.14) yields the same efficiency score as model (3.13), and also provides the frontier projections for DMUs.

The dual of model (3.14) is given below:

$$
\begin{aligned}
\max\ & \sum_{s=1}^{S} \eta'_s y_{sj_0} \\
\text{s.t.}\ & \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'^1_q z_{qj} \leq 0, j = 1, ..., N \\
& \sum_{q=1}^{Q} \gamma'^2_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} \leq 0, j = 1, ..., N \\
& \sum_{p=1}^{P} \upsilon'_p x_{pj_0} = 1 \\
& \gamma'^2_q - \gamma'^1_q \leq 0, q = 1, ..., Q \\
& \gamma'^1_q, \gamma'^2_q, \upsilon'_p, \eta'_s \geq 0, \quad \forall p, \forall q, \forall s.
\end{aligned}
\tag{3.15}
$$

In model (3.14) the constraints $\tilde{z}_{qj_0} \geq 0$ are redundant and can be omitted. Then, in model (3.15) the constraint $\gamma'^2_q - \gamma'^1_q \leq 0, q = 1, ..., Q$ becomes $\gamma'^1_q = \gamma'^2_q$, i.e. the intermediate products get the same weights both as outputs from the first stage and as inputs to the next stage. Therefore, model 3.15 becomes identical to model 3.6.

Another limitation of the multiplicative approach is that it can only be generalised to multi-stage series structures only when there are no stage specific inputs and outputs, i.e. it can only be applied in the case of Type 4 structures. Alternative multiplicative approaches have been developed to deal with different structures. The most common approach is the use of parametric techniques. Zha & Liang (2010), studied a two-stage structure with inputs shared between the two stages. Under the cooperative game condition, they defined the overall efficiency as the product of the stage efficiency scores. In order to solve the resulting, non-linear model, they calculated the lower and upper bounds of the stage efficiencies using a non-cooperative game approach, and then, treated the efficiency score of each stage as a parameter, and searched in the feasible region of the parameter to obtain the optimal solution.

3.2.2.2 The additive decomposition approach

The additive decomposition approach was introduced by Chen et al. (2009) for the case of a Type 4 structure with no stage specific inputs and outputs. If it is assumed that the aggregate outputs of the first stage are introduced unchanged to the second stage, then the overall efficiency can be decomposed as the weighted average of stage efficiencies as $\theta_j^0 = w_{1j}\theta_j^1 + w_{2j}\theta_j^2$, i.e., in the input orientation, under the CRS assumption:

$$\theta_j^0 = \frac{\sum_{s=1}^{S} \eta_s y_{sj} + \sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}} = w_{1j} \frac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj}} + w_{2j} \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \tag{3.16}$$

$$\text{and} \qquad w_{1j} + w_{2j} = 1. \tag{3.17}$$

The decomposition weights w_{1j} and w_{2j} can be defined endogenously by solving the system of equations (3.16) and (3.17). Then, for a $\text{DMU}_j, j = 1, ..., N$ we have:

$$w_{1j} = \frac{\sum_{p=1}^{P} v_p x_{pj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}}, \quad \text{and} \quad w_{2j} = \frac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}}. \tag{3.18}$$

Intuitively, as noted by Chen et al. (2009), the decomposition weights w_{1j} and w_{2j} represent the relative contribution of each stage to the overall efficiency. Therefore, in the input orientation, they are defined as the ratio of the stage-specific inputs to the overall process inputs.

Then, the input overall efficiency of DMU_{j_0} under CRS can be derived by solving the following fractional programme:

$$\begin{aligned}\theta_{j_0}^{0*} = \max\ [w_{1j_0}\theta_{j_0}^1 + w_{2j_0}\theta_{j_0}^2] = \frac{\sum_{s=1}^{S}\eta_s y_{sj_0} + \sum_{q=1}^{Q}\gamma_q z_{qj_0}}{\sum_{p=1}^{P}v_p x_{pj_0} + \sum_{q=1}^{Q}\gamma_q z_{qj_0}} \\ \text{s.t. } \theta_j^1 = \frac{\sum_{q=1}^{Q}\gamma_q z_{qj}}{\sum_{p=1}^{P}v_p x_{pj}} \leq 1, \quad j = 1, ..., N \\ \theta_j^2 = \frac{\sum_{s=1}^{S}\eta_s y_{sj}}{\sum_{q=1}^{Q}\gamma_q z_{qj}} \leq 1, \quad j = 1, ..., N \\ v_p, \gamma_q, \eta_s > 0, \forall p, \forall q, \forall.\end{aligned} \tag{3.19}$$

The constraint $\frac{\sum_{s=1}^{S}\eta_s y_{sj_0} + \sum_{q=1}^{Q}\gamma_q z_{qj_0}}{\sum_{p=1}^{P}v_p x_{pj_0} + \sum_{q=1}^{Q}\gamma_q z_{qj_0}} \leq 1$ is omitted, as it is implied by the two inequality constraints included in model 3.19, requiring the efficiency score of each stage not to exceed one.

Applying the Charnes-Cooper transformation, model 3.19 can be transformed into a linear one.

$$\begin{aligned}\theta_{j_0}^{0*} = \max\ & \sum_{q=1}^{Q}\gamma'_q z_{qj_0} + \sum_{s=1}^{S}\eta'_s y_{sj_0} \\ \text{s.t. } & \sum_{p=1}^{P}v'_p x_{pj_0} + \sum_{q=1}^{Q}\gamma'_q z_{qj_0} = 1, \\ & \sum_{q=1}^{Q}\gamma'_q z_{qj} - \sum_{p=1}^{P}v'_p x_{pj} \leq 0, \quad j = 1, ..., N \\ & \sum_{s=1}^{S}\eta'_s y_{sj} - \sum_{q=1}^{Q}\gamma'_q z_{qj} \leq 0, \quad j = 1, ..., N \\ & v'_p, \gamma'_q, \eta'_s > 0.\end{aligned} \tag{3.20}$$

Let $(\theta_{j_0}^{0*}, v_p'^*, \gamma_q'^*, \eta_s'^*)$ be the optimal solution of model (3.20). The optimal multipliers will be used in relations 3.18 to obtain the optimal decomposition weights for DMU_{j_0} as follows:

$$w_{1j_0}^* = \frac{\sum_{p=1}^{P}v_p'^* x_{pj_0}}{\sum_{p=1}^{P}v_p'^* x_{pj_0} + \sum_{q=1}^{Q}\gamma_q'^* z_{qj_0}}, \tag{3.21}$$

$$w_{2j_0}^* = \frac{\sum_{q=1}^{Q}\gamma_q'^* z_{qj_0}}{\sum_{p=1}^{P}v_p'^* x_{pj_0} + \sum_{q=1}^{Q}\gamma_q'^* z_{qj_0}}. \tag{3.22}$$

In order to obtain a unique efficiency decomposition, similarly to the multiplicative approach, when calculating the stage efficiency scores, one of the two stages will be given pre-emptive priority. The efficiency of the stage that is given pre-emptive priority is calculated first, while preserving the optimal overall efficiency level $\theta_{j_0}^{0*}$. Let θ_j^{kp*}, $k = 1, 2$ denote the optimal efficiency score of the priority stage. Suppose stage one is the priority stage, then, the following model is solved to calculate the efficiency score of the first stage:

$$\begin{aligned} \theta_{j_0}^{1p*} = \max\ & \theta_{j_0}^{1} \\ \text{s.t. } & \theta_j^1 \leq 1, \quad j = 1, ..., N \\ & \theta_j^2 \leq 1, \quad j = 1, ..., N \\ & \frac{\sum_{s=1}^{S} \eta_s y_{sj_0} + \sum_{q=1}^{Q} \gamma_q z_{qj_0}}{\sum_{p=1}^{P} v_p x_{pj_0} + \sum_{q=1}^{Q} \gamma_q z_{qj_0}} = \theta_{j_0}^{0*} \\ & v_p, \gamma_q, \eta_s > 0, \forall p, \forall q, \forall s. \end{aligned} \tag{3.23}$$

The equivalent linear model is

$$\begin{aligned} \theta_{j_0}^{1p*} = \ & \max \sum_{q=1}^{Q} \gamma'_q z_{qj_0} \\ & \text{s.t. } \sum_{p=1}^{P} v'_p x_{pj_0} = 1 \\ & \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} v'_p x_{pj} \leq 0, \quad j = 1, ..., N \\ & \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, \quad j = 1, ..., N \\ & (1 - \theta_{j_0}^{0*}) \sum_{q=1}^{Q} \gamma'_q z_{qj_0} + \sum_{s=1}^{S} \eta'_s y_{sj_0} = \theta_{j_0}^{0*} \\ & v'_p, \gamma'_q, \eta'_s > 0, \forall p, \forall q, \forall s. \end{aligned} \tag{3.24}$$

Then, from the decomposition equation (3.16), the efficiency of the second stage is calculated as

$$\theta_{j_0}^{2*} = \frac{\theta_{j_0}^{0*} - w_{1j_0}^{*} \theta_{j_0}^{1p*}}{w_{2j_0}^{*}}. \tag{3.25}$$

If the second stage was given pre-emptive priority, then the efficiency score of that stage would maximised first while preserving the overall efficiency level. The linearised model for the second stage is given below:

$$\begin{aligned}
\theta_{j_0}^{2p*} = \ & \max \sum_{s=1}^{S} \eta'_s y_{sj_0} \\
& \text{s.t.} \ \sum_{q=1}^{Q} \gamma'_q z_{qj_0} = 1 \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} v'_p x_{pj} \leq 0, \quad j = 1, ..., N \\
& \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, \quad j = 1, ..., N \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj_0} + \sum_{s=1}^{S} \eta'_s y_{sj_0} - \theta_{j_0}^{0*} \sum_{p=1}^{P} v'_p x_{pj_0} = \theta_{j_0}^{0*} \\
& v'_p, \gamma'_q, \eta'_s > 0.
\end{aligned} \tag{3.26}$$

Then, using the optimal decomposition weights given by (3.21) and (3.22), the efficiency of the first stage is calculated as

$$\theta_{j_0}^{1*} = \frac{\theta_{j_0}^{0*} - w_{2j_0}^{*} \theta_{j_0}^{2p*}}{w_{1j_0}^{*}}. \tag{3.27}$$

Lim & Zhu (2019) provided the primal-dual correspondences both for the multiplicative and the additive approaches, under the CRS and the VRS assumptions. Under the CRS assumption, the envelopment form of model (3.20) that can be used to obtain the frontier projections is given below.

$$\begin{aligned}
& \min \theta_{j_0} \\
& \text{s.t.} \ \sum_{j=1}^{N} \lambda_j x_{pj} \leq \theta_{j_0} x_{pj_0}, \quad p = 1, ..., P \\
& \sum_{j=1}^{N} \mu_{j_0} y_{sj} \geq y_{sj_0}, \quad s = 1, ..., S \\
& \sum_{j=1}^{N} (\lambda_j - \mu_j) z_{qj} \geq \tilde{z}_{qj_0}, \quad q = 1, ..., Q \\
& \lambda_j, \mu_j \geq 0.
\end{aligned} \tag{3.28}$$

In the output orientation, the decomposition weights are defined similarly, as the proportion of outputs produced from each stage as follows,

$$w_{1j} = \frac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{s=1}^{S} \eta_s y_{sj}}, \quad \text{and} \quad w_{2j} = \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{s=1}^{S} \eta_s y_{sj}}, \tag{3.29}$$

and the overall efficiency is decomposed as the weighted sum of stage efficiencies:

$$\phi_j^0 = w_{1j} \frac{\sum_{p=1}^{P} v_p x_{pj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} + w_{2j} \frac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{s=1}^{S} \eta_s y_{sj}}. \tag{3.30}$$

The output-oriented overall efficiency under the CRS assumption can be obtained through the following linear model

$$\begin{aligned}
\phi_{j_0}^{0*} = \min \ & \sum_{p=1}^{P} v'_p x_{pj_0} + \sum_{q=1}^{Q} \gamma'_q z_{qj_0} \\
\text{s.t.} \ & \sum_{q=1}^{Q} \gamma'_q z_{qj_0} + \sum_{s=1}^{S} \eta'_s y_{sj_0} = 1, \\
& \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, \quad \forall j \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} v'_p x_{pj} \leq 0, \quad \forall j \\
& v'_p, \gamma'_r, \eta'_s > 0, \forall p, r, s,
\end{aligned} \tag{3.31}$$

The linearised, output-oriented models for the overall and the stage efficiencies, when each stage is considered as priority stage, and under the CRS assumption are given in Table 3.1.

Table 3.1: The Output-oriented models in the additive decomposition approach

1st stage as priority stage	2nd stage as priority stage
CRS	
$$\begin{aligned} \phi_{j_0}^{1p*} &= \min \sum_{p=1}^{P} \upsilon'_p x_{pj_0} \\ \text{s.t. } & \sum_{q=1}^{Q} \gamma'_q z_{qj_0} = 1, \\ & \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, \quad \forall j \\ & \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} \leq 0, \quad \forall j \\ & \sum_{p=1}^{P} \upsilon'_p x_{pj_0} + \sum_{q=1}^{Q} \gamma'_q z_{qj_0} - \\ & \qquad - \theta^{0*} \sum_{s=1}^{S} \eta'_s y_{sj_0} = \theta_{j_0}^{0*} \\ & \upsilon'_p, \gamma'_r, \eta'_s > 0, \forall p, r, s, \end{aligned} \tag{3.32}$$	$$\begin{aligned} \phi_{j_0}^{2p*} &= \min \sum_{q=1}^{Q} \gamma'_q z_{qj_0} \\ \text{s.t. } & \sum_{s=1}^{S} \eta'_s y_{sj_0} = 1, \\ & \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, \quad \forall j \\ & \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} \leq 0, \quad \forall j \\ & \sum_{p=1}^{P} \upsilon'_p x_{pj_0} + (1-\theta_{j_0}^{0*}) \sum_{q=1}^{Q} \gamma'_q z_{qj} = \theta_{j_0}^{0*} \\ & \upsilon'_p, \gamma'_r, \eta'_s > 0, \forall p, r, s, \end{aligned} \tag{3.33}$$

Remark 3.2.1. *The optimisation process may result in $w_{1j}^* = 0$ or $w_{2j}^* = 0$ for some DMU_j. Then, restrictions can be imposed on the decomposition weights, i.e $w_{1j} \geq c$ and $w_{2j} \geq c$, where $c \in (0, 0.5]$, for $j = 1, 2, ..., N$, in model 3.19. Therefore, for DMU_{j_0} under evaluation, the following two constraints need to be added to the linear model (3.20) when the input orientation is used or to model (3.31) when the output orientation is used:*

$$\begin{aligned} & (c-1) \sum_{p=1}^{P} \upsilon'_p x_{pj_0} + c \sum_{q=1}^{Q} \gamma'_q z_{qj_0} \leq 0 \\ & c \sum_{p=1}^{P} \upsilon'_p x_{pj_0} + (c-1) \sum_{q=1}^{Q} \gamma'_q z_{qj_0} \leq 0 \\ & 0 < c \leq 0.5. \end{aligned}$$

Since by definition $w_{1j} + w_{2j} = 1$, when both decomposition weights are restricted to take values greater than a value c, then c can only take values in the interval $(0, 0.5]$.

Cook et al. (2010) extended the additive decomposition approach to general multi-stage series structures, as well as multi-stage parallel processes.

In the aforementioned models, the decomposition weights are defined endogenously. However, in other additive decomposition approaches, the decomposition weights are predetermined. If the decomposition weights are not defined endogenously, then model (3.19) remains in the following form

$$\begin{aligned} \theta_{j_0}^{0*} = \max\ & w_{1j_0} \frac{\sum_{q=1}^{Q} \gamma_q z_{qj_0}}{\sum_{p=1}^{P} v_p x_{pj_0}} + w_{2j_0} \frac{\sum_{s=1}^{S} \eta_s y_{sj_0}}{\sum_{q=1}^{Q} \gamma_q z_{qj_0}} \\ \text{s.t. } & \theta_j^1 = \frac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj}} \leq 1, \quad j = 1, ..., N \\ & \theta_j^2 = \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \leq 1, \quad j = 1, ..., N \\ & v_p, \gamma_q, \eta_s > 0, \forall p, \forall q, \forall, \end{aligned} \tag{3.34}$$

where $w_{1j_0} + w_{2j_0} = 1$. In this case, model (3.34) cannot be directly converted into a linear one using the Charnes-Cooper transformation, and heuristic algorithms have been developed to solve such models.

Liang et al. (2006) and Lim & Zhu (2013) developed an algorithm to convert model (3.34) into a parametric linear one, applying two Charnes-Cooper transformations simultaneously. For a fixed set of weights, a sequence of linear programs needs to be solved for different values of the parameter. Guo et al. (2017) showed that in this approach, the choice of the decomposition weights affects the overall efficiency score, and introduced a new model where the overall efficiency is only affected by changes in the stage efficiency scores.

As an other parametric approach, Ang & Chen (2016) suggested calculating the stage and overall efficiency scores through the following procedure. First, they suggested calculating the upper bounds of the stage efficiency scores $\bar{\theta}_{j_0}^1$ and $\bar{\theta}_{j_0}^2$, respectively, as

$$\begin{aligned} \bar{\theta}_{j_0}^1 = \max\ & \frac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj}}, \\ \text{s.t. } & \text{the constraints of (3.34)} \end{aligned} \tag{3.35a}$$

$$\begin{aligned} \bar{\theta}_{j_0}^2 = \max\ & \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}}, \\ \text{s.t. } & \text{the constraints of (3.34)} \end{aligned} \tag{3.35b}$$

Then, the lower bounds of the stage efficiency scores $\underline{\theta}_{j_0}^1$ and $\underline{\theta}_{j_0}^1$, respectively, will be obtained by maximising their efficiency score subject to the constraints of model (3.34), while keeping the upper bound level of the other stage's efficiency score unchanged.

Then, model (3.34) can be converted to the following parametric program, where $\theta_{j_0}^1$ is a parameter that belongs to the interval defined by the lower and upper bounds obtained in the previous step.

$$
\begin{aligned}
\max\ & w_{1j_0}\theta^1_{j_0} + w_{2j_0}\frac{\sum_{s=1}^{S}\eta_s y_{sj_0}}{\sum_{q=1}^{Q}\gamma_q z_{qj_0}} \\
\text{s.t.}\ & \frac{\sum_{q=1}^{Q}\gamma_q z_{qj}}{\sum_{p=1}^{P}\upsilon_p x_{pj}} \leq 1, \quad j = 1, ..., N \\
& \frac{\sum_{s=1}^{S}\eta_s y_{sj}}{\sum_{q=1}^{Q}\gamma_q z_{qj}} \leq 1, \quad j = 1, ..., N \\
& \frac{\sum_{q=1}^{Q}\gamma_q z_{qj}}{\sum_{p=1}^{P}\upsilon_p x_{pj}} = \theta^1_{j_0} \\
& \theta^1_{j_0} \in [\underline{\theta}^1_{j_0}, \overline{\theta}^1_{j_0}] \\
& \upsilon_p, \gamma_q, \eta_s > 0, \forall p, \forall q, \forall s.
\end{aligned} \tag{3.36}
$$

The optimal solutions to model will be found by searching within the interval defined by the lower and upper bound of the corresponding efficiency score. If $\theta^1_{j_0}$ is assigned a value, the efficiency score of the second stage is maximised. Therefore, Ang & Chen (2016) suggested approximating model (3.36) with model (3.37) below.

$$
\begin{aligned}
\max\ & \frac{\sum_{s=1}^{S}\eta_s y_{sj_0}}{\sum_{q=1}^{Q}\gamma_q z_{qj_0}} \\
\text{s.t.}\ & \frac{\sum_{q=1}^{Q}\gamma_q z_{qj}}{\sum_{p=1}^{P}\upsilon_p x_{pj}} \leq 1, \quad j = 1, ..., N \\
& \frac{\sum_{s=1}^{S}\eta_s y_{sj}}{\sum_{q=1}^{Q}\gamma_q z_{qj}} \leq 1, \quad j = 1, ..., N \\
& \frac{\sum_{q=1}^{Q}\gamma_q z_{qj}}{\sum_{p=1}^{P}\upsilon_p x_{pj}} = \theta^1_{j_0} \\
& \upsilon_p, \gamma_q, \eta_s > 0, \forall p, \forall q, \forall s.
\end{aligned} \tag{3.37}
$$

Model (3.37) can be converted into a linear one using the Charnes-Cooper transformation. To obtain the optimal efficiency score of the second stage model (3.37) is solved by setting $\theta^1_{j_0}$ equal to its upper bound $\bar{\theta}^1_{j_0}$ and gradually reducing its value by a small value ϵ, until the lower bound $\underline{\theta}^1_{j_0}$ is reached. The optimal overall efficiency score is then obtained as the maximum of the weighted sum of stage efficiency scores. Their approach can also be extended to general multi-stage structures. However, the use of predetermined decomposition weights significantly increases the computational burden compared to the approaches where the decomposition weights are defined endogenously.

Decomposition weight properties in the additive two-stage process, under the VRS

Ang & Chen (2016) and Despotis et al. (2016a) showed that under the CRS assumption, the endogenous definition of the decomposition weights results in a non-increasing relationship between them, i.e. $w_{1j} \geq w_{2j}$, for $j = 1, ..., N$, for some series network structures. This results in giving higher priority to the first stage in the efficiency decomposition. To overcome this problem, Ang & Chen (2016) suggested the use of constant decomposition weights instead, while Despotis et al. (2016a) introduced a novel overall efficiency composition approach that will be discussed in the next Section. In this section, we are going to show that under the VRS assumption, such a non-increasing relationship between the endogenous decomposition weights ($w_{1j} \geq w_{2j}, j = 1, ..., N$) cannot be established.

The four types of two-stage series structures of a DMU$_j$, $j = 1, ..., N$, are depicted in Figure 3.2. Under the CRS, for the decomposition weights of structures of type 1 and 3, in the input-oriented, additive decomposition model, Ang & Chen (2016) showed that $w_{1j} \geq w_{2j}$, for all $j = 1, ..., N$. The relation between the decomposition weights for the four types of two stage structures, under the VRS, is investigated below.

Firstly, we define the decomposition weights in all four types of two-stage structures depicted in Figure 3.2, under the VRS assumption, in the input orientation.

Table 3.2: Decomposition weights for the four two-stage structures, under the VRS assumption, in the input orientation

	Type 1/Type 3 structures	Type 2/Type 4 structures
w_{1j}	$\dfrac{\sum_{p=1}^{P} v_p x_{pj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}$	$\dfrac{\sum_{p=1}^{P} v_p x_{pj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}}$
w_{2j}	$\dfrac{\sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}$	$\dfrac{\sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}}$

The linear version of the model to obtain the overall efficiency in a Type 1 structure, under the VRS assumption, in the input orientation, is given in model (3.38) below. This is the most general version of a two-stage structure. The corresponding models for the other three structures can be actually obtained by removing the second stage specific inputs g_{tj} or the first stage specific output variable l_{rj} or both, to obtain the linear model for the overall efficiency for Type 2, 3 and 4 structure, respectively.

$$
\begin{aligned}
\theta_{j_0}^{0*} = \ \max \ & \sum_{q=1}^{Q} \gamma'_q z_{qj_0} + \sum_{r=1}^{R} \mu'_r l_{rj_0} + u^1 + \sum_{s=1}^{S} \eta'_s y_{sj_0} + u^2 \\
\text{s.t.} \ & \sum_{p=1}^{P} v'_p x_{pj_0} + \sum_{q=1}^{Q} \gamma'_q z_{qj_0} + \sum_{t=1}^{T} \pi'_t g_{tj_0} = 1, \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} + \sum_{r=1}^{R} \mu'_r l_{rj} - \sum_{p=1}^{P} v'_p x_{pj} + u^1 \leq 0, \quad j = 1, ..., N \\
& \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{t=1}^{T} \pi'_t g_{tj} + u^2 \leq 0, \quad j = 1, ..., N \\
& v'_p, \gamma'_r, \mu'_q, \pi'_t, \eta'_s > 0, \\
& u^1, u^2 \text{ free in sign.}
\end{aligned}
\tag{3.38}
$$

Below, the relationship between the decomposition weights in each type of structure, in the input orientation, is investigated.

Type 1 structure:

$$
w_{1j} - w_{2j} = \frac{\sum_{p=1}^{P} v_p x_{pj} - \sum_{q=1}^{Q} \gamma_q z_{qj} - \sum_{t=1}^{T} \pi_t g_{tj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}} \tag{3.39}
$$

From the first inequality constraint in model (3.38) that is derived from the requirement that the efficiency of the first stage should not exceed one, it follows that

$$
w_{1j} - w_{2j} \geq \frac{\sum_{r=1}^{R} \mu_r l_{rj_0} + u^1 - \sum_{t=1}^{T} \pi_t g_{tj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}. \tag{3.40}
$$

Type 2 structure:

$$
w_{1j} - w_{2j} = \frac{\sum_{p=1}^{P} v_p x_{pj} - \sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}} \tag{3.41}
$$

Similarly to the previous case, from the inequality constraint requiring that the efficiency of the first stage does not exceed one, it follows that

$$
w_{1j} - w_{2j} \geq \frac{\sum_{r=1}^{R} \mu'_r l_{rj_0} + u^1}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}}. \tag{3.42}
$$

Type 3 structure:

$$w_{1j} - w_{2j} = \frac{\sum_{p=1}^{P} v_p x_{pj} - \sum_{q=1}^{Q} \gamma_q z_{qj} - \sum_{t=1}^{T} \pi_t g_{tj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}$$
$$\geq \frac{u^1 - \sum_{t=1}^{T} \pi_t g_{tj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}. \tag{3.43}$$

Type 4 structure:

$$w_{1j} - w_{2j} = \frac{\sum_{p=1}^{P} v_p x_{pj} - \sum_{q=1}^{Q} \gamma_q z_{qj}}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}} \geq \frac{u^1}{\sum_{p=1}^{P} v_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj}}. \tag{3.44}$$

In all four cases, the denominator of the final fraction is positive, whereas the sign of the nominator varies, depending on the values of the optimal weights and the value and sign of the scalar u^1.

If the last fraction of any of the inequalities (3.39)-(3.44) was greater (less) than 0, then it would hold that $w_{1j} \geq w_{2j}$ ($w_{1j} \leq w_{2j}$), for all $j = 1, ..., N$ in that structure. That would imply that the relative importance of the stages in the efficiency evaluation of DMU$_j$ would be biased, in favour of the first (second) stage. Nevertheless, the sign of the last fraction in inequalities (3.39)-(3.44) varies among the different DMUs. Therefore, under the VRS assumption, no stage is favoured against the other by definition.

Remark 3.2.2. *In the input-oriented additive decomposition model, under the VRS assumption, the order relation between the endogenously defined decomposition weights is not fixed, but it depends on the optimal input output mix and the first stage scalar (u^1) of each DMU$_j$, $j = 1, ..., N$.*

Hence, unlike the CRS case, under the VRS assumption, the decomposition weights can be defined endogenously without introducing any bias in the production process.

3.2.3 Efficiency composition approaches

The efficiency composition approach was introduced by Despotis et al. (2016*a*) as an alternative to efficiency decomposition approaches, to overcome the non-unique decomposition issues and the bias introduced in the process under the CRS assumption, when the decomposition weight are defined endogenously. In this approach, Despotis et al. (2016*a*) suggested calculating the stage efficiency scores first, where the first stage efficiency score is obtained in the output orientation and the second stage efficiency score is calculated in the input orientation, using the standard DEA approach. The first and the second stage independent efficiency scores, in the output and input orientations, respectively, under the CRS assumption and for a Type 4 structure are given below, by models (3.45) and (3.46), respectively.

1st Stage (output-oriented)

$$\begin{aligned} \min\ & \frac{\sum_{p=1}^{P} v_p x_{pj_0}}{\sum_{q=1}^{Q} \gamma_q z_{qj_0}} \\ \text{s.t.}\ & \frac{\sum_{p=1}^{P} v_p x_{pj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \geq 1, \quad \forall j \\ & v_p, \gamma_q \geq 0, \forall p, q \end{aligned} \tag{3.45}$$

2nd Stage (Input-oriented)

$$\begin{aligned} \max\ & \frac{\sum_{s=1}^{S} \eta_s y_{sj_0}}{\sum_{q=1}^{Q} \gamma_q z_{qj_0}} \\ \text{s.t.}\ & \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \leq 1, \quad \forall j \\ & \gamma_q, \eta_s \geq 0, \forall q, s \end{aligned} \tag{3.46}$$

Combining the constraints of models (3.45) and (3.46), the following augmented models are obtained:

1st Stage (output-oriented)

$$\begin{aligned} \min\ & \frac{\sum_{p=1}^{P} v_p x_{pj_0}}{\sum_{q=1}^{Q} \gamma_q z_{qj_0}} \\ \text{s.t.}\ & \frac{\sum_{p=1}^{P} v_p x_{pj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \geq 1, \quad \forall j \\ & \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \leq 1, \quad \forall j \\ & v_p, \gamma_q, \eta_s \geq 0, \forall p, q, s \end{aligned} \tag{3.47}$$

2nd Stage (Input-oriented)

$$\begin{aligned} \max\ & \frac{\sum_{s=1}^{S} \eta_s y_{sj_0}}{\sum_{q=1}^{Q} \gamma_q z_{qj_0}} \\ \text{s.t.}\ & \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \leq 1, \quad \forall j \\ & \frac{\sum_{p=1}^{P} v_p x_{pj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \geq 1, \quad \forall j \\ & v_p, \gamma_q, \eta_s \geq 0, \forall p, q, s. \end{aligned} \tag{3.48}$$

Models (3.47) and (3.48) can be merged to the bi-objective model (3.49).

$$
\begin{aligned}
\min \quad & \frac{\sum_{p=1}^{P} v_p x_{pj_0}}{\sum_{q=1}^{Q} \gamma_q z_{qj_0}} \\
\max \quad & \frac{\sum_{s=1}^{S} \eta_s y_{sj_0}}{\sum_{q=1}^{Q} \gamma_q z_{qj_0}} \\
\text{s.t.} \quad & \frac{\sum_{p=1}^{P} v_p x_{pj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \geq 1, \quad \forall j \\
& \frac{\sum_{s=1}^{S} \eta_s y_{sj}}{\sum_{q=1}^{Q} \gamma_q z_{qj}} \leq 1, \quad \forall j \\
& v_p, \gamma_q, \eta_s \geq 0, \forall p, q, s.
\end{aligned}
\tag{3.49}
$$

Model (3.49) can be transformed into a MOLP by applying the Charnes-Cooper transformation. Setting $v'_p = \tau v_p$ $\gamma'_q = \tau \gamma_q$ and $\eta' = \tau \eta_s$, where τ is a scalar such that $\tau \gamma_q z_{qj_0} = 1$, results in the following MOLP.

$$
\begin{aligned}
\Theta_{j_0}^{1*} = \min \quad & \sum_{p=1}^{P} v'_p x_{pj_0} \\
\Theta_{j_0}^{2*} = \max \quad & \sum_{s=1}^{S} \eta'_s y_{sj_0} \\
\text{s.t.} \quad & \sum_{q=1}^{Q} \gamma'_q z_{qj_0} = 1 \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} v'_p x_{pj} \leq 0, \quad \forall j \\
& \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, \quad \forall j \\
& v'_p, \gamma'_q, \eta'_s \geq 0, \forall p, q, s.
\end{aligned}
\tag{3.50}
$$

If the first and the second objective function are optimised independently, the point $(\Theta_{j_0}^{1*}, \Theta_{j_0}^{2*})$ comprises the ideal point of model (3.50). Model (3.50) can be transformed into a LP with a single objective function using a scalarisation approach. Additive aggregation of the two objective functions is one scalarisation approach that can be implemented, and results in the following LP model.

$$\begin{aligned} \min \ & \sum_{p=1}^{P} \upsilon'_p x_{pj_0} - \sum_{s=1}^{S} \eta'_s y_{sj_0} \\ \text{s.t.} \ & \sum_{q=1}^{Q} \gamma'_q z_{qj_0} = 1 \\ & \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} \leq 0, \quad \forall j \\ & \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, \quad \forall j \\ & \upsilon'_p, \gamma'_q, \eta'_s \geq 0, \forall p, q, s. \end{aligned} \tag{3.51}$$

Another scalarisation approach implemented by Despotis et al. (2016*a*) to transform model (3.50) into a LP model was by using the L_∞ norm (Tchebycheff norm). This is used to formulate the following min-max model that detects a unique solution on the Pareto front, for which the deviations from the ideal point $(\Theta^{1*}_{j_0}, \Theta^{2*}_{j_0})$ are minimised and equal.

$$\begin{aligned} \min \ & \delta \\ \text{s.t.} \ & \sum_{p=1}^{P} \upsilon'_p x_{pj_0} - \Theta^{1*}_{j_0} \leq \delta \\ & \Theta^{2*}_{j_0} - \sum_{s=1}^{S} \eta'_s y_{sj_0} \leq \delta \\ & \sum_{q=1}^{Q} \gamma'_q z_{qj_0} = 1 \\ & \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} \leq 0, \quad \forall j \\ & \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \gamma'_q z_{qj} \leq 0, \quad \forall j \\ & \upsilon'_p, \gamma'_q, \eta'_s \geq 0, \forall p, q, s. \end{aligned} \tag{3.52}$$

After an optimal solution has been obtained, either by solving model (3.51) or model (3.52), the first and second stage efficiency scores for DMU_{j_0} can be calculated as

$$\theta_{j_0}^{1*} = \frac{\sum_{q=1}^{Q} \gamma_q'^* z_{qj_0}}{\sum_{p=1}^{P} \upsilon_p'^* x_{pj_0}} = \frac{1}{\sum_{p=1}^{P} \upsilon_p'^* x_{pj_0}} \qquad \text{and} \qquad \theta^{2*} = \frac{\sum_{s=1}^{S} \eta_s'^* y_{sj_0}}{\sum_{q=1}^{Q} \gamma_q'^* z_{qj_0}} = \sum_{s=1}^{S} \eta_s'^* y_{sj_0} \quad (3.53)$$

As it was discussed previously in this Chapter, in NDEA the duality does not hold, and projections of the inefficienct DMUs based on the efficiency scores obtained through the multiplier model do not necessarily reach the efficient frontier. Therefore, the envelopment model that is equivalent to model (3.51) will be used to calculate the frontier projections. The envelopment equivalent model is given below.

$$\begin{aligned}
\max\ & \varepsilon \sum_{p=1}^{P} \mathfrak{s}_p^- + \varepsilon \sum_{s=1}^{S} \mathfrak{s}_s^+ \\
\text{s.t.}\ & \sum_{j=1}^{N} \lambda_j x_{pj} + \mathfrak{s}_p^- = x_{pj_0} \quad \forall p, \\
& \sum_{j=1}^{N} \mu_j y_{sj} - \mathfrak{s}_s^+ = y_{sj_0}, \quad \forall s, \\
& \sum_{j=1}^{N} \lambda_j z_{qj} \geq \tilde{z}_{qj_0} \\
& \sum_{j=1}^{N} \mu_j z_{qj} \leq \tilde{z}_{qj_0} \\
& \lambda_j, \mu_j \mathfrak{s}_p^-, \mathfrak{s}_s^+ \geq 0, \forall j, p, s,
\end{aligned} \tag{3.54}$$

where $\tilde{z}_{qj_0} = z_{qj_0} + d_q$ and $d_q = (d_1, ..., d_Q)$ is free in sign.

Let $(\lambda^*, \mu^*, \tilde{z}_{qj_0}^*, \mathfrak{s}_s^{-*}, \mathfrak{s}_s^{+*})$ be the optimal solution to model (3.54). The efficient projections of a DMU_{j_0} are given as

$$\hat{x}_{pj_0} = x_{pj_0} - \mathfrak{s}_p^{-*}, \qquad \hat{y}_{sj_0} = y_{sj_0} + \mathfrak{s}_p^{+*}, \qquad \hat{z}_{qj_0} = \tilde{z}_{qj_0}^*. \tag{3.55}$$

The efficiency composition approach can also be applied under the VRS assumption. However, it can only be applied to Type 4 structures and cannot be directly extended into other two-stage structures due to the different orientations used to obtain the first and second stage efficiency scores. Despotis et al. (2016*b*) introduced a novel approach based on the weak-link concept, that allows the extension of the composition approach to different two-stage structures. In this approach, the overall and stage efficiency scores are calculated simultaneously in a two-phase method. Sahoo et al. (2021) suggested a simplified efficiency composition method. They introduced a linear NDEA model that provides the same efficiency assessment, while it is computationally easier than the the two-phase method, and can be directly applied to multi-stage series structures. They also extended their approach to the dynamic efficiency assessment.

3.2.4 Slacks-based measure approaches

The Network slacks-based measure (NSBM) was introduced by Tone & Tsutsui (2009) as an extension to the SBM and weighted SBM to deal with efficiency measurement in production processes with intermediate measures. Tone & Tsutsui (2009) defined the NSBM based on the production possibility set of k-stage processes, where $k = 1, ..., K$.

As it was defined in Section 3.1 let $x_j^k = \left(x_{1j}^k, ..., x_{P_kj}^k\right) \in \mathbb{R}_+^{P_k}$ denote the inputs to stage k, let $z_j^{(k,b)} = \left(z_{1j}^{(k,b)}, ..., z_{Q_kj}^{(k,b)}\right) \in \mathbb{R}_+^{Q_k}$ be the intermediate products of stage k that are used as inputs to another stage b, and finally, let $y_j^k = \left(y_{1j}^k, ..., y_{S_kj}^k\right) \in \mathbb{R}_+^{S_k}$ be the final outputs that exit the system at stage k. Then, the PPS under the VRS assumption can be defined as

$$
\begin{aligned}
&\sum_{j=1}^{N} \lambda_j^k x_{pj}^k \leq x_p, \forall p, k, \\
&\sum_{j=1}^{N} \lambda_j^k y_{sj}^k \geq y_s, \forall s, k \\
&\sum_{j=1}^{N} \lambda_j^k z_{qj}^{(k,b)} = z_q, \forall q, (k,b) \text{ (as outputs from k)} \\
&\sum_{j=1}^{N} \lambda_j^b z_{qj}^{(k,b)} = z_q, \forall q, (k,b) \text{ (as inputs to k)} \\
&\sum_{j=1}^{N} \lambda_j^k = 1 \\
&\lambda_j^k \geq 0, \forall j, k
\end{aligned}
\tag{3.56}
$$

where $\lambda_j^k \in \mathbb{R}_+^N$ is the intensity vector of stage k, $k = 1, ..., K$.

Tone & Tsutsui (2009) adopted the two following approaches to treat intermediate products:

(a) The "free" link approach, where the flow of the intermediate products is freely determined to preserve the continuity of being outputs of one stage and inputs to another stage:

$$\sum_{j=1}^{N} \lambda_j^b z_{qj}^{(k,b)} = \sum_{j=1}^{N} \lambda_j^k z_{qj}^{(k,b)}, \forall (k,b). \tag{3.57}$$

(b) The "fixed" link approach, where the intermediate flows are kept unchanged at the current level:

$$\begin{aligned} z_{qj_0}^{(k,b)} &= \sum_{j=1}^{N} \lambda_j^b z_{qj}^{(k,b)}, \forall (k,b), \\ z_{qj_0}^{(k,b)} &= \sum_{j=1}^{N} \lambda_j^k z_{qj}^{(k,b)}, \forall (k,b). \end{aligned} \tag{3.58}$$

Then, the DMU_{j_0} under evaluation can be represented as

$$\begin{aligned} x_{pj_0}^k &= \sum_{j=1}^{N} \lambda_j^k x_{pj}^k + \mathfrak{s}_s^{k-}, \quad \forall k \\ y_{sj_0}^k &= \sum_{j=1}^{N} \lambda_j^k y_{sj}^k - \mathfrak{s}_s^{k+}, \quad \forall k \\ &\sum_{j=1}^{N} \lambda_j^k = 1 \\ &\lambda_j^k, \mathfrak{s}_s^{k-}, \mathfrak{s}_s^{k+} \geq 0, \forall j, k \end{aligned} \tag{3.59}$$

and also (3.57) or (3.58) holds. The above representation refers to the VRS case. If the constraint $\sum_{j=1}^{N} \lambda_j^k = 1$ is omitted, then the DMU_{j_0} under the CRS assumption will be represented. However, under the CRS assumption, Tone & Tsutsui (2009) noted that in contrast to one-stage structures, their suggested NSBM may yield all the DMUs included in the set under evaluation as inefficient.

Similarly to the SBM for one-stage structures, Tone & Tsutsui (2009) suggested three types of NSBM, the input-oriented, the output-oriented and the non-oriented. Those three versions of the NSBM are briefly presented below.

Input orientation: The input-oriented overall efficiency score of DMUj_0 is calculated as the weighted arithmetic mean of stage efficiencies through the following model, using either the "free" link or the "fixed" link approach.

$$\theta_{j_0}^{*} = \min \sum_{k=1}^{K} w^k \left[1 - \frac{1}{P_k} \left(\sum_{p=1}^{P_k} \frac{\mathfrak{s}_p^{k-}}{x_{pj_0}^k} \right) \right]$$
$$\text{s.t. (3.59) and (3.57) or (3.58)} \tag{3.60}$$

where $\sum_{k=1}^{K} w^k = 1$, $w^k \geq 0$ represents the relative importance of stage k.

The optimal stage efficiency score can be obtain using the optimal input slacks $\mathfrak{s}_p^{k-*}$ from model (3.60) as

$$\theta_{j_0}^{k*} = 1 - \frac{1}{P_k} \left(\sum_{p=1}^{P_k} \frac{\mathfrak{s}_p^{k-*}}{x_{pj_0}^k} \right), \quad \forall k \tag{3.61}$$

Output orientation: Similarly, the output-oriented overall efficiency score of DMU j_0 is calculated as the weighted harmonic mean of stage efficiencies through the model presented below.

$$1/\phi_{j_0}^{*} = \max \sum_{k=1}^{K} w^k \left[1 + \frac{1}{S_k} \left(\sum_{s=1}^{S_k} \frac{\mathfrak{s}_s^{k+}}{y_{sj_0}^k} \right) \right]$$
$$\text{s.t. (3.59) and (3.57) or (3.58)} \tag{3.62}$$

where $\sum_{k=1}^{K} w^k = 1$, $w^k \geq 0$.

The optimal stage efficiency score can be obtain using the optimal output slacks $\mathfrak{s}_p^{k+*}$ from model (3.62) as

$$\phi_{j_0}^{k*} = \frac{1}{1 + \frac{1}{P_k} \left(\sum_{p=1}^{P_k} \frac{\mathfrak{s}_p^{k+*}}{x_{pj_0}^k} \right)}, \quad \forall k \tag{3.63}$$

Non-oriented case: In the non-oriented case, input and output slacks are considered simultaneously, as follows:

$$\rho_{j_0}^{*} = \min \frac{\sum_{k=1}^{K} w^k \left[1 - \frac{1}{P_k} \left(\sum_{p=1}^{P_k} \frac{\mathfrak{s}_p^{k-}}{x_{pj_0}^k} \right) \right]}{\sum_{k=1}^{K} w^k \left[1 + \frac{1}{S_k} \left(\sum_{s=1}^{S_k} \frac{\mathfrak{s}_s^{k+}}{y_{sj_0}^k} \right) \right]}$$
$$\text{s.t. (3.59) and (3.57) or (3.58)} \tag{3.64}$$

and $\sum_{k=1}^{K} w^k = 1$, $w^k \geq 0$.

Let ($\mathfrak{s}_p^{k-*}$,$\mathfrak{s}_p^{k+*}$) be the optimal input and output slacks obtained from model (3.64). The optimal stage efficiency score can be defined as

$$\rho_{j_0}^{*k} = \frac{1 - \frac{1}{P_k}\left(\sum_{p=1}^{P_k} \frac{\mathfrak{s}_p^{k-*}}{x_{pj_0}^k}\right)}{1 + \frac{1}{S_k}\left(\sum_{s=1}^{S_k} \frac{\mathfrak{s}_s^{k+*}}{y_{sj_0}^k}\right)}, \quad \forall k \tag{3.65}$$

All the above overall and stage efficiencies are independent from the units of measurement, i.e. they are units-invariant. In contrast to the input and output-oriented cases, in the non-oriented case the relationship between the overall and stage efficiency score cannot be established. Lu et al. (2014) introduced a modification of the non-oriented NSBM where the overall efficiency score is obtained as the arithmetic mean of the stage efficiencies.

Fukuyama & Mirdehghan (2012) noted that the models introduced by Tone & Tsutsui (2009) do not consider the slacks of intermediate products, and therefore, do not provide correct identification of the efficiency of a DMU. Chen et al. (2013) further remarked that such envelopment-based approaches should only be used to obtain the frontier projections of DMUs and multiplier-based approaches should be employed to calculate the efficiency scores.

The NSBM models introduced by Tone & Tsutsui (2009) assume cooperation between the stages. Defining the PPS in a similar way, where the different stages cooperate with each other, Lozano (2016) studied a general two-stage structure with undesirable products and Li et al. (2015) applied a NSBM model in a series structure. Other NSBM approaches are based on a PPS that can be separated into two independent parts that correspond to the different stages. Fukuyama & Mirdehghan (2012) used such an independent PPS to identify non-dominated DMUs in a two-phase procedure. Another group of studies assume a relational PPS, where the aggregated intermediate products of one stage are equal to the aggregated intermediate inputs of the next stage. For example, Lozano (2015) introduced a relational NSBM model where the slacks are not calculated at a stage level, but at a system level, i.e. considering the total consumption of inputs and production of outputs, and not within a specific stage. Kao (2018) provided a thorough review and classification of NSBM approaches.

3.2.5 Network approach

In the network approaches, the efficiency of DMUs is assessed considering that their production process consists of different sub-processes, but the efficiency of the different production stages is not examined. Färe (1991), Färe & Whittaker (1995) and Färe & Grosskopf (1996) adopted the network approach in the efficiency evaluation, in the presence of intermediate products.

The model used by Färe & Grosskopf (1996) for a Type 4 two-stage structure with no exogenous inputs/outputs, under the CRS assumption is similar to model (3.14), and its dual is given by model (3.15). Therefore, the network model suggested by Färe & Grosskopf (1996) is equivalent to the multiplicative model of Kao & Hwang (2008) and the game-theoretic model of Liang et al. (2008).

3.3 Conclusion

Various Network DEA models have been developed as an extension to conventional DEA models to assess the performance of DMUs when the production involves multiple stages. This Chapter has mainly focused on two-stage structures. The PPS of a general two-stage process has been defined and the main approaches to deal with efficiency measurement in two-stage production structures have been discussed.

In contrast to one-stage DEA models, in network DEA, the returns to scale properties do not always hold, as in some cases the CRS efficiency score of a DMU may be higher than its efficiency score under the VRS assumption. Furthermore, the duality between the envelopment and the multiplier model does not always hold, and the projections obtained with the multiplier model might not lie on the efficient frontier. Therefore, the multiplier version should be used to calculate the overall and stage efficiency scores, and the envelopment form should be employed to obtain the frontier projections.

Network DEA has the advantage over the conventional DEA that it can offer an insight into the production process and reveal which production stage comprises the main source of inefficiency. Among the different NDEA approaches, the efficiency decomposition approaches seem to be the most influential ones, with the additive decomposition approach offering the benefit of being applicable under general multi-stage structures and the VRS assumption. In the following Chapter, the additive decomposition approach is going to be implemented to assess the performance of the railway transport process in European countries.

4

Empirical study: Noise-pollution efficiency analysis of European railways

An efficient transport system is critical for attaining economic growth; it allows for the movement of people, goods and resources, provides access to services and facilities and enhances the quality of life. Governments need to invest in transport infrastructure while being considerate of sustainable development. During the last decades, the important share of transport in energy consumption and air pollution increased concerns about the impacts it will have on the natural ecosystem and climate change. Since railways are the most eco-friendly means of transport while demonstrating the lowest traffic congestion levels and high safety performance, they are becoming increasingly popular, and plans regarding their improvement and broader adoption are included in many governmental agendas.

However, railway noise generated by the wheel-rail contact, as well as aerodynamic noise were proved to create a major environmental problem. Specifically in the European Union (EU), railways are considered to be the second-highest source of noise pollution after road, both inside and outside urban areas (EEA Report No 22/2019 2020). Prolonged noise exposure is linked to well-being and health problems, such as sleep disturbances, annoyance and higher risk of cardiovascular diseases (EEA Report No 10/2014 2014). Railway noise is also related to economic costs, such as the depreciation of houses close to rail lines and productivity decrease of the employees, when railway noise exists in the workplace.

Efficiency evaluation of railways is very important in order to identify its sources of inefficiency and further improve its operation. It is crucial for the society and the economy to keep this sustainable mode of transport competitive and manage to form the modal split in its favour. DEA can be a useful tool in measuring the technical efficiency of DMUs relatively to an empirically constructed production frontier, which allows the comparison among the different DMUs. Its great advantage lies in its non-parametric nature which allows for the inclusion of multiple inputs and outputs in the production model. Therefore, there is a large number of studies using DEA models to assess the efficiency of railways in different geographical areas.

The purpose of this Chapter is to provide an evaluation of the efficiency of railways in the European countries during 2016-2017, considering the noise pollution generated, and its impact on humans. However, the railway transportation process can be considered to involve more than one stages, that are related to asset management and services provision, respectively. For this reason, a network DEA (NDEA) model with intermediate and undesirable outputs is formulated upon the assumption of variable returns to scale (VRS).

As distinct from the conventional one-stage DEA approaches, there is a quite limited number of studies in the DEA literature that consider the inner structure of railways' operation. In this study, we suggest that the final output of the railway transport process, i.e. passenger and freight movement, is a result of a two-stage process, the first one of which is related to asset management and the second one to the service offering. Ignoring the role of one of the two stages may result in misleading conclusions for the efficiency level of a country's railway system. The number of people exposed to high levels of railway noise is considered as an undesirable output. Despite the extended DEA literature, to the best of our knowledge, this is the first time that DEA is being used to incorporate noise effects on the operation of railway transport. Using the additive efficiency decomposition approach, it is possible to define the source of inefficiency for each country. Furthermore, a sensitivity analysis is performed to investigate how the different choices of stage weights affect the efficiency scores and rankings of the countries.

The rest of this Chapter is structured as follows. Section 4.1 is a review of the relevant DEA and NDEA literature on railway transport. In Section 4.2, the noise-pollution problem caused by railways in Europe is discussed. Then, in Section 4.3 the railway transport process is decomposed into two stages and in Section 4.4, a two-stage NDEA model with undesirable output is formulated. In Section 4.5, the results of the analysis are discussed. Finally, in the last Section, conclusions, main contributions of this study and some future research directions are provided. This Chapter is based on the published work by Michali et al. (2021).

4.1 Review of the DEA literature on railway transport

Many studies have aimed their attention at measuring the performance of the railway transport industry. The first studies that used DEA in this direction were in the 1990s, from Moesen (1994) and Oum & Yu (1994). Since then, many studies have assessed railway performance globally, measured their performance in specific regions. Below, some selected studies referring to different geographical regions are mentioned. The Section will then focus on studies that evaluated the efficiency of European railways, and their evolution in terms of efficiency since the late 1980's will be discussed. Finally, studies that consider undesirable outputs resulting from the railway operation and studies that take into account the inner structure of the railway transport process will be reviewed.

Graham (2008) assessed the efficiency and productivity of 200 urban railways globally, non-parametrically, using VRS and CRS DEA, and parametrically, by decomposing the total factor productivity (TFP) change. Yu (2008b) used directional distance functions and NDEA to measure the efficiency of 40 global railways. Yu & Lin (2008) measured passenger and freight services' efficiency and effectiveness of 20 railway companies using a multi-activity NDEA. Kutlar et al. (2013) evaluated the technical and allocative efficiency of 31 railway companies using CCR and BCC DEA models and used a second stage Tobit regression to test the impact of outputs on the efficiency scores. Focusing on the US, Chapin & Schmidt (1999) measured the performance of class I railroads using the CCR and BCC DEA models on panel data and Shi et al. (2011) using sequential DEA and Malmquist index (MI). Marchetti & Wanke (2017) used CCR and BCC DEA models to assess the efficiency of rail concessionaires in Brazil and a second stage bootstrap truncated regression to measure the impact of exogenous variables on the efficiency scores. Li et al. (2018) used CCR DEA and generalised DEA to measure the efficiency of Chinese railway administrations and Kuang (2018) applied BCC and super cross-efficiency DEA to assess the efficiency of China Railway Bureau. Jitsuzumi & Nakamura (2010) used BCC DEA to identify the sources of inefficiency in Japanese railways and calculate the optimal levels of subsidies. Mapapa (2004) applied CCR and BCC DEA and MI to evaluate the performance of sub-Saharan African railways. Mohajeri & Amin (2010) combined DEA and analytical hierarchy process (AHP) for the selection of the optimal railway station location in Mashhad. Rayeni & Saljooghi (2014) assessed and compared the efficiencies of Iranian railways over a 30-year period using cross-efficiency DEA. Azadeh et al. (2018) assessed the performance of Tehran-Karaj railway electrification system using BCC DEA and considering resilience engineering (RE). George & Rangaraj (2008) applied CCR and cross efficiency DEA to measure the performance of Indian railways. Bhanot & Singh (2014) used CCR and BCC DEA to compare indicators of business performance of Indian Railway container transport. Sharma et al. (2016) assessed the performance of 16 railway zones in India in terms of the services they provide applying BCC DEA and Malmquist Index. Kim et al. (2011) studied the modal shift to railways in Korea, as a more environmental means of transport. They measured railway freight transport efficiency applying CCR and BCC DEA models and made suggestions about how to expand the use of railways in freight transportation. Reorganisation, incorporation or privatization as well as passenger services, freight carriage, safety and energy consumption of railways are some common research topics in DEA literature. Mahmoudi et al. (2020) provided an extended review of DEA applications on the transportation and railway industry.

In the late 1980s, the need for increasing railways' eroded modal market share and coping with the new demands arising from globalisation sparked a series of reforms in European railways. That stimulated many studies to assess the performance of the railway system in Europe before and after such transformations, to extract useful conclusions towards its efficiency improvement. Oum & Yu (1994) assessed the efficiency of railways which were mainly focused on passenger services in 18 European countries and Japan, during the time period from 1978 to 1989. They suggested that managerial autonomy and less dependence on subsidies have a positive effect on the efficiency of a railway system. UK, Ireland, Netherlands, Spain, and Sweden had the most efficient railway systems during that period. According to Cantos et al. (1999), in the years that followed - from 1985 to 1995 - the financial and managerial autonomy continued to have a positive impact on the efficiency of railways and that reforms that took place, resulted in increasing productivity. During the same time period, smaller European railway companies seemed to have higher technical efficiency (De Jorge-Moreno & Garcia-Cebrian 1999).

Coelli & Perelman (1999) investigated the efficiency of 17 European railway firms from 1988 to 1993, using distance functions. De Jorge-Moreno & Suarez (2003) used quadratic functions to observe the efficiency of 19 railway firms in Europe for a long time period, from 1965 to 1998. They concluded that the separation of railway operations from railway infrastructure management - introduced in 1991 - and the reductions in personnel, affected the efficiency of firms. Hilmola (2007) studied the productivity and efficiency of freight railways in 31 European countries, from 1980 to 2003. During the 1990s there was an efficiency downfall of all the previously best-performing countries. Also, a high level of divergence in freight transport among the European countries was observed (Hilmola 2007; 2008; Savolainen & Hilmola 2009). Countries in the Baltic region, and notably Estonia and Latvia, were performing better in freight transport. However, in passenger transport, Netherlands, UK, Spain and Denmark were more efficient, while Eastern European countries were showing low performance (Hilmola 2008). Savolainen & Hilmola (2009) suggested that an associated development of railway and airline passenger transport would probably increase the efficiency. Growitsch & Wetzel (2009) used a DEA super-efficiency model with bootstrapping on 54 railway European firms during 2000-2004 and found that vertical integration favours the performance improvement in the majority of the railways included in the study. Cantos et al. (2010) studied the vertical and horizontal separation of railways in 16 European countries for the time period 1985-2005.

In more recent years, Sozen et al. (2012) and Sozen & Cipil (2018) compared Turkish railways to the railways of 23 EU member countries. Rotoli et al. (2015) considered accessibility among the European countries and suggested it could be improved by giving importance to the increase of railway speed. Rotoli et al. (2018) ranked the efficiency of Italian rail segments, from the standpoint of three different stakeholders; rail regulator, rail operator and the infrastructure manager. Khadem Sameni et al. (2016) were the first to implement DEA to assess the efficiency of 96 railway stations in Great Britain in terms of how well they manage train stops considering the existing station capacity.

Although railways are one of the safest means of transport, reduction of existing safety risks such as train collisions, derailments, level crossings or exposure of railway stuff to moving trains and electricity of high voltage can further improve its sustainability. Concerning railway safety in Europe, Noroozzadeh & Sadjadi (2013) measured the efficiency of 25 European passenger railways in 2008. Djordjević et al. (2018) used a non-radial DEA model to assess the efficiency of European railways regarding their level of safety in railway level crossings, during 2010-2012 and 2014. Roets et al. (2018) measured the efficiency of railway traffic control centres in Belgium in 2015, using cost allocation restrictions and a metafrontier approach. In such studies, the number of accidents, number of victims, surveillance staff, number of safety and non-safety interventions are some of the variables used to measure railway safety.

Within the global movement towards decreasing greenhouse gas emissions, some studies used DEA models to assess the energy-environmental efficiency of transport considering CO_2 emissions as an undesirable output. The majority of those studies refers to China (Chang et al. 2013; Cui & Li 2014; Zhou et al. 2014). Concerning railway transport in China, Liu et al. (2016) used a non-radial DEA model, window analysis and a second stage Tobit regression. Song et al. (2016) combined a non-radial DEA model with a second stage panel beta regression, and Liu et al. (2017) applied a SBM DEA model with parallel structure. Ha et al. (2011) measured the environmental inefficiency of railway companies in Japan, considering CO_2 emissions produced both during the train operation and the infrastructure construction. The environmental efficiency of railways in the EU countries during 2014-2016 was studied by Djordjević & Krmac (2019) using a non-radial DEA model. Emrouznejad, Marra, Yang & Michali (2022) provide a literature review and bibliometric analysis that shows the evolution in the DEA studies treating CO_2 as an undesirable output.

In most of the studies in the DEA transport literature, the production process is considered as a 'black box', where inputs are directly transformed to outputs. However, some studies model the production process considering its inner structure. Yu (2008b) measured the efficiency and effectiveness of 40 railways globally, during 2002, using a NDEA model with two sub-processes - production and consumption process - with shared, intermediate and exogenous inputs. Yu & Lin (2008) assessed the efficiency and effectiveness of 20 selected railway companies during the same year. In this study, the production stage was divided into two parallel processes - passenger and freight subprocesses - with stage-specific and shared inputs. Mallikarjun et al. (2014) used a non-oriented four-stage series NDEA model to study the performance of public railway transport in the US. Zhou & Hu (2017) used an additive two-stage NDEA model to measure the performance of railways in China, considering dust as an undesirable output of the second stage. Similarly, the two stages are production and service related respectively. Wanke et al. (2018) applied directional distance functions in a multi-stage NDEA model combining series and parallel structure, with an undesirable output - number of accidents - to evaluate the efficiency of Asian railways.

The number of studies considering the inner structure of the railway transport process is limited compared to the high volume of the NDEA literature and the one-stage DEA literature in railways. Furthermore, there is a lack of studies evaluating the efficiency of railways considering the major issue of noise-pollution that affects the people and the natural habitat in the surroundings of railway lines. Traffic noise-pollution is considered to be a serious problem mainly in Europe, where 113 million people are affected by long-term exposure to traffic noise levels of at least 55 decibels coming from different modes of transport (EEA 2021). In the following Section, the railway noise-pollution problem is discussed in more detail.

4.2 The railway noise-pollution problem in Europe

The most serious problem that railways cause to the environment is noise pollution. Notably in Europe, after road traffic noise, noise generated from railways is the second highest environmental health problem. According to 2017 estimations, about 22 million people were exposed to high levels of railway noise inside and outside urban areas (EEA Report No 22/2019 2020).

Railway noise mainly comes from the wheel and rail vibrations, which are generated by the contact of the rolling wheel with the rail (Kitagawa 2009, pg. 1). Bad rolling conditions originating from the poor maintenance of the rail lines or wheel flats, result in augmented noise levels. The braking technology that is used, also plays an important role; cast iron brake blocks corrugate the wheel surface resulting in higher rolling noise levels, while composite and sinter material blocks cause low roughness to the wheel, and thus, produce less noise (Pyrgidis 2016, pg. 428-429). Until the mid 2000s, cast iron brakes was the only brake technology used in freight wagons, due to their lower cost. On the other hand, the disc brake technology that is used in passenger trains, which are usually high-speed trains, generates lower rolling noise levels (Thomson & Gautier 2006, pg. 400). In this type of trains, aerodynamic noise seems to be the major problem (Thomson et al. 2015).

In contrast to passenger trains which are mainly operating during the day, the majority of freight wagons operate during the night hours, and are therefore considered, under the current brake technology, as the main source of noise pollution in Europe (ERA 006REC1072 Impact Assessment 2018). In 2006, the technical specifications for interoperability (TSI) which were introduced by the European Commission (EC), set noise emission limits for new wagons and implicitly prohibited the use of cast iron blocks (Commission Decision 2006/66/EC 2006). However, since the lifetime of wagons can be over

40 years, the renewal procedure would be very slow. Therefore, in 2008, the Commission announced new measures for noise emissions reduction, which suggested the retrofitting of the existing wagon fleet with composite brake blocks (Council Directive 2008/57/EC 2008). This would result in up to 10dB noise reduction.

Depending on its duration and intensity, noise can affect human health, causing from mild problems such as annoyance, sleeping disturbances or stress to the body to more serious problems such as increased blood pressure, insomnia and risk for cardiovascular diseases. Environmental noise in the classroom - coming from the road, rail and air traffic - is also related to children cognitive impairment (Clark & Stansfeld 2007; EEA Technical Report No 11/2010 2010).

The Environmental Noise Directive (END) (Council Directive 2002/49/EC 2002) defined L_{den} indicator to be used as a threshold against which human exposure to environmental noise is monitored. L_{den} is defined as the yearly average sound pressure level during all days, evenings and nights, where evening sound pressure value has a penalty of 5dB and night value has a penalty of 10 dB, where dB is considered as an A-weighting scale, used to measure loudness corresponding to the frequencies that human ear can perceive. L_{den} is calculated by the following formula:

$$L_{den} = 10 \cdot log \frac{1}{24} \left((\text{day hours}) \cdot 10^{\frac{L_{day}}{10}} + (\text{evening hours}) \cdot 10^{\frac{L_{evening}+5}{10}} + (\text{night hours}) \cdot 10^{\frac{L_{night}+10}{10}} \right),$$

where L_{day}, $L_{evening}$, and L_{night} are the yearly average sound pressure levels over day, evening and night hours, respectively.

According to the END, EU member states should keep environmental noise at levels where $L_{den} \leq 55$ dB and $L_{night} \leq 50$ dB. According to the World Health Organization (WHO) guidelines, noise should not exceed 40 dB during night (WHO 2009, pg. 108-109).

Next, the structure of the railway transport process is discussed.

4.3 The Railway transport process considering the impact of environmental noise

Railways is a capital-intensive industry that relies a lot on investments to maintain, improve and expand its assets, rolling stock and infrastructure, aiming to provide passenger and freight services of high quality and continue to be competitive in the modal market share. Therefore, from the operational perspective, the railway transport process is divided into two stages; the asset related and the services related. The first stage is related to the acquirement of rolling stock that satisfies EC standards and the development of a rail network with adequate line length. In the second stage, the quality of the services offered is evaluated by measuring the passenger and freight carriage as well as the impact that noise generated from the moving trains had on the population.

In this study, the infrastructure investment costs and the operating and maintenance costs are considered as inputs, and the number of wagons that are compliant with the noise standards specified in TSI, the total number of rail wagons (including compliant and non-compliant ones) and the length of operating lines are regarded as outputs of the first stage. The length of the operating lines and the total number of wagons are then introduced as inputs to the second stage, to produce two desirable outputs, million-tonne freight-kilometres (MT-km) and million passenger-kilometres (M-km), and one undesirable output, the total number of people exposed to high levels of railway noise ($L_{den} \geq$ 55 dB) inside and outside urban areas (see Figure 4.1)[1]. The passenger-km and tonne freight-km are calculated as the number of passenger journeys or tones of freight journeys multiplied by the average distance of all journeys, i.e. they represent the transport of one passenger or one tonne of freight, respectively, over one kilometre.

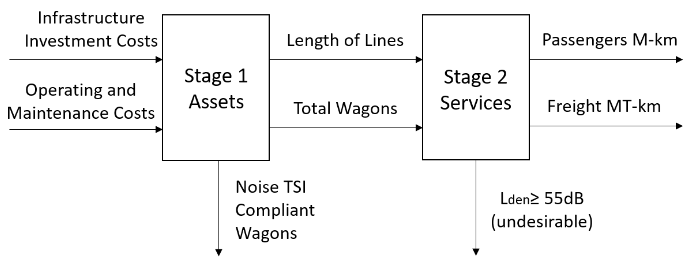


Figure 4.1: Railways model structure

This study assesses the environmental efficiency of railway systems in 22 European countries in 2016-2017. During that time, 20 of the countries under investigation were members of the European Union (EU) - United Kingdom left the EU on 31st January 2020. Switzerland and Norway are also included in the dataset, as they belong to the Schengen area. Concerning the rest of the EU members not included in this study, some of them have missing data and some others, such as Malta and Cyprus, have no railway system.

Data provided in Table 4.1, were collected from various sources. Infrastructure in-

[1] An alternative formulation of the second stage could separate wagons into noise TSI-compliant and noise TSI-non-compliant for a more accurate reflection of the railway transport process.

vestment and operational and maintenance costs were extracted from the 2019 European Commission report (Schroten et al. 2019, pg. 63) and are both measured in billion euros. The number of new and retrofitted wagons which are compliant with the TSI, as well as the total number of wagons in each country, were found in the 2018 report of the European Union Agency for the Railways (ERA) on the noise TSI (ERA 006REC1072 Impact Assessment 2018, pg. 23). The length of operating lines, passenger M-km and freight MT-km were extracted from the Eurostat (2016) database. The number of people exposed to noise levels higher than those established by the END as acceptable was found in Noise Country Fact Sheets 2019, in the European Environment Agency (EEA) webpage (EEA 2020). It should be noted that until 1st January 2019 - when common noise assessment methods (CNOSSOS-EU) started to be applied by all the EU member states - each country was using its own methods for noise pollution measurement, which involved the use of different parameters to capture meteorological conditions, ground absorption or population assignment to buildings (EEA Report No 22/2019 2020). All variables refer to 2016 measurements, except for the last one, which refers to 2017, since noise pollution impact measurements took place in 2007, 2012 and 2017. It should be noted that investment costs refer only to 2016 investments and do not capture any capital stock. Some descriptive statistics of the data set are provided in Table 4.2.

The non-parametric Spearman correlation analysis (Table 4.3) indicates that there is a positive relationship between the input and output variables. This can be interpreted as that an increase in the amount of inputs consumed results in a certain increase in the amount of outputs produced.

Because the efficiency scores calculated with the DEA methodology are relative and not absolute measures of the performance of a DMU, as the number of inputs and outputs increases, the discrimination power of the model diminishes and a number of inefficient DMUs may be falsely rated as efficient (see Section 2.4). There are several rules of thump in the DEA literature for relating the number of inputs and outputs to the sample size (Charles et al. 2019). Following the threshold $N \geq max\{p \times q \times s, 3(p+q+s)\}$, where N is the number of observations, p is the number of inputs, q the number of the first stage outputs, and s the number of second stage outputs, and given that in this case, due to data unavailability, the number of observations is not possible to be increased, there is a need to reduce the model dimensions. Here, the aggregation of the infrastructure investment costs and operation and maintenance costs by simple addition is applied, since these two variables are both measuring different types of costs and are highly correlated (Podinovski

& Thanassoulis 2007).

Table 4.1: Data Set

	DMU	Invest. Costs (bn €)	O&M Costs (bn €)	TSI Wagons	Total Wagons	Length of Lines (km)	Freight MT-km	Pass. M-km	$L_{den} \geq$ 55 dB
1	Austria	2.61	1.66	6511	23345	5491	21361	12497	1081900
2	Belgium	1.78	0.38	2312	12013	3607	0	10025	324400
3	Bulgaria	0.30	0.25	568	16915	4029	3434	1455	42300
4	Croatia	0.19	0.30	383	2274	2604	2160	827	26400
5	Czech Rep.	1.35	1.36	8000	42199	9564	15619	8738	268500
6	Denmark	0.39	0.13	225	366	2045	2616	6332	84300
7	Estonia	0.06	0.14	0	20849	1161	2340	316	6100
8	Finland	0.41	0.18	200	9942	5926	9456	3868	119400
9	France	5.09	3.67	8558	77678	28364	32569	90612	3780000
10	Germany	7.74	3.92	59626	165653	38623	126686	95465	6390500
11	Ireland	0.16	0.21	100	254	1931	101	1991	42600
12	Latvia	0.11	0.17	0	11888	1860	15873	584	40600
13	Lithuania	0.22	0.31	0	14828	1911	13790	280	11600
14	Netherlands	2.73	1.02	9000	21226	3058	6641	17483	312500
15	Poland	3.50	0.69	2750	83500	19132	50650	19067	419700
16	Portugal	0.71	0.26	3123	3313	2546	2774	4266	137100
17	Slovenia	0.23	0.18	226	3230	1209	4360	611	47600
18	Spain	5.23	0.73	6781	20833	16167	10550	26646	69300
19	Sweden	1.07	0.45	931	11000	10882	21406	12800	549400
20	UK	6.46	3.45	15467	18246	16253	17053	68010	1709400
21	Norway	0.52	0.48	516	1623	3895	3312	3695	123400
22	Switzerland	2.50	1.58	19236	21200	3650	12447	20657	482400

Table 4.2: Some descriptive statistics of the data set

	Invest. Costs	O&M Costs	TSI Wagons	Total Wagons	Length Lines	Freight MT-km	Pass. M-km	$L_{den} \geq$ 55 dB
Min.	0.0600	0.1300	0	254	1161	0	280	6100
1st Qu.	0.2475	0.2200	225.20	4970	2170	2908	1589	43850
Median	0.8900	0.4150	1621.5	15872	3772	10003	7535	130250
Mean	1.9709	0.9782	6568.8	26472	8359	17054	18465	730427
3rd Qu.	2.7000	1.2750	7695.2	21220	10552	16758	18671	466725
Max.	7.7400	3.9200	59626.0	165653	38623	126686	95465	6390500

Table 4.3: Non-parametric Spearman correlation matrix

	Invest. Costs	O&M Costs	TSI Wagons	Total Wagons	Length Lines	Freight MT-km	Pass. M-km	$L_{den} \geq$ 55 dB
Invest. Costs	1.000							
O&M Costs	0.864	1.000						
TSI Wagons	0.903	0.882	1.000					
Total Wagons	0.657	0.707	0.645	1.000				
Length Lines	0.840	0.766	0.715	0.586	1.000			
Fr. MT-km	0.597	0.637	0.479	0.689	0.674	1.000		
Pass. M-Km	0.951	0.810	0.888	0.591	0.831	0.561	1.000	
$L_{den} \geq$ 55dB	0.862	0.801	0.822	0.536	0.764	0.616	0.886	1.000

Linear regressions were also used to assess whether the inputs of each stage have a significant predictive value to each of the stage's outputs. Here the undesirable output of the second stage, i.e. the variable ($L_{den} \geq 55$), is considered as input to the second stage. This treatment of the undesirable output is further discussed in the following section, where the models are formulated.

1st stage:

i. (Noise TSI Compliant Wagons)= 3011·(Aggregated costs)−2311.03

ii. (Total Wagons)= 8386·(Aggregated costs)+1740.217

iii. (Length of lines)= 2551·(Aggregated costs)+836.510

2nd stage:

iv. (Passengers M-km)= 2.162·(Length of lines)−0.388· (Total Wagons) +0.013·($L_{den} \geq$ 55dB)+1052.099

v. (Freight MT-km)= −206·(Length of lines)+0.596· (Total Wagons) +0.004·($L_{den} \geq$ 55dB)+217.957

Table 4.4: Regression results

Regress. No.	R^2	significance F	coeff. 1 p-value	coeff. 2 p-value	coeff. 3 p-value
i.	0.613	0.000	0.000	-	-
ii.	0.553	0.000	0.000	-	-
iii.	0.771	0.000	0.000	-	-
iv.	0.925	0.000	0.000	0.004	0.000
v.	0.913	0.000	0.680	0.000	0.206

The $100R^2\%$ gives the percentage of the variability of the output variables that is explained from each regression model. The significance F that is less than 0.05 in all cases, indicates that all the above regressions are statistically significant, and therefore, they have a predictive effect. Although the coefficients 1 and 3 in regression v. are non-significant, the corresponding variables "Length of lines" and "$L_{den} \geq 55$dB" are still included in the model as their coefficients in regression iv. are significant.

In the following Section, the models that are going to be used for the efficiency evaluation of the railway process are provided.

4.4 A two stage Network DEA model with undesirable outputs

In this section, the DEA models that are going to be used to measure the efficiency of railways are formulated.

According to Peterson (1996), among the western European railway companies included in the study, the smallest operators showed increasing returns to scale (IRS), the medium sized operators showed CRS and the largest operators showed decreasing returns to scale (DRS). In this dataset, railways in different countries vary in size, and therefore, small operators co-exist with large ones. Chen et al. (2009, pg. 1170-1171) mention that in cases where large and small DMUs coexist in a dataset, a VRS assumption is required. Since there are CRS, IRS, and VRS operators, a VRS DEA model was adopted in this study to provide a more equitable efficiency analysis, regardless of the size of railway operators[2].

As it was discussed in Chapter 3 the great advantage of the additive decomposition approach against the multiplicative is that the first one can be used under VRS as in that case, the resulting models can be transformed into linear ones, without the need of using any parametric LP. Chen et al. (2009) introduced the additive decomposition method for a closed two-stage process with no external intermediate inputs or outputs (see Section 3.2.2.2). In this Section, the additive approach is going to be applied to the production process discussed in the previous section (see Figure 4.1), which is an open two-stage production process, with external intermediate outputs and undesirable outputs. The assumption introduced by Kao & Hwang (2008), that the optimal level of

[2]For a review of the different models used to assess the efficiency of railways under different returns to scale assumptions see Mahmoudi et al. (2020).

outputs resulting from the first stage is introduced unchanged to the second stage is also adopted. The formulation of the models is done in the input orientation, and under the assumption of VRS.

Consider the railway transport process described in the previous Section, which consists of two serially connected stages (see Figure 4.2). Consider the general case where there are N DMUs. Each DMU_j, $j = 1, ..., N$ consumes P inputs x_{pj}, $p = 1, ..., P$ in the first stage to produce R final outputs l_{rj} and Q intermediate products z_{qj}, $q = 1, ..., Q$, which are then used as inputs in the second stage. From the second stage S good outputs y_{sj}, $s = 1, ..., S$ and D bad outputs $(y_b)_{dj}$, $d = 1, ..., D$ are produced.

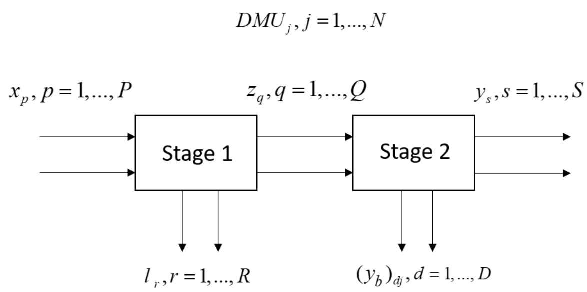


Figure 4.2: Two-stage process with undesirable outputs

If undesirable outputs are treated as normal outputs, then a DMU with lower undesirable products would be falsely considered as less efficient. In this approach, bad outputs produced from the second stage are treated as normal inputs to this stage, and thus, through the optimisation process, the aim is to proportionally decrease inputs to the second stage and undesirable outputs simultaneously. The first and second stage input-oriented efficiency scores of the DMU_{j0}, under the VRS assumption, can be calculated independently one from another as

$$\begin{aligned} \max\ \theta^1_{j0} &= \frac{\sum_{q=1}^{Q} \gamma^A_q z_{qj0} + \sum_{r=1}^{R} \mu_r l_{rj0} + u^A}{\sum_{p=1}^{P} v_p x_{pj0}} \\ \text{s.t. } & \theta^1_j \leq 1, \quad j = 1, ..., N \qquad (4.1a) \\ & v_p, \mu_r, \gamma^A_q > 0, \\ & u^A \text{ free in sign} \end{aligned}$$

$$\begin{aligned} \max\ \theta^2_{j0} &= \frac{\sum_{s=1}^{S} \eta_s y_{sj0} + u^B}{\sum_{q=1}^{Q} \gamma^B_q z_{qj0} + \sum_{d=1}^{D} \xi_t y_{b_{dj0}}} \\ \text{s.t. } & \theta^2_j \leq 1, \quad j = 1, ..., N \qquad (4.1b) \\ & \gamma^B_q, \xi_t, \eta_s > 0, \\ & u^B \text{ free in sign.} \end{aligned}$$

In order to link the two stages, as in Kao & Hwang (2008), it is assumed that for the multipliers of the intermediate products, $\gamma^A_q = \gamma^B_q = \gamma_q$.

Then for a DMU_j, the decomposition weights w_{1j}, w_{2j} can be defined as the proportion of inputs consumed by each stage, as

$$w_{1j} = \frac{\sum_{p=1}^{P} \upsilon_p x_{pj}}{\sum_{p=1}^{P} \upsilon_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{d=1}^{D} \xi_d (y_b)_{dj}}, \tag{4.2}$$

$$w_{2j} = \frac{\sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{d=1}^{D} \xi_d (y_b)_{dj}}{\sum_{p=1}^{P} \upsilon_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{d=1}^{D} \xi_d (y_b)_{dj}}. \tag{4.3}$$

Treating intermediate products as outputs and inputs at the same time, the overall efficiency of DMU$_j$ under VRS, is defined and additively decomposed as

$$\theta_j^0 = w_{1j}\theta_j^1 + w_{2j}\theta_j^2 \tag{4.4}$$

$$= \frac{\sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{r=1}^{R} \mu_r l_{rj} + u^A + \sum_{s=1}^{S} \eta_s y_{sj} + u^B}{\sum_{p=1}^{P} \upsilon_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{d=1}^{D} \xi_d (y_b)_{dj}}, \tag{4.5}$$

where for the decomposition weights, it holds that $w_{1j} + w_{2j} = 1$, $j = 1, 2, ..., N$.

The linearised model for the overall efficiency score is provided below:

$$\begin{aligned}
\theta_{j_0}^{0*} = \ \max \ & \sum_{q=1}^{Q} \gamma_q' z_{qj_0} + \sum_{r=1}^{R} \mu_r' l_{rj_0} + u^1 + \sum_{s=1}^{S} \eta_s' y_{sj_0} + u^2 \\
\text{s.t.} \ & \sum_{p=1}^{P} \upsilon_p' x_{pj_0} + \sum_{q=1}^{Q} \gamma_q' z_{qj_0} + \sum_{d=1}^{D} \xi_d' (y_b)_{dj_0} = 1, \\
& \sum_{q=1}^{Q} \gamma_q' z_{qj} + \sum_{r=1}^{R} \mu_r' l_{rj} - \sum_{p=1}^{P} \upsilon_p' x_{pj} + u^1 \leq 0, \quad j = 1, ..., N \\
& \sum_{s=1}^{S} \eta_s' y_{sj} - \sum_{q=1}^{Q} \gamma_q' z_{qj} - \sum_{d=1}^{D} \xi_d' (y_b)_{dj} + u^2 \leq 0, \quad j = 1, ..., N \\
& (c_0 - 1) \sum_{p=1}^{P} \upsilon_p' x_{pj_0} + c_0 \sum_{q=1}^{Q} \gamma_q' z_{qj_0} + c_0 \sum_{d=1}^{D} \xi_d' (y_b)_{dj_0} \leq 0 \\
& c_0 \sum_{p=1}^{P} \upsilon_p' x_{pj_0} + (c_0 - 1) \sum_{q=1}^{Q} \gamma_q' z_{qj_0} + (c_0 - 1) \sum_{d=1}^{D} \xi_d' (y_b)_{dj_0} \leq 0 \\
& \upsilon_p', \gamma_r', \mu_q', \eta_s', \xi_d' > 0, \\
& u^1, u^2 \text{ free in sign.}
\end{aligned} \tag{4.6}$$

where the last two constraints in model (4.6) are the linearised equivalents of $w_{1j_0}, w_{2j_0} \geq c_0$, for some chosen $c_0 \in (0, 0.5]$. As it was discussed in Remark 3.2.1, these two constraints are included to ensure that model (4.6) will not assign zero decomposition weights to any stage. Regarding the choice of the value c_0, further investigation and discussion is provided in the following Section. The primal model of model (4.6) and the variable-constraint correspondences between the primal and the dual are provided in the Appendix.

Let $(\theta_{j_0}^{0*}, v_p'^*, \gamma_q'^*, \mu_l'^*, \eta_s'^*, \xi_d'^*)$ be the optimal solution to model (4.6). The optimal decomposition weights $w_{1j_0}^*$ and $w_{2j_0}^*$ for DMU_{j_0} are calculated substituting the optimal multipliers into relations (4.2) and (4.3).

If stage one is considered as the priority stage, then, the first stage efficiency of DMU_{j_0} is calculated by maximising $\theta_{j_0}^{1p}$, while maintaining optimal overall efficiency $\theta_{j_0}^{0*}$. The linear model for obtaining the first stage priority score is as follows:

$$
\begin{aligned}
\theta_{j_0}^{1p*} = \ & \max \ \sum_{q=1}^{Q} \gamma_q' z_{qj_0} + \sum_{r=1}^{R} \mu_r' l_{rj_0} + u^1 \\
& \text{s.t.} \ \sum_{p=1}^{P} v_p' x_{pj_0} = 1 \\
& \quad \sum_{q=1}^{Q} \gamma_q' z_{qj} + \sum_{r=1}^{R} \mu_r' l_{rj} + u^1 - \sum_{p=1}^{P} v_p' x_{pj} \leq 0, \quad j = 1, ..., N \\
& \quad \sum_{s=1}^{S} \eta_s' y_{sj} + u^2 - \sum_{q=1}^{Q} \gamma_q' z_{qj} - \sum_{d=1}^{D} \xi_d' (y_b)_{dj_0} \leq 0, \quad j = 1, ..., N \\
& \quad (1 - \theta_{j_0}^{0*}) \sum_{q=1}^{Q} \gamma_q' z_{qj_0} - \theta_{j_0}^{0*} \sum_{d=1}^{D} \xi_d' (y_b)_{dj_0} + \sum_{r=1}^{R} \mu_r' l_{rj_0} + \sum_{s=1}^{S} \eta_s' y_{sj_0} + u^1 + u^2 = \theta_{j_0}^{0*} \\
& \quad v_p', \gamma_r', \mu_q', \xi_d', \eta_s' > 0, \\
& \quad u^1, u^2 \text{ free in sign.}
\end{aligned}
\tag{4.7}
$$

The second stage efficiency score of DMU_{j_0} is obtained by substituting the optimal decomposition weights and $\theta_{j_0}^{1p*}$ in equation (3.25).

Similarly, if stage two is considered as the priority stage, then

$$
\begin{aligned}
\theta_{j_0}^{2p*} = \ \max \ & \sum_{s=1}^{S} \eta'_s y_{sj_0} + u^2 \\
\text{s.t.} \ & \sum_{q=1}^{Q} \gamma'_q z_{qj_0} + \sum_{d=1}^{D} \xi'_d (y_b)_{dj_0} = 1 \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj} + \sum_{r=1}^{R} \mu'_r l_{rj_0} + u^1 - \sum_{p=1}^{P} v'_p x_{pj} \leq 0, \quad j = 1, ..., N \\
& \sum_{s=1}^{S} \eta'_s y_{sj} + u^2 - \sum_{q=1}^{Q} \gamma'_q z_{qj} - \sum_{d=1}^{D} \xi'_d (y_b)_{dj} \leq 0, \quad j = 1, ..., N \\
& \sum_{q=1}^{Q} \gamma'_q z_{qj_0} + \sum_{r=1}^{R} \mu'_r l_{rj_0} + u^1 + \sum_{s=1}^{S} \eta'_s y_{sj_0} + u^2 - \theta_{j_0}^{0*} \sum_{p=1}^{P} v'_p x_{pj_0} = \theta_{j_0}^{0*} \\
& v'_p, \gamma'_r, \mu'_q, \xi'_d, \eta'_s > 0, \\
& u^1, u^2 \text{ free in sign.}
\end{aligned}
\tag{4.8}
$$

Then, using the optimal decomposition weights $w_{1j_0}^*$ and $w_{2j_0}^*$ and the second stage optimum efficiency level, the efficiency of the first stage is calculated through equation (3.27).

4.5 Efficiency analysis

In the model described above, the first stage measures the performance of European countries in building and maintaining their railway infrastructure and rolling stock, while the second stage efficiency measures their performance in providing passenger and freight services considering the less possible environmental noise impact on humans. The efficiency of the whole process and the two sub-processes is evaluated using the additive decomposition methodology elaborated in the previous section while assuming that railways operate under the VRS technology.

If the last two constraints in model (4.6) are omitted, through the optimisation process, for five countries, namely France, Lithuania, Poland, Spain and the UK, the optimal decomposition weights take values $w_1^* = 0$ and $w_2^* = 1$, whereas for Portugal and Switzerland it results that $w_1^* = 1$ and $w_2^* = 0$. That means that for these countries, one of the two stages' contribution to the overall process is ignored. Therefore, the decomposition weight restrictions are incorporated into model (4.6). In order to investigate above what level the decomposition weights should be restricted to lie, a sensitivity analysis of the overall efficiency scores to the decomposition weight restrictions was conducted. Then, efficiency scores were calculated for the chosen decomposition weights' threshold.

4.5.1 Sensitivity Analysis

Sensitivity analysis of the overall efficiency scores was performed for different given values of the lowest allowed level c_0 of decomposition weights, i.e. $w_{ij}, w_{2j} \geq c_0$, $c_0 \in S$, $S = \{0.01, 0.02, 0.03, ..., 0.48, 0.49, 0.5\}$. Results reveal that although for some countries overall efficiency has a very slight downward tendency as c_0 increases, for the majority of countries, overall efficiency scores and optimal decomposition weights are generally stable. Furthermore, for all countries, stage efficiency scores are stable.

Efficiency scores start to be more sensitive to changes of c_0, as c_0 exceeds some threshold, and they are completely destabilised when $c_0 = 0.5$. For $c_0 = 0.5$, Austria and the Netherlands show infeasibility in the stage efficiency models.

Rankings based on the overall efficiency score seem not to be significantly affected for most of the countries, even for large values of c_0. Portugal, Switzerland, Latvia, UK, Bulgaria and Finland are the most sensitive to weight restrictions. Estonia and Germany are overall efficient for all $c_0 \in S$, and Poland is overall efficient for all $c_0 \in S \backslash \{0.47, 0.48, 0.49, 0.5\}$. For space-saving, rankings of the countries for half of the c_0 values - with c_0 step change being 0.02 - are given in Table 4.5.

Portugal and Switzerland are the only countries in the set, for which, for all the different weight restrictions, efficiency decomposition is not unique and changing the priority stage yields different stage efficiency scores. As c_0 increases and restrictions on the decomposition weights become more severe, the optimisation process is forced to assign greater optimal values to some decomposition weights. Therefore, the optimal values of the decomposition weights tend to coincide with the values of c_0 and $1-c_0$ for a growing number of countries. In this analysis, this upturn starts happening for $c_0 \geq 0.14$. Therefore, above that threshold, for some countries the relative contribution of each stage to the overall process is forced to change. However, for most countries, this does not affect their rankings significantly. Nevertheless, as c_0 increases, the number of countries for which it is not possible to have a unique efficiency decomposition rises. For example, four countries do not have unique efficiency decomposition for $c_0 = 0.2$, and ten countries for c_0=0.5.

For the cases when decomposition weight restrictions are needed, there is no rule for choosing a value for c_0 and the choice depends on what the managerial preferences are.

Table 4.5: Overall efficiency rankings for different decomposition weight restrictions

c_0 \ DMU	Austria	Belgium	Bulgaria	Croatia	Czech Rep.	Denmark	Estonia	Finland	France	Germany	Ireland	Latvia	Lithuania	Netherl.	Poland	Portugal	Slovenia	Spain	Sweden	UK	Norway	Switzer.
0.02	21	22	15	17	16	12	1	14	4	1	13	8	6	19	1	10	18	5	11	7	20	9
0.04	21	22	15	17	16	12	1	14	4	1	13	7	6	19	1	10	18	5	11	8	20	9
0.06	21	22	15	17	16	12	1	14	4	1	13	5	7	19	1	11	18	6	10	8	20	9
0.08	21	22	15	17	16	12	1	14	4	1	13	5	7	19	1	11	18	6	10	8	20	9
0.10	21	22	15	17	16	12	1	14	4	1	13	5	7	19	1	11	18	6	10	8	20	9
0.12	21	22	15	17	16	12	1	14	4	1	13	5	7	19	1	11	18	6	10	8	20	9
0.14	21	22	15	17	16	12	1	14	4	1	13	5	7	19	1	11	18	6	10	8	20	9
0.16	21	22	15	17	16	12	1	14	4	1	13	5	7	19	1	11	18	6	10	8	20	9
0.18	21	22	15	17	16	11	1	14	4	1	13	5	7	19	1	12	18	6	10	8	20	9
0.20	21	22	15	17	16	11	1	14	5	1	13	4	7	19	1	12	18	6	10	8	20	9
0.22	21	22	15	17	16	11	1	14	5	1	13	4	7	19	1	12	18	6	10	8	20	9
0.24	21	22	15	17	16	11	1	14	5	1	13	4	7	19	1	12	18	6	9	8	20	10
0.26	21	22	15	17	16	11	1	14	5	1	12	4	7	19	1	13	18	6	9	8	20	10
0.28	21	22	15	17	16	11	1	14	5	1	12	4	7	19	1	13	18	6	9	8	20	10
0.30	21	22	15	17	16	10	1	14	5	1	12	4	7	19	1	13	18	6	9	8	20	11
0.32	21	22	14	17	16	10	1	13	5	1	12	4	7	19	1	15	18	6	8	9	20	11
0.34	21	22	14	17	16	10	1	13	5	1	11	4	7	19	1	15	18	6	8	9	20	12
0.36	21	22	13	17	16	10	1	12	5	1	11	4	7	19	1	15	18	6	8	9	20	14
0.38	21	22	13	17	16	10	1	12	5	1	11	4	7	19	1	15	18	6	8	9	20	14
0.39	21	22	13	17	16	9	1	12	5	1	11	4	8	19	1	15	18	6	7	10	20	14
0.40	21	22	13	17	16	9	1	12	5	1	11	4	8	19	1	15	18	6	7	10	20	14
0.42	21	22	13	16	14	9	1	12	5	1	11	4	8	19	1	17	18	6	7	10	20	15
0.44	21	22	13	15	14	9	1	12	5	1	11	4	8	19	1	17	18	6	7	10	20	16
0.46	21	22	13	15	14	9	1	12	5	1	11	4	8	19	1	17	18	6	7	10	20	16
0.48	21	22	12	15	14	9	1	10	5	1	13	4	8	19	3	17	18	6	7	11	20	16
0.50	21	22	11	15	14	9	1	10	5	1	13	3	7	19	4	17	18	6	8	12	20	16

Large values of c_0 may be too restrictive, impacting on the efficiency scores and sometimes resulting in infeasibility problems. Therefore, there is a range of smaller c_0 values for which efficiency scores and optimal decomposition weight values are not significantly affected. That means that for this range of c_0 values, efficiency scores show low volatility.

In order to deduce the sensitivity threshold, a volatility measure of the overall efficiency scores was evaluated as follows[3]:

1. We calculate the overall efficiency scores of each DMU$_j$, $j = 1, ..., N$, for all $c_0 \in S$

2. For a small $r \in \mathbb{Z}^+$ we calculate the volatility index as the sum of the standard deviations of the overall efficiency scores of each DMU, i.e. $V_{c_0} = \sum_{j=1}^{N} sd\{\theta^{0*}_{j,c_0-r}, ..., \theta^{0*}_{j,c_0+r}\}$

3. We choose the range of c_0 values which minimise V_{c_0}.

Here, the above algorithm is repeated for $r = 1, 2, 3, 4$. Lower values of the volatility index indicate greater stability of the efficiency scores. The resulting volatility indices are presented in Table A2 in the Appendix. As it is shown in Table A2, the volatility index is low and stable for all $c_0 \leq 0.15$ for all $r = 1, 2, 3, 4$. Volatility increases for $c_0 \geq 0.19$,

[3] This algorithm is based on the algorithm suggested by Politis et al. (2001) for the selection of subsample size when applying subsampling bootstrap.

$c_0 \geq 0.18$, $c_0 \geq 0.17$ and $c_0 \geq 0.16$ for $r = 1, 2, 3, 4$ respectively. Therefore, $c_0 = 0.15$ is deduced as an overall sensitivity threshold in this analysis.

However, if c_0 is too small, for some DMUs one stage will be assigned a very low contribution to the overall process. There is a range of c_0 values which are not too restrictive, but also ensure that no stage will be ignored. Table 4.6 shows the overall and stage efficiency scores, as well as the decomposition weights for the case when $c_0 = 0.1$, where the exponent p indicates the priority stage. By imposing $w_{1j}, w_{2j} \geq 0.1$, we prevent one of the two stages to undertake the weight of the whole process, and secondly, $c_0 = 0.1$ lies below the defined sensitivity threshold.

4.5.2 Efficiency scores

According to the results, four countries, Estonia, Finland, Germany and Poland are first-stage efficient. These countries are also efficient in the second stage, except for Finland which shows a relatively low performance in the second stage. In total, 11 out of 22 countries are efficient in the second stage - without including Switzerland. These countries constitute half of the sample, which seems to be a great difference to the number of first-stage efficient countries. However, performing Wilcoxon signed-rank test for the efficiency scores of the two stages, we fail to reject the null hypothesis that the scores of the two stages do not differ significantly, for any level of significance. Also, the Spearman correlation between the stage efficiency scores is zero, indicating that an increase (decrease) in one stage's efficiency score does not imply an increase (decrease) in the other stage's score.

Bulgaria, Croatia, Czech Republic, Finland, Portugal, Sweden and Switzerland - considering the first stage as priority stage - have significantly lower second stage efficiency score. To investigate whether their lower second stage efficiency is due to the number of people affected by noise, L_{den} variable is excluded from the model. In this case, Switzerland and the Netherlands show infeasibility when the first stage is considered as the priority stage. According to the results (see Table A3 in the Appendix for the case when $w_{1j}, w_{2j} \geq 0.1$), overall efficiency scores are lower for all countries except for Estonia and Germany, which remain overall efficient. However, it is not possible to extract a safe conclusion about whether this happens because countries perform relatively well in terms of the number of people affected by railway noise or because the reduction of the model dimensions results in increasing its discrimination power. Similarly, for the majority of

Table 4.6: Efficiency scores and optimal decomposition weights, when $w_{1j}, w_{2j} \geq 0.1$

	DMU	θ^{0*}	w^*_{1j}	w^*_{2j}	θ^{1p*}	θ^{2*}	θ^{1*}	θ^{2p*}
1	Austria	0.4336	0.6219	0.3781	0.2748	0.6949	0.2748	0.6949
2	Belgium	0.3974	0.6892	0.3108	0.3332	0.5397	0.3332	0.5397
3	Bulgaria	0.7038	0.6206	0.3794	0.9088	0.3685	0.9088	0.3685
4	Croatia	0.6653	0.7025	0.2975	0.7270	0.5195	0.7270	0.5195
5	Czech Rep.	0.6673	0.6950	0.3050	0.7356	0.5116	0.7356	0.5116
6	Denmark	0.7629	0.5592	0.4408	0.5760	1	0.5760	1
7	Estonia	1	0.4310	0.5690	1	1	1	1
8	Finland	0.7051	0.5351	0.4649	1	0.3657	1	0.3657
9	France	0.9882	0.1000	0.9000	0.8822	1	0.8822	1
10	Germany	1	0.8531	0.1469	1	1	1	1
11	Ireland	0.7313	0.6844	0.3156	0.7373	0.7181	0.7373	0.7181
12	Latvia	0.9773	0.2785	0.7215	0.9186	1	0.9186	1
13	Lithuania	0.9493	0.1000	0.9000	0.4932	1	0.4932	1
14	Netherlands	0.5480	0.8053	0.1947	0.4387	1	0.4387	1
15	Poland	1	0.4198	0.5802	1	1	1	1
16	Portugal	0.7873	0.9000	0.1000	0.8250	0.4480	0.7995	0.6769
17	Slovenia	0.6600	0.8027	0.1973	0.5764	1	0.5764	1
18	Spain	0.9559	0.1000	0.9000	0.5594	1	0.5594	1
19	Sweden	0.8058	0.3903	0.6097	0.8628	0.7692	0.8628	0.7692
20	UK	0.9380	0.1000	0.9000	0.3804	1	0.3804	1
21	Norway	0.4676	0.7176	0.2824	0.4818	0.4317	0.4818	0.4317
22	Switzerland	0.8936	0.9000	0.1000	0.9552	0.3398	0.8818	1

the countries, the second stage efficiency scores are the same or lower than those when L_{den} is included in the model. Austria and Belgium are the only countries whose second stage efficiency increases when L_{den} variable is omitted.

4.5.3 Policy Implications

Based on the optimal decomposition weights obtained, it is possible to specify which stage is of the highest relative importance for each country. In other words, the optimal decomposition weights can be used by the countries included in the data set as guidance about defining the optimal portion of inputs that they should devote to each stage.

According to the results, 13 out of 22 countries included in the study, namely Portugal, Switzerland, Germany, the Netherlands, Slovenia, Norway, Croatia, Czech Republic, Belgium, Ireland, Austria and Bulgaria, should give more importance to their assets investment, operation and maintenance to improve their efficiency, since, for these countries, the contribution of the first stage to the overall process is higher.

On the other hand, railway industries in France, Spain, the UK and Lithuania, should focus their operation management almost completely on the services they provide, aiming to optimise their freight and passenger carriage, while reducing its noise effects on the environment. Railway operation in Latvia and Sweden should also be more services-related, while Denmark, Finland, Estonia and Poland should give approximately the same balance in their assets and services operation.

Considering the interoperability framework in which European railways operate, to limit the railway noise pollution problem and improve environmental efficiency, changes and measures should be planned and adopted in a cross-country context. Abatement of the railway noise sources in a single country would not resolve the problem and could even harm the competitiveness of railways against other means of transport. Therefore, the common standards set by the European Commission through the Directives can help in this direction. Cooperation and exchange of expertise among the European countries could further foster efficiency improvement of the railway sector.

Furthermore, in reducing railway noise, countries should also focus both on the good maintenance of rail tracks and the increase of the number of wagons that are compliant with the EC standards to achieve the maximum possible noise reduction.

The multiplier model, which was formulated in the previous sections, is used to calculate the efficiency scores of DMUs. In NDEA, it is also possible to provide targets for the input/output variables of each DMU by solving the envelopment form of the model, which is based on the PPS. The envelopment equivalent of model (4.6) is provided in the Appendix (model(1)). However, because of the use of the decomposition weight restrictions that affect the efficiency scores, in this case it is not possible to get the frontier projections even with the envelopment model. Therefore, further study needs to be done on how to get the frontier projections of the DMUs in the cases when the decomposition weight restrictions are necessary.

4.6 Conclusion

Railways have unarguably many advantages, such as higher safety, less energy consumption, less pollution and less traffic congestion, compared to other means of transport. While recognising that the development and maintenance of railways should be given priority, it is vital to take into consideration the impact that railways have on the environment in order to be able to mitigate it. Acknowledging that noise pollution is a major environmental problem caused by railways, this study focused on incorporating it in the efficiency evaluation of the railway transport process.

The railway industry is capital-intensive, and its purpose is to optimise its passenger and freight services. For this reason, the railway transport process was divided into two stages, assets and services-related. The problem of noise pollution is linked to both stages. In the asset stage, good maintenance of the rail lines and retrofitting of the rail wagons with more silent, composite brake technology can mitigate the noise generation. On the other hand, high-quality railway services should entail the minimisation of the number of people affected by railway noise. Therefore, both these factors were taken into account when building the model.

The additive decomposition approach was adjusted to account for intermediate and undesirable outputs. This allowed us to have a better insight into the railways' operation, detect which part of the production process is the main source of inefficiency, and which stage has the highest relative importance for each country.

The performance of railways in 22 European countries during 2016-2017 was studied since the railways' pollution problem seemed to be more significant in this area. The asset, services and overall efficiency scores obtained, revealed that there was no significant difference in the performance of European railways in total, between the two stages. An interesting result is also that, except for Finland, countries which show efficient performance in the asset stage are also efficient in services provision. However, although many countries seemed to be efficient in the second stage, they got a low asset efficiency score, indicating that the inverse relationship did not hold.

The overall efficiency rankings were not significantly affected by imposing different constraints on the decomposition weights of each stage. Consequently, changing the relative importance of each stage, in general, did not affect its relative performance significantly.

This research can be extended by using DEA models to study the railway noise pollution problem in different regions other than Europe. Furthermore, a future study could distinguish between the noise generated by the passenger high-speed trains, and freight wagons or between the impact that railway noise has inside and outside urban areas. Finally, another future research may consider the impact of railway noise on wildlife.

A limitation of this study is that due to decomposition weight restrictions that need to be used to avoid the assignment of zero relative importance in one stage, it is not possible to obtain the frontier projections, i.e. the targets, for the countries, even when the envelopment form of the models is being used. This is an issue of the additive decomposition algorithm in NDEA that needs to be studied further.

Another limitation is that due to data unavailability, the collected variables refer to consecutive years, and this has probably affected the accuracy of the results reported. Furthermore, due to missing data, some European countries were not included in the data set. Since DEA provides relative efficiency measurement, the inclusion or omission of DMUs impacts the efficiency scores of the sample. Therefore, the obtained efficiency scores can only be indicative of the real noise-pollution picture in European railways, as the complete data set of European countries would be needed to have a more accurate efficiency measurement. In most of the cases, the whole population of DMUs is not available, and the efficient frontier is formulated by the available sample of DMUs. Therefore, DEA fails to consider any sampling variation, which would in practice affect the efficiency scores of DMUs.

In general, conventional DEA lacks any consideration of noise in the efficiency measurement process, such as sampling noise, specification or measurement errors. Furthermore, the distribution of the deviations from the frontier is not examined. Stochastic DEA approaches have been developed to deal with these shortcomings, and provide a statistical framework to efficiency measurement. In the following Chapter, the main extensions of the deterministic DEA are discussed.

5

Bootstrapping in DEA

Conventional DEA was developed as a deterministic method of efficiency measurement, without considering any measurement or specification errors, sample noise or data variations. Extensions of the deterministic DEA based on axioms from production theory and statistics or including distributional assumptions have been developed to deal with such shortcomings.

Extensions of the deterministic DEA have been classified by Olesen & Petersen (2016) into three main groups that aim to deal with the aforementioned deficiencies of DEA: (i) approaches that consider sampling noise by treating inefficiencies as random one-side deviations from the frontier, (ii) approaches that consider random noise - coming from measurement or specification errors - and sampling noise simultaneously, by considering random two-sided deviations from the frontier, and (iii) approaches that treat stochastic inputs/outputs by defining a random PPS.

In the first category, the observed set of DMUs is considered a random sample drawn independently and uniformly from an underlying population. Therefore, the true efficient frontier is unknown, and the set of observed inefficiencies is just a random draw from an unknown distribution. Banker (1993) was the first one that interpreted the DEA scores as maximum likelihood estimators. The estimation of random inefficiencies is fully non-parametric since both the structure of the frontier and the error term are specified through postulates.

In the second direction, the error term of the model consists of both random noise and random inefficiencies, and the difficulty in these approaches is how to disentangle those two terms. In the semi-parametric approach introduced by Banker & Thrall (1992) the frontier is specified as a monotone and concave function, but no functional form is assumed. The model residuals consist of a composite term of random noise and random inefficiency. These two terms are assumed to be independent of each other and the inputs and are also assumed to come from specific distributional structures. Then, random noise and random inefficiency estimates can be obtained by maximising the likelihood function reflecting the presumed structure.

The third group of studies allows for random disturbances in the data, such as measurement errors, and it is mainly based on the theory of chance-constraints introduced by Charnes & Cooper (1959). In chance-constrained programming (CCP), some of the coefficients of the LP are random variables following a specific distribution and the expected loss of the criterion is minimised subject to the requirement that the probability that any constraint is violated is bounded. Therefore, it can be said that the efficient frontier envelops the observed DMUs most of the time. Efficiency measures are computed based on probabilistic comparisons among DMUs and DMUs are no more characterised as "efficient" or "inefficient", but as "probably efficient" or "probably inefficient" (Cooper et al. 2004, pg. 229). Again, in these approaches, no functional form is required for the efficient frontier, but parametric assumptions need to be made for the variation of the input and output variables.

There is a high volume of DEA literature in all three directions. This Chapter only focuses on approaches related to the first direction that deals with sampling noise. Since the true frontier is unknown, estimations need to be made based on the observed sample of DMUs. DEA was proven to be a consistent but biased estimator of the true frontier and the asymptotic properties of the DEA estimator have been widely studied. Although results about the asymptotic distribution of the DEA estimator have been extracted, in the case of multiple inputs and outputs the asymptotic distribution can only be approximated by applying bootstrapping techniques. Since the observed sample is the only sample available, it cannot provide direct estimates of the true efficiency scores. The main assumption in bootstrapping is that the observed sample mimics the underlying population. Therefore, by generating many bootstrap samples from the observed sample, we create a bootstrap world that mimics the true world. In bootstrapping, the originally observed sample takes the role of the population and the bootstrap sample is now considered the original sample. If bootstrapping is consistent, asymptotically will provide good approximations of the true efficiencies.

Many DEA studies have focused on the methodological development or the application of bootstrapping techniques. A review of the most influential studies in all three directions can be found Olesen & Petersen (2016). Simar & Wilson (2015) provided a discussion that focuses on the main statistical approaches to DEA, such as bootstrapping techniques and partial frontier approaches. Moradi-Motlagh & Emrouznejad (2022) provided a bibliometric analysis of the bootstrap DEA literature, as well as developments on the relevant software, and an extensive overview of the most impactful articles on the field.

The rest of the Chapter is structured as follows: In Section 5.1, some important concepts that are being used throughout the Chapter are defined. In Section 5.2, the underlying data generating process is defined. Next, in Section 5.3, the properties of the DEA estimator are discussed and the results on its convergence rates are detailed. In Section 5.4, the main principles of the bootstrapping methodology in the DEA context are discussed. In Section 5.5, the bias-corrected estimates are provided, and in Section 5.6, the procedure for constructing the confidence interval estimates is explained. In Section 5.7, the main bootstrap approaches developed in the DEA context are reviewed and consistency issues are discussed in detail. Some extensions of the use of bootstrap in hypothesis testing and some example applications of bootstrapping in the DEA literature are provided in Section 5.8. Finally, in Section 5.9 conclusions are given.

5.1 Preliminaries

In practice, the true PPS and hence, the true efficiency scores are unknown. Based on an observed sample of DMUs, the DEA estimator is used to approximate the true efficient frontier. There are two main properties that a well-defined estimator should have: unbiasedness and consistency. These two properties are defined below.

Let $\widehat{\theta}_n$ be a set of random variables that is used as an estimator of a parameter θ. The bias of the estimator $\widehat{\theta}$ is defined as the difference between the expected value of the estimator and the the true parameter, i.e.,

$$Bias(\widehat{\theta}_n) = E(\widehat{\theta}_n) - \theta \tag{5.1}$$

Definition 5.1.1. ***(unbiasedness)*** *$\widehat{\theta}_n$ is called an unbiased estimator of θ iff $E(\widehat{\theta}_n) = \theta$.*

Regarding the second property of consistency, an estimator has to be consistent in order to give meaningful results, i.e. as the sample size tends to infinity, the estimator needs to converge to the true parameter. There are different types of convergence. In order to define a consistent estimator, first, we need to define two types of convergence of random variables; convergence in probability and almost sure convergence.

Let $X_n, n \in \mathbb{N}$ be a sequence of random variables defined on a sample space Ω.

- X_n converges in probability to the random variable X if $\underset{n\to\infty}{lim} P(|X_n - X| > \epsilon) = 0$, $\forall \epsilon > 0$, and this type of convergence is denoted as $X_n \overset{P}{\to} X$.

- X_n converges almost surely to the random variable X if $P(\underset{n\to\infty}{lim} |X_n - X| = 0) = 1$, and this type of convergence is denoted as $X_n \overset{a.s.}{\to} X$.

Also, if $X_n \overset{a.s.}{\to} X$, then it is implied that $X_n \overset{P}{\to} X$.

Below, the weak and the strong consistency of an estimator are defined.

Definition 5.1.2. ***(consistency)*** *$\widehat{\theta}_n$ is a weakly consistent estimator of θ if $\widehat{\theta}_n \overset{P}{\to} \theta$. $\widehat{\theta}_n$ is a strongly consistent estimator of θ if $\widehat{\theta}_n \overset{a.s.}{\to} \theta$.*

Sometimes it may be difficult to prove the convergence of an estimator and obtain its rate of convergence. Alternatively, common approach of demonstrating the convergence of an estimator is by proving that the estimator is bounded in probability (stochastically bounded). It is considered that $X_n = O_p(c_n)$, if for any $\epsilon > 0$ there exist $0 < M < \infty$ and m, such that for all $n > m$ it holds that $P(|X_n| \geq Mc_n) \leq \epsilon$. If in the aforementioned definition it is set that $c_n = 1$, $\forall n$, then $X_n = O_p(1)$ implies that X_n is bounded in probability.

Definition 5.1.3. ***(stochastic boundedness)*** *A set of random variables X_n is bounded in probability if there exist $0 < M < \infty$ and m such that for all $n > m$ it holds that $P(X_n \geq M) \leq \epsilon$, for any $\epsilon > 0$.*

Note that if $X_n = O_p(c_n)$ then that means that $X_n/c_n = O_p(1)$, and therefore we can say that $X_n c_n^{-1}$ is bounded in probability.

The rate of convergence represents the rate at which the estimation error decreases as the sample size increases. The definition of the convergence rate is given below.

Definition 5.1.4. ***(convergence rate)*** *An estimator $\widehat{\theta}_n$ converges at a rate $n^{-\tau}$, $\tau > 0$ if τ is the largest positive number such that $|\widehat{\theta}_n - \theta| = O_p(n^{-\tau})$.*

Finally, another important concept that is mentioned in this chapter is that of the asymptotic distribution of an estimator.

Definition 5.1.5. ***(asymptotic distribution)*** *Let X_n be a sequence of random variables with distributions F_{X_n}. Then, X_n converges in distribution to a random variable X with distribution F_X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, at all points x at which $F_X(x)$ is continuous.*

Convergence in distribution in denoted by $X_n \overset{D}{\to} X$. Note that both almost sure convergence and convergence in probability imply convergence in distribution.

5.2 The Data Generating Process

Consider a production process where P inputs are consumed to produce S outputs, and let $S_N = \{(x_j, y_j)|j = 1, ..., N\}$ be an observed set of DMUs. Below, the input orientation is considered, but translation to the output orientation is straightforward.

The observed set of DMUs is considered as a sample drawn from a population of DMUs -with unknown distribution $f(x, y)$- through a data generating process $\mathcal{P} = (T, f(\cdot, \cdot))$, where T is the PPS of the population. Therefore, $\mathcal{P}$ is fully characterised by the assumptions made about T, the distribution of inputs and outputs $f(x, y)$, and the sampling process according to which the elements of S_N were drawn. The true PPS T and, as a result, the true efficiency score $\theta(x, y) = inf\{\theta|\theta x \in X(y)\}$ are unknown (for the definition of $X(y)$ see in Section 2.2).

The DGP is founded on the following axioms:

(i) T is a compact convex set

(ii) $(x, y) \notin T$ if $x = 0$ and $y \geq 0$, i.e. all production requires some inputs

(iii) If $(x, y) \in T$ and $x' \geq x$, $y' \leq y$, where $x, x' \in \mathbb{R}^m, y, y' \in \mathbb{R}^r$, then $(x', y') \in T$, i.e. inputs and outputs are strongly disposable

(iv) All observations in S_N are identically independently distributed (i.i.d.) random variables in T

(v) $\theta(x, y|\mathcal{P})$ is differentiable for all $(x, y) \in T$, i.e. the frontier is smooth (usually, $\theta(x, y|\mathcal{P})$ is assumed to be twice continuously differentiable)

(vi) $f(x, y)$ has support T, is continuous on T, positive close to the boundary and strictly positive on the boundary of T.

Now, consider a DMU$_{j_0} \in S_N$ that consumes some inputs x_{j_0} to produce some outputs y_{j_0}. The DEA estimators applied on the sample S_N can be used to provide an estimate

$\widehat{\theta}(x_{j_0}, y_{j_0})$ of $\theta(x_{j_0}, y_{j_0})$. More specifically, what is required is an estimation of the distribution $\widehat{\theta}(x_{j_0}, y_{j_0}) - \theta(x_{j_0}, y_{j_0})$, but with S_N being the only sample available, this distribution cannot be obtained directly.

In the next Section, a review of the main studies on the asymptotic properties of the CRS and VRS estimators is provided.

5.3 Review on the properties of the DEA estimator

Banker (1993) was among the first researchers to consider DEA as a maximum likelihood estimator (MLE) of the efficient frontier. Since then, several studies have focused on investigating the statistical properties of the CRS and VRS DEA estimators in different dimensions and established their rates of convergence. Below, some of the most influential studies in the field are discussed.

For the case of one input and multiple outputs, Banker (1993) proved that if the efficient frontier is considered as a monotonically increasing and concave production function and the one-side deviations from the frontier, i.e. the inefficiencies, are considered to be distributed independently of the inputs and have a monotonically decreasing density function, then DEA is a weakly consistent estimator of the true efficient frontier. That means that as the number of DMUs included in the sample increases, the sample efficiency scores of DMUs tend to the true efficiency values with some probability. However, he showed that for finite samples of DMUs, DEA is a biased estimator, and very large samples are required to achieve a bias that tends to zero.

The convergence rate of an estimator depends on how the difference between the estimator and the true parameter is measured. For the two-dimensional case, Korostelev et al. (1995) measured the divergence between the estimator an the true production set as the Lebesgue measure of their symmetric difference, and proved that DEA is a consistent estimator with optimal rates of convergence. Kneip et al. (1998) extended the results to the multi-dimensional case. Assuming the frontier is smooth and under the consistency requirement that the input-output density is positive close to the frontier and strictly positive on the frontier, they proved that the convergence rate of the VRS-DEA point estimator depends on the smoothness of the frontier.

Let $\widehat{\theta}_{VRS}(x_{j_0}, y_{j_0})$ be the estimator of the true efficiency score $\theta(x_{j_0}, y_{j_0})$ of DMU$_{j_0}$, where $\widehat{\theta}_{VRS}(x_{j_0}, y_{j_0})$ is obtained under the VRS assumption by solving a BCC model (see Section 2.3.2). For the case when the frontier is twice differentiable, Kneip et al. (1998)

proved that for the efficiency score estimate of DMU_{j_0} (fixed point (x_{j_0}, y_{j_0})) it holds that

$$\widehat{\theta}_{VRS}(x_{j_0}, y_{j_0}) - \theta(x_{j_0}, y_{j_0}) = O_p(n^{-\frac{2}{P+S+1}}), \tag{5.2}$$

where P and S are the number of inputs and outputs respectively. Note that the PPS defined by the sample is always a subsample of the population. Therefore, the distance of the DMU under evaluation from the sample best practice is always lower than its distance from the true best practice of the population. In other words, the true efficiency score is always lower than the sample efficiency score. Under the global CRS assumption, Park et al. (2010) adopting the same methodology as Kneip et al. (1998) proved that DEA estimator converges at a faster rate, i.e. under the CRS assumption it holds that

$$\widehat{\theta}_{CRS}(x_{j_0}, y_{j_0}) - \theta(x_{j_0}, y_{j_0}) = O_p(n^{-\frac{2}{P+S}}). \tag{5.3}$$

In all cases, the rate at which the DEA estimator converges to the true frontier depends on the number of inputs and outputs; as the dimensions of the model increase, the number of data records should increase exponentially in order to achieve the same rate of convergence. As an example, consider that the assumptions made in Kneip et al. (1998) hold (a twice differentiable smooth frontier, positive density function close to the frontier, VRS) and the single input -single output case with a set of $N = 100$ DMUs being available. In this case, the rate of convergence would be $100^{-2/3} = 0.0464$. Consider now the case of $P = 2$ inputs and $S = 3$ outputs. In this case, a sample of $N = 100$ DMUs would have a much slower rate of convergence $100^{-2/6} = 0.2154$. In order to achieve the same rate of convergence as in the single input-output case, a sample of $N = 10000$ DMUs would be needed. Simar & Wilson (2008) provided a more detailed discussion on the curse of dimensionality of DEA estimators and a comparison with parametric estimators. Those usually have optimal rate of convergence $N^{-1/2}$, which in the case of multiple inputs/outputs is faster than that of non-parametric estimators. In cases when the size of the sample is not possible to be increased, a reduction in the input/output dimensions is necessary (see Section 2.4 for a reference to some dimension reduction approaches).

For the case of one input and one output, under the VRS assumption, Gijbels et al. (1999) proved that

$$N^{2/3}(\widehat{\theta}_{VRS}(x_{j_0}, y_{j_0}) - \theta(x_{j_0}, y_{j_0})) \xrightarrow{D} F_U(\cdot), \tag{5.4}$$

and derived the analytical form of the distribution $F(\cdot)$. This analytical form depends on two constants. Gijbels et al. (1999) defined the estimators of those constants using the method of moments. This is the only case where the asymptotic distribution can be used in practice to make inference. Jeong & Park (2006) extended their work to higher output dimensions. In the general case of multiple inputs and outputs, it is more difficult to derive a closed expression for the asymptotic distribution of the DEA estimator. For the multivariate case, Kneip et al. (2008) proved that

$$N^{2/P+S+1}\left(\frac{\widehat{\theta}_{VRS}(x_{j_0}, y_{j_0})}{\theta(x_{j_0}, y_{j_0})}\right) \xrightarrow{D} F_M(\cdot), \tag{5.5}$$

and found that the asymptotic distribution $F_M(\cdot)$ is stochastically dominated by a cumulative exponential distribution. However, it is difficult to estimate the distribution's parameters and thus, in practice, this result cannot be used for making inference. Park et al. (2010) extended the asymptotic results to the CRS case, and under the global CRS assumption they found that the DEA estimator follows an exponential distribution. For the special case when $P = S = 1$,

$$n(\widehat{\theta}_{VRS}(x_{j_0}, y_{j_0}) - \theta(x_{j_0}, y_{j_0})) \xrightarrow{D} 1 - e^{-\lambda w}, \tag{5.6}$$

for some parameter λ and for all $w \geq 0$.

5.4 The bootstrap setting

In practice, except for the bivariate case, the only way to get the sampling distribution of $\widehat{\theta}_*(x_{j_0}, y_{j_0}) - \theta(x_{j_0}, y_{j_0})$ -or equivalently of $\frac{\widehat{\theta}_*(x_{j_0}, y_{j_0})}{\theta(x_{j_0}, y_{j_0})}$ - in higher dimensions is by using bootstrapping techniques, where $*$ denotes that the estimators are obtained either under the CRS or the VRS assumption. To simplify the notation, the returns to scale specification for the estimated frontier will be omitted from now on, since what is discussed holds under both the CRS and the VRS assumption.

Bootstrap was first suggested by Efron (1979). Given a random sample $X_n = \{X1, ..., X_n\}$ drawn from an unknown probability distribution f, bootstrapping was introduced as a method to obtain the sampling distribution of a random variable $Y(X_n, f) = \theta_{\widehat{f}} - \theta_f$ based on the observed sample, through simulations. Since the observed sample is the only sample available, the bootstrap methodology is being used for generating bootstrap samples from the original sample by repeatedly simulating the data generating process

(DGP) of the original sample. The main assumption that is made, is that the original sample mimics the population that it comes from. Therefore, by repeatedly drawing samples with replacement from the original sample would be like drawing samples from the population itself. For every bootstrap sample, an estimate $\widehat{Y}$ of the random variable Y is obtained, and when the number of bootstrap samples becomes large, the sampling distribution of Y can be approximated. In what follows, the main principles of the bootstrap methodology within the DEA context are presented and the main bootstrapping approaches are discussed.

As it was mentioned before, the main assumption in bootstrapping is that the original sample S_N generated through the unknown DGP $\mathcal{P}$, mimics the underlying population that it comes from. Therefore, a bootstrap sample $S_N^* = \{(x_j^*, y_j^*)|j = 1, ..., N\}$ generated from the original sample S_N through a known DGP $\hat{\mathcal{P}} = \mathcal{P}(\hat{\Psi}, \hat{f}(\cdot, \cdot))$ can be used to estimate the unknown sampling distribution of $\widehat{\theta}(x_{j_0}, y_{j_0})$, i.e. $\widehat{\widehat{\theta}}(x_{j_0}, y_{j_0})$ - which is obtained from a bootstrap sample - is assumed to be an estimate of $\widehat{\theta}(x_{j_0}, y_{j_0})$. In other words, $\hat{\theta}(x_{j_0}, y_{j_0})$ is an estimator of $\theta(x_{j_0}, y_{j_0})$ obtained from the sample S_N through $\mathcal{P}$, and $\widehat{\widehat{\theta}}(x_{j_0}, y_{j_0})$ is an estimator of $\hat{\theta}(x_{j_0}, y_{j_0})$ obtained from the bootstrap sample S_N^* generated through $\hat{\mathcal{P}}$. If $\hat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$, i.e. $\hat{\mathcal{P}}$ converges to $\mathcal{P}$, for a given DMU_{j_0} it holds that

$$\left(\widehat{\widehat{\theta}}(x_{j_0}, y_{j_0}) - \widehat{\theta}(x_{j_0}, y_{j_0})\right)\bigg|\hat{\mathcal{P}} \sim \left(\widehat{\theta}(x_{j_0}, y_{j_0}) - \theta(x_{j_0}, y_{j_0})\right)\bigg|\mathcal{P}, \tag{5.7}$$

or equivalently,

$$\frac{\widehat{\theta}(x_{j_0}, y_{j_0})}{\widehat{\widehat{\theta}}(x_{j_0}, y_{j_0})}\bigg|\hat{\mathcal{P}} \sim \frac{\theta(x_{j_0}, y_{j_0})}{\widehat{\theta}(x_{j_0}, y_{j_0})}\bigg|\mathcal{P}. \tag{5.8}$$

The distributions of the right-hand side in relations (5.7) and (5.8) are unknown, but Monte Carlo simulations of the left-hand side can provide approximations of them. By generating a sufficiently large number B of bootstrap samples and applying the DEA estimator to each one of those, a set of B estimates $\widehat{\widehat{\theta}}(x_{j_0}, y_{j_0})$ can be obtained[1]. These can be used to derive the distribution of the left-hand side in relations (5.7) and (5.8). Given that $\widehat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$, as $B, N \to \infty$, the approximation of the right

[1]Normally, the notation that would be used for a bootstrap estimate would be $\widehat{\theta}^*(x_{j_0}, y_{j_0})$. Here, the double-hat notation is being used instead of a star to denote the bootstrap estimates to avoid any confusion with the notation used in the previous chapters, where star denotes the optimal value of a parameter. However, the star notation is used as usual for the bootstrap sample.

hand side becomes accurate. Later in this Chapter, the consistency of $\mathcal{P}$ will be discussed in more detail.

5.5 Bias-corrected estimates

As it was mentioned before, DEA was proved to be a biased estimator of the true frontier. To obtain the bias of the DEA estimator, its expected value $E(\widehat{\theta})$ needs to be calculated. This value is unknown, but it can be obtained empirically, by calculating the mean efficiency of the bootstrap samples. Therefore, the bias of the DEA estimator can be obtained as

$$\widehat{Bias}_B(\widehat{\theta}(x_{j_0}, y_{j_0})) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\widehat{\theta}}_b(x_{j_0}, y_{j_0}) - \widehat{\theta}(x_{j_0}, y_{j_0}) \tag{5.9}$$

And the bias-corrected estimator will be

$$\begin{aligned}\widehat{\theta}^{bc}(x_{j_0}, y_{j_0}) &= \widehat{\theta}((x_{j_0}, y_{j_0}) - \widehat{Bias}_B(\widehat{\theta}(x_{j_0}, y_{j_0})) \\ &= 2\widehat{\theta}(x_{j_0}, y_{j_0}) - \frac{1}{B}\sum_{b=1}^{B}\widehat{\widehat{\theta}}_b(x_{j_0}, y_{j_0}).\end{aligned} \tag{5.10}$$

An estimate of the variance of $\widehat{\theta}(x_{j_0}, y_{j_0})$ can be obtained as the variance of the bootstrap samples as

$$\widehat{\sigma}^2 = \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\widehat{\theta}}_b(x_{j_0}, y_{j_0}) - \frac{1}{B}\sum_{b=1}^{B}\widehat{\widehat{\theta}}_b(x_{j_0}, y_{j_0})\right)^2 \tag{5.11}$$

Therefore, if the last term in relation 5.10 is ignored, the variance of the bias-corrected estimator $\widehat{\theta}^{bc}(x_{j_0}, y_{j_0})$ is approximately $4\widehat{\sigma}^2$.

As it was noted by Efron (1993), the bias correction introduces extra noise into the estimation. Therefore, Efron (1993) suggested that the bias-corrected estimators should not be used if

$$\frac{|\widehat{Bias}_B(\widehat{\theta}(x_{j_0}, y_{j_0}))|}{\widehat{\sigma}} \leq \frac{1}{4}. \tag{5.12}$$

5.6 Confidence interval estimates

Construction of the confidence interval estimates for the true efficiency scores is of interest. The approach suggested by Simar & Wilson (1998) for constructing the confidence intervals, uses the bias-corrected estimations. However, as it was mentioned in the previous section, bias correction introduces extra noise into the estimation process. Simar & Wilson (1999) suggested an approach for constructing the confidence interval estimates without using the bias estimations.

For a reason related to the boundary constraints that will be discussed in the next Section, in some cases, instead of estimating θ it is preferred to use the Shephard (1970) distance functions. Let $\delta(x, y) = \frac{1}{\theta(x,y)}$ denote the input distance function, and let $\widehat{\delta}(x, y) = \frac{1}{\widehat{\theta}(x,y)}$ be the input distance function estimate of $\delta(x, y)$. Then, for a confidence level $a \in (0, 1)$ the a-th quantiles $c_{a/2}, c_{1-a/2}$ of the sampling distribution of $\widehat{\delta}(x, y) - \delta(x, y)$ would be obtained by

$$P(c_{a/2} \leq \widehat{\delta}(x, y) - \delta(x, y) \leq c_{1-a/2}) = 1 - a. \tag{5.13}$$

However the distribution $\widehat{\delta}(x, y) - \delta(x, y)$ is unknown and therefore, the empirical distribution $\widehat{\widehat{\delta}}(x, y) - \widehat{\delta}(x, y)$ will be used to estimate the quantiles, i.e.

$$P(\widehat{c}_{a/2} \leq \widehat{\widehat{\delta}}(x, y) - \widehat{\delta}(x, y) \leq \widehat{c}_{1-a/2} | S_N) = 1 - a. \tag{5.14}$$

The quantiles can be obtained by sorting the values $\widehat{\widehat{\delta}}(x, y) - \widehat{\delta}(x, y)$ in ascending order and then setting $c_{a/2,m}$ equal to the first value that exceeds $(a/2)100\%$ of the observations and $c_{1-a/2,m}$ equal to the value that is less than $(1 - a/2)100\%$ of the observations.

Then, $\delta(x, y)$ will lie within the confidence interval

$$[\widehat{\delta}(x, y) - \widehat{c}_{1-a/2}, \widehat{\delta}(x, y) - \widehat{c}_{a/2}], \tag{5.15}$$

and therefore the $(1 - a)100\%$ confidence interval estimate for $\theta(x, y)$ will be

$$\left[\frac{1}{\widehat{\delta}(x, y) - \widehat{c}_{a/2}}, \frac{1}{\widehat{\delta}(x, y) - \widehat{c}_{1-a/2}}\right]. \tag{5.16}$$

5.7 Some bootstrap approaches and consistency issues

As it was mentioned in Section 5.4, in order for the relations (5.7) or (5.8) to hold, it is necessary that $\hat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$. The simplest way to generate a bootstrap sample is by drawing with replacement from the original sample, a bootstrap sample of the same size. This is known as naïve bootstrap. However, it is proven that naïve bootstrapping does not yield consistent boundary estimations (see Bickel & Freedman (1981, pg. 1210), counter-example 2). Because the DEA efficiency score is bounded ($0 \leq \theta \leq 1$ in the input orientation and $\theta \geq 1$ in the output orientation) it follows that the DEA estimates obtained with naïve bootstrapping are inconsistent close to the boundary.

To further explain this, note that the true PPS always includes the PPS defined by the sample, i.e. $\hat{\Psi} \subset \Psi$. That implies that for a DMU$_{j_0}$, the true efficiency score is always lower than the sample efficiency score, i.e. $\theta(x_{j_0}, y_{j_0}) < \widehat{\theta}(x_{j_0}, y_{j_0})$. Therefore, a consistent $\mathcal{P}$ should preserve the same relationship between the sample efficiency score and the bootstrap efficiency score, i.e. it should hold that $\widehat{\theta}(x_{j_0}, y_{j_0}) < \widehat{\widehat{\theta}}(x_{j_0}, y_{j_0})$. However, in naïve bootstrapping, the bootstrap estimate will equal with the sample estimate with non-zero probability, whereas the probability that the sample estimate equals with the true parameter is zero.

Different bootstrap methodologies have been developed to overcome the inconsistency issue of the naïve bootstrapping. In the following subsections, the main bootstrap approaches within the DEA framework and how they can provide consistent estimates are discussed.

Consider a sample $S_N = \{(x_j, y_j), j = 1, ..., N\}$ which is considered as N realisations of the i.i.d. random input and output variables $(X, Y) \in \mathbb{R}_+^{P+S}$ on the convex PPS $\widehat{T}$, and let $\widehat{\theta}_j \equiv \widehat{\theta}(x_j, y_j)$, $j = 1, ..., N$ denote the sample efficiency scores obtained by applying a DEA estimator.

First, due to the radial nature of the Farrell's efficiency measure, it is more intuitive that the inputs (in the input orientation) of the Cartesian coordinates $(x_j, y_j), j = 1, ..., N$ are transformed into polar coordinates, i.e. a point (x, y) is transformed into (ω, η, y), where (ω, η) are the polar coordinates of $x \in \mathbb{R}_+^P, P \geq 2$. The modulus $\omega = \sqrt{x_1^2 + ... + x_P^2} \in \mathbb{R}_+$ denotes the distance of vector x from the origin and each element i of the vector $\eta \in [0, \frac{\pi}{2}]^{P-1}$ is calculated as $\eta_i = arctan(\frac{x_{i+1}}{x_1}), i = 1, ..., P-1, x_1 > 0$, or $\eta_i = \frac{\pi}{2}$ if $x_1 = 0$, and denotes the angle of ω to the axes defined by the inputs.

The transformation into polar coordinates will also be reflected onto the distribution function. After the transformation to the polar coordinates, the DGP will now be characterised by the density function $f(\omega, \eta, y)$. Using the Bayes' law, this density function can be decomposed into

$$f(\omega, \eta, y) = f(\omega|\eta, y)f(\eta|y)f(y). \tag{5.17}$$

From the above decomposition it can be seen that decision is made on the output levels first, then, based on the output levels, the input mix η is decided and finally the amount of input levels is set.

The efficiency score can also be defined with respect to the polar coordinates. Let $x^{\delta}(y) = \theta(x, y)x$ be the efficient level given the outputs and a specific input mix, and the modulus of the efficient projection is $\omega(x^{\delta}(y)) = inf\{\omega \in \mathbb{R}_{+}|f(\omega|y, \eta) > 0\}$. Therefore, the Farrell efficiency score is defined as

$$\theta(x, y) = \frac{\omega(x^{\delta}(y))}{\omega(x)}. \tag{5.18}$$

In terms of the density function of $\theta(x, y)$, from relation (5.18) the density $f(\omega|\eta, y)$ with support on $[\omega(x^{\delta}(y)), \infty]$ results in the density $f(\theta|\eta, y)$ on $[0, 1]$. However, the support of the new density function is bounded from both sides, and as it was previously mentioned, boundary estimations may be problematic. In this case, due to the boundary in zero, it may happen that the resulting lower bounds of the confidence interval estimates are negative. Using the Shephard (1970) input distance function $\delta(x, y) = \frac{1}{\theta(x,y)}$ instead, the density $f(\delta|y, \eta)$ has now support on $[1, \infty]$ which is only bounded from below. Therefore, the estimation of $f(\delta, y, \eta)$ will now be based on the sample $S_{N,\widehat{\delta}} = \{(\widehat{\delta}_j, \eta_j, y_j)|j = 1, ..., N\}$.

5.7.1 A homogeneous non-smooth bootstrap approach

In the homogeneous bootstrap, it is assumed that the distribution of inefficiencies is homogeneous among the different DMUs. Therefore, it is implied that

$$f(\delta|y, \eta) = f(\delta), \tag{5.19}$$

i.e. it is assumed that the density distribution of the length of the input vector is independent from the decision on the input mix that is going into the production. As it will be further discussed below, although in practice this assumption is very convenient- and even necessary in most of the cases when DMU-specific distributions are not available- it is a quite restrictive one.

From the axiom (vi) it is implied that $f(\omega|\eta, y)$ is strictly positive on the boundary, i.e. $f(\omega(x^e(y))|\eta, y) > 0$. In most of the cases, this is simply considered as a regularity condition. However, Olesen & Petersen (2016) noted that this condition does not always hold. They brought the example of some bank branches, which they all use the same amount of an input and where the modulus ω follows a normal distribution which is truncated to an interval the end points of which depend on the talent of the bank manager. They showed that the management talent affects the support of the inefficiency distribution, and as a result, not every DMU has a positive probability to reach the efficient frontier. Making the assumption that the distribution of inefficiencies is homogeneous across DMUs may partially overcome the issue - as it puts positive probability mass on the frontier for all DMUs, but at the cost of ignoring the management talent variable. As an alternative, Olesen & Petersen (2016) suggested conditioning the distribution of inefficiencies on an environmental variable that captures the management talent, but as they noted, measuring a management talent variable is very difficult in practice. In terms of the incorporation of environmental variables in the production process, the consideration of environmental factors in the efficiency measurement involves a semi-parametric two stage procedure introduced by Simar & Wilson (2007).

Consider now the original sample S_N, and let $\widehat{f}(\delta)$ be the empirical distribution of the inverse efficiency scores $\widehat{\delta}_j$, $j = 1, ..., N$, based on the sample $S_{N,\widehat{\delta}}$. A bootstrap sample $S_N^* = \{(x_j^*, y_j^*)|j = 1, ..., N\}$ can be obtained through the two following steps:

1. Each observation $(x_j, y_j) \in S_N$ is first projected to the efficient frontier $\partial\widehat{X}(y)$ that is estimated by the sample. Therefore, the projected points will be $(\widehat{\delta}_j^{-1}x_j, y_j), j = 1, ..., N$.

2. Each efficient point is projected back by a value $\widehat{\widehat{\delta}}_j$, where $\widehat{\widehat{\delta}}_j, j = 1, ..., N$ are random draws from the empirical distribution $\widehat{f}(\delta)$. This last projection away from the frontier results to the points $(\widehat{\widehat{\delta}}_j\widehat{\delta}_j^{-1}x_j, y_j)$ in Cartesian coordianates that formulate the bootstrap sample S_N^*.

Therefore, the bootstrap sample is defined as $S_N^* = \{(x_j^*, y_j^*)|j = 1, ..., N\}$, where $x_j^* =$

$\widehat{\widehat{\delta}}_j \widehat{\delta}_j^{-1} x_j$, $j = 1, ..., N$. By repeating the above process $B >> 0$ times, B bootstrap samples can be obtained and used to get the distribution of $\widehat{\delta}(x_{j_0}, y_{j_0}) - \widehat{\widehat{\delta}}(x_{j_0}, y_{j_0})$ or $\widehat{\theta}(x_{j_0}, y_{j_0}) - \widehat{\widehat{\theta}}(x_{j_0}, y_{j_0})$.

However, in naïve bootstrapping, the bootstrap inverse efficiency scores $\widehat{\widehat{\delta}}$ are drawn from the empirical distribution $\widehat{f}(\delta)$, which puts probability mass $1/N$ at each $\widehat{\delta}$. Let D_N denote the sample of inverse efficiency estimates $\widehat{\delta}$ obtained by applying a DEA estimator to the sample S_N. Then, in naïve bootstrap, $\widehat{f}(\delta)$ is defined as

$$\widehat{f}(\delta) = \begin{cases} 1/N, & \text{if } \delta \in D_N \\ 0, & \text{otherwise.} \end{cases} \tag{5.20}$$

Therefore, if the random draws in step 2 are done with replacement, the probability of selecting a specific $\widehat{\delta}_{j_0} \equiv \widehat{\delta}(x_{j_0}, y_{j_0})$ in a draw is $1/N$, and the probability of not selecting $\widehat{\delta}_{j_0}$ is $1 - 1/N$. When the process is repeated N times to obtain the bootstrap sample S_N^*, the probability of not drawing $\widehat{\delta}_{j_0}$ is $(1 - 1/N)^N$. Thus, the probability of selecting $\widehat{\delta}_{j_0}$ is $1 - (1 - 1/N)^N$. Therefore, asymptotically,

$$\underset{N \to \infty}{lim} P(\widehat{\widehat{\delta}}_{j_0} = \widehat{\delta}_{j_0}) = \underset{N \to \infty}{lim} \left[1 - \left(1 - \frac{1}{N}\right)^N\right] = 1 - e^{-1} \approx 0.632. \tag{5.21}$$

However, from axiom (vi) it is assumed that the true density function $f(\delta)$ is continuous, and therefore, the true probability that $P(\widehat{\delta}_{j_0} = \delta_{j_0}) = 0$, which explains why the naïve bootstrap is inconsistent.

Smoothing of the empirical distribution function or the use of subsampling techniques have been proven to overcome the issue discussed above and to give consistent estimations of the production frontier. In the following Sections, these approaches are reviewed.

5.7.2 Bootstrap with smoothing

The first study that applied a bootstrap technique to approximate the distribution of the DEA estimator under the VRS assumption, was by Simar & Wilson (1998). To overcome the inconsistency problem of the naïve bootstrap discussed above, Simar & Wilson (1998) introduced a homogeneous smooth bootstrap. Although the homogeneity assumption is restrictive, this method can give good estimations even with a relatively small data set. Later, Simar & Wilson (2000) suggested a smooth bootstrap that was not based on the homogeneity assumption. However, due to the high computational burden of the last approach, the homogeneous bootstrap is the most often used one. Next, both approaches are discussed.

In smooth bootstrap, the random draws $\widehat{\widehat{\delta}}_j$ are generated from a smooth version of $\widehat{f}(\delta|y,\eta)$. Under the homogeneity assumption, a smooth density function $\widehat{f}_h(\delta)$ can be obtained by smoothing the empirical density function using a kernel density estimation as follows:

$$\widehat{f}_h(\delta) = \frac{1}{Nh}\sum_{j=1}^{N} K\Big(\frac{\delta - \widehat{\delta}_j}{h}\Big), \tag{5.22}$$

where $K(\cdot)$ denotes a symmetric kernel density function with zero mean, i.e. $\int_{-\infty}^{\infty} K(\delta)d\delta = 1$, $K(-\delta) = K(\delta)$ and $\int_{-\infty}^{\infty} \delta K(\delta)d\delta = 0$, and h is a smoothing parameter. More specifically, in a frequency histogram determined by the sample, h is considered as the width of the histogram bins. The length of this histogram will be equal to Nh. Then, $\widehat{f}_h(\delta)$ can be considered as the average of N density functions, whose domain has been stretched by a value h (amount of variation of the N densities) and then centred around the observed sample of $\widehat{\delta}_j$.

However, $\widehat{f}_h(\delta)$ does not take into account the boundary condition that $\delta \geq 1$, and therefore, there is a positive probability that some of the random draws $\widehat{\widehat{\delta}}_j$, $j = 1, ..., N$ are less than one, although the support of the true density function $f(\delta)$ is $[1, \infty)$. A solution to this could be to obtain the estimate $\widehat{f}_h(\delta)$ ignoring the boundary conditions, and then set $\widehat{f}_h(\delta) = 0$ for all $\delta < 1$. However, in that case, the resulting density function estimate $\widehat{f}_h(\delta)$ would no more add to unity.

Therefore, to overcome this inconsistency caused by the boundary condition, kernel smoothing can be used together with reflection techniques (Silverman 1986). A reflection around a point $x_0 \in \mathbb{R}$ is a function $r_{x_0} : \mathbb{R} \to \mathbb{R}$ given by

$$r_{x_0}(x) = 2x_0 - x. \tag{5.23}$$

In this case, the observed efficiencies $\widehat{\delta}_j, j = 1..., N$ need to be reflected around the extreme point given by the boundary condition, i.e. around $x_0 = 1$. Therefore, we get N reflected efficiency estimates $2 - \widehat{\delta}_j, j = 1, ..., N$. If D_N is the original sample of inverse efficiency estimates, the reflection results in the augmented set $D_{2N} = D_N \sqcup \{2 - \widehat{\delta}_j | j = 1, ..., N\}$, where $\sqcup$ denotes the disjoint union of two sets. That means if $\{x_1, ..., x_n\}$ and $\{y_1, ..., y_n\}$ are two sets, their disjoint union is given as $\{x_1, ..., x_n, y_1, ..., y_n\}$. The density of the augmented set D_{2N} is obtained using the kernel density estimator given in 5.22, which in this case will take the form

$$\widehat{g}_h(\delta) = \frac{1}{2Nh} \sum_{j=1}^{N} \left[K\left(\frac{\delta - \widehat{\delta}_j}{h}\right) + K\left(\frac{\delta - (2 - \widehat{\delta}_j)}{h}\right) \right]. \tag{5.24}$$

Then, $\widehat{f}_h(\delta)$ is defined as the truncated $\widehat{g}_h$ as follows:

$$\widehat{f}_h(\delta) = \begin{cases} 2\widehat{g}_h(\delta), & \text{if } \delta \geq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{5.25}$$

Now, $\widehat{f}_h(\delta)$ adds to unity and has support over $[1, \infty)$.

Usually, the Gaussian kernel density function is being used. However as noted by Simar & Wilson (2008), the choice of the scaling parameter h is of more importance. Different rules for choosing h exist in the literature. One of the methods discussed in Silverman (1986), suggests choosing the value of h that minimises the approximate mean integrated square error (MISE), where $MISE_{\widehat{g}}(h) = E\Big[\int (g_h(\delta) - \widehat{g}_h(\delta))^2 d\delta\Big]$. If a Gaussian kernel function is being used, minimising the MISE results in

$$h = 1.06 \min\,(\widehat{\sigma}, \widehat{R}/1.34) N^{-1/5}, \tag{5.26}$$

where $\widehat{\sigma}$ and $\widehat{R}$ are the estimated standard deviation and interquartile range, respectively, of $\widehat{\delta}_j$, $j = 1, ..., N$.

If $\widehat{f}_h(\delta)$ is also assumed to be a normal density function, then the optimal value of h becomes

$$h = 1.06\widehat{\sigma} N^{-1/5}. \tag{5.27}$$

More rules and further discussion on the choice of the optimal smoothing parameter can be found in Scott (2015).

The inconsistency issue of the homogeneous smooth bootstrap and the application of the reflection technique is further explained through the following illustrative example.

Example:

Consider a sample consisting of ten inverse efficiency estimates $\widehat{\delta}_j, j = 1, ..., 10$, $D_{10} =$

$\{1, 1, 1, 1, 1.1, 1.2, 1.25, 1.3, 1.35, 1.4\}$. The empirical distribution of the sample can be smoothed using a gaussian kernel density. In Figure 5.1 below, the discrete, empirical distribution function is depicted in red color, and the blue thick line depicts the Gaussian kernel density estimate $\widehat{f}(\delta)$, where the smoothing parameter h is defined by Silverman's rule (Silverman 1986). Similarly, in Figure 5.2, the empirical and the kernel cumulative distribution estimations are provided.

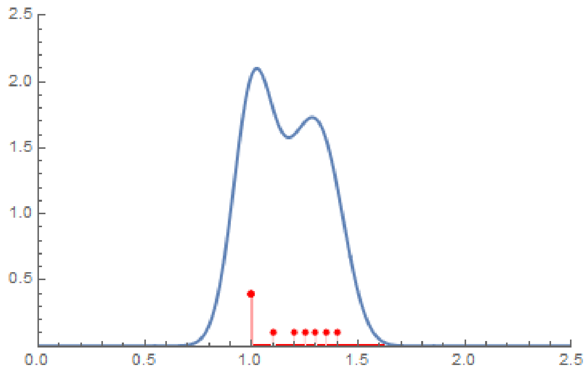


Figure 5.1: *In red:* Empirical distribution $\widehat{f}(\delta)$, based on the sample D_{10}. *In blue:* Gaussian kernel density estimate $\widehat{f}(\delta)$, with bandwidth $h = 0.0906$.

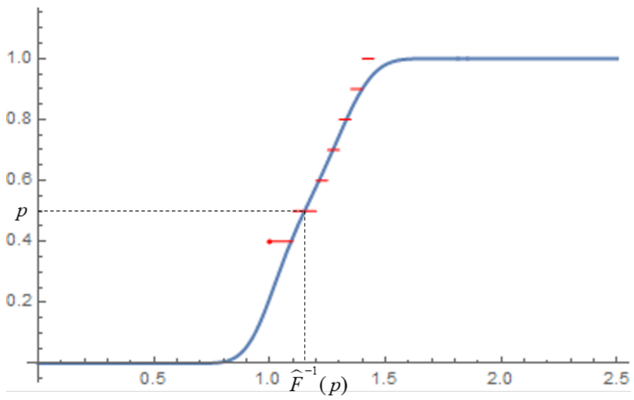


Figure 5.2: *In red:* Empirical cumulative distribution $\widehat{F}(\delta)$, based on the sample D_{10}. *In blue:* Gaussian kernel cumulative distribution estimate $\widehat{F}(\delta)$.

From Figure 5.1 it can be seen that the smoothed estimate of $f(\delta)$ ignores the boundary condition $\delta \geq 1$, and therefore, there is a positive probability that some of the random draws $\widehat{\widehat{\delta}}_j, j = 1, ..., 10$ will be less than one. A way to obtain the random draws $\widehat{\widehat{\delta}}_j$, $j = 1, ..., 10$ from $\widehat{f}(\delta)$ is by using the inverse kernel cumulative distribution $\widehat{F}^{-1}(p)$ through the following procedure:

- First, we generate a random sample P_N of ten probabilities $p_j \in [0, 1]$, $j = 1, ..., N$. For example, the sample $P_N = \{0.1631681, 0.5291093, 0.3951027, 0.3513978, 0.3189761, 0.2194194, 0.8379487, 0.9727054, 0.2068039, 0.5119396\}$ consists of ten independent and random draws form the uniform distribution $U[0, 1]$.

- Then, $\widehat{\widehat{\delta}}_j = \widehat{F}^{-1}(p_j)$, as illustrated in Figure 5.2.

  In this case, the resulting sample of $\widehat{\widehat{\delta}}_j, j = 1, ..., 10$ is $\{0.973663, 1.16801, 1.08851, 1.06581, 1.04977, 1.00206, 1.35429, 1.48408, 0.995894, 1.15718\}$. It can be seen that $\widehat{\widehat{\delta}}_1, \widehat{\widehat{\delta}}_9 < 1$.

If we apply the reflection technique, the augmented sample that includes the original and the reflected efficiencies will be $D_{20} = \{1, 1, 1, 1, 1.1, 1.2, 1.25, 1.3, 1.35, 1.4\} \sqcup \{1, 1, 1, 1, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6\}$. In Figure 5.3 the Kernel density function estimate $\widehat{g}_h(\delta)$ of the augmented set D_{20} is depicted with a blue thick line. Then, the truncated density estimate $\widehat{f}_h(\delta)$ is depicted in Figure 5.4.

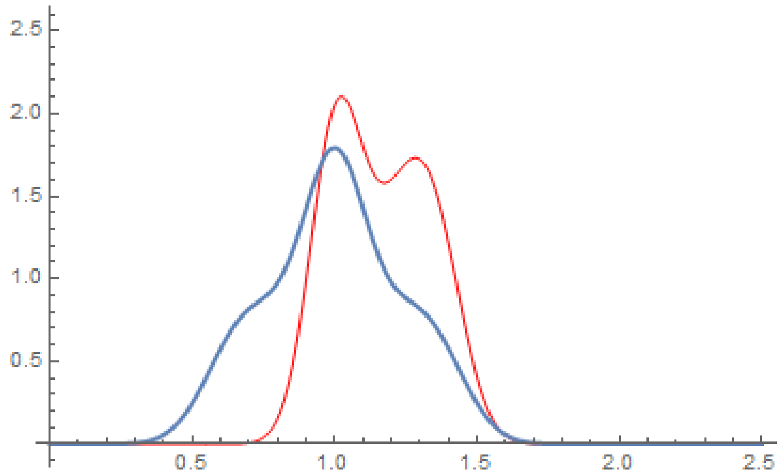


Figure 5.3: *In blue:* Gaussian kernel density estimate $\widehat{g}_h(\delta)$, with bandwidth $h = 0.111$, based on the augmented sample D_{20}. *In red:* the estimate $\widehat{f}_h(\delta)$, with bandwidth $h = 0.0906$, based on the original sample D_{10}.

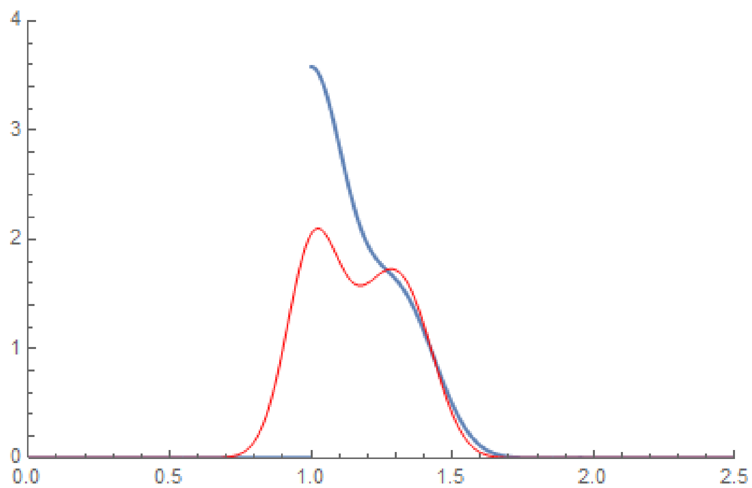


Figure 5.4: *In blue:* Truncated Gaussian kernel density estimate $\widehat{f}(\delta)$, based on the augmented sample D_{20}. *In red:* the estimate $\widehat{f}_h(\delta)$, with bandwidth $h = 0.0906$, based on the original sample D_{10}.

In Figure 5.5, the truncated Kernel cumulative distribution estimate $\widehat{F}_h(\delta)$ is depicted in a blue thick line. The inverse CDF $F_h^{-1}(\delta)$ that will be used to draw the back projections $\hat{\hat{\delta}}_j$, $j = 1, ..., 10$ is provided in Figure 5.6.

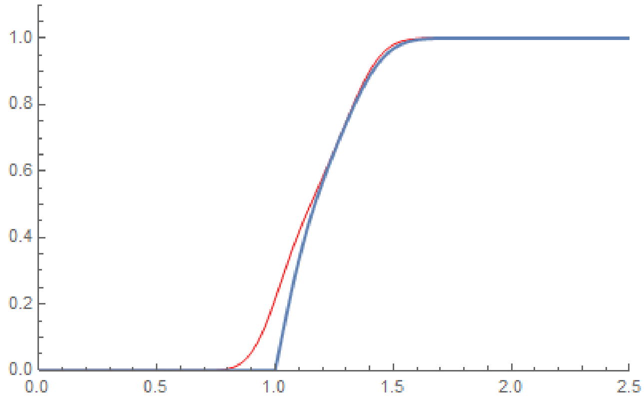


Figure 5.5: *In red:* Gaussian kernel cumulative distribution estimate $\widehat{F}(\delta)$ based on the sample D_{10}. *In blue:* Truncated Gaussian kernel cumulative distribution estimate $\widehat{F}(\delta)$, based on D_{20}.

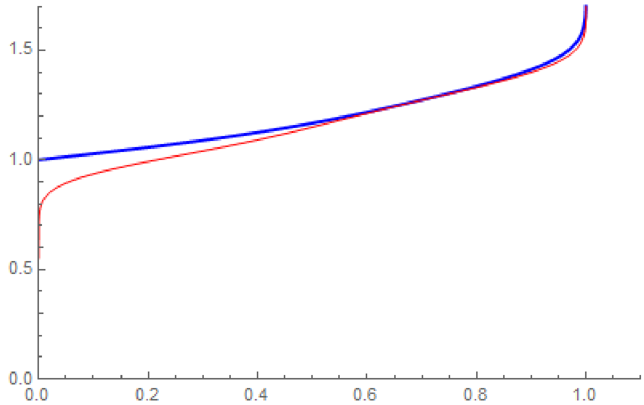


Figure 5.6: *In red:* Gaussian kernel inverse CDF estimate $\widehat{F}_h^{-1}(\delta)$ based on the sample D_{10}. *In blue:* Truncated Gaussian kernel inverse CDF estimate $\widehat{F}_h^{-1}(\delta)$, based on D_{20}.

It can be seen that for any probability $p \in [0, 1]$, the value of the Kernel inverse CDF will be $\widehat{F}_h^{-1}(p) \geq 1$. In this example, using the previously drawn random sample of probabilities P_N, the resulting sample of the inverse bootstrap efficiency estimates $\hat{\hat{\delta}}_j$, $j = 1, ..., N$ will be $\{1.04636, 1.1815, 1.12334, 1.10703, 1.0956, 1.06327, 1.36194, 1.51056, 1.05941, 1.17329\}$. □

In the above example, $\widehat{F}_h^{-1}(p)$ was used to draw the sample of $\hat{\hat{\delta}}_j, j = 1, ..., N$. However, the PDF estimate $\widehat{f}_h(\delta)$ or the corresponding inverse CDF estimate $\widehat{F}_h^{-1}(p)$ does not need to be computed, and only decision on the Kernel function and the smoothing parameter h are needed to draw the back projections $\hat{\hat{\delta}}_j$. Let $D_{2N} = \{\widehat{\delta}_1, ..., \widehat{\delta}_N, 2 - \widehat{\delta}_1, ..., 2 - \widehat{\delta}_N\}$. Then, $\hat{\hat{\delta}}_j$ can be obtained through the following steps (Simar & Wilson 2008, pg. 51):

1. Draw a naïve bootstrap sample β_j^*, $j = 1, ..., N$ from D_{2N}.

2. Calculate

$$\beta_j^{**} = \beta_j^* + h\epsilon_i^*, \tag{5.28}$$

where ϵ^* are independent draws from the Kernel density function $K(\cdot)$, and h is the smoothing parameter.

3. Correct the mean and variance of the random sample of β^{**} to be the same as the mean and variance of D_{2N} by calculating

$$\beta_j^{***} = \bar{\beta}^* + \frac{\beta_j^{**} - \bar{\beta}^*}{\sqrt{1 + \frac{h^2\sigma_K^2}{\sigma_\beta^2}}}, \tag{5.29}$$

where $\bar{\beta}^* = (1/N)\sum_{j=1}^{N} \beta_j^*$ is the mean of the sample β_j^*, σ_K^2 is the variance of the kernel density function, and $\sigma_\beta^2 = (1/N)\sum_{j=1}^{N}(\beta_j^* - \bar{\beta}^*)^2$.

4. Finally, $\widehat{\widehat{\delta}}_j$ are obtained by reflecting the draws β_j^{***} that are less than one, i.e.

$$\widehat{\widehat{\delta}}_j = \begin{cases} 2 - \beta_j^{***}, \text{ if } \beta_j^{***} < 1 \\ \beta_j^{***}, \text{ otherwise} \end{cases} \tag{5.30}$$

After obtaining $\widehat{\widehat{\delta}}_j, j = 1, ..., N$ the bootstrap sample can be generated as it was described in step 2 of the two-step procedure provided in the previous Section, as $S_N^* = \{(\widehat{\widehat{\delta}}_j \widehat{\delta}_j^{-1} x_j, y_j)|j = 1, ...N\}$.

Although the homogeneous smooth bootstrap seems to be the most easy to implement and therefore the most commonly used one, other smooth bootstrap approaches have been suggested in the DEA literature. Simar & Wilson (2000) extended the homogeneous smooth bootstrap to a heterogeneous one, where the distribution of inefficiencies varies across the DMUs. In the heterogeneous bootstrap, we need to draw N random and independent points $(\widehat{\delta}, \eta, y)$ from the multivariate kernel density function estimate $\widehat{f}_h(\delta, \eta, y)$. However, in practice, the heterogeneous bootstrap is very difficult to implement, as it requires a great amount of data that is usually not available.

Kneip et al. (2008) proved the consistency of a double-smooth bootstrap technique. In the double-smooth bootstrap, in addition to smoothing the empirical distribution of the efficiency estimates, the efficient frontier estimate needs to be smoothed as well. In this case, the bootstrap sample is generated by projecting back the efficient projections, away from a smoothed version of the frontier. Smoothing of the frontier requires the selection of a second smoothing parameter. Kneip et al. (2008) suggest a rule for the selection of the smoothing parameter (see (Kneip et al. 2008, pg. 16-18)).

Although this double-smooth method proposed is consistent, in practice it is computationally very demanding. Kneip et al. (2011) introduced a consistent, simplified version of the original method, which is computationally easier. What they suggested is to use the naïve bootstrap to draw observations that are "far" from the frontier and fill the bootstrap sample by drawing from a smooth, uniform distribution with support "near" the frontier. The terms "near" and "far" are controlled by another smoothing parameter.

5.7.3 Subsampling bootstrap

Another bootstrap approach that was proved by Kneip et al. (2008) to provide consistent estimates is bootstrapping with subsampling. Subsampling bootstrap - originally suggested by Swanepoel (1986) - consists of drawing $m = N^{\kappa}$ observations, usually with replacement, for $\kappa \in (0, 1)$. Although subsampling seems very similar to the naïve bootstrap, by drawing $m < N$ observations, the frequency at which the sample maximum is drawn is reduced, and under the conditions that $m \to \infty$ and $m/N \to 0$ as $N \to \infty$, this approach provides good approximations of the true efficiency scores. In other words, m needs to tend to infinity at rate slower than N. Then,

$$m^{\frac{2}{P+S+1}} \left(\frac{\widehat{\widehat{\theta}}_{j_0}}{\widehat{\theta}_{j_0}} - 1 \right) \overset{approx}{\sim} N^{\frac{2}{P+S+1}} \left(\frac{\widehat{\theta}_{j_0}}{\theta_{j_0}} - 1 \right), \tag{5.31}$$

where P the number of inputs, S the number of outputs, and $\widehat{\theta}_{j_0} \equiv \widehat{\theta}(x_{j_0}, y_{j_0})$.

In order to obtain the bias of subsampling bootstrap estimations, the only difference from relation 5.9 is that it needs to be multiplied by the term $\frac{m}{N}^{\frac{2}{P+S+1}}$ to consider for the difference in the size of the original and the bootstrap sample.

The fact that no estimation of the density function of the efficiency estimates is required, makes subsampling bootstrap computationally easier than the smooth bootstrap, and avoids any restrictive assumption of homogeneity in the distribution of the efficiency estimates among DMUs.

Let $S_N = \{(x_j, y_j) | j = 1, ..., N\}$ be the originally observed sample of DMUs. The bootstrap subsample $S^*_m = \{(x^*_j, y^*_j)|j = 1, ..., m\}$, is generated by drawing randomly and independently with replacement (or without replacement) $m < N$ observations from S_N. The DEA estimator based on the subsample S^*_m is used to calculate the bootstrap efficiency estimates $\widehat{\widehat{\theta}}(x_{j_0}, y_{j_0})$. This process is repeated $B >> 0$ times. The bootstrap efficiency estimates obtained on the subsamples S_m are used to approximate the distribution $N^{\frac{2}{P+S+1}}\left(\frac{\widehat{\theta}_{j_0}}{\theta_{j_0}} - 1\right)$. The a-quantile estimations $\widehat{c}_{a/2}$, $\widehat{c}_{1-a/2}$ are given by

$$P\left(\widehat{c}_{a/2} \leq m^{\frac{2}{P+S+1}}\left(\frac{\widehat{\widehat{\theta}}_{j_0}}{\widehat{\theta}_{j_0}} - 1\right) \leq \widehat{c}_{1-a/2,m}|S_N\right) = 1 - a \tag{5.32}$$

The estimated quantiles $\widehat{c}_{a/2}, \widehat{c}_{1-a/2}$ are used to get the approximation

$$P\left(\widehat{c}_{a/2} \leq N^{\frac{2}{P+S+1}}\left(\frac{\widehat{\theta}_{j_0}}{\theta_{j_0}} - 1\right) \leq \widehat{c}_{1-a/2,m}\right) \approx 1 - a. \tag{5.33}$$

Then, the confidence interval for $\theta(x_{j_0}, y_{j_0})$ is

$$\left[\frac{\widehat{\theta}_{j_0}}{1 + N^{-\frac{2}{P+S+1}}\widehat{c}_{1-a/2,m}}, \frac{\widehat{\theta}_{j_0}}{1 + N^{-\frac{2}{P+S+1}}\widehat{c}_{a/2,m}}\right]. \tag{5.34}$$

Below, an example is provided to demonstrate how drawing a subsample from the original sample S_N overcomes the inconsistency problem of the naïve bootstrap. As it was mentioned before, the true frontier is assumed to be smooth and therefore, there is zero probability that the efficiency estimates can be equal to the true efficiency score of a point (x_{j_0}, y_{j_0}). Therefore, if the bootstrapping is consistent, the probability that a bootstrap estimate $\widehat{\widehat{\theta}}_{j_0}$ is equal to the sample estimate $\widehat{\theta}_{j_0}$ is zero.

For simplicity reasons, consider the production process with one input and one output, i.e $x \in \mathbb{R}_+$ and $y \in \mathbb{R}_+$. Then, the efficient frontier is defined by a piece-wise linear curve. Let $(x_{j_0}, y_{j_0}) \in S_N$ for which the efficient projection $(\widehat{\theta}_{j_0}x_{j_0}, y_{j_0})$ lies on a linear segment P_1P_2 of the frontier, where $P_1 = (x_{k_0}, y_{k_0})$, $P_2 = (x_{\lambda_0}, y_{\lambda_0}), k_0, \lambda_0 \neq j_0$. The probability that $\widehat{\widehat{\theta}}_{j_0} = \widehat{\theta}_{j_0}$ is equal to the probability that the bootstrap sample includes both the points P_1 and P_2.

Let $\mathcal{A}$ denote the event that the bootstrap subsample S^*_m includes P_1, $\mathcal{B}$ denote the event that S^*_m includes P_2, and let $\mathcal{A}'$, $\mathcal{B}'$ be the complement events of $\mathcal{A}$ and $\mathcal{B}$ that S^*_m

does not include P_1, and that S^*_m does not include P_2, respectively. Then,

$$P(\widehat{\widehat{\theta}}_{j_0} = \widehat{\theta}_{j_0}) = P(\mathcal{A} \cap \mathcal{B}) = 1 - P((\mathcal{A} \cap \mathcal{B})') = 1 - P(\mathcal{A}' \cup \mathcal{B}')$$
$$\stackrel{\substack{\text{addition}\\ \text{law}}}{=} 1 - \big(P(\mathcal{A}') + P(\mathcal{B}') - P(\mathcal{A}' \cap \mathcal{B}')\big). \quad (5.35)$$

There are two possible ways of drawing the subsample: with or without replacement. First, consider the case of drawing with replacement. When subsampling is done with replacement, the probability that in one draw P_1 will be selected is $1/N$, and the probability that it will not be selected is $1-1/N$. When this process is repeated m times, the probability that P_1 is included in the subsample is $P(\mathcal{A}) = (1 - 1/N)^m$. Similarly, $P(\mathcal{B}) = P(\mathcal{A})$. The probability that we will draw either P_1 or P_2 is $2/N$, and the probability that we will not draw neither P_1 nor P_2 is $1 - 2/N$. If this process is repeated m times, the probability that neither P_1 nor P_2 are included in the subsample is $P(\mathcal{A}' \cap \mathcal{B}') = (1 - 2/N)^m$. Therefore, substituting the above probabilities in relation 5.35 results in

$$P(\widehat{\widehat{\theta}}_{j_0} = \widehat{\theta}_{j_0}) = 1 - 2\left(1 - \frac{1}{N}\right)^m + \left(1 - \frac{2}{N}\right)^m. \quad (5.36)$$

When the subsampling is done without replacement, the probability that the point P_i, $i = 1, 2$ is drawn in the first draw will be $1/N$, in the second draw the probability will be $1/(N-1)$, etc. and similarly, point P_1 will not be drawn in the first draw with probability $1 - 1/N$, in the second draw with probability $1 - 1/(N-1)$, etc., and in the m-th draw with probability $1 - 1/(N-(m-1))$. Therefore, the probability that P_1 is not included in the subsample will be $P(\mathcal{A}') = \prod_{j=1}^{m} \big(1 - 1/(N+1-j)\big)$. In a similar way, $P(\mathcal{B}) = P(\mathcal{A})$, and $P(\mathcal{A}' \cap \mathcal{B}') = \prod_{j=1}^{m} \big(1 - 2/(N+1-j)\big)$. Therefore, for subsampling without replacement, from 5.35 it results that

$$P(\widehat{\widehat{\theta}}_{j_0} = \widehat{\theta}_{j_0}) = 1 - 2\prod_{j=1}^{m}\left(1 - \frac{1}{N+1-j}\right) + \prod_{j=1}^{m}\left(1 - \frac{2}{N+1-j}\right). \quad (5.37)$$

Asymptotically, under the condition that m tends to infinity at a slower rate than N, both probabilities given in 5.36 and 5.37 tend to zero. Further discussion and an example that demonstrates the consistency of subsampling bootstrap can be found in Kneip et al. (2011, pg. 11-12).

The above can be generalised to more dimensions. In the case $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$ we would need the probability of drawing the $(p+q)$ efficient points that form the frontier facet that (x_{j_0}, y_{j_0}) is projected (see Simar & Wilson (2000, pg. 786)).

In some cases, a bootstrap subsample S_m may yield no feasible solutions. This happens when $y_{j_0} \overset{\geq}{\neq} \max\{y_j|(x_j, y_j) \in S_m\}$ and/or $x_{j_0} \overset{\leq}{\neq} \min\{x_j|(x_j, y_j) \in S_m\}$. In this case, the DMU under evaluation can be added into the subsample, i.e. $\widehat{\widehat{\theta}}_{j_0} = 1$. Based on the definition of the DGP, the probability of this happening tends to to zero, and therefore, including the DMU under evaluation into the subsample does not affect the asymptotic properties of the bootstrap estimator. However, the choice of the subsample size seems to be crucial.

Although subsampling is easy to implement, its performance is sensitive to the choice of the subsample size m. Based on the minimum volatility criterion introduced by Politis et al. (2001) for the subsample size selection, Simar & Wilson (2010) suggested the following algorithm to choose an optimal subsample size in the DEA context. Let $I_{m,low}$ and $I_{m,up}$ be the lower and the upper bounds of a confidence interval estimate for a DMU (x_{j_0}, y_{j_0}), resulting from subsampling bootstrap, with subsample size m. Then,

1. Obtain the confidence interval estimates $[I_{m,low}, I_{m,up}]$, for various subsample sizes m.

2. For a small $r \in \mathbb{Z}^+$, calculate the volatility index
   $V_m = \sum_{m=m_{small}}^{m_{big}} sd\{I_{m-r,low}, ..., I_{m+r,low}\} + \sum_{m=m_{small}}^{m_{big}} sd\{I_{m-r,up}, ..., I_{m+r,up}\}$

3. Choose the subsample size m that corresponds to the minimum volatility V_m.

Although the above algorithm provides a rule of thumb, up to date, there is no robust method for the selection of the optimal subsample size.

Simar & Wilson (2010) studied both subsampling with and without replacement and found that the first one yields better results.

5.8 Extensions and applications of bootstrapping

The use of bootstrapping techniques in the DEA context is not limited to the efficiency score estimation. Various studies have developed hypothesis tests for different assumptions, and bootstrap approaches have been applied to obtain the critical value of the test statistic, in most cases using the homogeneous smooth bootstrap or subsampling bootstrap. Up to the mid-2010s results were solely based on Monte Carlo simulations, and it is during the last years that the theoretical foundations are set based on the central limit theorem.

Simar & Wilson (2001) developed hypothesis test on whether the inputs and outputs can be aggregated, and Simar & Wilson (2002) tested the returns to scale. Simar & Zelenyuk(2006; 2007) used hypothesis tests to compare the efficiency distributions and their means between two samples. Simar & Wilson (2010) tested the convexity assumption of the PPS and the returns to scale using subsampling bootstrap. Kneip et al. (2016) based on the central limit theorem (Kneip et al. 2015), provided theoretical results on the consistency of the hypothesis tests. Daraio et al. (2015) based on the theoretical results of Kneip et al. (2016) and Kneip et al. (2015) developed a hypothesis test of the separability condition[2], both for non-conditional and conditional efficiencies. The approach used by Kneip et al. (2016) and Daraio et al. (2015) requires randomly splitting the original sample into two independent subsamples. However, this approach involves some uncertainty regarding the split of the sample. To eliminate this ambiguity, Simar & Wilson (2020) suggested repeating the random splits a large number of times.

Various bootstrapping techniques have been widely applied in the DEA context to obtain bias-corrected estimates and confidence intervals for the efficiency scores. For example, Wanke (2012) used a Gaussian kernel to draw bootstrap samples and provide confidence interval estimates for the BCC efficiency scores of 68 Brazilian airports, and for testing the returns-to-scale. Marchetti & Wanke (2017) applied a second stage bootstrap truncated regression to assess the impact of contextual variables in the performance of the Brazilian rail concessionaires. In Nwaogbe et al. (2017), the impact of contextual variables on 30 major Nigerian airports was assessed by combining first stage bootstrap efficiency estimations with a second stage censored quantile regression. Li et al. (2022) applied the homogeneous smooth bootstrap to estimate the eco-efficiency of the Chinese hotel industry and to perform hypothesis test about the returns to scale and for comparing the eco-efficiency means between two hotel clusters. It is interesting to note that although different review studies on the methodological developments of bootstrapping are available, there is a lack of a bibliometric analysis of the studies that apply bootstrapping techniques in the DEA context.

[2]The separability condition needs to be satisfied in order to regress the efficiency estimates on some environmental variable $\mathcal{Z} \in \mathbb{R}^r$. The separability condition states that in order to use the second stage regression, the shape of the conditional support of inputs and outputs (conditioned on $\mathcal{Z}$) needs to be independent of the environmental factor $\mathcal{Z}$. See Simar & Wilson (2007) for more details.

5.9 Conclusion

DEA provides an empirical estimation of the production frontier based on an observed sample of DMUs, and sampling noise is not taken into account. However, alterations in the observed sample can affect the shape of the efficient frontier and therefore, the efficiency scores of DMUs. If the observed set of DMUs is considered as a random sample drawn independently and uniformly from an underlying population through an unknown data generating process, then the true efficient frontier is unknown, and DEA can only provide estimates of the true efficiency scores. It has been proven that DEA is a consistent but biased estimator of the true efficiency scores.

In the multivariate setting, the asymptotic distribution of the DEA estimator cannot be obtained explicitly, and the only way to approximate its asymptotic distribution is through bootstrapping techniques. In this chapter, the asymptotic properties of the DEA estimator were discussed and the main bootstrap approaches that are being used in the DEA context were reviewed. Due to the boundary conditions, naïve bootstrapping is proved to provide inconsistent estimates, and smoothing techniques or subsampling bootstrap should be used instead. Among the different bootstrapping techniques, subsampling bootstrap is the computationally easiest one as it does not require any smoothing of the empirical distribution of inefficiencies. Furthermore, it does not require any restrictive assumption about the distribution of inefficiencies among DMUs, allowing for heterogeneity in the efficiency estimation.

Although the application of bootstrapping approaches has been studied extensively in the one-stage DEA literature, there are very limited attempts for making statistical inference in NDEA. In the following chapter, bootstrapping is extended to cases where the production process has a two-stage structure.

6

Bootstrapping in Network DEA

Until now, studies on making statistical inference about DEA have been limited to production processes with a one-stage structure. However, there is a high volume of studies in the DEA field, that have developed the Network DEA (NDEA) models to measure the efficiency of DMUs with more complex production processes that involve more than one stages to produce the final outputs.

Despite the great number of applications of NDEA, there are very limited attempts for making statistical inference in NDEA. Trinh & Zelenuyk (2015), based on the work of Simar & Wilson (2002), developed hypothesis tests to examine whether the difference between the first moments and the difference between the density distributions of the efficiency scores in one-stage DEA and NDEA is significant. Bostian et al. (2018) suggested a statistical approach to make inference about NDEA based on a parametric Bayesian approach. Dia et al. (2020) applied a kernel smoothing bootstrap in a three-stage NDEA assessing the efficiency of Canadian credit unions, where the CRS and VRS efficiency scores of each stage are calculated independently. The overall efficiency is then calculated as the average or as the product of the stage efficiency scores. To the best of our knowledge, there is no study investigating the construction of confidence interval estimates for the overall and stage efficiency scores in network production structures taking into account the connection between the stages.

The aim of this Chapter is to address this deficiency in the DEA literature by studying the performance of subsampling bootstrap in NDEA, through Monte Carlo simulations. Among the different bootstrapping approaches, subsampling bootstrap does not require any restrictive assumptions about the distribution of inefficiencies, while is computationally easier compared to the heterogeneous smooth bootstrap. Therefore, it was chosen as the most appropriate for NDEA, especially since in network structures the dimensions of the model are usually higher compared to one-stage structures.

In this Chapter, the general two-stage structure is being studied and the stage efficiency estimates are calculated using the additive decomposition approach, upon the assumption of VRS. The input orientation is being examined, although the extension to the output orientation is straightforward. Coverage probabilities of the confidence intervals for a fixed point coming from two data generating processes (DGPs) - defined on a five and a seven-dimensional input-output space, respectively - are calculated. We show that in NDEA, coverage probabilities, i.e. the true probabilities that the estimated confidence intervals include the true efficiency scores, are more sensitive to the choice of subsample size than in one-stage structures. Finally, the subsampling methodology is applied to the data set referring to European railways that was discussed in Chapter 4 to get confidence interval estimates about the true environmental efficiency of European railways and demonstrate the performance of subsampling bootstrap in real-world cases. Some practical rules to overcome the issues that arise in the estimation of the confidence intervals due to the peculiarities of the additive decomposition algorithm are also suggested.

The remainder of this Chapter is structured as follows. In Section 6.1, the true frontier in the general two-stage structure is defined. In Section 6.2, the Data Generating Process is axiomatically defined. In Section 6.3, the corresponding DEA estimator and its properties are discussed. In Section 6.4 subsampling methodology is adapted to NDEA. Section 6.5 includes details about the Monte Carlo simulations and a discussion of the results. Section 6.6 is an illustration of subsampling to a data set, where the production model has a network structure. Finally, in Section 6.7, conclusions, limitations and future directions of this study are provided. A list of the notations that are being used in this Chapter can be found in the Appendix. This Chapter is based on the published work by Michali et al. (2022).

6.1 The frontier models in the general two-stage structure

Consider the general two-stage series structure depicted in Figure 6.1. Suppose that a Decision Making Unit (DMU) in a first stage consumes P inputs $x_p = (x_1, ..., x_P) \in \mathbb{R}^P_+$ to produce R final first stage outputs $l_r = (l_1, ..., l_R) \in \mathbb{R}^R_+$ and Q intermediate outputs $z_q = (z_1, ..., z_Q) \in \mathbb{R}^Q_+$. In the second stage, intermediate products obtained from the first stage, and external second stage inputs $g_t = (g_1, ..., g_T) \in \mathbb{R}^T_+$ are consumed to produce S final outputs $y_s = (y_1, ..., y_S) \in \mathbb{R}^S_+$.

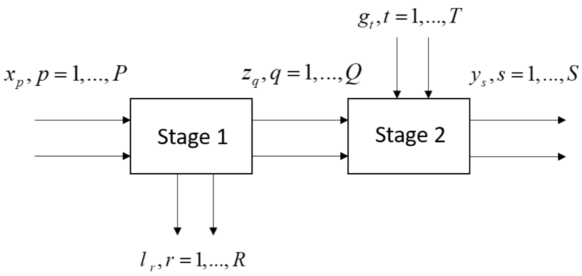


Figure 6.1: General two-stage series network structure of a DMU

In this network structure, the true production possibility set (PPS) of the overall production process is defined as

$$T = \left\{ (x, l, z, g, y) \in \mathbb{R}_+^{P+R+Q+T+S} \middle| x \text{ can produce } l, \text{ and } z, z \text{ and } g \text{ can produce } y \right\}.$$

The PPS T is unknown and is defined based on the following assumptions discussed in Shephard (1970) and Banker et al. (1984)

(i) T is closed and convex

(ii) $(x, l, z, g, y) \notin T$ if $x = 0$ and $(l, z, g, y) \overset{\neq}{\geq} 0$ or if $z = g = 0$ and $y \overset{\neq}{\geq} 0$, i.e. all production requires some inputs,

(iii) if $(x, l, z, g, y) \in T$ and $\bar{x} \geq x, \bar{l} \leq l$ and $\bar{g} \geq g, \bar{y} \leq y$ then $(\bar{x}, \bar{l}, \bar{z}, \bar{g}, \bar{y}) \in T$ (strong disposability of inputs and outputs).

Consider the decomposition of the production process into its component stages and let T_1 and T_2 denote the PPSs of the first and second stage respectively. Then,

$$T_1 = \{(x, l, z) \in \mathbb{R}_+^{P+R+Q} | \exists (g, y) \in \mathbb{R}_+^{T+S} : (x, l, z, g, y) \in T\}, \tag{6.1}$$

$$T_2 = \{(z, g, y) \in \mathbb{R}_+^{Q+T+S} | \exists (x, l) \in \mathbb{R}_+^{P+R} : (x, l, z, g, y) \in T\}. \tag{6.2}$$

T, T_1 and T_2 can be described by their input or output correspondence sets, which inherit their properties. The input possibility sets for the overall process, the first and the second stage, respectively are

$$X(l, y) = \{(x, z, g) \in \mathbb{R}_+^{P+Q+T} | (x, l, z, g, y) \in T\}, \tag{6.3}$$

$$X_1(l, z) = \{x \in \mathbb{R}_+^{P} | (x, l, z) \in T_1\}, \tag{6.4}$$

$$X_2(y) = \{(z, g) \in \mathbb{R}_+^{Q+T} | (z, g, y) \in T_2\}, \tag{6.5}$$

and the output possibility sets of the overall process, the first and second stage, respectively, are

$$Y(x, z, g) = \{(l, y) \in \mathbb{R}_+^{R+Q+S} | (x, l, z, g, y) \in T\}, \tag{6.6}$$

$$Y_1(x) = \{(l, z) \in \mathbb{R}_+^{R+Q} | (x, l, z) \in T_1\}, \tag{6.7}$$

$$Y_2(z, g) = \{y \in \mathbb{R}_+^{S} | (z, g, y) \in T_2\}. \tag{6.8}$$

The efficient boundaries of the input possibility sets are defined below.

$$\partial X(l, y) = \{(x, z, g) | (x, z, g) \in X(l, y), \theta^0 (x, z, g) \notin X(l, y), \forall 0 < \theta^0 < 1\}, \tag{6.9}$$

$$\partial X_1(l, z) = \{x | x \in X(l, z), \theta^1 x \notin X(l, z, y), \forall 0 < \theta^1 < 1\}, \tag{6.10}$$

$$\partial X_2(y) = \{(z, g) | (z, g) \in X(y), \theta^2 (z, g) \notin X(y), \forall 0 < \theta^2 < 1\}. \tag{6.11}$$

Similarly, the efficient boundaries of the output possibility sets are defined as

$$\partial Y(x, z, g) = \{(l, y) | (l, y) \in Y(x, z, g), \lambda^0 (l, y) \notin Y(x, z, g), \forall \lambda^0 > 1\}, \tag{6.12}$$

$$\partial Y_1(x) = \{(l, z) | (l, z) \in Y_1(x), \lambda^1 (l, z) \notin Y_1(x)\}, \tag{6.13}$$

$$\partial Y_2(z, g) = \{y | y \in Y_2(z, g), \lambda^2 y \notin Y_2(z, g)\}. \tag{6.14}$$

Let DMU_{j_0} denote a DMU under evaluation. Then, the Farrell (1957) input efficiency measure of DMU$_{j_0}$ for the overall process and the two stages, respectively can be defined as

$$\theta^0(x,l,z,g,y) = \inf\{\theta^0|\theta^0(x,z,g) \in X(l,y)\}, \tag{6.15}$$

$$\theta^1(x,l,z) = \inf\{\theta^1|\theta^1 x \in X_1(l,z)\}, \tag{6.16}$$

$$\theta^2(z,g,y) = \inf\{\theta^2|\theta^2(z,g) \in X_2(y)\}. \tag{6.17}$$

The output efficiency measure of DMU$_{j_0}$ for the overall process and the two stages, respectively is defined as

$$\lambda^0(x,l,z,g,y) = \sup\{\lambda^0|\lambda^0(l,y) \in Y(x,z,g)\}, \tag{6.18}$$

$$\lambda^1(x,l,z) = \sup\{\lambda^1|\lambda^1(l,z) \in Y_1(x)\}, \tag{6.19}$$

$$\lambda^2(z,g,y) = \sup\{\lambda^2|\lambda^2(y) \in Y_2(z,g)\}. \tag{6.20}$$

Here, the input efficiency measure is going to be used, and for simplicity of notation let $\theta(x,z,g) \equiv \theta^0$, $\theta(x) \equiv \theta^1$ and $\theta(z,g) \equiv \theta^2$.

The Farrell input efficiency measure $\theta^{\phi}_{j_0}$, $\phi = 0,1,2$ of the DMU$_{j_0}$ under evaluation is the euclidean distance of the point $(x_{j_0}, l_{j_0}, z_{j_0}, g_{j_0}, y_{j_0})$ from its projection to the overall or stage-efficient boundary, respectively, in direction parallel to the subspace defined by the overall or stage-specific input coordinates, respectively, while keeping the level of outputs fixed. Note that the above formulations refer to the specific network structure, but they can be easily adapted to define the efficiency measure when the production process has a different structure.

6.2 The data generating process

Let $S_N = \{(x_j, l_j, z_j, g_j, y_j)|j = 1, ..., N\}$ be a random sample of N DMUs that is assumed to be generated by an unknown data generating process (DGP), $\mathcal{P}$. Similarly to one-stage structures, in the general two-stage structure, the unknown DGP $\mathcal{P}$ is defined on the following assumptions:

(i) All observations in S_N are independent and indentically distributed (i.i.d.) random variables in T, and $(x_j, l_j, z_j) \in T_1$, $(z_j, g_j, y_j) \in T_2$, for $j = 1, ..., N$.

(ii) $f(x,l,z,g,y)$ has support T, is continuous on T, and $f(\theta^0 x, \theta^0 z, l, \theta^0 g, y) > 0$,

$f(x, z, l)$ has support T_1, is continuous on T_1 and $f(\theta^1 x, z, l) > 0$,

$f(z, g, y)$ has support T_2, is continuous on T_2 and $f(\theta^2 z, \theta^2 g, y) > 0$.

(iii) $\theta(x, l, z, g, y|\mathcal{P})$ is twice continuously differentiable $\forall (x, l, z, g, y) \in T$

$\theta(x, z, l|\mathcal{P})$ is twice continuously differentiable $\forall (x, z, l) \in T_1$

$\theta(z, g, y|\mathcal{P})$ is twice continuously differentiable $\forall (z, g, y) \in T_2$

Assumption (ii) is required to ensure that DMUs will be observed on the frontier of each process, whereas assumption (iii) requires that the frontier of each process is smooth.

6.3 The estimator

In practice, the true production sets T, T_1, T_2 are unknown, and therefore, the efficiency scores $\theta^{\phi}_{j_0}$, $\phi = 0, 1, 2$ of a DMU_{j_0} need to be estimated based on the observed sample of DMUs.

Let $\widehat{\theta}^0_j$, $\widehat{\theta}^1_j$ and $\widehat{\theta}^2_j$ denote the estimators of the overall, the first stage and the second stage efficiency, respectively, for DMU_j, $j = 1, 2, .., N$, with respect to the observed sample S_N.

The efficiency estimate for the overall process, under the VRS assumption can be obtained using model (4.6) provided in Chapter 4, Section 4.4, where instead of the undesirable outputs y_b the second stage inputs g are used. Similarly, when stage 1 is the priority stage, model (4.7) can be used to obtain the first stage efficiency estimates, replacing y_b with g, and then use relation (3.25) for the second stage efficiency estimate. If stage 2 is given pre-emptive priority, then, model (4.8) is being used replacing y_b with g and the efficiency estimate for the first stage is obtained from the decomposition equation (3.27).

Note that the true PPS always includes the sample PPS, i.e $\widehat{T} \subset T$, $\widehat{T_1} \subset T_1$ and $\widehat{T_2} \subset T_2$. Therefore, it holds that $\widehat{\theta}^0 \geq \theta^0$, for all points $(x, l, z, g, y) \in \widehat{T}$, $\widehat{\theta}^1 \geq \theta^1$, for all $(x, l, z) \in \widehat{T_1}$, and $\widehat{\theta}^2 \geq \theta^2$, for all $(z, g, y) \in \widehat{T_2}$. As it was discussed in the previous Chapter, Kneip et al. (1998) proved that the rate at which the estimator converges depends on the number of input and output variables and the degree of the true frontier smoothness. If it is assumed that the true frontier is twice differentiable then, the rate of convergence in the one-stage structure is $N^{-\frac{2}{(p-1)+q+2}}$, where N is the size of the sample and p, q are the number of inputs and outputs, respectively. This result can be easily extended to series network structures. The convergence rates of the overall and stage processes, respectively, are given in the following Proposition.

Proposition 6.3.1. *For a two-stage process with final first stage inputs and second stage outputs, under the assumptions 6.1 (i)-(iii) and 6.2 (i)-(iii), it holds that*

$$\widehat{\theta}^0 - \theta^0 = O_p(N^{-\frac{2}{P+Q+R+T+S+1}}) \tag{6.21}$$

$$\widehat{\theta}^1 - \theta^1 = O_p(N^{-\frac{2}{P+Q+R+1}}) \tag{6.22}$$

$$\widehat{\theta}^2 - \theta^2 = O_p(N^{-\frac{2}{Q+T+S+1}}) \tag{6.23}$$

Proof. The proof for the convergence rate for each stage results from the proof of the convergence rate in the one-stage structure provided by Kneip et al. (1998, pg. 7-9) considering the input and output possibility sets of each stage and the overall process, as those are defined in relations (6.3)-(6.5) and (6.6)-(6.8). □

In the next section, the subsampling bootstrap methodology used for the approximation of $\theta^\phi, \phi = 0, 1, 2$ is discussed.

6.4 Bootstrapping with Subsampling in NDEA

The subsampling methodology suggested by Kneip et al. (2008) and Simar & Wilson (2010) can be adapted to the NDEA case, by considering the inner structure of DMUs. Kneip et al. proved that drawing pseudo-samples of size $m = N^k$, for $k \in (0, 1)$, where N is the original sample size, allows for consistent inference.

Consider the general two-stage network structure depicted in Figure 6.1 and let $\mathcal{S}_N$ be the original sample of N DMUs. In each replication, a bootstrap subsample

$$\mathcal{S}^*_m = \{(x^*_j, l^*_j, z^*_j, g^*_j, y^*_j)|j = 1, ..., m\}$$

consists of $m < N$ independent and identically distributed (iid) draws with replacement from the original sample.

For each bootstrap subsample, for a DMU$_{j_0}$, models (4.6), (4.7) (or model (4.8)) and relation (3.25) (or relation (3.27)) are used to get the overall and stage bootstrap efficiency estimates $\widehat{\theta}^{0*}_{j_0}, \widehat{\theta}^{1*}_{j_0}$ and $\widehat{\theta}^{2*}_{j_0}$, respectively.

Let $A = 2/(P + R + Q + T + S + 1)$, $B = 2/(P + R + Q + 1)$ and $\Gamma = 2/(Q + T + S + 1)$. Adjusting the result of Kneip et al. (2008) to the specific general network structure, it holds that

$$m^{\tau}\left(\frac{\widehat{\widehat{\theta}}_{j_0}^{\phi}}{\widehat{\theta}_{j_0}^{\phi}}-1\right) \overset{approx}{\sim} N^{\tau}\left(\frac{\widehat{\theta}_{j_0}^{\phi}}{\theta_{j_0}^{\phi}}-1\right), (\tau,\phi) \in \{(A,0),(B,1),(\Gamma,2)\}. \tag{6.24}$$

Therefore, for each DMU_j, $j = 1, ..., N$ in the sample, the bootstrap estimates of the overall and stage efficiencies, respectively, can be used to approximate the right-hand side in relations (6.24) for each process $\phi = 0, 1, 2$, and construct the confidence intervals for the true overall and stage efficiencies.

Similarly to one-stage structures, the quantiles $c_{a/2,m,\phi}$ and $c_{1-a/2,m,\phi}$ of the unknown distributions $N^{\tau}\left(\frac{\widehat{\theta}_{j_0}^{\phi}}{\theta_{j_0}^{\phi}}-1\right)$ of a DMU_{j_0}, for a chosen confidence level $a \in (0,1)$, can be estimated by the quantiles $\widehat{c}_{a/2,m,\phi}$ and $\widehat{c}_{1-a/2,m,\phi}$ of the bootstrap distributions $m^{\tau}\left(\frac{\widehat{\widehat{\theta}}_{j_0}^{\phi}}{\widehat{\theta}_{j_0}^{\phi}}-1\right)$, where $(\tau,\phi) \in \{(A,0),(B,1),(\Gamma,2)\}$.

Finally, the true overall and stage efficiencies will lie in the confidence intervals

$$\left[\frac{\widehat{\theta}_{j_0}^{\phi}}{(1+N^{-\tau}\widehat{c}_{1-a/2,m,\phi})}, \frac{\widehat{\theta}_{j_0}^{\phi}}{(1+N^{-\tau}\widehat{c}_{a/2,m,\phi})}\right], \tag{6.25}$$

for $(\tau,\phi) \in \{(A,0),(B,1),(\Gamma,2)\}$, respectively.

6.5 Monte Carlo Simulations

In order to examine the performance of subsampling bootstrap in network structures, Monte Carlo simulations were performed. Different sets of experiments were conducted for the general network structure (see Figure 6.1) using a five-dimensional and a seven-dimensional DGP, respectively, i.e. for $P = R = Q = T = S = 1$ and for $P = R = T = 1$, $Q = S = 2$. For each DGP, two sets of experiments were performed, one to estimate the efficiency of the overall process and another set of experiments to estimate the efficiencies of the stage processes.

Each experiment consists of 1000 Monte Carlo trials with 2000 bootstrap replications in each trial. Also, experiments were conducted for four sample sizes $N \in \{25, 50, 100, 200\}$ and twelve subsample sizes $m = [N^k]$, for $k \in \{0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$, where $[N^k]$ denotes the integer part of N^k.

In all the experiments, the confidence intervals for the true overall and stage efficiencies of a fixed point were estimated. Due to sampling noise, the true probability that the confidence interval contains the true efficiency score may differ from the nominal probability that the confidence level represents. Therefore, through the Monte Carlo trials, the true probability of the confidence interval containing the true parameter is approximated by calculating the coverage probabilities, i.e. the actual proportion of the estimated confidence intervals that include the true efficiency score.

In set of experiments referring to the five-dimensional case, an efficient first stage input x_e is drawn from the uniform distribution $U[5, 20]$. The efficient input is used to generate one final first stage output $l = (x_e)^\beta$ and one intermediate first stage output $z = (x_e)^\gamma$, where $\beta, \gamma > 0$. Second stage-specific efficient input g_e is also drawn from the uniform distribution $U[5, 20]$. The second stage efficient input and the intermediate product from the first stage, are used to generate one final output $y = \left((g_e)^\zeta z^\xi e^{-0.2|\epsilon|}\right)^\nu$, where $\epsilon \sim N(0, 1)$ and $\zeta, \xi, \nu > 0$. The term $e^{-0.2|\epsilon|}$ is added to the DGP to better reflect real-world scenarios by adding some stochasticity to the data.

For the experiments of the seven-dimensional case, a similar DGP is used. An efficient first stage input x_e and a second stage-specific input g_e are drawn from the uniform distribution $U[5, 20]$. One final first stage output is given by $l = (x_e)^\beta$ and the two intermediate first stage outputs are given by $z_1 = (x_e)^\gamma$ and $z_2 = (x_e)^\delta$, respectively, for some $\beta, \gamma, \delta > 0$. The two final outputs are generated by $y_1 = \left((g_e)^\zeta z_1^\xi z_2^\lambda e^{-0.2|\epsilon|}\right)^\nu$ and $y_2 = \left((g_e)^\zeta z_1^\lambda z_2^\xi e^{-0.2|\epsilon|}\right)^\nu$, where $\epsilon \sim N(0, 1)$ and $\zeta, \xi, \lambda, \nu > 0$. Both in the five and the seven-dimensional case, the DGP ensures that at each stage, as well as in the overall process, an increase in stage-specific inputs results in a non-proportional increase in the stage outputs, to account for VRS.

As it was mentioned before, estimates of the overall and the stage efficiencies, respectively, were obtained through different sets of experiments. In the set of experiments that were performed to estimate the overall efficiency, inefficiency was added to the first stage inputs $x = x_e e^{0.2|\epsilon|}$. The inefficient inputs were then used to calculate the sample overall efficiency of the fixed point under investigation. In the set of experiments concerning the true stage efficiencies, inefficiency was introduced to each stage's specific inputs, i.e., $x = x_e e^{0.2|\epsilon|}$ and $g = g_e e^{0.2|\epsilon|}$, and these inefficient first and second stage-specific inputs were used to calculate the first and second stage sample efficiencies of the fixed point. Since the input orientation is considered, the true overall inefficiency can be defined as the proportion of the input going to the production that exceeds the efficient input. Therefore, the true first and second stage efficiencies are defined as x_e/x and g_e/g, respectively.

In the case of multiple inputs, since radial projections to the efficient frontier are being used, all inputs are reduced by the same proportion, and the true efficiency score would be defined in a similar way.

Experiments were performed for $\beta = 0.6, \gamma = 0.7, \delta = 0.8, \zeta = 0.1, \xi = 0.3, \lambda = 0.5$ and $\nu = 0.8$, and for the case when $x_e = g_e$. All the efficiency scores were obtained for the fixed point with $(x_{e0}, g_{e0}) = (13, 13)$ and $\epsilon_0 = 1$, and with true overall and stage efficiency score $\theta_0^{\phi} = 0.8187$, for $\phi = 0, 1, 2$. This point lies about in the middle of the cloud of the generated points.

In many subsamples, the fixed point shows infeasibility in one of the two stages. Infeasibility problems occur when one (or more) of the outputs of the fixed point (overall or stage-specific output) is (are) greater than the respective maximum output(s) of the points included in the subsample, and/or when one or more of the inputs of the fixed point is (are) less than the corresponding minimum input(s) of the points belonging to the subsample. Infeasibility problems in the subsamples occur more often in NDEA compared to one stage DEA because of the higher dimensions. Therefore, in this study, in all bootstrap replications the fixed point of interest was being added to the subsample.

Another way to treat infeasibilities is to set the efficiency score of the infeasible DMU equal to one. In the additive efficiency decomposition approach, in some cases the algorithm may give zero optimal multipliers and therefore, zero efficiency scores. For the overall process this happens in cases where the DMU is overall efficient. Therefore, for DMUs with zero overall efficiency score, it can be set that their overall efficiency score is equal to one. This is an approach also adopted by Chen et al. (2009). However, it may also happen a DMU to be assigned a non-zero overall efficiency score, and get a zero efficiency score for the one of the two stages when a stage is considered as priority stage, but get a non-zero efficiency score when the priority stage changes. So, if the DMU under evaluation is not included in the subsample, zero overall efficiencies need to be set equal to one, and the stage efficiency estimates need to be obtained by the priority stage models. On the other hand, including the DMU under evaluation in the subsample significantly reduces the times when this happens and therefore, provides a partial solution to the problem.

Confidence interval estimates for the true efficiency scores were obtained for three levels of significance, $\alpha \in \{0.1, 0.05, 0.01\}$ to get the $90\%, 95\%$ and 99% confidence intervals. Coverage probabilities represent the proportion of confidence intervals, i.e. the proportion of Monte Carlo trials, that the true efficiency score is included in the estimated confidence interval. The resulting coverage probabilities of the confidence intervals for the overall and stage efficiencies, for the five and seven-dimensional cases, respectively, when the DMU under evaluation is included in the subsample are given in Tables 6.1 and 6.2.

Results from Monte Carlo simulations when the DMU under evaluation is not included in the subsample, the efficiency score of infeasible DMUs is set equal to one and both stages are considered alternately as priority stages are provided in Tables 6.3 and 6.4. In most of the cases, the confidence intervals obtained with this approach have higher range compared to those obtained when the DMU under evaluation is included in the subsample. Furthermore, from the coverage probabilities reported in 6.3 and 6.4, it seems that changing the priority stage does not affect the coverage probabilities.

Table 6.1: Coverage probabilities for the confidence interval estimates when $P = R = Q = T = S = 1$ and the DMU under evaluation is included in the subsample

		Overall			1st Stage			2nd Stage		
		1-α			1-α			1-α		
N	k	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	0.40	0.938	0.976	0.983	0.930	0.965	0.976	0.980	0.984	0.985
	0.45	0.984	0.988	0.988	0.965	0.982	0.985	0.986	0.987	0.987
	0.50	0.989	0.989	0.989	0.978	0.985	0.985	0.991	0.991	0.991
	0.55	0.989	0.989	0.989	0.979	0.983	0.985	0.990	0.991	0.991
	0.60	0.990	0.990	0.990	0.981	0.988	0.988	0.990	0.995	0.995
	0.65	0.984	0.992	0.992	0.981	0.990	0.993	0.959	0.985	0.996
	0.70	0.974	0.994	0.995	0.972	0.990	0.993	0.930	0.970	0.995
	0.75	0.943	0.975	0.996	0.938	0.973	0.992	0.837	0.917	0.988
	0.80	0.889	0.951	0.993	0.903	0.951	0.983	0.740	0.861	0.969
	0.85	0.833	0.910	0.978	0.879	0.923	0.977	0.647	0.783	0.944
	0.90	0.742	0.841	0.964	0.818	0.880	0.962	0.517	0.670	0.878
	0.95	0.646	0.787	0.924	0.766	0.857	0.931	0.410	0.576	0.808
50	0.40	0.986	0.999	0.999	0.920	0.986	1.000	0.999	0.999	0.999
	0.45	1.000	1.000	1.000	0.941	0.993	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	0.978	1.000	1.000	0.999	1.000	1.000
	0.55	1.000	1.000	1.000	0.987	1.000	1.000	0.995	0.999	1.000
	0.60	0.995	1.000	1.000	0.997	1.000	1.000	0.969	0.994	1.000

	0.65	0.987	0.998	1.000	0.994	0.999	1.000	0.906	0.973	0.999
	0.70	0.958	0.982	0.999	0.983	0.996	1.000	0.767	0.905	0.992
	0.75	0.910	0.966	0.994	0.969	0.987	1.000	0.618	0.787	0.968
	0.80	0.826	0.915	0.980	0.937	0.974	0.995	0.434	0.643	0.910
	0.85	0.707	0.826	0.954	0.900	0.942	0.984	0.296	0.466	0.771
	0.90	0.556	0.723	0.905	0.855	0.909	0.975	0.180	0.320	0.643
	0.95	0.412	0.568	0.821	0.769	0.859	0.949	0.104	0.205	0.481
100	0.40	1.000	1.000	1.000	0.868	0.967	1.000	1.000	1.000	1.000
	0.45	1.000	1.000	1.000	0.900	0.977	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	0.958	0.993	1.000	0.989	0.998	1.000
	0.55	1.000	1.000	1.000	0.980	0.996	1.000	0.945	0.987	0.999
	0.60	0.996	1.000	1.000	0.988	0.997	1.000	0.791	0.941	0.999
	0.65	0.973	0.995	1.000	0.995	0.999	1.000	0.539	0.792	0.983
	0.70	0.903	0.959	0.997	0.988	0.996	1.000	0.261	0.488	0.884
	0.75	0.786	0.899	0.986	0.975	0.991	0.999	0.130	0.300	0.730
	0.80	0.563	0.789	0.947	0.944	0.978	0.996	0.047	0.140	0.489
	0.85	0.361	0.556	0.868	0.903	0.946	0.988	0.015	0.059	0.273
	0.90	0.210	0.363	0.717	0.839	0.908	0.973	0.007	0.025	0.131
	0.95	0.106	0.211	0.515	0.773	0.850	0.940	0.006	0.009	0.069
200	0.40	1.000	1.000	1.000	0.737	0.929	0.999	1.000	1.000	1.000
	0.45	1.000	1.000	1.000	0.819	0.958	0.999	0.995	1.000	1.000
	0.50	1.000	1.000	1.000	0.920	0.982	0.999	0.817	0.976	1.000
	0.55	0.998	1.000	1.000	0.958	0.994	1.000	0.430	0.791	0.995
	0.60	0.971	0.996	1.000	0.985	0.997	1.000	0.098	0.365	0.907
	0.65	0.868	0.964	0.999	0.995	0.997	1.000	0.014	0.103	0.621
	0.70	0.628	0.841	0.985	0.994	1.000	1.000	0.001	0.015	0.288
	0.75	0.333	0.569	0.903	0.987	0.996	1.000	0.000	0.000	0.077
	0.80	0.136	0.310	0.731	0.967	0.988	0.999	0.000	0.000	0.017
	0.85	0.037	0.124	0.442	0.925	0.964	0.997	0.000	0.000	0.002
	0.90	0.011	0.036	0.226	0.861	0.927	0.984	0.000	0.000	0.001
	0.95	0.007	0.013	0.087	0.773	0.846	0.949	0.000	0.000	0.001

Table 6.2: Coverage probabilities for the confidence interval estimates when $P = R = T = 1$, $Q = S = 2$ and the DMU under evaluation is included in the subsample.

		Overall			1st Stage			2nd Stage		
		1-α			1-α			1-α		
N	k	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	0.40	0.728	0.872	0.978	0.948	0.977	0.988	0.171	0.331	0.769
	0.45	0.882	0.957	0.996	0.970	0.985	0.989	0.312	0.517	0.901
	0.50	0.950	0.987	0.997	0.982	0.990	0.990	0.460	0.756	0.962
	0.55	0.945	0.985	0.997	0.982	0.989	0.990	0.450	0.751	0.964
	0.60	0.977	0.995	0.997	0.983	0.990	0.990	0.557	0.854	0.988
	0.65	0.993	0.998	0.998	0.986	0.992	0.992	0.872	0.957	1.000
	0.70	0.993	0.997	0.998	0.977	0.992	0.993	0.917	0.969	1.000
	0.75	0.972	0.993	0.998	0.958	0.982	0.992	0.960	1.000	1.000
	0.80	0.939	0.974	0.997	0.930	0.964	0.988	1.000	1.000	1.000
	0.85	0.908	0.951	0.994	0.895	0.942	0.980	1.000	1.000	1.000
	0.90	0.849	0.914	0.980	0.817	0.901	0.966	0.999	1.000	1.000
	0.95	0.774	0.871	0.952	0.748	0.845	0.943	0.996	0.998	1.000
50	0.40	0.598	0.808	0.980	0.908	0.984	1.000	0.056	0.193	0.585
	0.45	0.761	0.903	0.995	0.935	0.989	1.000	0.147	0.320	0.748
	0.50	0.910	0.978	0.999	0.977	1.000	1.000	0.303	0.491	0.893
	0.55	0.941	0.990	1.000	0.983	0.999	1.000	0.375	0.570	0.930
	0.60	0.980	0.996	1.000	0.995	1.000	1.000	0.497	0.785	0.965
	0.65	0.987	0.999	1.000	0.996	1.000	1.000	0.597	0.891	0.998
	0.70	0.990	0.996	1.000	0.991	0.997	1.000	0.847	0.946	1.000
	0.75	0.977	0.990	0.999	0.972	0.989	1.000	0.917	0.980	1.000
	0.80	0.947	0.976	0.996	0.935	0.973	0.997	0.954	1.000	1.000
	0.85	0.907	0.949	0.985	0.890	0.940	0.986	1.000	1.000	1.000
	0.90	0.851	0.910	0.969	0.830	0.898	0.964	1.000	1.000	1.000
	0.95	0.753	0.842	0.943	0.751	0.831	0.935	0.998	1.000	1.000
100	0.40	0.556	0.770	0.985	0.816	0.956	0.999	0.048	0.170	0.543
	0.45	0.652	0.845	0.992	0.858	0.972	1.000	0.075	0.242	0.618

	0.50	0.863	0.970	0.998	0.940	0.983	1.000	0.229	0.418	0.814
	0.55	0.929	0.986	1.000	0.969	0.995	1.000	0.317	0.508	0.887
	0.60	0.972	0.994	1.000	0.987	0.998	1.000	0.419	0.634	0.943
	0.65	0.991	0.998	1.000	0.992	1.000	1.000	0.566	0.791	0.976
	0.70	0.991	0.999	0.999	0.988	0.998	1.000	0.697	0.915	0.999
	0.75	0.983	0.993	0.999	0.976	0.990	0.999	0.886	0.962	1.000
	0.80	0.949	0.979	0.998	0.944	0.975	0.996	0.954	0.997	1.000
	0.85	0.894	0.943	0.991	0.899	0.943	0.984	0.996	1.000	1.000
	0.90	0.820	0.887	0.964	0.832	0.896	0.963	0.999	1.000	1.000
	0.95	0.726	0.808	0.923	0.746	0.832	0.928	0.995	0.999	1.000
200	0.40	0.385	0.642	0.969	0.627	0.868	0.997	0.020	0.111	0.497
	0.45	0.536	0.794	0.988	0.734	0.910	0.998	0.051	0.197	0.600
	0.50	0.753	0.918	0.999	0.860	0.970	1.000	0.160	0.370	0.733
	0.55	0.894	0.978	1.000	0.935	0.990	1.000	0.281	0.518	0.854
	0.60	0.964	0.991	1.000	0.974	0.994	1.000	0.449	0.642	0.926
	0.65	0.986	0.998	1.000	0.987	1.000	1.000	0.581	0.733	0.969
	0.70	0.993	0.998	1.000	0.993	0.998	1.000	0.690	0.880	0.992
	0.75	0.983	0.994	0.999	0.977	0.993	0.999	0.848	0.953	1.000
	0.80	0.962	0.984	0.997	0.944	0.976	0.994	0.934	0.987	1.000
	0.85	0.901	0.959	0.989	0.880	0.938	0.983	0.975	1.000	1.000
	0.90	0.819	0.887	0.970	0.828	0.876	0.956	0.999	1.000	1.000
	0.95	0.697	0.788	0.913	0.749	0.823	0.901	0.988	0.999	1.000

The results indicate that the choice of the subsample size is crucial for getting high coverage probabilities, irrespective of the original sample size. However, as the sample size increases, sensitivity on the subsample size seems to increase. Coverage probabilities of the confidence intervals for the true overall and first stage efficiency seem to rise as k increases and then, in most cases, after some point they show a downturn. These conclusions seem to be in accordance with those for one stage structures (see Kneip et al. (2008, pg. 1682-1683)). According to the results, one of the two stages shows greater sensitivity to the subsample size, and for some subsample sizes confidence interval estimates can even have zero coverage probabilities. However, this sensitivity is not related to the choice of priority stage. Monte Carlo simulations were performed both by treating first stage as the priority stage, and then by treating the second stage as the priority stage, and coverage probabilities were not affected. Results for stage one being the priority stage and then for stage two being the priority stage, when the DMU under evaluation is not included in the subsample can be found in the Appendix. The issue lies in the additive decomposition algorithm, which in some cases assigns zero or even negative efficiency scores, though less often.[1]. If this happens at a stage level, and not to the overall efficiency, then this mainly affects the coverage probabilities of that stage. The number of bootstrap replications certainly affects the coverage probabilities in NDEA more than in conventional DEA due to the lower convergence rate. However, it is computationally difficult to further increase the replications to a number significantly higher than 2000.

In the five-dimensional case (see Table 6.1), coverage probabilities for the overall and the second stage efficiency are very high for lower values of k, but as k increases they get very poor and even tend to zero as the original sample size increases. In the cases when the estimations of the overall efficiency scores are poor, the coverage probabilities of the second stage seem to be also affected. For the first stage, $k = 0.60$ or $k = 0.65$ seems to result in higher coverages.

In the seven-dimensional case (see Table 6.2), for the overall and the first stage, subsample sizes resulting from $k = 0.65$ or $k = 0.70$, in most of the cases, yield higher coverage probabilities, however, for a wide range of k, coverage probabilities are still very high. Coverage probabilities for the second stage true efficiency are very poor for the first half values of k, especially for the larger sample sizes. Nonetheless, as k increases coverage probabilities for the second stage become very high; for this stage, in most of the cases, a value of k around 0.9 gives the best coverage probabilities. The difference between the results in the five and the seven-dimensional cases indicate that the coverage probabilities and the optimal subsample sizes strongly depend on the DGP.

Wrong choice of subsample size may result in totally misleading confidence interval estimates. Politis et al. (2001) suggested a minimum volatility criterion for the selection of the optimal subsample size. Let $I_{m,low}$ and $I_{m,up}$ be the lower and the upper bounds of a confidence interval estimate, resulting from subsampling bootstrap, with subsample size m. For a small $r \in \mathbb{Z}^+$, Politis et al. suggested calculating the volatility index $V_m = \sum_{m=m_{small}}^{m_{big}} sd\{I_{m-r,low}, ..., I_{m+r,low}\} + \sum_{m=m_{small}}^{m_{big}} sd\{I_{m-r,up}, ..., I_{m+r,up}\}$ and then choosing the subsample size m that corresponds to the minimum V_m.

[1]Similar issues seem to occur with the multiplicative decomposition algorithm. Peyrache & Silva (2022) provide an example with negative efficiency scores in parallel structures.

Table 6.3: Coverage probabilities for the confidence interval estimates when $P = R = Q = T = S = 1.$ and the DMU under evaluation is not included in the subsample.

		Overall			1st Stage						2nd Stage					
					Stage 1 Priority			Stage 2 Priority			Stage 1 Priority			Stage 2 Priority		
		1-α			1-α			1-α			1-α			1-α		
N	k	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	0.40	0.967	0.994	1.000	0.991	0.994	0.998	0.985	0.991	0.998	0.993	0.997	1.000	0.989	0.997	1.000
	0.45	0.996	1.000	1.000	0.991	0.993	0.998	0.982	0.987	0.997	0.976	0.994	0.999	0.974	0.991	0.999
	0.50	0.998	1.000	1.000	0.990	0.994	0.999	0.983	0.989	0.998	0.959	0.985	0.998	0.958	0.983	0.998
	0.55	0.999	1.000	1.000	0.989	0.994	0.998	0.984	0.990	0.997	0.959	0.986	0.998	0.957	0.983	0.998
	0.60	0.998	1.000	1.000	0.983	0.993	0.997	0.974	0.988	0.996	0.917	0.979	0.998	0.914	0.975	0.998
	0.65	0.989	0.997	1.000	0.948	0.987	0.996	0.943	0.983	0.995	0.821	0.920	0.996	0.820	0.919	0.995
	0.70	0.971	0.993	1.000	0.928	0.978	0.994	0.924	0.970	0.993	0.761	0.883	0.995	0.764	0.885	0.994
	0.75	0.935	0.975	0.999	0.893	0.948	0.994	0.891	0.945	0.993	0.661	0.810	0.967	0.670	0.814	0.965
	0.80	0.891	0.945	0.993	0.873	0.915	0.990	0.870	0.913	0.988	0.572	0.727	0.938	0.572	0.733	0.940
	0.85	0.830	0.909	0.983	0.833	0.896	0.979	0.831	0.894	0.977	0.492	0.654	0.904	0.490	0.659	0.904
	0.90	0.733	0.851	0.954	0.786	0.865	0.958	0.786	0.865	0.957	0.399	0.560	0.820	0.390	0.558	0.816
	0.95	0.636	0.775	0.922	0.738	0.825	0.933	0.736	0.824	0.932	0.336	0.488	0.744	0.331	0.479	0.748
50	0.40	0.994	1.000	1.000	0.970	0.999	1.000	0.971	0.999	1.000	0.990	0.999	1.000	0.988	0.999	1.000
	0.45	1.000	1.000	1.000	0.975	0.997	1.000	0.973	0.997	1.000	0.974	0.998	1.000	0.972	0.998	1.000
	0.50	1.000	1.000	1.000	0.981	0.998	1.000	0.979	0.997	1.000	0.875	0.971	0.999	0.878	0.971	1.000
	0.55	0.999	1.000	1.000	0.978	0.995	1.000	0.977	0.995	1.000	0.816	0.943	0.997	0.825	0.947	0.997
	0.60	0.995	1.000	1.000	0.970	0.985	1.000	0.969	0.985	1.000	0.679	0.859	0.985	0.685	0.865	0.985
	0.65	0.985	0.996	1.000	0.960	0.980	0.998	0.962	0.980	0.998	0.559	0.771	0.959	0.578	0.780	0.966
	0.70	0.962	0.986	1.000	0.936	0.969	0.996	0.936	0.969	0.996	0.418	0.627	0.922	0.431	0.635	0.920
	0.75	0.917	0.968	0.994	0.912	0.955	0.988	0.911	0.955	0.988	0.321	0.522	0.853	0.327	0.526	0.861
	0.80	0.844	0.924	0.987	0.879	0.929	0.979	0.878	0.931	0.979	0.235	0.403	0.751	0.237	0.409	0.757
	0.85	0.717	0.855	0.964	0.842	0.903	0.967	0.842	0.903	0.969	0.168	0.302	0.654	0.172	0.302	0.658
	0.90	0.587	0.744	0.920	0.801	0.873	0.946	0.800	0.871	0.946	0.115	0.230	0.532	0.121	0.227	0.531

	0.95	0.427	0.599	0.839	0.749	0.823	0.919	0.747	0.822	0.919	0.075	0.174	0.430	0.066	0.168	0.425
100	0.40	1.000	1.000	1.000	0.975	0.996	1.000	0.976	0.997	1.000	0.845	0.984	1.000	0.851	0.984	1.000
	0.45	1.000	1.000	1.000	0.982	0.998	1.000	0.983	0.998	1.000	0.732	0.942	1.000	0.747	0.943	1.000
	0.50	1.000	1.000	1.000	0.986	0.996	1.000	0.985	0.996	1.000	0.363	0.704	0.986	0.387	0.735	0.989
	0.55	0.999	1.000	1.000	0.979	0.994	0.998	0.980	0.994	0.998	0.212	0.541	0.943	0.230	0.561	0.952
	0.60	0.997	0.999	1.000	0.977	0.989	0.997	0.977	0.989	0.997	0.110	0.326	0.849	0.123	0.346	0.861
	0.65	0.972	0.996	1.000	0.965	0.981	0.997	0.966	0.981	0.997	0.056	0.175	0.668	0.059	0.186	0.690
	0.70	0.904	0.961	0.997	0.947	0.971	0.992	0.947	0.971	0.992	0.018	0.090	0.463	0.019	0.096	0.480
	0.75	0.781	0.910	0.989	0.933	0.959	0.985	0.933	0.959	0.985	0.009	0.050	0.315	0.009	0.053	0.323
	0.80	0.567	0.772	0.957	0.900	0.946	0.981	0.898	0.946	0.981	0.008	0.031	0.189	0.007	0.029	0.199
	0.85	0.355	0.555	0.872	0.854	0.912	0.964	0.853	0.909	0.964	0.005	0.014	0.118	0.004	0.014	0.118
	0.90	0.205	0.367	0.700	0.796	0.871	0.953	0.796	0.870	0.953	0.002	0.012	0.077	0.001	0.009	0.076
	0.95	0.104	0.211	0.511	0.734	0.828	0.929	0.734	0.826	0.928	0.001	0.009	0.055	0.002	0.006	0.053
200	0.40	1.000	1.000	1.000	0.936	0.993	1.000	0.936	0.993	1.000	0.012	0.269	0.982	0.020	0.294	0.982
	0.45	1.000	1.000	1.000	0.957	0.996	1.000	0.957	0.996	1.000	0.004	0.056	0.833	0.004	0.071	0.846
	0.50	1.000	1.000	1.000	0.982	0.997	1.000	0.982	0.997	1.000	0.000	0.003	0.325	0.001	0.005	0.357
	0.55	0.996	0.999	1.000	0.986	0.997	1.000	0.986	0.997	1.000	0.000	0.003	0.104	0.000	0.002	0.128
	0.60	0.970	0.993	1.000	0.978	0.997	1.000	0.978	0.996	1.000	0.000	0.001	0.025	0.000	0.001	0.029
	0.65	0.858	0.965	0.998	0.974	0.990	0.999	0.973	0.990	0.999	0.000	0.000	0.005	0.000	0.000	0.009
	0.70	0.613	0.842	0.988	0.960	0.981	0.998	0.960	0.981	0.998	0.000	0.000	0.004	0.000	0.000	0.003
	0.75	0.303	0.567	0.914	0.929	0.971	0.995	0.929	0.970	0.995	0.000	0.000	0.000	0.000	0.000	0.000
	0.80	0.106	0.279	0.721	0.898	0.945	0.988	0.897	0.943	0.988	0.000	0.000	0.000	0.000	0.000	0.000
	0.85	0.041	0.103	0.442	0.848	0.906	0.977	0.847	0.904	0.976	0.000	0.000	0.000	0.000	0.000	0.000
	0.90	0.022	0.048	0.200	0.780	0.865	0.957	0.780	0.864	0.957	0.000	0.000	0.000	0.000	0.000	0.000
	0.95	0.010	0.024	0.078	0.714	0.799	0.915	0.714	0.799	0.914	0.000	0.000	0.000	0.000	0.000	0.000

Table 6.4: Coverage probabilities for the confidence interval estimates when $P = R = T = 1, Q = S = 2$ and the DMU under evaluation is not included in the subsample.

		Overall			1st Stage						2nd Stage					
					Stage 1 Priority			Stage 2 Priority			Stage 1 Priority			Stage 2 Priority		
		1-α			1-α			1-α			1-α			1-α		
N	k	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	0.40	0.810	0.931	0.995	0.929	0.951	0.955	0.929	0.951	0.955	0.911	0.964	0.988	0.938	0.980	0.987
	0.45	0.930	0.982	1.000	0.957	0.965	0.968	0.957	0.965	0.968	0.963	0.982	0.991	0.979	0.988	0.990
	0.50	0.972	0.996	1.000	0.967	0.972	0.972	0.967	0.972	0.972	0.976	0.993	0.993	0.989	0.992	0.992
	0.55	0.971	0.995	1.000	0.967	0.970	0.972	0.967	0.970	0.972	0.976	0.990	0.993	0.987	0.990	0.992
	0.60	0.982	0.998	0.999	0.970	0.976	0.976	0.971	0.976	0.976	0.981	0.995	0.996	0.991	0.994	0.995
	0.65	0.989	0.996	0.999	0.968	0.984	0.986	0.968	0.984	0.986	0.990	0.994	0.997	0.988	0.993	0.996
	0.70	0.980	0.994	0.999	0.957	0.980	0.986	0.957	0.981	0.987	0.971	0.993	0.998	0.972	0.993	0.997
	0.75	0.956	0.982	0.996	0.934	0.962	0.987	0.935	0.962	0.987	0.949	0.979	0.998	0.947	0.978	0.997
	0.80	0.921	0.958	0.994	0.902	0.949	0.985	0.902	0.950	0.985	0.923	0.954	0.993	0.922	0.958	0.995
	0.85	0.877	0.935	0.987	0.871	0.930	0.973	0.872	0.930	0.973	0.891	0.938	0.985	0.890	0.936	0.981
	0.90	0.810	0.887	0.965	0.817	0.885	0.960	0.815	0.885	0.960	0.838	0.901	0.965	0.832	0.904	0.965
	0.95	0.746	0.835	0.944	0.740	0.850	0.941	0.739	0.849	0.942	0.788	0.869	0.954	0.776	0.861	0.954
50	0.40	0.696	0.859	0.989	0.960	0.995	1.000	0.972	0.995	1.000	0.933	0.987	0.998	0.955	0.997	1.000
	0.45	0.818	0.932	0.999	0.972	0.995	1.000	0.976	0.996	1.000	0.974	0.993	1.000	0.991	1.000	1.000
	0.50	0.936	0.987	1.000	0.983	0.997	1.000	0.984	0.998	1.000	0.988	0.998	1.000	0.997	1.000	1.000
	0.55	0.962	0.996	1.000	0.986	0.998	1.000	0.985	0.997	1.000	0.991	0.997	1.000	0.997	0.999	1.000
	0.60	0.984	0.998	1.000	0.984	0.994	1.000	0.984	0.994	1.000	0.989	0.999	1.000	0.993	0.999	1.000
	0.65	0.990	0.997	1.000	0.971	0.991	1.000	0.971	0.990	1.000	0.977	0.995	0.999	0.982	0.995	0.999
	0.70	0.986	0.993	0.997	0.958	0.981	0.996	0.959	0.980	0.995	0.961	0.979	0.999	0.957	0.982	0.999
	0.75	0.972	0.988	0.997	0.928	0.966	0.991	0.927	0.967	0.991	0.943	0.968	0.992	0.937	0.967	0.993
	0.80	0.939	0.974	0.996	0.885	0.943	0.985	0.885	0.943	0.986	0.893	0.949	0.984	0.896	0.944	0.986
	0.85	0.893	0.938	0.985	0.835	0.903	0.973	0.835	0.901	0.973	0.837	0.905	0.972	0.829	0.904	0.973
	0.90	0.840	0.900	0.970	0.785	0.853	0.956	0.785	0.853	0.956	0.767	0.854	0.955	0.763	0.851	0.953

	0.95	0.742	0.839	0.940	0.710	0.805	0.924	0.711	0.805	0.925	0.691	0.782	0.912	0.684	0.782	0.912
100	0.40	0.682	0.850	0.991	0.947	0.986	1.000	0.946	0.987	1.000	0.970	0.996	1.000	0.987	0.999	1.000
	0.45	0.753	0.913	0.995	0.953	0.990	1.000	0.953	0.990	1.000	0.983	0.997	1.000	0.995	1.000	1.000
	0.50	0.902	0.976	0.999	0.980	0.995	1.000	0.980	0.995	1.000	0.993	1.000	1.000	0.997	1.000	1.000
	0.55	0.955	0.989	1.000	0.988	0.995	1.000	0.988	0.995	1.000	0.994	1.000	1.000	0.998	0.999	1.000
	0.60	0.977	0.994	1.000	0.988	0.996	1.000	0.988	0.996	1.000	0.990	0.998	1.000	0.993	0.998	1.000
	0.65	0.987	0.999	1.000	0.976	0.994	1.000	0.975	0.994	1.000	0.984	0.993	1.000	0.985	0.993	1.000
	0.70	0.988	0.994	1.000	0.951	0.982	0.997	0.952	0.981	0.997	0.957	0.984	0.997	0.953	0.984	0.998
	0.75	0.975	0.989	0.999	0.914	0.961	0.994	0.915	0.961	0.994	0.916	0.962	0.996	0.923	0.960	0.996
	0.80	0.938	0.974	0.994	0.880	0.924	0.986	0.880	0.923	0.986	0.862	0.929	0.985	0.862	0.932	0.986
	0.85	0.873	0.933	0.983	0.830	0.882	0.969	0.831	0.883	0.970	0.785	0.871	0.963	0.781	0.866	0.964
	0.90	0.800	0.871	0.961	0.764	0.843	0.929	0.765	0.843	0.929	0.696	0.793	0.929	0.688	0.794	0.934
	0.95	0.697	0.803	0.918	0.689	0.781	0.894	0.684	0.781	0.893	0.610	0.724	0.879	0.588	0.705	0.876
200	0.40	0.557	0.789	0.986	0.869	0.984	1.000	0.871	0.984	1.000	0.986	0.999	1.000	0.997	1.000	1.000
	0.45	0.686	0.874	0.994	0.918	0.994	1.000	0.918	0.994	1.000	0.996	1.000	1.000	1.000	1.000	1.000
	0.50	0.860	0.955	1.000	0.973	0.998	1.000	0.973	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.55	0.926	0.986	1.000	0.987	0.998	1.000	0.987	0.998	1.000	0.997	1.000	1.000	0.999	1.000	1.000
	0.60	0.971	0.997	1.000	0.989	0.998	1.000	0.991	0.998	1.000	0.992	0.998	1.000	0.993	0.999	1.000
	0.65	0.989	0.997	1.000	0.983	0.997	1.000	0.983	0.997	1.000	0.965	0.990	1.000	0.970	0.991	1.000
	0.70	0.990	0.996	0.999	0.979	0.990	0.998	0.977	0.990	0.998	0.928	0.970	0.997	0.933	0.974	0.998
	0.75	0.972	0.989	0.996	0.949	0.978	0.997	0.949	0.978	0.997	0.872	0.932	0.992	0.885	0.936	0.993
	0.80	0.941	0.969	0.994	0.900	0.957	0.988	0.901	0.955	0.988	0.816	0.886	0.972	0.823	0.889	0.971
	0.85	0.882	0.938	0.982	0.869	0.905	0.979	0.869	0.906	0.979	0.740	0.825	0.930	0.743	0.829	0.933
	0.90	0.787	0.871	0.955	0.806	0.878	0.954	0.805	0.879	0.953	0.642	0.742	0.892	0.641	0.748	0.894
	0.95	0.664	0.777	0.902	0.743	0.803	0.917	0.744	0.803	0.918	0.553	0.658	0.822	0.538	0.656	0.822

Results from applying the above algorithm to the confidence interval estimates of each Monte Carlo trial are given in Tables 6.5 and 6.6, for the five and seven-dimensional case, respectively.

Table 6.5: Mode, mean and range of the optimal subsample size values among the different Monte Carlo trials, using the minimum volatility algorithm, in the five-dimensional case.

		Overall			1st Stage			2nd Stage		
			1-α			1-α			1-α	
N		0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	mode	5	5	5	5	5	5	5	5	5
	mean	8.149	7.874	7.297	9.423	9.144	9.126	9.835	8.458	6.952
	range	13	13	13	14	13	13	13	13	13
50	mode	33	33	8	33	33	33	33	33	8
	mean	25	21.5	15.3	23.67	22.67	20.56	25.67	24.33	19.74
	range	28	28	26	28	28	26	28	28	26
100	mode	63	63	63	63	63	63	63	63	63
	mean	55.4	50.04	41.54	48.79	45.32	44.94	51.04	50.28	41.48
	range	56	56	56	56	56	56	51	56	56
200	mode	117	117	117	117	117	117	117	117	117
	mean	105.6	104.4	96.71	91.26	86.74	84.37	97.59	96.99	91.59
	range	93	93	107	99	99	107	93	93	99

Table 6.6: Mode, mean and range of the optimal subsample size among the different Monte Carlo trials, using the minimum volatility algorithm, in the seven-dimensional case.

		Overall			1st Stage			2nd Stage		
			1-α			1-α			1-α	
N		0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	mode	5	5	5	5	5	5	5	5	5
	mean	6.926	7.010	7.617	7.913	7.827	8.67	10.1	10.63	10.8
	range	13	14	13	13	14	13	13	13	13
50	mode	33	33	8	33	33	8	8	8	12
	mean	24.41	19.28	12.06	21.90	20.54	16.94	15.15	16.5	17.67
	range	28	28	26	28	28	26	28	28	26
100	mode	63	63	63	63	63	63	12	12	12
	mean	55.08	51.66	34.63	52.16	47.91	42.93	22.92	21.05	23.47
	range	46	46	46	56	56	56	56	56	56
200	mode	117	117	117	117	117	117	10	10	10
	mean	101.7	100.8	95.59	96.43	95.84	93.6	39.97	34.07	27.09
	range	93	86	107	93	86	107	107	107	107

According to the results, in the five-dimensional DGP (see Table 6.5), in almost all the cases, the mode for the optimal subsample size is the same for obtaining the overall and stage efficiency estimations and does not change for the different levels of significance. However, from the range of the optimal subsample values, it can be seen that the optimal subsample size can vary significantly among the different Monte Carlo trials, although all samples are generated through the same DGP, and it should be expected that the algorithm would yield similar values for the optimal subsample size.

Furthermore, as it was mentioned previously, for this five-dimensional DGP, values of $k = 0.60$ and $k = 0.65$ yielded the highest coverage probabilities in most of the cases. These values correspond to subsample sizes $[25^{0.60}] = 6$ and $[25^{0.65}] = 8$, respectively, when $N = 25$, subsample sizes 10 and 12 when $N = 50$, 15 and 19 when $N = 100$ and 24 and 31 respectively, when $N = 200$. These subsample sizes, however, are much lower than those indicated by the algorithm, especially when the sample size increases.

For the seven-dimensional DGP, there is a significant difference between the optimal subsample size for obtaining the overall and first stage efficiency score, compared to the optimal subsample size for the second stage. That raises an important issue, as all the estimations, for the overall, first and second stage efficiencies, are based on the same subsample size, and it is not possible to use the stage-specific optimal subsample size without affecting the consistency of the results. This difference between the overall-first stage and the second stage optimal subsample sizes in this DGP, is also reflected in the coverage probabilities reported in Table 6.2, where the value of k with higher coverage probabilities for the second stage ($k = 0.90$), was significantly different compared to that for the overall and first stage estimations ($k = 0.65, 0.70$). Similarly to the five-dimensional case, the subsample sizes that resulted in higher coverage probabilities in Table 6.2, do not coincide with the optimal subsample sizes yielded by the minimum volatility algorithm.

These issues indicate the great sensitivity of the algorithm, and demonstrate the need for defining a more robust way for the subsample size selection in NDEA. This variability of the optimal subsample size among the different Monte Carlo trials and the different stages can be imputed to the zero-efficiency score issue of the NDEA algorithm that mainly affects the upper bounds of the confidence intervals. This issue is further discussed in the following Section were the subsampling bootstrap is applied in a real dataset, and some suggestions for dealing with it are provided.

6.6 Application

In this section, the subsampling methodology is applied in the efficiency analysis of railways in 22 European countries that was discussed in Chapter 4. A limitation of this data set is that it does not include information about all the European countries, as countries with missing data were excluded. As in DEA the efficient frontier is empirically constructed from the available set of DMUs, omission of DMUs may have an impact on its shape and result in some DMUs being assigned a higher efficiency score. Since the true frontier is unknown, the subsampling methodology can be used to provide estimations of the true efficiency scores of European railways.

Confidence interval estimates were obtained for different values of k, i.e. for various subsample sizes. Initially, the DMU under evaluation was included in the subsample. For very small subsample sizes, for many countries in this data set, the upper bound of the confidence interval estimate for some stage efficiency score was either above one, or negative. The negative bounds result very rarely, and are due to the boundary condition of $\theta^{\phi}, \phi = 0, 1, 2$ on zero. An upper bound above one occurs in the cases when in a significant number of bootstrap replications, the algorithm returns zero values. This is an issue that can occur in efficiency decomposition approaches, but becomes more common when applying the subsampling bootstrap, because of the smaller size of the sample. If this happens in more than $(a/2)100\%$ of the bootstrap replications, the upper bound of the confidence interval is above one. This was more common in the cases when the subsample size was very small, but even for larger subsample sizes there were still some DMUs for which the upper bound exceeds one.

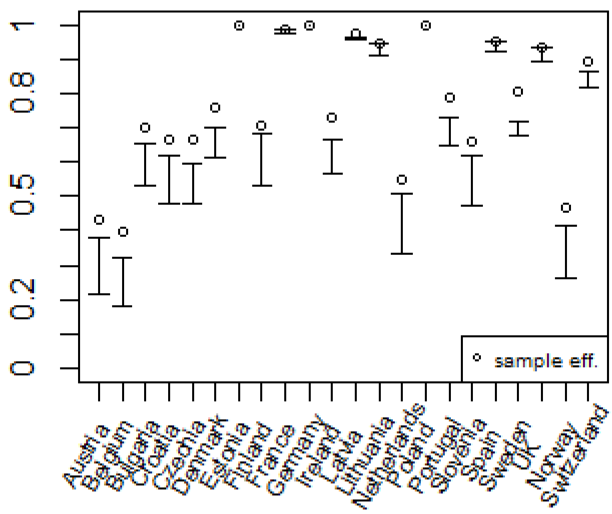
In some cases, obtaining the bootstrapped stage efficiency scores while treating both stages as priority stages, and use the decomposition equation (3.25) - or the equivalent equation for the first stage- to obtain the stage efficiency score when the priority-stage-model yields infeasibility, overcomes the zero efficiency estimate issue. However, this usually does not offer a solution to the problem, as both priority stages may yield zero bootstrap estimates. Therefore, removing the zero bootstrap estimates seems to be the only way to prevent them from distorting the upper bound of the confidence interval. Although this reduces the size of the bootstrap efficiency sample, if the number of bootstrap replications is large enough, it should not affect the quality of the results.

The subsample size that was used was $m = 7$, for $k = 0.65$. The choice of the subsample size was based on the coverage probabilities obtained from the Monte Carlo simulations and the results from the minimum volatility algorithm of Politis et al. (2001) for the overall process.

The model was implemented in R. In order to minimise the computational time, after drawing the 2000 bootstrap subsamples for a specific k, i.e. for a specific subsample size, the confidence intervals for each DMU were calculated separately and not in one *for* loop- but based on the same subsamples. In this way, it is possible to use parallel processing and reduce the computational cost. Although this dataset is small, in larger datasets this approach can make a significant difference in the computational time.

Figure 6.2 reports the sample overall and stage efficiency scores and their corresponding confidence interval estimates, for $k = 0.65$ and subsample size $m = [22^{0.65}] = 7$, for the significance level $a = 0.1$. The original sample efficiency scores when stage one is the priority stage are also depicted in Figure 6.2 with small circles. The blue circles denote the efficiency estimate of the countries for which the decomposition is not unique, when stage two is the priority stage. The exact values of the confidence interval estimates when the DMU under evaluation is included in the subsample are provided in Table A4 in the Appendix.

(a)



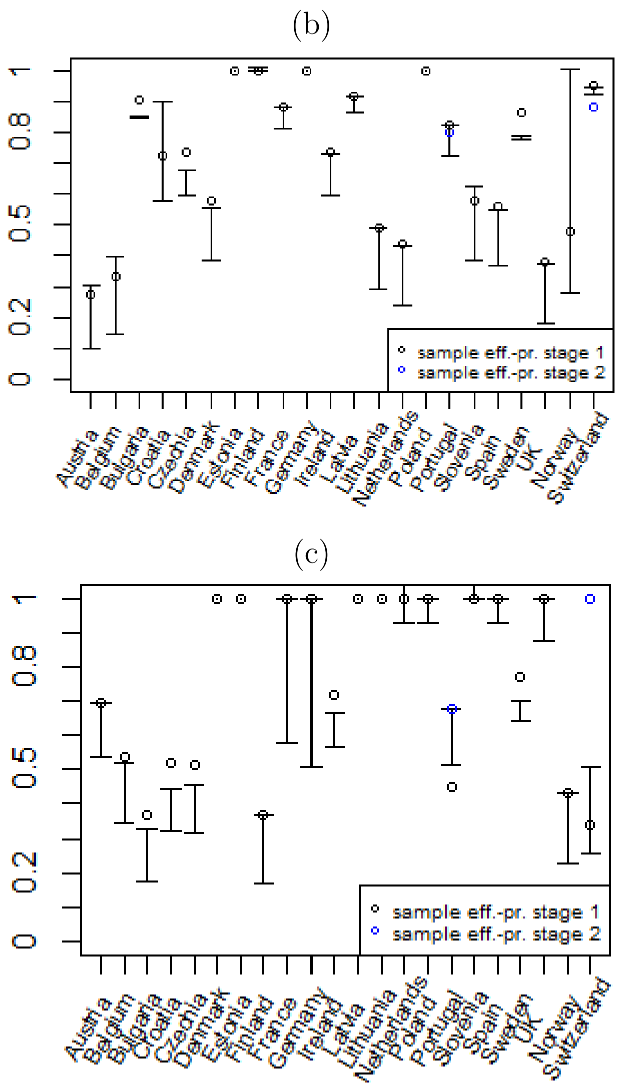


Figure 6.2: Confidence interval estimates for (a) the overall, (b) the first stage, and (c) the second stage efficiency scores of European railways, for $k = 0.65$, $m = 7$, and $a = 0.1$.

For some efficient countries, the confidence interval estimates converge to a single point, as in the majority of the bootstrap replications these countries yield an efficiency score equal to one, even for different subsample sizes.

According to the results, in this data set, the specific subsample size yields better overall efficiency estimates compared to stage efficiency scores, and estimations about the second stage are in most cases better compared to the first stage estimations. This is in accordance with Monte Carlo simulation results, where for one of the two stages coverage probabilities were more sensitive to the subsample size selection.

Confidence interval estimates reveal where the true efficiency score for each country lies. Although some efficiency scores appear to be very high, the lower bound estimation for their true efficiency is much lower, and sometimes a DMU with higher efficiency score from another might in reality be less efficient. For example, based only on the original sample, France and Germany are second stage efficient, whereas Sweden has lower efficiency score. However, the lower bound estimation for France and Germany is lower than that of Sweden. That means in reality, there is a chance that those two countries are less efficient in the second stage than Sweden. In this data set, confidence interval estimation is particularly useful in the second stage, where without the bootstrap estimations, 11 countries appear to be efficient. However, confidence interval estimates reveal that there might be differences in their true efficiency scores. For the second stage efficient countries Netherlands, Poland and Spain, for example, the lower bound remains above 0.9, whereas for the UK, the lower bound is about 0.8. For Germany and France, the lower bounds lie even lower, around 0.5. Therefore, confidence interval estimates should be considered from the countries to get a better insight into what is the main source of inefficiency for their railway network, to be able to form an effective improvement agenda. As it was indicated by the results, considering these estimates, provides higher discrimination among the different countries' railway-efficiency level.

Confidence interval estimates for $k = 0.65$ and $a = 0.1$ when the DMU under evaluation is not included in the subsample, are given in Figure 6.3 below.

As it was discussed in the previous section, not including the DMU under evaluation in the subsample, increases the range of the confidence intervals. In this dataset, this approach resulted in upper bounds above one for the majority of DMUs. The following adaptations resulted in avoiding the high upper bound, and significantly reduced the range of the confidence intervals: (i) setting the zero overall efficiency scores equal to one, (ii) treating both stages as priority stage and (iii) using the decomposition equation when the priority-stage-model yields zero efficiency score values. Alternatively, (i) setting the zero overall efficiency scores equal to one, and (ii) removing the stage bootstrap efficiency estimates that are equal to zero. However, even after following the aforementioned steps, the resulting confidence intervals in most of the cases are too wide to offer any insight about the true efficiency scores. Therefore, the including the DMU under evaluation in the subsample seems to be a preferred approach.

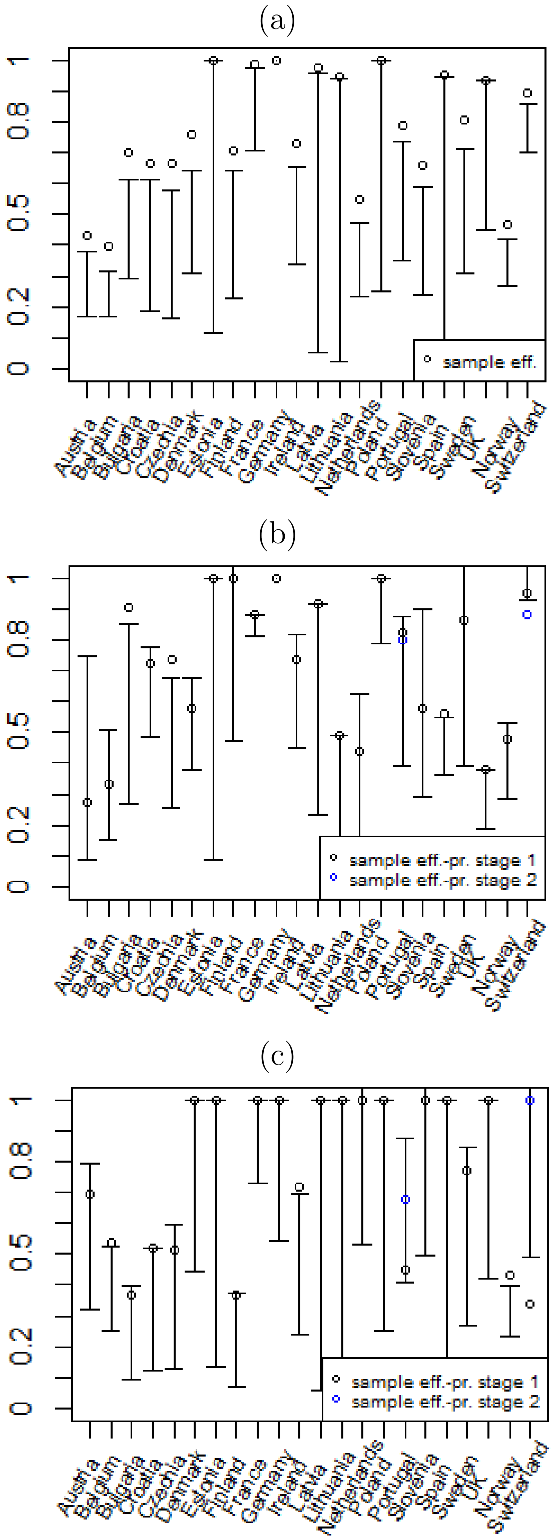


Figure 6.3: Confidence interval estimates for (a) the overall, (b) the first stage, and (c) the second stage efficiency scores of European railways, for $k = 0.65$, $m = 7$, and $a = 0.1$, when the DMU under evaluation is not included in the subsample.

The exact values of the confidence interval estimates when the DMU under evaluation is not included in the subsample are provided in Table A5 in the Appendix.

6.7 Conclusion

The DEA approach, where the production frontier is constructed empirically, does not consider for sampling noise. Bootstrapping techniques are now well-established in the DEA literature, and are broadly used to make statistical inference about the efficiency scores in one stage production processes. However, in many cases, the production structure of DMUs involves sub-stages which need to be considered in the efficiency measurement. This Chapter examined the applicability and performance of bootstrapping in general two-stage structures, where the additive decomposition approach is used to calculate the overall and stage efficiency scores, and the VRS is assumed, and the subsampling bootstrap was applied to obtain confidence interval estimates for the environmental efficiency of European railways.

Bootstrapping can be computationally demanding, especially in high-dimensional models. For this reason, among the different bootstrapping techniques, the subsampling methodology was considered in this study, as it does not require any kernel smoothing and also allows for heterogeneity in the efficiency distributions among the different DMUs.

Monte Carlo simulations were performed, based on samples obtained through two DGPs, defined for a five and a seven dimensional two-stage series structure, respectively. According to the results, in network structures, the coverage probabilities are more sensitive to the DGP compared to single-stage structures. Similarly to one-stage processes, coverage probabilities are very sensitive both to the sample and subsample sizes, and get lower as the sample size increases. However, in contrast to conventional DEA, in NDEA for some subsample sizes, coverage probabilities tend to zero. That means that the subsample size should be selected very carefully, notably when the size of the original sample is large, as in any other case, the resulting confidence intervals could be misleading.

The selection of the subsample size is a limitation of applying subsampling bootstrap in NDEA. The algorithm suggested by Politis et al. (2001) may offer a rule of thumb, especially in one-stage structures, but in NDEA, it was shown that the optimal subsample size resulting from the minimum volatility algorithm is not always the one that yields the highest coverage probabilities. This issue was further investigated and it seems that a peculiarity of the additive efficiency decomposition approach that assigns zero optimal weights to some DMUs when considering some stage as priority stage affects the performance of the minimum volatility algorithm. Therefore, in this study it was suggested to use the priority-stage model to get the efficiency score of both stages and use the decomposition equation in the cases when the priority-stage-model yields zero optimal results.

A limitation of applying subsampling bootstrap in the specific data set was that small, medium and large scale operators coexisted in the sample. As a result, a subsample frontier may differ significantly from the original sample frontier, distorting the confidence interval estimates. Furthermore, due to this heterogeneity, for some of the sample-efficient countries the confidence interval estimate converged to a single point. If the size of the original sample was larger, a solution to this could be to divide the data set into different clusters and get confidence interval estimates in each cluster separately. However, if clustering is applied in this data set then the size of the original sample would be too small to provide any discrimination among DMUs. Therefore, in this case conditioning the efficiency of DMUs on some environmental factors could offer a solution to the problem.

Future steps could focus on the consideration of environmental factors, defining a reliable method for the subsample size selection, as well as studying the performance of bootstrapping in other network structures, under different returns to scale assumptions, and/or considering substitution effects. Extension to the output orientation can also be considered, although the model orientation should not affect the results.

7

Conclusion

Conventional DEA does not consider the inner structure of DMUs or any sampling noise in the efficiency evaluation process. Therefore, a big branch of the DEA literature has focused on the evaluation of the efficient frontier in the cases of multi-stage production processes. Furthermore, bootstrapping techniques have been used to account for sampling noise, and various studies have focused on deriving the properties of the DEA estimator. In this thesis, the application of bootstrapping techniques was extended to network DEA models, and the performance of these techniques was examined through Monte Carlo simulations. The suggested approach was implemented to obtain confidence interval estimations about the efficiency scores of the railway transport process in Europe.

First, we revised the main DEA models and provided a review and comparison of the main approaches used in network DEA to obtain the efficiency scores of the overall and stage processes in two-stage series structures. Among the different network DEA approaches, we used the additive efficiency decomposition approach to study the efficiency of European railways. Railways are a capital-intensive industry, which aims to provide passenger and freight carriage. Therefore, in order to have an insight into the railway operations process and better identify any sources of inefficiency, the railway transport process was divided into two stages; the first one was related to the railway assets and the second one to the railway service provision.

Although railways are the most environmentally friendly means of transport, noise pollution generated from the contact of the rolling wheel with the rail lines, as well aerodynamic noise, seems to be a major environmental issue, especially in Europe. Regarding the impact of railway noise on humans, this can cause mild to severe health issues depending on the duration and intensity of noise. The European Commission has set noise emission limits and has defined common noise assessment methods for all EU member states. To achieve noise-emission goals, countries need to gradually retrofit their wagon fleet with more silent braking technology, as well as keep the rail line network well-maintained. Therefore, in this study, the number of wagons that are compliant with the noise emission limits, as well as the number of people exposed to high levels of railway noise in each country were considered in the efficiency measurement.

As in Europe there are small and large-scale railway operators, the VRS assumption was adopted. Although in the additive decomposition algorithm the endogenous definition of the decomposition weights results in a non-increasing relationship between them, favouring the first stage, we showed that under the VRS assumption such a non-increasing relationship cannot be established. However, decomposition weight restrictions had to be used to prevent the optimisation process from assigning a zero weight to one of the two stages, as in that case, the contribution of that stage to the overall process would be ignored. The decomposition weight restrictions were defined based on sensitivity analysis results. Sensitivity analysis revealed that although any use of decomposition weight restrictions affects the efficiency scores, for a wide range of stepwise changes in the restrictions, efficiency scores are only very slightly affected, and rankings are robust to alterations. For the cases when the stakeholders do not have specific suggestions regarding the relative importance of each stage, and so there are no specific weight restrictions that should be applied, we suggested an algorithm for the choice of the decomposition weight restrictions so that their impact on the efficiency scores is minimised.

Poland, Netherlands and Germany were found to be overall efficient. Based on the results, except for Finland, asset-efficient countries are also service efficient. However, countries that appear to be efficient in the service provision stage, seem to be able to apply further improvements to their asset management.

A limitation of the dataset that was used is that for many European countries, data were not available. As a result, the efficient frontier and the efficiency scores of DMUs were affected. In any case, the available set of DMUs is a sample drawn from an underlying population. Therefore, next, we provided a review of the bootstrap approaches used in the one-stage DEA context to account for sampling noise. We discussed the assumptions made in each approach and we highlighted any inconsistency or computational issues using illustrative examples. In this way, we affirmed the advantage of subsampling bootstrap, which does not rely on any restrictive homogeneity assumptions about the distribution of inefficiencies among DMUs and is also computationally easier than other consistent heterogeneous bootstrap approaches.

Although bootstrap approaches have been widely studied and implemented in one-stage production structures, there is a very limited number of studies that make statistical inference in multi-stage structures, and usually, in these studies, the relation between the different stages is not considered. In this thesis, we set a statistical framework in Network DEA. The frontier model was defined for general two-stage series structures, but definitions can be extended to different network structures.

We run Monte Carlo simulations in R to assess the performance of subsampling bootstrap in two-stage series structures. Data were generated for two different structures, a five, and a seven-dimensional one, under the VRS assumption. For a fixed point with a known true efficiency score, experiments were performed for a range of different sample and subsample sizes, with the fixed point under evaluation being included and not being included in the subsample. The probabilities that the true efficiency score lies within the estimated confidence intervals were calculated.

Based on the resulting coverage probabilities, the performance of subsampling is more sensitive to the sample and subsample size selection compared to one-stage structures, due to the higher dimensions of the models. Usually, for larger samples (e.g. 200 DMUs) the overall performance was lower. For subsamples that include neither a very small nor very large proportion of the original sample DMUs, coverage probabilities were usually very high. Furthermore, we showed that the choice of the priority stage does not affect the resulting coverage probabilities.

The subsample size selection is a limitation of implementing subsampling bootstrap. We revisited the minimum volatility algorithm that is being used in one-stage DEA. Based on the findings, due to some peculiar stage efficiency estimates resulting from the additive decomposition algorithm, the minimum volatility algorithm should preferably be applied to the confidence interval estimates for the overall efficiency and not to the estimates for the stage efficiencies.

The peculiarity of the additive decomposition algorithm lies in the return of some negative or zero efficiency estimates for the stage efficiencies. These results impact the upper bounds of the confidence interval estimates, and in some cases may result in upper bounds that are above one. We suggested that one solution to this could be to set the zero overall efficiency estimates equal to one and remove those discrepant stage efficiency estimates from the bootstrap sample. However, this would be at the cost of reducing the size of the bootstrap efficiency sample, and it should only be used only in the cases when the number of bootstrap replications is high enough so that the quality of the results is not affected. Alternatively, it is suggested to set the zero overall efficiency estimates equal to one, obtain the efficiency estimate for each stage while it is considered a priority stage and if a zero estimate is obtained, use the decomposition equation to calculate this efficiency estimate instead. This approach, also reduces significantly the times that upper bounds above one occur.

In the case when the DMU under evaluation was not included in the subsample, the range of the estimated confidence intervals was significantly larger than when the DMU under evaluation was included in the subsample. Such large confidence intervals could not offer any useful insight into the true inefficiency of DMUs. Therefore, the subsampling methodology was implemented to get the confidence interval estimates for the efficiency score of European railways, including each time the DMU under evaluation in the subsample. For many countries that appeared to be efficient in one of the two stages, the resulting confidence interval estimates revealed that their true efficiency score is probably much lower. Sometimes, for a country that had a higher efficiency than score another, the confidence interval estimates revealed that in reality, it might be less efficient.

To conclude, the above research can be extended in many directions. First, regarding the efficiency measurement in the railway transport process, future studies could distinguish between different railway noise sources, such as noise generated by passenger high-speed trains, or noise generated by freight wagons. Also, variables that consider the impact of railway noise inside and outside urban areas separately, or on the wildlife could be included in the model.

Furthermore, in this study targets for countries were not provided. Although in network DEA, the envelopment model is said to provide frontier projections, in this case, decomposition weights were restricted, and as a result, DMUs were not projected on the efficient frontier even with the envelopment model. Therefore, future studies could investigate the impact of the decomposition weight restrictions on the frontier projections, and suggest alternative, unbiased approaches for efficiency evaluation.

Since the duality between the envelopment and the multiplier model does not hold in network DEA, the difference between the projections yielded by the two models could be studied asymptotically. Furthermore, projections given by different network DEA approaches could be studied asymptotically.

The performance of bootstrapping in other network structures could also be investigated. Finally, considering that the returns to scale properties in network DEA do not hold as in one-stage DEA, another research direction could focus on how hypothesis tests about the returns to scale in network DEA could be developed.

Appendix

The notation that is being used in Chapters 3 and 6 is given below.

List of notations in Chapter 3:

j	Index of DMUs.
j_0	DMU under evaluation.
N	Number of DMUs in the sample.
$x_{pj} = (x_{1j}, ..., x_{Pj})$	Vector of 1st stage inputs for DMU$_j$.
$l_{rj} = (l_{1j}, ..., l_{Rj})$	Vector of 1st stage outputs for DMU$_j$.
$z_{qj} = (z_{1j}, ..., z_{Qj})$	Vector of intermediate products for DMU$_j$.
$g_{tj} = (g_{1j}, ..., g_{Tj})$	Vector of 2nd stage inputs for DMU$_j$.
$y_{sj} = (y_{1j}, ..., y_{Sj})$	Vector of 2nd stage outputs for DMU$_j$.
$v_p = (v_1, ..., v_P)$	Vector of multipliers for the 1st stage inputs in the fractional dual model.
$\mu_r = (\mu_1, ..., \mu_R)$	Vector of multipliers for the 1st stage outputs in the fractional dual model.
$\gamma_q = (\gamma_1, ..., \gamma_Q)$	Vector of multipliers for the intermediate products in the fractional dual model.
$\pi_t = (\pi_1, ..., \pi_T)$	Vector of multipliers for the 2nd stage inputs in the fractional dual model.
$\eta_s = (\eta_1, ..., \eta_S)$	Vector of multipliers for the 2nd stage outputs in the fractional dual model.
$\xi_d = (\xi_1, ..., \xi_D)$	Vector of multipliers for the undesirable outputs in the fractional dual model.
$v'_p = (v'_1, ..., v'_P)$	Vector of multipliers for the 1st stage inputs in the linear dual model.
$\mu'_r = (\mu'_1, ..., \mu'_R)$	Vector of multipliers for the 1st stage outputs in the linear dual model.
$\gamma'_q = (\gamma'_1, ..., \gamma'_Q)$	Vector of multipliers for the intermediate products in the linear dual model.
$\pi'_t = (\pi'_1, ..., \pi'_T)$	Vector of multipliers for the 2nd stage inputs in the linear dual model.
$\eta'_s = (\eta'_1, ..., \eta'_S$	Vector of multipliers for the 2nd stage outputs in the linear dual model.
$\xi'_d = (\xi'_1, ..., \xi'_D)$	Vector of multipliers for the undesirable outputs in the linear dual model.
w_{1j}	Decomposition weight of the 1st stage.
w_{2j}	Decomposition weight of the 2nd stage.
θ_j^0	Overall input efficiency score of DMU$_j$.
θ_j^1	1st stage input efficiency score of DMU$_j$.
θ_j^2	2nd stage input efficiency score of DMU$_j$.
ϕ_j^0	Overall output efficiency score of DMU$_j$.
ϕ_j^1	1st stage output efficiency score of DMU$_j$.
ϕ_j^2	2nd stage output efficiency score of DMU$_j$.
k	Stage indicator.
p	Priority stage indicator.
λ_j	Intensity vector of the first stage in the envelopment model.
μ_j	Intensity vector of the second stage in the envelopment model.

List of notations in Chapter 6:

j	Index of DMUs.
j_0	DMU under evaluation.
N	Number of DMUs in the sample.
m	Number of DMUs in the subsample.
k	exponent used to define the subsample size
x_e	Vector of first stage efficient inputs
g_e	Vector of second stage efficient inputs
$x_{pj} = (x_{1j}, ..., x_{Pj})$	Vector of 1st stage inputs for DMU$_j$.
$l_{rj} = (l_{1j}, ..., l_{Rj})$	Vector of 1st stage outputs for DMU$_j$.
$z_{qj} = (z_{1j}, ..., z_{Qj})$	Vector of intermediate products for DMU$_j$.
$g_{tj} = (g_{1j}, ..., g_{Tj})$	Vector of 2nd stage inputs for DMU$_j$.
$y_{sj} = (y_{1j}, ..., y_{Sj})$	Vector of 2nd stage outputs for DMU$_j$.
$v_p = (v_1, ..., v_P)$	Vector of multipliers for the 1st stage inputs in the fractional dual model.
$\mu_r = (\mu_1, ..., \mu_R)$	Vector of multipliers for the 1st stage outputs in the fractional dual model.
$\gamma_q = (\gamma_1, ..., \gamma_Q)$	Vector of multipliers for the intermediate products in the fractional dual model.
$\pi_t = (\pi_1, ..., \pi_T)$	Vector of multipliers for the 2nd stage inputs in the fractional dual model.
$\eta_s = (\eta_1, ..., \eta_S)$	Vector of multipliers for the 2nd stage outputs in the fractional dual model.
$\xi_d = (\xi_1, ..., \xi_D)$	Vector of multipliers for the undesirable outputs in the fractional dual model.
$v'_p = (v'_1, ..., v'_P)$	Vector of multipliers for the 1st stage inputs in the linear dual model.
$\mu'_r = (\mu'_1, ..., \mu'_R)$	Vector of multipliers for the 1st stage outputs in the linear dual model.
$\gamma'_q = (\gamma'_1, ..., \gamma'_Q)$	Vector of multipliers for the intermediate products in the linear dual model.
$\pi'_t = (\pi'_1, ..., \pi'_T)$	Vector of multipliers for the 2nd stage inputs in the linear dual model.
$\eta'_s = (\eta'_1, ..., \eta'_S)$	Vector of multipliers for the 2nd stage outputs in the linear dual model.
$\xi'_d = (\xi'_1, ..., \xi'_D)$	Vector of multipliers for the undesirable outputs in the linear dual model.
w_{1j}	Decomposition weight of the 1st stage.
w_{2j}	Decomposition weight of the 2nd stage.
θ_j^0	True overall efficiency score of DMU$_j$.
θ_j^1	True first stage efficiency score of DMU$_j$.
θ_j^2	True second stage efficiency score of DMU$_j$.
$\widehat{\theta^0}_j$	Estimation of the overall efficiency score of DMU$_j$.
$\widehat{\theta^1}_j$	Estimation of the first stage efficiency score of DMU$_j$.
$\widehat{\theta^2}_j$	Estimation of the second stage efficiency score of DMU$_j$.
$\widehat{\widehat{\theta^0}}_j$	Bootstrap estimation of the overall efficiency score of DMU$_j$.
$\widehat{\widehat{\theta^1}}_j$	Bootstrap estimation of the first stage efficiency score of DMU$_j$.
$\widehat{\widehat{\theta^2}}_j$	Bootstrap estimation of the second stage efficiency score of DMU$_j$.
ϕ	Stage indicator.
p	Priority stage indicator.
a	confidence level
$I_{m_{low}}$	Lower bound of the confidence interval estimate obtained with subsample size m
$I_{m_{up}}$	Upper bound of the confidence interval estimate obtained with subsample size m

As it was discussed in Chapters 3 and 4, in NDEA, the multiplier model should be used to obtain the efficiency scores of DMUs and the envelopment model to obtain the frontier projections. However, as it is mentioned in Section 4.5.3 when decomposition weight restricitons were used, the envelopment-primal model cannot provide the frontier projections. Model (1) is the primal of model (4.6), and Table A1 includes the variable-constraint correspondences between the two models.

$$\begin{aligned}
&\min \theta_{j_0} \\
&\text{s.t.} \ -\sum_{j=1}^{N}\lambda_j x_{pj} + \theta_{j_0}x_{pj_0} + (c_0-1)\pi_{j_0}x_{pj_0} + c_0\rho_{j_0}x_{pj_0} \geq 0, \quad \forall p \\
&\sum_{j=1}^{N}\lambda_j l_{rj} \geq l_{rj_0}, \quad \forall r \\
&\sum_{j=1}^{N}\lambda_j z_{qj} - \sum_{j=1}^{N}\mu_j z_{qj} + \theta_{j_0}z_{qj_0} + \pi_{j_0}c_0 z_{qj_0} + \rho_{j_0}(c_0-1)z_{qj_0} \geq z_{qj_0}, \quad \forall q \\
&-\sum_{j=1}^{N}\mu_j (y_b)_{dj} + \theta_{j_0}(y_b)_{dj_0} + \pi_{j_0}c_0(y_b)_{dj_0} + \rho_{j_0}(c_0-1)(y_b)_{dj_0} \geq 0, \quad \forall d \qquad (1) \\
&\sum_{j=1}^{N}\mu_j y_{sj} \geq y_{sj_0}, \quad \forall s \\
&\sum_{j=1}^{N}\lambda_j = 1 \\
&\sum_{j=1}^{N}\mu_j = 1 \\
&\lambda_j, \mu_j \geq 0, \forall j, \pi_{j_0}, \rho_{j_0} \geq 0
\end{aligned}$$

Table A1: Primal-dual correspondences between models (4.6) and (1).

duals \ primal	λ_j	μ_j	θ_{j_0}	π_{j_0}	ρ_{j_0}		primal RHS/ dual obj.
v'_p	$-x_{pj}$	0	x_{pj_0}	$(c_0-1)x_{pj_0}$	$c_0x_{pj_0}$	$\geq$	0
μ'_r	l_{rj}	0	0	0	0	$\geq$	l_{j_0}
γ'_q	z_{qj}	$-z_{qj}$	z_{qj_0}	$c_0z_{qj_0}$	$(c_0-1)z_{qj_0}$	$\geq$	z_{j_0}
ξ'_d	0	$-(y_b)_{dj}$	$(y_b)_{dj_0}$	$c_0(y_b)_{dj_0}$	$(c_0-1)(y_b)_{dj_0}$	$\geq$	0
η'_s	0	y_{sj}	0	0	0	$\geq$	y_{sj_0}
u^1	1	0	0	0	0	$=$	1
u^2	0	1	0	0	0	$=$	1
	$\leq$	$\leq$	$=$	$\leq$	$\leq$		
prim. obj./ dual RHS	0	0	1	0	0		

In Section 4.5.1, a volatility index was calculated to choose the decomposition weight restrictions c_0. In Table A2, the values of the volatility index V_{c_0} for different values of c_0

are provided. The lower the value of V_{c_0}, the less the impact of the decomposition weight restrictions to the efficiency scores.

Table A2: Volatility indices for the overall efficiency scores

	V_{c_0}			
c_0	r=1	r=2	r=3	r=4
0.01	-	-	-	-
0.02	0.02677	-	-	-
0.03	0.02677	0,04233	-	-
0.04	0.02677	0.04233	0.05784	-
0.05	0.02677	0.04233	0.05784	0.07332
0.06	0.02677	0.04233	0.05784	0.07332
0.07	0.02677	0.04233	0.05784	0.07332
0.08	0.02677	0.04233	0.05784	0.07332
0.09	0.02677	0.04233	0.05784	0.07332
0.10	0.02677	0.04233	0.05784	0.07332
0.11	0.02677	0.04233	0.05784	0.07332
0.12	0.02677	0.04233	0.05784	0.07332
0.13	0.02677	0.04233	0.05784	0.07332
0.14	0.02677	0.04233	0.05784	0.07332
0.15	0.02677	0.04233	0.05784	0.07332
0.16	0.02677	0.04233	0.05784	0.07402
0.17	0.02677	0.04233	0.05863	0.07590
0.18	0.02677	0.04328	0.06073	0.07837
0.19	0.02799	0.04568	0.06343	0.08152
0.20	0.03076	0.04861	0.06679	0.08501
0.21	0.03313	0.05189	0.07025	0.08857
0.22	0.03466	0.05437	0.07335	0.09201
0.23	0.03547	0.05569	0.07564	0.09498
0.24	0.03565	0.05651	0.07696	0.09710
0.25	0.03606	0.05695	0.07780	0.09877
0.26	0.03636	0.05738	0.07875	0.10020
0.27	0.03647	0.05822	0.07986	0.10183
0.28	0.03727	0.05915	0.08139	0.10451
0.29	0.03817	0.06060	0.08401	0.10747
0.30	0.03923	0.06326	0.08690	0.11042
0.31	0.04167	0.06576	0.08964	0.11324
0.32	0.04344	0.06782	0.09186	0.11582
0.33	0.04381	0.06903	0.09364	0.11816
0.34	0.04381	0.06941	0.09507	0.12042
0.35	0.04404	0.07007	0.09624	0.12344
0.36	0.04493	0.07123	0.09882	0.12848
0.37	0.04602	0.07416	0.10432	0.13448
0.38	0.04895	0.08015	0.11065	0.14099
0.39	0.05523	0.08615	0.11691	0.14779
0.40	0.05894	0.09112	0.12280	0.15471
0.41	0.05940	0.09439	0.12823	0.16276
0.42	0.06059	0.09649	0.13406	0.17116
0.43	0.06270	0.10101	0.13990	0.18142
0.44	0.06707	0.10686	0.14910	0.19375
0.45	0.07175	0.11558	0.16147	0.20728
0.46	0.07844	0.12697	0.17421	0.22148
0.47	0.08775	0.13726	0.18658	-
0.48	0.09380	0.14618	-	-
0.49	0.09634	-	-	-
0.50	-	-	-	-

In Section 4.5.2, to investigate whether for some first stage efficient countries, the lower second stage efficiency score is due to the number of people affected by noise, L_{den} variable was excluded from the model. The resulting efficiency scores are provided in Table A3 below.

Table A3: Efficiency scores and optimal decomposition weights, when L_{den} is omitted from the model, and $w_{1j}, w_{2j} \geq 0.1$

	DMU	θ^{0*}	w^*_{1j}	w^*_{2j}	θ^{1p*}	θ^{2*}	θ^{1*}	θ^{2p*}
1	Austria	0.4204	0.7703	0.2297	0.2748	0.9089	0.2748	0.9089
2	Belgium	0.3655	0.8872	0.1128	0.3328	0.6226	0.3328	0.6226
3	Bulgaria	0.6985	0.6430	0.3570	0.9081	0.3210	0.9081	0.3210
4	Croatia	0.6650	0.7123	0.2877	0.7421	0.4742	0.7421	0.4742
5	Czech Rep.	0.6431	0.7890	0.2110	0.7356	0.2970	0.7356	0.2970
6	Denmark	0.6490	0.7700	0.2300	0.5735	0.9016	0.5735	0.9016
7	Estonia	1	0.9000	0.1000	1	1	1	1
8	Finland	0.7006	0.5678	0.4322	1	0.3072	1	0.3072
9	France	0.9474	0.4462	0.5538	0.8822	1	0.8822	1
10	Germany	1	0.9000	0.1000	1	1	1	1
11	Ireland	0.7261	0.7161	0.2839	0.7339	0.7065	0.7339	0.7065
12	Latvia	0.9473	0.6478	0.3522	0.9186	1	0.9186	1
13	Lithuania	0.5885	0.7721	0.2279	0.4932	0.9117	0.4932	0.9117
14	Netherlands	0.4454	0.9000	0.1000	0.5146	-0.1774	0.3838	1
15	Poland	0.8495	0.4829	0.5171	1	0.7089	1	0.7089
16	Portugal	0.7826	0.9000	0.1000	0.8250	0.4008	0.7943	0.6769
17	Slovenia	0.6537	0.8174	0.1826	0.5764	1	0.5764	1
18	Spain	0.4528	0.6516	0.3484	0.4851	0.3926	0.4851	0.3926
19	Sweden	0.7696	0.4717	0.5283	0.8628	0.6863	0.8628	0.6863
20	UK	0.5788	0.6567	0.3433	0.3586	1	0.3586	1
21	Norway	0.4620	0.7717	0.2283	0.4808	0.3985	0.4808	0.3985
22	Switzerland	0.7207	0.9000	0.1000	0.9552	-1.3899	0.6896	1

The exact values of the confidence interval estimates for the overall and stage efficiency scores in the cases when the DMU under evaluation is included and is not included in the subsample are provided in Tables A4 and A5, respectively.

Table A4: Sample efficiency scores and bootstrap confidence interval estimates when stage 1 is the priority stage, for $a = 0.1$, and when the DMU under evaluation is included in the subsample.

	DMU	θ^{0*}	lower b.	upper b.	θ^{1p*}	lower b.	upper b.	θ^{2*}	lower b.	upper b.
1	Austria	0.433637	0.208357	0.383818	0.274788	0.124054	0.302411	0.694873	0.394349	0.769616
2	Belgium	0.397365	0.200504	0.338427	0.333166	0.184933	0.474321	0.539735	0.307296	0.525567
3	Bulgaria	0.703788	0.326014	0.630012	0.908831	0.431459	0.853548	0.368453	0.159534	0.356542
4	Croatia	0.665272	0.479040	0.628639	0.727019	0.561044	0.898322	0.519453	0.310693	0.510360
5	Czech Rep.	0.667277	0.273828	0.605546	0.735596	0.374623	0.720089	0.511593	0.237643	0.502487
6	Denmark	0.762902	0.611630	0.714738	0.576000	0.371754	0.562920	1	0.555620	1
7	Estonia	1	1	1	1	1	1	1	0.999998	1
8	Finland	0.705130	0.531244	0.681369	1	0.999274	1	0.365703	0.203466	0.376088
9	France	0.988221	0.978487	0.988221	0.882214	0.809538	0.882214	1	0.512199	1
10	Germany	1	0.999999	1	1	0.999999	1	1	0.247326	1
11	Ireland	0.731261	0.566912	0.692171	0.737318	0.575343	0.737318	0.718122	0.548835	0.709402
12	Latvia	0.977333	0.958806	0.965361	0.918610	0.866319	0.918610	1	0.916369	1
13	Lithuania	0.949317	0.909080	0.949317	0.493179	0.318361	0.493179	1	0.916369	1
14	Netherlands	0.547971	0.340820	0.519539	0.438678	0.266457	0.558130	1	0.916369	1.001652
15	Poland	1	1	1	1	0.999921	1	1	0.620001	1
16	Portugal	0.787282	0.647227	0.765360	0.824983	0.704233	0.824983	0.447969	0.241157	0.404513
17	Slovenia	0.659973	0.472300	0.631500	0.576376	0.405605	0.584509	1	0.999993	1.040584
18	Spain	0.955938	0.922253	0.954360	0.559388	0.387003	0.546031	1	0.999998	1
19	Sweden	0.805756	0.680584	0.722400	0.862833	0.780057	0.850267	0.769217	0.645058	0.693513
20	UK	0.938041	0.889435	0.937427	0.380410	0.215562	0.375286	1	0.799631	1
21	Norway	0.467649	0.274073	0.425356	0.481784	0.307165	0.522290	0.431738	0.226559	0.431629
22	Switzerland	0.893634	0.814142	0.871282	0.955176	0.925232	0.955176	0.339756	0.082077	0.306216

Table A5: Sample efficiency scores and bootstrap confidence interval estimates when stage1 is the priority stage, for $a = 0.1$, when the DMU under evaluation is not included in the subsample.

	DMU	θ^{0*}	lower b.	upper b.	θ^{1p*}	lower b.	upper b.	θ^{2*}	lower b.	upper b.
1	Austria	0.433637	0.169129	0.380566	0.274788	0.085673	0.747752	0.694873	0.320606	0.794757
2	Belgium	0.397365	0.170588	0.315284	0.333166	0.151864	0.510512	0.539735	0.248719	0.528217
3	Bulgaria	0.703788	0.288663	0.614996	0.908831	0.266616	0.851818	0.368453	0.092900	0.394161
4	Croatia	0.665272	0.187660	0.611379	0.727019	0.476124	0.759991	0.519453	0.118708	0.518712
5	Czech Rep.	0.667277	0.163166	0.578460	0.735596	0.255138	0.677112	0.511593	0.129482	0.598083
6	Denmark	0.762902	0.309675	0.645281	0.576000	0.376728	0.676997	1	0.443666	1
7	Estonia	1	0.113483	1	1	0.085119	1	1	0.135359	1
8	Finland	0.705130	0.225920	0.644926	1	0.470908	1.077505	0.365703	0.067574	0.374695
9	France	0.988221	0.704528	0.979453	0.882214	0.813515	0.882214	1	0.728649	1
10	Germany	1	1	1	1	1	1	1	0.545027	1
11	Ireland	0.731261	0.339602	0.657388	0.737318	0.448851	0.818779	0.718122	0.241490	0.698535
12	Latvia	0.977333	0.052632	0.960600	0.918610	0.231303	0.918610	1	0.056902	1
13	Lithuania	0.949317	0.022188	0.939561	0.493179	0.097823	0.493179	1	0.026116	1
14	Netherlands	0.547971	0.234199	0.470156	0.438678	0.025540	0.624109	1	0.532986	1.207284
15	Poland	1	0.252828	1	1	0.788225	1	1	0.251806	1
16	Portugal	0.787282	0.347551	0.736736	0.824983	0.391342	0.878788	0.447969	0.410544	0.875853
17	Slovenia	0.659973	0.239808	0.590949	0.576376	0.293321	0.900588	1	0.494209	1.109359
18	Spain	0.955938	0.088945	0.948399	0.559388	0.361779	0.548354	1	0.099149	1
19	Sweden	0.805756	0.306352	0.710857	0.862833	0.389731	1.077505	0.769217	0.266748	0.848623
20	UK	0.938041	0.448779	0.935648	0.380410	0.187375	0.380411	1	0.421112	1
21	Norway	0.467649	0.265462	0.421282	0.481784	0.286718	0.528833	0.431738	0.235597	0.396026
22	Switzerland	0.893634	0.699329	0.858995	0.955176	0.927642	1.039189	0.339756	0.489470	1.530992

Bibliography

Adler, N., Liebert, V. & Yazhemsky, E. (2013), 'Benchmarking airports from a managerial perspective', *Omega* **41**, 442–458.

Ang, S. & Chen, C. (2016), 'Pitfalls of decomposition weights in the additive multi-stage DEA model', *Omega* **58**, 139–153.

Azadeh, A., Salehi, V. & Kianpour, M. (2018), 'Performance evaluation of rail transportation systems by considering resilience engineering factors: Tehran railway electrification system', *Transportation Letters* **10**(1), 12–25.

Banker, R. (1984), 'Estimating most productive scale size using data envelopment analysis', *European Journal of Operational Research* **17**, 35–44.

Banker, R. (1993), 'Maximum likelihood, consistency and data envelopment analysis: a statistical foundation', *Management Science* **39**(10), 1265–1273.

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

Banker, R. & Thrall, R. (1992), 'Estimation of returns to scale using data envelopment analysis', *European Journal of Operational Research* **62**, 74–84.

Bhanot, N. & Singh, H. (2014), 'Benchmarking the performance indicators of Indian Railway container business using data envelopment analysis', *Benchmarking* **21**(1), 101–120.

Bickel, P. & Freedman, D. (1981), 'Some Asymptotic Theory for the Bootstrap', *Ann. Statist.* **9**(6), 1196–1217.

Bostian, M., Daraio, C., Färe, R., Grosskopf, S., Izzo, M., Leuzzi, L., Ruocco, G. & Weber, W. (2018), Inference for Nonparametric Productivity Networks: A Pseudo-likelihood Approach, Technical Report 6, Sapienza, Università di Romà.

Cantos, P., Pastor, J. & Serrano, L. (1999), 'Productivity, efficiency and technical change in the European railways: a non-parametric approach', *Transportation* **26**(4), 337–357.

Cantos, P., Pastor, J. & Serrano, L. (2010), 'Vertical and horizontal separation in the European railway sector and its effects on productivity', *Journal of Transport Economics and Policy* **44**(2), 139–160.

Castelli, L., Pesenti, R. & Ukovich, W. (2010), 'A classification of DEA models when the internal structure of the decision making units is considered', *Annals of Operations Research* **173**, 207–235.

Chang, Y., Zhang, N., Danao, D. & Zhang, N. (2013), 'Environmental efficiency analysis of transportation system in China: a non-radial dea approach', *Energy Policy* **58**, 277–283.

Chapin, A. & Schmidt, S. (1999), 'Do mergers improve efficiency? Evidence from deregulated rail freight', *Journal of Transport Economics and Policy* **33**(2), 147–162.

Charles, V., Aparicio, J. & Zhu, J. (2019), 'The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis', *European Journal of Operational Research* **279**, 929–940.

Charnes, A. & Cooper, W. (1959), 'Chance-constrained programming', *Management Science, INFORMS* **6**(1), 73–79.

Charnes, A. & Cooper, W. (1962), 'Programming with linear fractional functionals', *European Journal of Operational Research* **9**(3-4), 181–186.

Charnes, A., Cooper, W., Golany, B., Halek, R., Klopp, G., Schmitz, E. & Thomas, D. (1986), 'Two phase data envelopment analysis approach to policy evaluation and management of army recruiting activities: tradeoffs between joint services and army advertising', *Research Report CCS no. 532, Center for Cybernetic Studies* .

Charnes, A., Cooper, W., Golany, B., Seiford, L. & Stutz, J. (1985), 'Foundations of data envelopment analysis and pareto–koopmans empirical production functions', *Journal of Econometrics* **30**, 91–107.

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

Chen, Y., Cook, W., Kao, C. & Zhu, J. (2013), 'Network DEA pitfalls: Divisional efficiency and frontier projection under general network structures', *European Journal of Operational Research* **226**(3), 507–515.

Chen, Y., Cook, W. & Zhu, J. (2010), 'Deriving the DEA frontier for two-stage processes', *European Journal of Operational Research* **202**(1), 138–142.

Chen, , Cook, W., Li, N. & Zhu, J. (2009), 'Additive efficiency decomposition in two-stage DEA', *European Journal of Operational Research* **196**(3), 1170–1176.

Chilingerian, J. & Sherman, H. (1990), 'Managing physician efficiency and effectiveness in providing hospital services', *Health Services Management Research* **13**(1), 3–15.

Clark, C. & Stansfeld, S. (2007), 'The effect of transportation noise on health and cognitive development: A review of recent evidence', *International Journal of Comparative Psychology* **20**(2-3), 145–158.

Coelli, T. & Perelman, S. (1999), 'Comparison of parametric and non-parametric distance functions: With application to European railways', *European Journal of Operational Research* **117**(2), 326–339.

Commission Decision 2006/66/EC (2006), 'concerning the technical specification for interoperability relating to the subsystem 'rolling stock-noise' of the trans-European conventional rail system', *Official Journal of the European Union* **L 37/1**.

Cook, W. & Zhu, J. (2014), *Data Envelopment Analysis: A Handbook of Modeling Internal Structure and Network*, Vol. 208, Springer US, New York.

Cook, W., Zhu, J., Bi, G. & Yang, F. (2010), 'Network DEA: Additive efficiency decomposition', *European Journal of Operational Research* **207**(2), 1122–1129.

Cooper, W., Huang, Z. & Li, S. (2004), Chance constrained dea, *in* W. Cooper, L. Seiford & J. Zhu, eds, 'Handbook on Data Envelopment Analysis', Kluwer Academic Publishers.

Cooper, W., Seiford, L. & Tone, K. (2007), *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2 edn, Springer, New York.

Cooper, W., Seiford, L. & Zhu, J. (2011), Data envelopment analysis: History, models, and interpretations, *in* W. Cooper, L. Seiford & J. Zhu, eds, 'Handbook on Data Envelopment Analysis. International Series in Operations Research & Management Science', Vol. 164, Springer, Boston, MA.

Council Directive 2002/49/EC (2002), 'relating to the assessment and management of environmental noise', *Official Journal of the European Union* **L 189/12**.

Council Directive 2008/57/EC (2008), 'on the interoperability of the rail system within the Community', *Official Journal of the European Union* **L 191/1**.

Cui, Q. & Li, Y. (2014), 'The evaluation of transportation energy efficiency: an application of three-stage virtual frontier DEA', *Transportation Research Part D* **29**, 1–11.

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

Daraio, C., Simar, L. & Wilson, P. (2015), 'Testing the "separability" condition in two-stage nonparametric models of production', *LEM Papers Series 2015/21, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy* .

De Jorge-Moreno, J. & Garcia-Cebrian, L. (1999), 'Measuring of production efficiency in the European railways', *European Business Review* **99**(5), 332–344.

De Jorge-Moreno, J. & Suarez, C. (2003), 'Has the efficiency of European railway companies been improved?', *European Business Review* **15**(4), 213–220.

Debreu, G. (1951), 'The coefficient of resource utilization', *Econometrica* **19**(3), 273–292.

Despotis, K., Koronakos, G. & Sotiros, D. (2016*a*), 'Composition versus decomposition in two-stage network DEA: a reverse approach', *Journal of Productivity Analysis* **45**, 71–87.

Despotis, K., Koronakos, G. & Sotiros, D. (2016*b*), 'The "weak-link" approach to network dea for two-stage', *European Journal of Operational Research* **254**(2), 481–492.

Dia, M., Takouda, P. & Golmohammadi, A. (2020), 'Assessing the performance of Canadian credit unions using a three-stage network bootstrap DEA', *Annals of Operations Research* .

Djordjević, B. & Krmac, E. (2019), 'Evaluation of energy-environment efficiency of European transport sectors: Non-radial DEA and TOPSIS approach', *Energies* **12**(15).

Djordjević, B., Krmac, E. & Mlinarić, T. (2018), 'Non-radial DEA model: A new approach to evaluation of safety at railway level crossings', *Safety Science* **103**, 234–246.

Dyson, R., Allen, R., Camanho, A., Podinovski, V., Sarrico, C. & Shale, E. (2001), 'Pitfalls and protocols in dea', *European Journal of Operational Research* **132**(2), 245–259.

EEA (2019*a*), 'Final energy consumption in europe by mode of transport'. Accessed: 2022-07-20.
URL: *https://www.eea.europa.eu/data-and-maps/indicators/transport-final-energy-consumption-by-mode/assessment-10*

EEA (2019*b*), 'Greenhouse gas emissions from transport in europe'. Accessed: 2022-07-20.
URL: *https://www.eea.europa.eu/data-and-maps/indicators/transport-emissions-of-greenhouse-gases/transport-emissions-of-greenhouse-gases-12*

EEA (2020), 'Noise country fact sheets 2019'.
URL: *https://www.eea.europa.eu/themes/human/noise/noise-fact-sheets*

EEA (2021), 'Noise pollution is still widespread across europe, but there are ways to reduce the volume'. Accessed: 2022-05-05.
URL: *https://www.eea.europa.eu/signals/signals-2020/articles/noise-pollution-is-still-widespread*

EEA Report No 10/2014 (2014), 'Noise in Europe 2014', *Luxembourg: Publications Office of the European Union* .
URL: *https://www.eea.europa.eu/publications/noise-in-europe-2014*

EEA Report No 22/2019 (2020), 'Environmental noise in Europe-2020', *Luxembourg: Publications Office of the European Union* .
URL: *https://www.eea.europa.eu/publications/environmental-noise-in-europe*

EEA Technical Report No 11/2010 (2010), Good Practice Guide on Noise Exposure and Potential Health Effects, Technical report.

Efron, B. (1979), 'Bootstrap methods: another look at the jackknife', *The Annals of Statistics* **7**(1), 1–26.

Efron, B. (1993), *An Introduction to the Bootstrap*, Chapman Hall, Inc., New York.

Emrouznejad, A., Marra, M., Yang, G. & Michali, M. (2022), 'Eco-efficiency considering netzero and data envelopment analysis: A critical and structured literature review', *IMA Journal of Management Mathematics* **in 2nd revision**.

Emrouznejad, A. & Yang, G. (2018), 'A survey and analysis of the first 40 years of scholarly literature in dea: 1978–2016', *Socio-Economic Planning Sciences* **61**(1), 1–5.

Emrouznejad, A., Yang, G., Khoveyni, M. & Michali, M. (2022), Data envelopment analysis: Recent developments and challenges, *in* S. Salhi & J. Boylan, eds, 'The Palgrave Handbook of Operations Research', Palgrave Macmillan, Cham.

ERA 006REC1072 Impact Assessment (2018), 'Revision of the Noise TSI: Application of Noise TSI requirements to existing freight wagons'.

Eurostat (2016), 'Railway transport'.
URL: *https://ec.europa.eu/eurostat/data/database*

Farrell, M. (1957), 'The measurement of productive efficiency', *Journal of the Royal Statistical Society, Series A (General)* **120**(3), 253–281.

Fukuyama, H. & Mirdehghan, S. (2012), 'Identifying the efficiency status in network dea', *European Journal of Operational Research* **220**, 85–92.

Färe, R. (1991), 'Measuring farrell efficiency for a firm with intermediate inputs', *Academia Economic Papers* **19**, 329–340.

Färe, R., Grabowski, R., Grosskopf, S. & Kraft, S. (1997), 'Efficiency of a fixed but allocatable input: A non-parametric approach', *Economics Letters* **56**, 187–193.

Färe, R. & Grosskopf, S. (1996), 'Productivity and intermediate products: A frontier approach', *Economics Letters* **50**(1), 65–70.

Färe, R. & Grosskopf, S. (2000), 'Network dea', *Socio-Economic Planning Sciences* **34**(1), 35–49.

Färe, R. & Primont, D. (1984), 'Efficiency measures for multiplant firms', *Operations Research Letters* **3**, 257–260.

Färe, R. & Whittaker, G. (1995), 'An intermediate input model of dairy production using complex survey data', *J. of Agricultural Economics* **46**(2), 201–213.

George, S. & Rangaraj, N. (2008), 'A performance benchmarking study of Indian Railway zones', *Benchmarking* **15**, 599–617.

Gijbels, I., Mammen, E., Park, B. & Simar, L. (1999), 'On estimation of monotone and concave frontier functions', *J. Am. Stat. Assoc.* **94**(445), 220–228.

Graham, D. (2008), 'Productivity and efficiency in urban railways: parametric and non-parametric estimates', *Transportation Research Part E: Logistics and Transportation Review* **44**(1), 84–99.

Growitsch, C. & Wetzel, H. (2009), 'Testing for economies of scope in European railways', *Journal of Transport Economics and Policy* **43**(1), 1–24.

Guo, C., Shureshjani, R., Foroughi, A. & Zhu, J. (2017), 'Decomposition weights and overall efficiency in two-stage additive network dea', *European Journal of Operational Research* **257**(3), 896–906.

Ha, H.-K., Yoshida, Y. & Zhang, A. (2011), 'Social efficiency benchmarking of Japanese domestic transport services: A comparison of rail and air', *Transportation Research Part D* **16**, 554–561.

Hilmola, O. (2007), 'European railway freight transportation and adaptation to demand decline – efficiency and partial productivity analysis from period of 1980–2003', *International Journal of Productivity and Performance Management* **56**(3), 205–225.

Hilmola, O. (2008), 'Railway efficiency analysis from larger Europe during period of 1994–2003', *International Journal of Operational Research* **3**(3), 255–280.

Homburg, C. (2001), 'Using data envelopment analysis to benchmark activities', *International Journal of Production Economics* **73**(1), 51–58.

Jeong, S. & Park, B. (2006), 'Large sample approximation of the distribution for convex-hull estimators of boundaries', *Scand. J. Stat* **33**(1), 139–151.

Jitsuzumi, T. & Nakamura, A. (2010), 'Causes of inefficiency in Japanese railways: Application of DEA for managers and policymakers', *Socio-economic planning sciences* **44**(3), 161–173.

Kao, C. (2014), 'Network data envelopment analysis: A review', *European Journal of Operational Research* **239**, 1–16.

Kao, C. (2018), 'A classification of slacks-based efficiency measures in network data envelopment analysis with an analysis of the properties possessed', *European Journal of Operational Research* **270**(3), 1109–1121.

Kao, C. & Hwang, S. (2008), 'Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan', *European Journal of Operational Research* **185**(1), 418–429.

Kao, C. & Hwang, S. (2010), 'Efficiency measurement for network systems: It impact on firm performance', *Decision Support Systems* **48**, 437–446.

Kao, C. & Hwang, S. (2011), 'Decomposition of technical and scale efficiencies in two-stage production systems', *European Journal of Operational Research* **211**, 515–519.

Khadem Sameni, M., Preston, J. & Khadem Sameni, M. (2016), 'Evaluating efficiency of passenger railway stations: A DEA approach', *Research in Transportation Business and Management* **20**, 33–38.

Kim, H., Choi, C., Woo, J., Choi, Y., Kim, K. & Wu, D. (2011), 'Efficiency of the modal shift and environmental policy on the Korean railroad', *Stochastic Environmental Research and Risk Assessment* **25**(3), 305–322.

Kitagawa, T. (2009), 'The influence of wheel and track parameters on rolling noise', *Quarterly Review of RTRI* **50**(1), 32–38.

Kneip, A., Park, B. & Simar, L. (1998), 'A note on the convergence of nonparametric DEA efficiency measures', *Econometric Theory* **14**(6), 783–793.

Kneip, A., Simar, L. & Wilson, P. (2008), 'Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models', *Econometric Theory* **24**(6), 1663–1697.

Kneip, A., Simar, L. & Wilson, P. (2011), 'A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators', *Comput Econ* **38**(4), 483–515.

Kneip, A., Simar, L. & Wilson, P. (2015), 'When bias kills the variance: central limit theorems for DEA and FDH efficiency scores', *Econometric Theory* **31**, 394–422.

Kneip, A., Simar, L. & Wilson, P. (2016), 'Testing hypotheses in nonparametric models of production', *Journal of Business & Economic Statistics* **34**(3), 435–456.

Koopmans, T. C. (1951), An analysis of production as an efficient combination of activities, *in* T. C. Koopmans, ed., 'Activity Analysis of Production and Allocation', Cowles Commission Monograph No. 13., John Wiley and Sons Inc., pp. 33–97.

Koronakos, G. (2017), The composition and the weak-link approaches to Network Data Envelopment Analysis, PhD thesis, University of Piraeus, Piraeus.

Korostelev, A., Simar, L. & Tsybakov, A. (1995), 'On estimation of monotone and convex boundaries', *Publications de l'Institut de Statistique de l'Université de Paris XXXIX* **1**, 3–18.

Kuang, X. (2018), Evaluation of railway transportation efficiency based on super-cross efficiency, *in* 'IOP Conference Series: Earth and Environmental Science', Vol. 108.

Kutlar, A., Kabasakal, A. & Sarikaya, M. (2013), 'Determination of the efficiency of the world railway companies by method of DEA and comparison of their efficiency by Tobit analysis', *Quality and Quantity* **47**(6), 3575–3602.

Lewis, H. & Sexton, T. (2004), 'Network dea: Efficiency analysis of organizations with complex internal structure', *Computers and Operations Research* **31**(9), 1365–1410.

Li, Y., Lei, X., Dai, Q. & Liang, L. (2015), 'Performance evaluation of participating nations at the 2012 london summer olympics by a two-stage data envelopment analysis', *European Journal of Operational Research* **243**(3), 964–973.

Li, Y., Li, X. & Khalid, M. (2018), 'Measuring technical efficiency of Chinese railway administrations by DEA method', *Journal of Interdisciplinary Mathematics* **21**(4), 825–836.

Li, Y., Liu, A., Yu, Y., Zhang, Y., Zhan, Y. & Lin, W. (2022), 'Bootstrapped dea and clustering analysis of eco-efficiency in china's hotel industry', *Sustainability* **14**(5), 2925.

Liang, L., Cook, W. & Zhu, J. (2008), 'Dea models for two-stage processes: Game approach and efficiency decomposition', *Naval Research Logistics* **55**, 643–653.

Liang, L., Yang, F., Cook, W. & Zhu, J. (2006), 'Dea models for supply chain efficiency evaluation', *Annals of Operations Research* **145**(1), 35–49.

Lim, S. & Zhu, J. (2013), 'Integrated data envelopment analysis: Global vs local optimum', *European Journal of Operational Research* **229**(1).

Lim, S. & Zhu, J. (2019), 'Primal-dual correspondence and frontier projections in two-stage network dea models', *Omega* **83**, 236–248.

Liu, H., Zhang, Y., Zhu, Q. & Chu, J. (2017), 'Environmental efficiency of land transportation in China: A parallel slack-based measure for regional and temporal analysis', *Journal of Cleaner Production* **142**, 867–876.

Liu, Z., Qin, C. & Zhang, Y. (2016), 'The energy-environment efficiency of road and railway sectors in China: Evidence from the provincial level', *Ecological Indicators* **69**, 559–570.

Lozano, S. (2015), 'Alternative sbm model for network dea', *Computers and Industrial Engineering* **32**, 33–40.

Lozano, S. (2016), 'Slacks-based inefficiency approach for general networks with bad outputs: an application to the banking sector', *Omega* **60**, 73–84.

Lu, W., Kweh, Q. & Huang, C. (2014), 'Intellectual capital and national innovation systems performance', *Knowledge-Based Systems* **71**, 201–210.

Mahmoudi, R., Emrouznejad, A., Shetab-Boushehri, S. & Hejazi, S. (2020), 'The origins, development and future directions of data envelopment analysis approach in transportation systems', *Socio-Economic Planning Sciences* **69**.

Mallikarjun, S., Lewis, H. & Sexton, T. (2014), 'Operational performance of U.S. public rail transit and implications for public policy', *Socio-Economic Planning Sciences* **48**(1), 74–88.

Mapapa, A. (2004), 'Management of railways in Sub-Saharan Africa', *Rail International* **35**(8), 13–21.

Marchetti, D. & Wanke, P. (2017), 'Brazil's rail freight transport: Efficiency analysis using two-stage DEA and cluster-driven public policies', *Socio-economic planning sciences* **59**, 26–42.

Michali, M., Emrouznejad, A., Dehnokhalaji, A. & Clegg, B. (2021), 'Noise-pollution efficiency analysis of European railways: A network DEA model', *Transportation Research Part D: Transport and Environment* **98**, 102980.

Michali, M., Emrouznejad, A., Dehnokhalaji, A. & Clegg, B. (2022), 'Subsamplng bootstrap in Network DEA', *European Journal of Operational Research (Accepted)* .

Moesen, W. (1994), 'The need for performance auditing in the public sector and the best-practice frontier', *European Journal of Law and Economics* **1**(4), 263–274.

Mohajeri, N. & Amin, G. (2010), 'Railway station site selection using analytical hierarchy process and data envelopment analysis', *Computers and Industrial Engineering* **59**(1), 107–114.

Moradi-Motlagh, A. & Emrouznejad, A. (2022), 'The origins and development of statistical approaches in non-parametric frontier models: A survey of the first two decades of scholarly literature (1998-2020).', *Annals of Operations Research* .

Moullin, M. (2002), *Delivering Excellence in Health and Social Care*, Open University Press, Buckingham.

Noroozzadeh, A. & Sadjadi, S. (2013), 'A new approach to evaluate railways efficiency considering safety measures', *Decision Science Letters* **2**(2), 71–80.

Ntovantzi, M., Bertoldi, P., Calvete, S., Economidou, M., Kona, A., Serrenho, T. & Morton, C. (2015), Do we have effective energy efficiency policies for the transport sector? results and recommendations from an analysis of the national and sustainable energy action plans, *in* 'Proceedings of ECEEE 2015 Summer Study on Energy Efficiency', Toulon/Hyères, France.

Nwaogbe, O., Wanke, P., Barros, I. & Azad, A. (2017), 'Efficiency Driver in Nigerian Airports: A Bootstrap DEA–Censored Quantile Regression Approach', *Socio-Economic Planning Sciences* **59**, 26–42.

Olesen, O. & Petersen, N. (2016), 'Stochastic data envelopment analysis- A review', *European Journal of Operational Research* **251**, 2–21.

Oum, T. & Yu, C. (1994), 'Economic efficiency of railways and implications for public policy', *J Transport Economics and Policy* **28**(2), 121–138.

Park, B., Jeong, S. & Simar, L. (2010), 'Asymptotic distribution of conical-hull estimators of directional edges', *Ann. Stat* **38**(3), 1320–1340.

Peterson, J. (1996), 'The economics of British rail privatization: an assessment', *Transport Reviews* **16**(1), 1–21.

Peyrache, A. & Silva, M. C. A. (2022), A comment on decomposition of efficiency in network production models, CEPA Working Papers Series WP0742022, University of Queensland, School of Economics.

Podinovski, V. & Thanassoulis, E. (2007), 'Improving discrimination in data envelopment analysis: some practical suggestions', *Journal of Productivity Analysis* **28**(1-2), 117–126.

Politis, D., Romano, J. & Wolf, M. (2001), 'On the asymptotic theory of subsampling', *Statistica Sinica* **11**, 1105–1124.

Pyrgidis, C. (2016), *Railway transportation systems: design, construction and operation*, CRC Press.

Raab, R. & Lichty, R. (2002), 'Identifying sub-areas that comprise a greater metropolitan area: The criterion of country relative efficiency', *Journal of Regional Science* **42**(3), 579–594.

Ray, S. (2004), *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*, Cambridge University Press, UK.

Rayeni, M. & Saljooghi, F. (2014), 'Ranking and measuring efficiency using secondary goals of cross-efficiency evaluation -a study of railway efficiency in Iran', *International Journal of Services and Operations Management* **17**(1), 1–16.

Roets, B., Verschelde, M. & Christiaens, J. (2018), 'Multi-output efficiency and operational safety: An analysis of railway traffic control centre performance', *European Journal of Operational Research* **271**(1), 224–237.

Rotoli, F., Navajas Cawood, E. & Christidis, P. (2015), 'A data envelopment analysis approach for accessibility measures: Simulating operational enhancement scenarios for railway across Europe', *European Transport Research Review* **7**(2).

Rotoli, F., Valeri, E., Ricci, S., Rizzetto, L. & Malavasi, G. (2018), 'An analysis of the railway access charges regime in the Italian context', *Transport Policy* **64**, 20–28.

Sahoo, B., Saleh, H., Shafiee, M., Tone, K. & Zhu, J. (2021), 'An alternative approach to dealing with the composition approach for series network production processes', *Asia-Pacific Journal of Operational Research* **38**(6), 1–27.

Savolainen, V. & Hilmola, O. (2009), 'The relative technical efficiency of European transportation systems concerning air transport and railways', *International Journal of Business Performance Management* **11**(1-2), 19–42.

Schroten, A., van Wijngaarden, L., Brambilla, M., Gatto, M., Trosky, F., Kramer, H., Monden, K., Bertschmann, D., Killer, M., Lambla, V., El Beyrouty, K. & Amaral (2019), 'Overview of transport infrastructure expenditures and costs', *European Commission* (4K83), 175.

Scott, D. (2015), *Multivariate Density Estimation (Theory, Practice, and Visualization)*, John Wiley and Sons, New Jersey.

Seiford, L. & Zhu, J. (1999), 'Profitability and marketability of the top 55 US commercial banks', *Management Science* **45**(9), 1270–1288.

Sexton, T. & Lewis, H. (2003), 'Two–stage dea: An application to major league baseball', *Journal of Productivity Analysis* **19**(2), 227–249.

Sharma, M., Debnath, R., Oloruntoba, R. & Sharma, S. (2016), 'Benchmarking of rail transport service performance through DEA for Indian railways', *International Journal of Logistics Management* **27**(3), 629–649.

Shephard, R. (1970), *The Theory of Cost and Production Functions*, Princeton University Press, Princeton.

Shi, F., Lim, S. & Chi, J. (2011), 'Railroad productivity analysis: Case of the American Class I railroads', *International Journal of Productivity and Performance Management* **60**(4), 372–386.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Simar, L. & Wilson, P. (1998), 'Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models', *Management Science* **44**(1), 49–61.

Simar, L. & Wilson, P. (1999), 'Estimating and bootstrapping Malmquist indices', *Eur J. Oper Res* **115**(3), 459–471.

Simar, L. & Wilson, P. (2000), 'A general methodology for bootstrapping in nonparametric frontier models', *J. Appl Stat* **27**(6), 779–802.

Simar, L. & Wilson, P. (2001), 'Testing restrictions in nonparametric efficiency models', *Communications in Statistics Part B: Simulation and Computation* **30**, 159–184.

Simar, L. & Wilson, P. (2002), 'Non-parametric tests of returns to scale', *European Journal of Operational Research* **139**(1), 115 – 132.

Simar, L. & Wilson, P. (2007), 'Estimation and inference in two-stage, semi-parametric models of productive efficiency', *Journal of Econometrics* **136**, 31–64.

Simar, L. & Wilson, P. (2008), Statistical inference in Nonparametric Frontier Models: Recent Developments and Perspectives, *in* H. Fried, C. A. K. Lovell & S. Schmidt, eds, 'The Measurement of Productive Efficiency and Productivity Change', Oxford Scholarship Online.

Simar, L. & Wilson, P. (2010), 'Inference by the m out of n bootstrap in nonparametric frontier models', *J. Prod Anal* **36**(1), 33–53.

Simar, L. & Wilson, P. (2015), 'Statistical approaches for non-parametric frontier models: A guided tour', *International Statistical Review* **83**(1), 77–110.

Simar, L. & Wilson, P. (2020), 'Hypothesis testing in nonparametric models of production using multiple sample splits', *J. Prod Anal* **53**, 287–303.

Simar, L. & Zelenyuk, V. (2006), 'On testing the equality of two distribution functions of efficiency scores estimated from DEA', *Econometric Reviews* **24**, 497–522.

Simar, L. & Zelenyuk, V. (2007), 'Statistical inference for aggregates of farrell-type efficiencies', *Journal of Applied Econometrics* **27**, 1367–1394.

Song, M., Zhang, G., Zeng, W., Liu, J. & Fang, K. (2016), 'Railway transportation and environmental efficiency in China', *Transportation Research Part D* **48**, 488–498.

Sozen, A., Alp, I. & Kurt, B. (2012), 'Total factor efficiency of Turkish state railways', *Energy Education Science and Technology Part A: Energy Science and Research* **29**(1), 663–678.

Sozen, A. & Cipil, F. (2018), Efficiency analysis of Turkey's transportation system using decision support model: Data envelopment method, *in* 'Intelligent Transportation and Planning: Breakthroughs in Research and Practice', p. 480–509.

Swanepoel, J. (1986), 'A note on proving that the (modified) bootstrap works', *Communications in Statistics: Theory Methods* **15**, 3193–3203.

Thanassoulis, E. (2001), *Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text with Integrated Software*, NewYork: Springer.

Thomson, D. & Gautier, P. (2006), 'Review of research into wheel/rail rolling noise reduction', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* **220**, 385–408.

Thomson, D., Latorre Iglesias, E., Liu, X. & Zhu, J.and Hu, Z. (2015), 'Recent developments in the prediction and control of aerodynamic noise from high-speed trains', *International Journal of Rail Transportation* **3**, 119–150.

Tone, K. (1996), 'A simple characterization of returns to scale in dea', *Journal of the Operational Research Society* **39**(4), 604–613.

Tone, K. (1997), Several algorithms to determine multipliers for use in cone-ratio envelopment approaches to efficiency evaluations in dea, *in* H. Amman, B. Rustem & A. Whinston, eds, 'Computational Approaches to Economic Problems. Advances in Computational Economics', Vol. 6, Springer, Boston, MA.

Tone, K. (2001), 'A slacks-based measure of efficiency in data envelopment analysis', *European journal of operational Research* **130**(3), 498–509.

Tone, K. & Tsutsui, M. (2009), 'Network DEA: A slacks based measure approach', *European Journal of Operational Research* **197**(1), 243–252.

Trinh, K. & Zelenuyk, V. (2015), Bootstrap-based testing for network DEA: Some Theory and Applications, Working Paper Series WP05/2015, School of Economics, University of Queensland, Australia.

UIC & CER (2012), 'Moving towards sustainable mobility. A strategy for 2030 and beyond for the European railway sector', *UIC Communications Department, Paris* .

Wang, C., Gopal, R. & Zionts, S. (1997), 'Use of data envelopment analysis in assessing information technology impact on firm performance', *Annals of Operations Research* **73**, 191–213.

Wanke, P. (2012), 'Capacity shortfall and efficiency determinants in Brazilian airports: evidence from bootstrapped DEA estimates', *Socio-Economic Planning Sciences* **46**, 216–229.

Wanke, P., Chen, Z., Liu, W., Antunes, J. & Azad, M. A. (2018), 'Investigating the drivers of railway performance: Evidence from selected Asian countries', *Habitat International* **80**, 49–69.

WHO (2009), 'Night noise guidelines for europe', *WHO Regional Office for Europe, Copenhagen* .

Yu, M. (2008b), 'Assessing the technical efficiency, service effectiveness, and technical effectiveness of the world's railways through NDEA analysis', *Transportation Research Part A: Policy and Practice* **42**(10), 1283–1294.

Yu, M. & Lin, E. (2008), 'Efficiency and effectiveness in railway performance using a multi-activity network DEA model', *Omega* **36**(6), 1005–1017.

Zha, Y. & Liang, L. (2010), 'Two-stage cooperation model with input freely distributed among the stages', *European Journal of Operational Research* **205**(2), 332–338.

Zhou, G., Chung, W. & Zhang, Y. (2014), 'Measuring energy efficiency performance of China's transport sector: a data envelopment analysis approach', *Expert Syst. Appl* **41**, 709–722.

Zhou, H. & Hu, H. (2017), 'Sustainability evaluation of railways in China using a two-stage network DEA model with undesirable outputs and shared resources', *Sustainability (Switzerland)* **9**(1), 1–23.

Zhu, J. (2000), 'Multi-factor performance measure model with an application to fortune 500 companies', *European Journal of Operational Research* **123**(1), 105–124.

Zhu, J. (2016), *Data Envelopment Analysis: A Handbook of Empirical Studies and Applications*, Springer, New York.