CARDIFF UNIVERSITY

ESSAYS ON EFFICIENCY AND PRODUCTIVITY: THE GREEK BANKING CASE

by Panagiotis Tziogkidis

A Thesis Submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy of Cardiff University

2014

ECONOMICS DEPARTMENT, CARDIFF BUSINESS SCHOOL

This page was intentionally left blank

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ………………………………………………………. (candidate)

Date …………………………

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of …………………………(insert MCh, Md, MPhil, PhD etc, as appropriate)

Signed ………………………………………………………. (candidate)

Date …………………………

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources are acknowledged by footnotes giving explicit references.

Signed ………………………………………………………. (candidate)

Date …………………………

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ………………………………………………………. (candidate)

Date …………………………

Abstract

Bootstrap DEA is a valuable tool for gauging the sensitivity of DEA scores towards sampling variations, hence allowing for statistical inference. However, it is associated with generous assumptions while evidence on its performance is limited. This thesis begins with the evaluation of the performance of bootstrap DEA in small samples through a variety of Monte Carlo simulations. The results indicate cases under which bootstrap DEA may underperform and it shown how the violation of the fundamental assumption of equal bootstrap and DEA biases may affect confidence intervals and cause the evidenced underperformance. An alternative approach, which utilises the Pearson system random number generator, seems to perform well towards this respect. In particular, coverage probabilities converge to the nominal ones for samples as small as 120 observations and the bootstrap biases are very close to the DEA ones. In the presence of technological heterogeneity, though, poor performance is observed in all cases, which is not surprising as even the applicability of simple DEA is questionable.

Using an illustrative example from the deregulation of the Greek banking sector during late 80s, potential differences arising from the various approaches are discussed. In particular, the theoretical explorations are extended to the case of the Global Malmquist productivity index, which is used to examine the productivity change of Greek banks during (de)regulation. Some differences are observed on the magnitudes of the estimated quantities of interest and on the probability masses at the tails of the relevant bootstrap distributions. Qualitatively, though, the overall conclusions are very similar; the provision of commercial freedoms enhanced the productivity of commercial banks whereas the imposition of prudential controls had the opposite effect. This result is of topical interest as the European Supervisory Mechanism, which recently assumed duties, will closely supervise “significant institutions” which includes the 4 biggest Greek banks and their banking subsidiaries.

Keywords: *efficiency, productivity, DEA, bootstrap DEA, Global Malmquist index, hypothesis testing, Monte Carlo simulations, banking, deregulation*

JEL Classifications: *C12, C14, C15, C61, C67, G21, G28, L25*

Acknowledgements

I would like to thank my supervisors Prof Kent Matthews and Prof Patrick Minford for their support and guidance throughout the PhD, as well as the two examiners. Their comments have certainly improved the structure and the content of the thesis. I would also like to thank the staff and my colleagues at Cardiff Business School for their useful comments on my work presented at Cardiff Economics PhD Workshops. Research funding from the Economic and Social Research Council (ESRC), Cardiff Business School and the Julian Hodge Institute of Applied Macroeconomics is gratefully acknowledged.

I have also benefited from comments from various participants at the International Data Envelopment Analysis Society Conference (Thessaloniki, 2011), the Quantitative Economics Doctorate Meeting (Copenhagen, 2011), the Financial Engineering and Banking Society Conference (Paris, 2013) and the European Workshop on Efficiency and Productivity Analysis Workshop (Helsinki, 2013). I would particularly like to thank Prof Leopold Simar and Prof Paul Wilson for their constructive feedback on my paper presented at the EWEPA conference. Their comments have significantly improved the quality of my work while their encouragement for my future research plans is deeply appreciated. Moreover, Prof Mike Tsionas, Prof Robin Sickles and Prof Kris Kerstens have also provided useful suggestions for my work. Thanks also go to Dr Yiannis Kouropalatis who has advised me on presentational and other peripheral, though important, aspects of my work.

Special thanks also go to Prof Costas Siriopoulos who has been my mentor since my undergraduate studies at the University of Patras and who has also provided valuable comments and useful advice during my PhD. Last but not least, I would like to thank my family for supporting me all these years.

My warmest gratitude, though, is reserved for my partner, Anna Ziouti, who has been patient, caring and understanding throughout my PhD life and to whom I devote this work.

List of Abbreviations

AEC:	Adjusted Efficiency Change
AIC:	Akaike information criterion
AMISE:	Asymptotic Mean Integrated Square Error
ASE:	Athens Stock Exchange
BCV:	Biased Cross-Validation
CRS:	Constant Returns to Scale
DEA:	Data Envelopment Analysis
DGP:	Data Generating Process
DMU:	Decision Making Unit
DRS:	Decreasing Returns to Scale
FDH:	Free Disposable Hull
GAS:	Greek Accounting Standards
IAS:	International Accounting Standards
IRS:	Increasing Returns to Scale
ISE:	Integrated Square Error
LCV:	Likelihood Cross-Validation
LSCV:	Least Squares Cross-Validation
M&As:	Mergers and Acquisitions
MISE:	Mean Integrated Square Error
MPSS:	Most Productive Scale Size
OLS:	Ordinary Least Squares
RTS:	Returns to Scale
SFA:	Stochastic Frontier Analysis
SJPI or SJ:	Sheather-Jones Plug-In method
SW1998:	Simar and Wilson’s (1998) confidence intervals
SW2000:	Simar and Wilson’s (2000a) confidence intervals
TFA:	Thick Frontier Approach
VRS:	Variable Returns to Scale

Contents

List of tables

List of figures

1 Introduction

The analysis of efficiency and productivity is an area of research interest and practical importance for various fields. The motivation behind such analyses is associated with the evaluation of certain management practices or the effects of firm-level or economy-wide events on firm performance. The development of appropriate models for the measurement of efficiency and productivity of various *decision making units* (DMUs) traces back to the theoretical works of Debreu (1951), Koopmans (1951) and Farrell (1957), while various techniques have been developed since then.

Empirical studies in the literature have explored various datasets, for different time periods and have employed a range of efficiency measurement techniques. Not surprisingly, owing to this diversity it is possible to obtain different results, even when applying efficiency and productivity analysis methods on the same group of firms. In fact, it is possible to obtain different results even when focusing on a certain industry and using a certain technique, if we vary the group of firms considered in the analysis (perhaps due to the choice of different time periods or the inclusion or exclusion of certain firms from the analysis).

The sensitivity of results towards sampling variations is certainly relevant to linear programming techniques of efficiency measurement, such as Data Envelopment Analysis (DEA) which is used in this thesis, as efficiency frontiers are constructed from the data. The different shapes of the frontier that may result from data variations highlights two issues: (i) that sample selection needs careful consideration as it may affect results, and that (ii) there is a “true” population frontier which is unobservable within which all observations are enveloped. Hence, sample efficiency scores are only estimators of their underlying population values, the detection of which is a challenge worth pursuing. One way to perform this task is by using a technique called bootstrap DEA which allows constructing confidence intervals for these “true” efficiency scores.

This thesis examines through simulations the extent to which bootstrap DEA is successful towards covering the aforementioned “true” efficiency scores. Indicating cases where the performance might not be satisfactory, we propose a variation of the original technique which seems to perform well in small samples. After suggesting guidelines for the implementation of bootstrap DEA, we perform an empirical illustration on the Greek banking sector during the reforms of the late 80s.

The current chapter serves as a preface of the thesis, outlining the research questions, presenting the preliminary findings and motivating the topics examined. The remainder of the chapter is structured as follows: section 1.1 outlines the purpose of the thesis and succinctly presents the major findings; section 1.2 justifies the focus of the empirical application on the Greek banking sector; section 1.3 states the motivation and contribution of the thesis, while section 1.4 outlines its structure.

1.1 Purpose of study and preliminary results

The measurement of efficiency and productivity can be performed either with parametric or non-parametric models; in each case there are strengths and weaknesses. Perhaps the most popular non-parametric model is data envelopment analysis (DEA), which empirically constructs an efficiency frontier from the data. One of its attractive features is that there is no need to specify a production function, but at the same time the lack of a parametric specification makes statistical inference challenging. One relatively recent development is the implementation of bootstrap methods to construct confidence intervals for the efficiency score of each DMU where its "true" value is expected to lie.

The initial contribution by Simar and Wilson (1998) has led to further developments and extensions of bootstrap DEA such as the bootstrap Malmquist index (Simar and Wilson, 1999), the introduction of bootstrap tests on returns to scale (Simar and Wilson, 2002), the implementation of two-stage bootstrap DEA to account for environmental variables (Simar and Wilson, 2007) and others. However attractive these developments may be, there are no clear guidelines in the literature on sample size requirements; even in the works of Simar and Wilson (we will elaborate on this issue in section 2.1). The fact that the literature has also investigated alternatives to the initial version of bootstrap DEA (see section 4.1), and in particular of the smoothing techniques applied in the first steps of the algorithm, indicates that the required sample size is still a concern and that there is room for further improvement. However, since the most recent developments seem to require 1000 observations or more, and due to the fact that all kernel density estimation methods introduce additional variability (Simar and Wilson, 2002), it may be a good idea to explore alternative approaches to kernel density estimation which could be applicable to small samples which are often met in the empirical DEA literature.

The purpose of the thesis is to theoretically explore the behaviour of bootstrap DEA, to assess its performance through Monte Carlo simulations and to propose an alternative approach that is applicable in smaller samples. The theoretical explorations focus on the limitations of the existing approaches and on cases under which these approaches may underperform. We show that the assumption of equal bootstrap and DEA (or model) biases is central for the performance of these methods and that their violation may result in confidence intervals which overestimate or underestimate the "true" efficiency scores. The implication is that hypothesis testing may lead to wrong decisions and it should be therefore used with care.

The literature is not rich in simulation evidence on bootstrap DEA (see section 2.7) and the Monte Carlo experiments in this thesis are by far more extensive compared to other papers; yet not exhaustive. As Silverman and Young (1987) suggested, when kernel smoothing techniques are used, the performance of the bootstrap procedures should be evaluated under various setups and data generating processes; therefore, the author believes that there may still be room for further explorations. In our simulations we find that bootstrap DEA (and even simple DEA in fact) should not be used if the firms of the sample exhibit substantial technological differences; this may result in distributions with a thin tail towards 1, dominated by the firms with access to superior technology. On the other hand, when the firms in the sample do not exhibit such heterogeneities, bootstrap DEA yields better results. However, in our simulations we find that, although bootstrap DEA has nice asymptotic properties, it is not safe to be used with small samples; at least not in its initial form. An interesting "by-product" of our investigations is that if all the firms in the sample (and the underlying population) have almost identical production processes (and are therefore technologically homogeneous) the sampling variations almost disappear after a certain sample size and the resulting scores are approximately equal to the population ones; this suggests that a simple application of DEA would be adequate, if the sample is larger "enough".

Taking into account the comment by Simar and Wilson (2002) that kernel density estimation methods may introduce additional variability in bootstrap DEA, we proposed an alternative approach. The "moments bootstrap", as we named it, uses the first four moments of the empirical distribution of DEA scores to construct a pseudo-population from which draws can be performed within the context of bootstrap DEA. Effectively, our approach replaces the kernel density estimation step in the original paper of Simar and Wilson (1998) with pseudo-population generation from sample moments. Simulation evidence suggests that the resulting confidence intervals yield coverage probabilities that converge to the nominal ones for sample sizes as small as 120 DMUs in a 2-input/2-output setup.

The lessons learned from our theoretical investigations are summarised in a succinct manual-type chapter (chapter 5) where we suggest guidelines on the application of bootstrap DEA and hypothesis testing. We also perform an illustrative application (chapter 6) on the Greek banking sector during the period of sector reforms of the late 80s. In particular, we use a global frontier to compute the global technical efficiency scores of Greek banks and we show how our proposed approaches can be extended to the case of the Global Malmquist productivity index of Pastor and Lovell (2005). Although we observe some differences in the proposed approaches (mainly with respect to the shape of the bootstrap distributions, the associated confidence intervals and the rejection rates of the null hypothesis of no change in productivity), we find that we would reach the same qualitative conclusion with each approach. In particular, we find that the provision of commercial freedoms enhances the productivity of Greek banks the following year while the imposition of prudential controls has the opposite effect, which is in line with banking theory. Our empirical findings also indicate that the overall behaviour of the Greek banking sector is driven by big banks, which may carry implications for the current situation as the 4 biggest banks in Greece entered the ESM on the 4th of November 2014 and they will be more closely supervised.

1.2 Why Greece?

The empirical application of the thesis concerns one of the most interesting periods in Greek banking which could be termed as the “modernization” period. This term is justified by the introduction of a series of “Europeanization” laws in banking and the abolishment of other outdated ones through a directive for the restructuring and modernization of the Greek banking sector in the view of the forthcoming Single Market. In particular, in 1987 a framework of sector reforms was introduced with a 5 year implementation period which aimed at the deregulation of the previously heavily regulated Greek banking sector. By 1993 Greek banks enjoyed more commercial freedoms but this was followed by the imposition of prudential controls, mainly aiming at the capital adequacy of banks through the adoption of Basel I.

During the years that followed and until the entrance of Greece in the Eurozone in 2001, the macroeconomic conditions had been improving while a wave of mergers and acquisitions was observed during the last few years. The latter probably served the strategic goals of banks as size was an important aspect of the heavily concentrated Greek banking sector, but it could be also considered as a “preparatory” step before the entrance to the Eurozone which would open up possibilities for expansion abroad but could also attract competition from other EU member states.

During these last few years before the entrance to the Eurozone, but mainly after 2001, we observe banks moving towards a universal banking model, offering a wider range of banking and other financial services, which was also evident in the substantial increase of their off-balance-sheet (OBS) activities. In addition, new entrants appear in the market, reducing concentration and increasing competition. The biggest Greek banks expand their activities to the relatively unexploited Balkan, Eastern European and Turkish financial markets. At the same time, the access to the cheap funds from the ECB meant that Greek banks could offer loans and mortgages at historically low interest rates which led to rapid growth of baking operations and which also increased the investment activity in Greece. The Greek banking sector had changed in structure and conduct of business and we consider the entrance to the EU as a turning point for Greek banks.

In 2009, Greece was severely hit by a debt crisis which was the result of accumulating deficits and significant operational and cost inefficiencies in the public sector. Greece entered an agreement with Troika (IMF, European Commission and European Central Bank) to introduce austerity measures and enhance its finances. The negative outlook of Greece led to a panic of depositors and investors and to a subsequent fall in banking revenues and deposits, making the survival of most Greek banks questionable. Especially after the 53.5% "haircut" of debt in 2012, Greek banks, which held most of Greek bonds, were obliged to report losses of many billions of Euros, which was mainly financed by equity, leading to unforeseen negative equity for 4 big banks. This resulted in a consolidation wave and two recapitalisations that Greek banks had to undergo in order to gain access to liquidity funds. The Greek banking sector is still in a transition process while the recent inclusion of the 4 biggest Greek banks under the direct supervision of the ESM poses challenges on their efficient operation.

The aforementioned events show that the Greek banking sector has an interesting history. The examination of its deregulation and reregulation period might be relevant today and could carry implications about the effect of tighter controls imposed in an already turbulent period for Greek banks. The fact that the deregulation and reregulation occurred in a period when the Greek banking sector was highly concentrated and the macroeconomic outlook of Greece was in a bad shape (similar to the current situation), suggests that the lessons of the past could be used to draw implications for the present.

1.3 Motivation and contribution

The initial motivation for this research project related to the recent Greek debt crisis which led to the situation described above. To draw implications for the current situation, we decided to use data from the past due to the aforementioned similarities. In fact, there is no study in the literature that covers the whole period of reforms from 1987 to 1994 as we do (while we also extend it until 1999 to capture longer term effects). Moreover, evidence from recent studies on the Greek banking sector indicates that significant destabilising events have had a negative impact on banks' efficiency (Siriopoulos and Tziogkidis, 2010), which motivated us to examine whether this was also true for the reforms of the late 80s.

Addressing this question required the use of an appropriate methodology of efficiency and productivity assessment which would offer meaningful results. The best candidate approaches to test such hypotheses were those of Simar and Wilson (1998,1999, 2000a); however, we were concerned about their compatibility with small samples, as in our case. This motivated the theoretical explorations of the thesis, which preceded our empirical analysis and became the main focus of this monograph.

The contributions of the thesis are the following: (i) it assesses the performance of bootstrap DEA under a range of Monte Carlo simulations which are the most extensive compared to others in the literature, (ii) it indicates cases where bootstrap DEA may underperform and explains the possible sources of this underperformance and its implications for confidence interval construction and hypothesis testing, (iii) it proposes an alternative method to smoothing (the *moments bootstrap*) that seems to perform well in small samples, (iv) it provides suggested guidelines for the application of bootstrap DEA and uses data from the unexplored Greek banking (de)regulation era to perform an empirical illustration.

1.4 Structure of the thesis

The thesis begins with the theoretical explorations on bootstrap DEA and the development of the alternative approach to smoothing and it continues with the suggestions on the application of bootstrap DEA and an empirical illustration of the methods discussed. The structure of the thesis is as follows: chapter 2 introduces, discusses and evaluates the performance of bootstrap DEA; chapter 3 explains how hypotheses could be tested using bootstrap DEA and explains the implications of the violation of fundamental assumptions for the applicability of the hypothesis tests; chapter 4 introduces the *moments-bootstrap* as an alternative method to the smooth bootstrap of Simar and Wilson; chapter 5 suggests guidelines for the implementation of bootstrap DEA and hypothesis testing; chapter 6 performs an empirical illustration on Greek banking, while chapter 7 concludes the thesis, summarises its limitations and proposes areas for future research.

2 Small Samples and Bootstrap DEA: a Monte Carlo Analysis[1]

2.1 Introduction

The analysis of efficiency and productivity can be performed by using either parametric or non-parametric models. Non-parametric models such as data envelopment analysis (DEA) are more flexible since they are free of assumptions about the functional form of the production function or the distribution of inefficiency. In particular, in DEA the user just needs to specify a reasonable input-output system which adequately captures the underlying production processes in the dataset used. On the other hand, it is not possible to apply statistical inference on DEA since it is deterministic (there is no random error to introduce unexplained variability). Recently, Simar and Wilson (1998) addressed this issue by applying the bootstrap on DEA scores. The idea in bootstrapping DEA scores is to evaluate the sensitivity of a decision making unit (DMU) towards changes of the reference set against which its efficiency score is assessed. Hence, a distribution of efficiency scores can be generated for each DMU and it can be used for statistical inference and hypothesis testing. Since Simar and Wilson's (1998) seminal paper, many

[1] This chapter is a revised version of a previous one which was amended according to comments received by Prof L. Simar and Prof P. Wilson at the 13th European Workshop on Efficiency and Productivity Analysis (EWEPA) in Helsinki (17-20 June 2013). All concerns raised by Simar and Wilson have been addressed while mathematical proofs are provided were necessary to illustrate the validity of the approach followed here. I would like to cordially thank both Prof L. Simar and Prof P. Wilson for their valuable feedback and suggestions which significantly enhanced the quality of this chapter and which carry transferable implications for the rest of the thesis. Of course, any remaining errors are the author’s responsibility.

works have followed, involving extensions of the original approach or implementations of the bootstrap on other DEA models.

Bootstrap DEA, as most bootstrap applications, is asymptotically consistent. That is, as the sample size approaches the population size (or theoretically infinity) then all assumptions that make use of the asymptotic properties of the bootstrap are valid. The assumption which is most commonly used in bootstrap DEA[2] is that the bootstrap bias is asymptotically equal to the DEA bias (or model bias)[3]. Based on this assumption, bootstrap DEA could be used to uncover the population or "true" efficiency score of any DMU by correcting twice for bootstrap bias (Simar and Wilson, 1998) or to construct low-variance confidence intervals that centre this "true" efficiency score (Simar and Wilson, 2000a, 1998). In practice the two biases are different and arguably there is no guarantee that this difference is negligible. Sample size can affect the magnitude of the biases and it is therefore worthwhile exploring the performance of bootstrap DEA across various sample sizes: especially smaller ones which are observed in many empirical applications.

Despite the fact that numerous papers have applied these methods (and therefore make use of these assumptions), there is no clear indication of what is considered to be an adequate sample size for various dimensions (number of inputs and outputs). In fact in some applied works of Simar and Wilson there is no comment on whether the sample size meets some "size criteria"; in all cases, though, their sample size at least satisfies

[2] Actually in its implementation: that is, after the bootstrap values have been generated.
[3] The concepts of bootstrap and DEA bias will be properly introduced later in this chapter, along with the required formality.

the “rule of thumb” for DEA applications[4] while in other cases it well exceeds it. For example, Simar and Wilson (1998) include an “illustrative example” in their paper which employs the Färe et al. (1989) data on 19 electric power utilities under a 3-input / 1-output specification. Other examples provided in Simar and Wilson (2008)[5] include the program follow-through application of Charnes et al. (1981) which uses data from 70 schools in a 5-input / 3-output model, and the study of Mouchart and Simar (2002) on European air traffic controllers, which includes 37 units that use one aggregated input variable (resulting from 2 inputs) and one aggregated output variable (resulting from 4 outputs).

The motivation for the examination of the finite sample behaviour of bootstrap DEA can be found in the analysis of these two latter examples. In particular, Simar and Wilson (2008) state in the analysis of the program follow-through study:

> “*Despite the fact that the sample size is rather small in this high-dimensional problem, the confidence intervals are of moderate length.”*
> (Simar and Wilson, 2008; page 467)

Moreover, Simar and Wilson (2008) state for the analysis of the air traffic controllers study:

> *“Due to the small number of observations… inputs were aggregated into a single measure… Outputs were also aggregated into a single measure…”*
> (Simar and Wilson, 2008; page 463)

The authors in these examples seem to acknowledge the issue of the finite sample performance of the bootstrap since the sample sizes were well-above the required ones

[4] The “rule of thumb” states that in order to overcome the issue of dimensionality in DEA, the minimum number of DMUs to be included in the sample should exceed the sum of inputs and outputs multiplied by 3. For example if the total input and output variables are 4, then the minimum sample size is 12.

[5] See page 463, section 4.3.5.5 “Examples” in the referenced book chapter.

implied by the “rule of thumb”. It is therefore important to examine what would be an acceptable sample size under different scenarios.

In this chapter we explore the plausibility of bootstrap DEA in small samples since they are most often met in empirical studies. The preference of DEA over parametric models (such as SFA) when dealing with very small samples is well-known in the literature since DEA performs better in these situations and simulations have shown it (Krüger, 2012; Van Biesebroeck, 2007). We therefore perform Monte Carlo experiments over various dimensions and data generating processes in order to associate minimum sample requirements with specific cases that the applied researcher might deal with. We proceed by exploring the extent to which the aforementioned assumption of equal bootstrap biases applies, we evaluate the performance of bootstrap DEA on the basis of coverage probabilities while we examine the behaviour of the bootstrap distribution and of the associated confidence intervals.

The results of this exercise indicate some cases where bootstrap DEA cannot be safely implemented, especially in finite samples. In particular, we find that in smaller samples the assumption of equal bootstrap and DEA biases is a generous one, while coverage probabilities do not always converge “fast enough” to their nominal values. In larger samples, coverage probabilities do not necessarily increase, but exhibit a clear asymptotic tendency to converge. Comparing the coverage of the confidence intervals of Simar and Wilson (1998) and Simar and Wilson (2000a) under weak conditions the latter perform better only in cases which are not in accordance with good DEA practice. This carries implications for models which make use of these intervals such as the bootstrap Malmquist Index (Simar and Wilson, 1999), tests of returns to scale using bootstrap DEA (Simar and Wilson, 2002), or the more recent and well-known two-stage bootstrap DEA (Simar and Wilson, 2007).

We also find interesting the observation that the width of confidence intervals becomes narrow quite fast; even for a sample size of 200 DMUs. In fact, in larger samples the intervals become so narrow that they almost converge to a certain value. This suggests that the value added in applying bootstrap DEA to test hypotheses in large samples is limited given that we would expect most null hypotheses to be rejected. On the other hand this suggests that DEA scores become more robust towards sampling variations. This further motivates our interest in the small sample behaviour of bootstrap DEA.

In the sections that follow we proceed step by step in introducing the concepts of efficiency and bootstrap DEA (section 2.2) and we provide formal foundations of the theory involved in efficiency analysis (section 2.3) and the methods used to estimate efficiency (section 2.4). Having established the essential knowledge on efficiency analysis we explain the bootstrap in a general setup (section 2.5) and then proceed by analysing bootstrap DEA and its associated technicalities (section 2.6). We then provide general information about Monte Carlo simulations and discuss previous findings (section 2.7), we analyse the methodological aspects of the simulations that we perform (section 2.8), we present the results of the Monte Carlo simulations (section 2.9) while we also perform the same exercise using large samples (section 2.10). Finally, we conclude the chapter, discussing the implications of our results and suggesting areas for future research (section 2.11).

2.2 General concepts

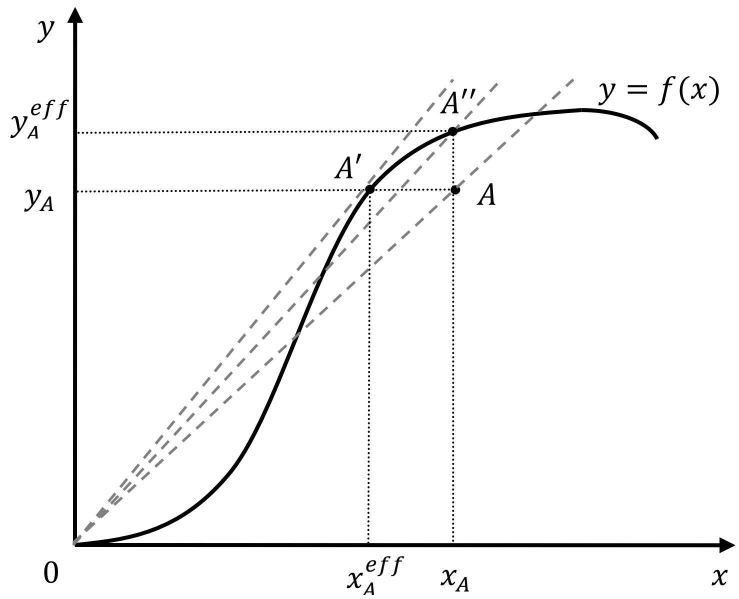
Before we begin our analysis, some informal definitions and discussion are necessary to ease the exposition of the technical material that follows. An excellent introduction to the concepts of efficiency and productivity can be found in Coelli et al. (2005) which we follow in this section. The discussion will employ Figure 2.1, which resembles Figure 1.2 in Coelli et al (2005; pp 5) and which presents a production frontier in a one-input/one-output setup. The intuitive interpretation of the frontier is that it suggests the maximum possible output (y) that can be produced using a certain level of input (x) and with the available production processes and technology captured by the production function $(f(\cdot))$. All the input-output combinations on and below the frontier comprise the feasible set, whereas combinations above the frontier are not technologically feasible.

Figure 2.1 also illustrates firm A which operates below the frontier and is therefore "technically" inefficient. To become efficient (and hence operate on the frontier) it could "technically" contract its input towards point A' (input orientation) or expand its output towards point A'' (output orientation). In this example, technical efficiency in input (output) orientation can be measured as the ratio of the efficient level of input (output) divided by the actual input (output).

Productivity is defined as the ratio of outputs over inputs, which is also known as the average product. Graphically, it is represented by the slope of the ray from the origin to any point of interest, which is depicted by the dashed lines in Figure 2.1. This also shows that changing the scale of operations leads to different levels of productivity. In fact, point A' is associated with the maximum average product (maximum productivity) in this example, which is known as the most productive scale size (MPSS) or the point of technically optimal scale (TOPS). It is worthwhile noting that under output orientation, the projection to point A'' is not associated with MPSS, suggesting that there is room for further improvement in the productivity of firm A by exploiting scale economies. This leads to an important clarification: technical efficiency does not necessarily imply scale efficiency (the extent to which a firm operates under the MPSS). It also suggests that the operations of a firm can be improved by both becoming more technically inefficient and by exploiting scale economies (at least in this example) [6].

[6] For more information on the issue of productivity and its association with RTS and scale of operations see Banker et al. (2004).

Figure 2.1. Technical efficiency, productivity and scale operations



Source: adopted and extended Figure 1.2 in Coelli et al (2005; pp 5)

There is also a time component in the analysis of efficiency and productivity, which has not been mentioned thus far. This relates to the fact that over time technical efficiency, scale of operations and technology might change, leading to respective changes in productivity. The first two sources of productivity change are known as efficiency change and scale efficiency change whereas the last one is known as technical change and it is associated with shifts of the frontier (technical progress or regress). Index number approaches (such as the Malmquist index) have been developed to measure changes in productivity and its components.

In a one-input/one-output setup one could perform computations related to efficiency and productivity even manually. However, when the dimensions increase it is necessary to employ appropriate methods to aggregate inputs in a single "index of inputs" and outputs in a single "index of outputs" to perform the necessary computations (Coelli et al., 2005). These methods are both parametric and non-parametric with the most popular ones being "Stochastic Frontier Analysis" (SFA) from the parametric family and "Data Envelopment Analysis" (DEA) from the non-parametric one. DEA is the method that is employed throughout this study while its technical details are discussed in section 2.4.3.

DEA was introduced by Charnes et al. (1978) and uses linear programming principles. In the original paper Charnes et al. (1978) propose as a measure of "technical efficiency":

> "*the maximum of a ratio of weighted outputs to weighted inputs subject to the condition that the similar ratios for every DMU be less than or equal to unity*" (Charnes et al., 1978; pp.430)

They then transform this fractional program into two linear ones (one being the dual to the other) known as the "envelopment" and "multiplier" forms and which will be discussed in more detail later. The intuition in DEA is that the technical efficiency of a DMU is computed with respect to a piece-wise linear frontier which is constructed using the available data and it is therefore a measure of relative efficiency (relative to the DMUs in the sample). Perhaps the greatest advantage of DEA is that it does not require the specification of a functional form of the production function, though at the cost of being deterministic and therefore not suitable for statistical inference.

Applying the bootstrap on DEA (Simar and Wilson, 1998), or bootstrap DEA as it is commonly called, offers a solution to this issue. The DEA score of a DMU is deemed as a sample "estimate" of its population value (or "true" as termed here), suggesting that the estimated DEA score is sensitive towards sampling variations. The random resampling in the bootstrap DEA process can be considered as simulating these sampling variations. This allows extracting a distribution of bootstrapped efficiency scores for each DMU which can be used to construct confidence intervals where their “true” (or population) efficiency scores lie. The bootstrap confidence intervals can be used to test various hypotheses. For example, in the illustrative example in Simar and Wilson (1998), the authors use the constructed confidence intervals to compare the technical efficiency between electric utility firms by observing the overlap of the constructed intervals.

2.3 Theoretical foundations

In this section we formally introduce some concepts relevant to efficiency and productivity analysis. Several authors have provided an excellent and rigorous treatment of these concepts (Fried et al., 2008; Mas-Colell et al., 1995; Shepard, 1970; Varian, 1992) on which we base our exposition here, while maintaining where possible the same notation as in Simar and Wilson (1998).

The starting point is the definition of a feasible set (or production set, or technology set) which is the set of possible input-output combinations with a given technology (Mass-Colell et al., 1995; Fried et al., 2008). Let us denote with x the vector of p inputs and with y the vector of q outputs. The feasible set Ψ is then:

$$\Psi = \left\{(x, y) \in \mathbb{R}_{+}^{p+q} | \text{x can produce y}\right\} \tag{2.1}$$

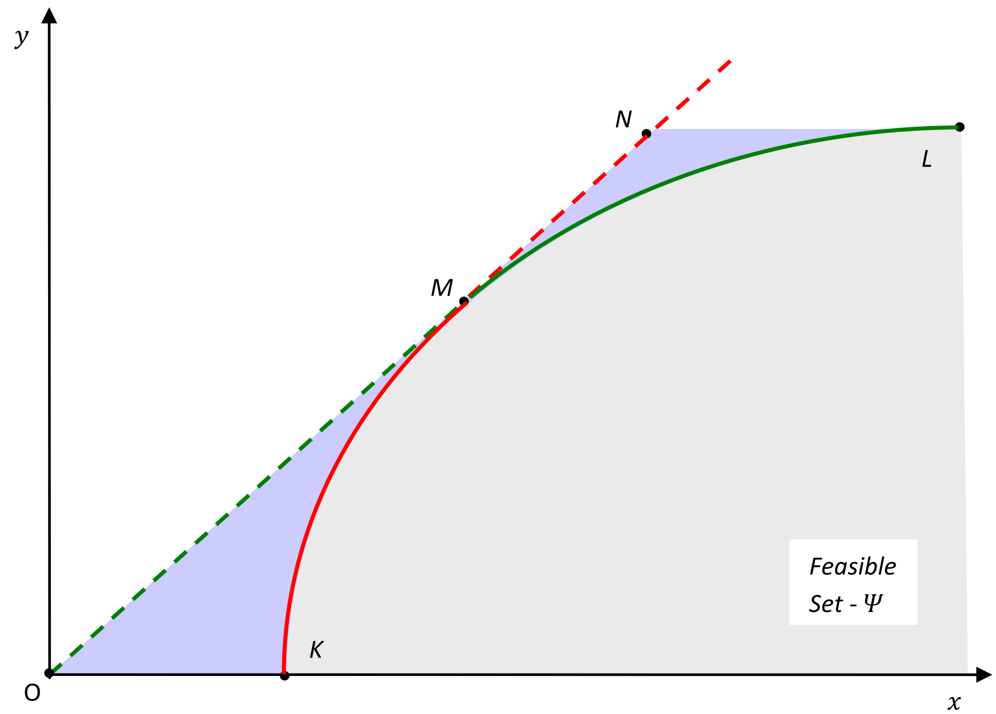
An elaborate, yet not exhaustive, account of the properties of production sets can be found in Mas-Colell et al. (1995)[7]. We highlight the importance of the convexity assumption which suggests that a linear combination between any two points should lie within the feasible set. With reference to the simple example in Figure 2.1 the feasible set can only include the combinations on and below the concave part of the production function. Moreover, the assumption of free disposal implies that more inputs can be used without any reduction in outputs: otherwise the extra inputs (or outputs) would be disposed of at no cost. Again with reference to Figure 2.1, the part of the frontier that bends backwards violates the assumption of free disposal.

Figure 2.2 below represents what Coelli et al. (2005) refer to as "the economically feasible region of production" under 4 different assumptions on technology, while being consistent with the aforementioned properties of feasible sets. In particular, in this simple 1-input/1-output setup, section OMN presents a production frontier that exhibits constant returns to scale (CRS) while the section KML presents a frontier associated with variable returns to scale (VRS). Finally, the sections OML and KMN correspond to frontiers that exhibit non-increasing (NIRS) and non-decreasing (NDRS) returns to scale, respectively. The areas on and below these sections determine the

[7] In particular they list 12 assumptions commonly used but not in combination as some may be mutually exclusive. Also some of the assumptions can be dropped depending on the analysis. The interested reader may refer to pages 130-135 in Mas-Colell et al. (1995) for a full description of these properties, which we also list here for reference: (i) the set is non-empty, (ii) the set is closed, (iii) no free lunch, (iv) possibility of inaction, (v) free disposal, (vi) irreversibility, (vii) non-increasing returns to scale, (viii) non-decreasing returns to scale, (ix) constant returns to scale, (x) additivity, (xi) convexity, and (xii) the set is a convex cone. Shepard (1970) and Varian (1992) also provide an account of these properties.

feasible set in each case. It is quite straightforward to see that the feasible set serves, among others, as a representation of the production technology.

Figure 2.2. Economically feasible sets



An alternative representation of technology is through what is known as the transformation function. The transformation function $T(x, y)$ has the property (Mas-Colell et al., 1995):

$$\Psi = \left\{(x, y) \in \mathbb{R}_{+}^{p+q} \mid T(x, y) \leq 0\right\} \tag{2.2}$$

If $T(x, y) = 0$ then the corresponding input/output combinations would lie on the "transformation frontier", while a special case of the transformation frontier is the production function or frontier for $q = 1$, that is one output (Coelli et al. 2005).

We can now define technical efficiency with respect to the feasible set[8]. Koopmans (1951) stated that a firm is technically efficient if an increase in any output requires the reduction of at least another output or the increase of at least one input. Also, a firm is technically efficient if a reduction in one input is necessarily accompanied by an increase in at least another input or a reduction in at least one output. Debreu (1951) and Farrell (1957) proposed a radial measure of technical efficiency. In particular, in input orientation technical inefficiency is the proportional reduction of all inputs that would set a firm technically efficient (keeping outputs fixed), while in output orientation it is the required proportional expansion of all outputs (keeping inputs fixed).

Two alternative representations of the feasible set which are associated with the input and output orientations are those of the input requirement set and of the output correspondence set. The input requirement set includes the vector of inputs required to produce a certain level of outputs:

$$X(y) = \{x \in \mathbb{R}_{+}^{p} \,|\, (x, y) \in \Psi\} \tag{2.3}$$

while the output correspondence set includes the vector of outputs that are possible to be produced by (or correspond to) a certain vector of inputs:

$$Y(x) = \{y \in \mathbb{R}_{+}^{q} \,|\, (x, y) \in \Psi\} \tag{2.4}$$

The boundaries of $X(y)$ and $Y(x)$ are in fact the same, but the movement towards the frontier invites different interpretations, with regards to the two orientations.

Using the notation in Simar and Wilson (1998), who follow the analysis of Shepard (1970), we could define the Debreu-Farrell boundary of $X(y)$ as follows:

[8] We follow Fried et al. (2008) here – see page 20-21.

$$\partial X(y) = \{x | x \in X(y); (\theta x, y) \notin X(y) \;\; \forall\, \theta \in [0,1)\} \tag{2.5}$$

and the boundary of $Y(x)$ as:

$$\partial Y(x) = \{y | y \in Y(x); (x, \eta y) \notin Y(x) \;\; \forall\, \eta \in (1, \infty)\} \tag{2.6}$$

A moment's reflection will make clear that $\partial X(y)$ represents an isoquant while $\partial Y(x)$ represents a production possibility frontier (Fried et al., 2008). The intuition behind the notation for $\partial X(y)$ ($\partial Y(x)$) is that any radial contraction (expansion) of inputs (outputs) with the same output (input) levels would not be a member of these boundary sets.

Focusing on input orientation, the Debreu-Farrell technical efficiency for firm k is defined as:

$$\theta_k = \theta(x_k, y_k) = min\{\theta |\, \theta x_k \in X(y_k)\} \tag{2.7}$$

while the efficient level of input is determined by:

$$x^{\partial}(x_k | y_k) = \theta_k x_k \tag{2.8}$$

It is straightforward that if firm k is technically efficient, then $\theta_k = 1$ while if it is technically inefficient then $0 < \theta_k < 1$. Also from (2.8) we see that if firm k is technically inefficient it should use a fraction θ_k of its inputs (or contract its inputs by $1 - \theta_k$) in order to become technically efficient in the Debreu-Farrell sense.

To calculate technical efficiency scores various methods have been proposed and developed; parametric and non-parametric. These are reviewed in the next section, but the main focus is on DEA which is employed in this study.

2.4 Estimation of technical efficiency

The calculation of technical efficiency is straightforward in the simple case of a single input and a single output. However, in higher dimensions these computations can only be performed with the use of relevant parametric and non-parametric techniques.

2.4.1 Parametric approaches

Parametric models involve specifying a production function while inefficiency for each firm is estimated by the appropriate decomposition of the error term of the estimated function (most commonly a cost function) into a random component and an inefficiency component. In the case of multiple outputs, aggregators or appropriate distance functions are used, initially outlined by Shepard (1970). Despite the restrictions imposed by the specification of a production function, parametric models have the advantage of distinguishing the various sources of randomness (measurement error, specification error, etc) from inefficiency (Bauer et al., 1998). The most common parametric models used include the stochastic frontier approach (SFA), the thick frontier approach (TFA) and the distribution free approach (DFA).

In the stochastic frontier approach (SFA), which was introduced by Aigner et al. (1977) and Meeusen and Van den Broeck (1977), the random component of the error term is assumed to follow a symmetric distribution while the inefficiency-related component is assumed to follow an asymmetric distribution. However, as it is pointed out in Bauer et al. (1998) and the therein references, the inefficiencies calculated are sensitive towards the choice of the latter distribution. They also argue that:

> "*…any distributional assumptions simply imposed without basis in fact are quite arbitrary and could lead to significant error in estimating individual firm efficiencies.*" (Bauer et al, 1998; pp.94)

The thick frontier approach (TFA) was proposed by Berger and Humphrey (1992) to measure the efficiency of US commercial banks. TFA uses the same functional form for the frontier as SFA, but the regression is based on the firms with the lowest average costs for each (predetermined) size class. Differences among firms within the same size class are perceived to be random while differences among groups are perceived as inefficiency. The major disadvantage of this method, apart from the ones that apply to SFA and are common, is that the results are not inefficiency scores but estimated values of inefficiency differences.

The distribution-free approach (DFA), introduced by Berger (1993), uses a functional form as with SFA and TFA but without imposing restrictions on the distribution of the random error or inefficiency. It is based on panel data techniques where a constant level of efficiency is assumed for each firm over time and any deviations about this average level are attributed to randomness. DFA shares the same disadvantages with SFA, plus the fact that, due to the nature of panel data analysis, the efficiency estimates concern the entire period under consideration and not each year separately.

2.4.2 Non-parametric approaches

Non-parametric models benefit from being flexible as there is no need to specify a functional form for the production function. The user assumes an unobserved transformation or production process where a set of inputs produces a set of outputs and the frontier is constructed on the basis of the observed data. The disadvantage of non-parametric models is that any measurement or specification errors are incorporated in the estimated inefficiency, which explains the lower scores of non-parametric models compared to parametric ones (Bauer et al., 1998).

The two most popular techniques are the data envelopment analysis (DEA), introduced by Charnes et al. (1978), and the free disposal hull (FDH), introduced by Deprins et al. (1984), both of which belong to the broad category of non-parametric hull models. The fundamental difference between the two methods lies in the convexity assumption used by DEA, which is not adopted in FDH.

2.4.3 Data envelopment analysis

The definition of a Pareto-Koopmans efficient firm or *decision making unit* (DMU) under the scope of DEA is:

> "*A DMU is fully efficient if and only if it is not possible to improve any input or output without worsening some other input or output*" (Cooper et al., 2006; pp.45)

Data envelopment analysis (DEA), as already mentioned, is a non-parametric technique introduced by Charnes et al. (1978), which uses linear programming principles to compute efficiency scores of *decision making units* (DMUs). Their initial proposed measure of technical efficiency for DMU k (see pp.42 in section 2.2 for definition) could be described by the following fractional program:

$$\hat{z}_k = \max\left\{ z = \frac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{s=1}^{p} v_s x_{sk}} \middle| \frac{\sum_{s=1}^{p} v_s x_{si}}{\sum_{r=1}^{q} u_r y_{ri}} \geq 1;\ v_s, u_r \geq 0;\ \ \forall\, i = 1, \dots, n \right\} \tag{2.9}$$

where p is the number of inputs (x) that DMU k uses and q the number of outputs (y), while v_s and u_r are the weights on the s^{th} input and r^{th} output which will be determined by the solution of this problem and which will be used to compute the technical efficiency score of DMU k.

Charnes et al. (1978) transformed the fractional program in (2.9) into a linear one as follows:

$$\hat{\theta}_k = \max\left\{ \theta = \sum_{r=1}^{q} \mu_r y_{rk} \,|\, \sum_{r=1}^{q} \mu_r y_{ri} \leq \sum_{s=1}^{p} \nu_s x_{si}\,;\ \sum_{s=1}^{p} \nu_s x_{sk} = 1;\ \nu_s, \mu_r \geq 0;\ \ \forall\, i = 1, \dots, n \right\} \tag{2.10}$$

where $\mu_r = \left(\sum_{s=1}^{p} \nu_s x_{sk}\right)^{-1} u_r$ and $\nu_r = \left(\sum_{s=1}^{p} \nu_s x_{sk}\right)^{-1} v_r$. The linear program (2.10) computes the input oriented technical efficiency score for DMU k and it is also known as the "multiplier form". Its dual linear program is:

$$\hat{\theta}_k = \min\left\{ \theta \middle| y_k \leq \sum_{i=1}^{n} \lambda_i y_i\,;\ \theta x_k \geq \sum_{i=1}^{n} \lambda_i x_i\,;\ \theta > 0; \lambda_i \geq 0,\ \ \forall\ i = 1, \dots, n \right\} \tag{2.11}$$

which returns the same result as in (2.10) and it is known as the "envelopment form".

These two linear programs are known as the CCR model (from the initials of the authors Charnes, Cooper and Rhodes), while they are also known as the CRS (constant returns to scale) model[9]. The latter is due to the fact that the resulting boundary facets (the frontier) form a convex cone on which only (efficient) firms which exhibit CRS lie. For example, in the simple 1-input/1-output case the frontier is a straight line from the origin and through the DMU with the highest average product (output to input ratio), which is also deemed as exhibiting CRS.

It seems useful to provide a graphical illustration of how DEA works in input orientation (an assumption adopted throughout this study) and how the multiplier form is related to the envelopment one. Perhaps the best way to do this is to consider the example in Figure 2.3 which is an extension of Fried et al (2008; pp.48). In this 2-input/1-output example each DMU uses inputs $x_1$and x_2 to produce 1 unit of output y (let us denote it y_0). DMUs B, C, D and E use input vectors x^B, x^C, x^D, x^E to produce y_0, forming a piecewise linear frontier (which in fact is an isoquant). DMU A is inefficient as it uses vector x^A to produce y_0, which involves proportionately more inputs. To be efficient, DMU A should be producing y_0 using $\theta^A x^A$, $\theta^A \in (0,1)$. This radial reduction in inputs is graphically represented in Figure 2.3 by the projection of x^A onto the frontier along the ray from the origin to x^A. This projection intersects the frontier

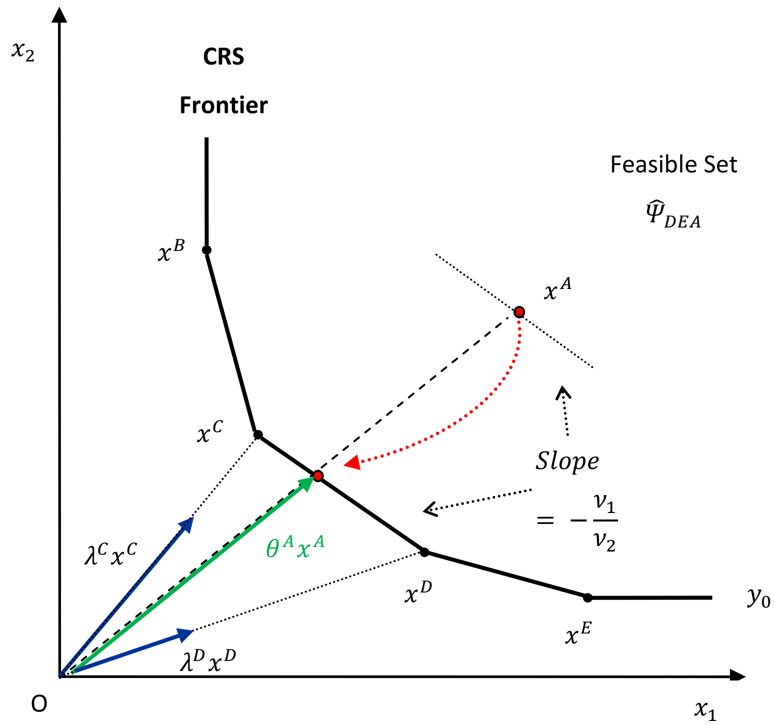
[9] Banker et al. (1984) developed what is known as the BCC or VRS model which allows for DMUs on the frontier to exhibit variable returns to scale. In this case the multiplier form becomes $\hat{\theta}_k = max\ \{\theta = \sum_{r=1}^{q} \mu_r y_{rk} - \mu_k | \sum_{r=1}^{q} \mu_r y_{ri} \leq \sum_{s=1}^{p} v_s x_{si} - u_k;\ \sum_{s=1}^{p} v_s x_{sk} = 1;\ v_s, \mu_r \geq 0;\ \ \forall\, i = 1, \dots, n\}$, where μ_k is called the slope parameter and introduces concavity on the frontier. For the envelopment form one just needs to add the following convexity constraint in (2.11): $\sum_{i=1}^{n} \lambda_i = 1$.

through the linear section defined by x^C and x^D, suggesting that DMUs C and D serve as benchmarks for DMU A in this example.

The multiplier and envelopment weights both have an economic interpretation. Let us first consider the multiplier model in (2.10) and in particular the constraint $\sum_{s=1}^{p} v_s x_{sk} = 1$, which, if adapted to our case for DMU $k = A$, we have: $v_1 x_{1A} + v_2 x_{2A} = 1$. It is straightforward to graphically represent this constraint as $x_{2A} = \frac{1}{v_2} - \frac{v_1}{v_2} x_{1A}$ which is parallel to the frontier section defined by x^C and x^D, which are the benchmarks for DMU A. Since $-\frac{v_1}{v_2}$ is the slope of the isoquant/frontier, it can be interpreted as the marginal rate of technical substitution between inputs x_1 and x_2 at the projection of DMU A on the frontier.

The envelopment form in (2.11) determines the exact position of $\theta^A x^A$ on the frontier by using a convex combination of x^C and x^D, so that $\lambda^C x^C + \lambda^D x^D = \theta^A x_A$. This vector is represented in Figure 2.3 by the green arrow. The envelopment weights λ_C and λ_D can be thought of as the proportion of the inputs of DMU C (λ_C) and DMU D (λ_D) that DMU A needs to use in order to become technically efficient. Given that C and D are the benchmarks, then $\lambda_A = \lambda_B = \lambda_E = 0$. Also if, for example, $\lambda_C = 1$ and $\lambda_D = 0$, then vectors x^A, $\theta^A x_A$ and x^C would necessarily lie on the same ray. We could state that for DMU A the multiplier weights have defined the slope of (the section of) the frontier against which it is benchmarked, while the envelopment weights have defined the exact position of its projection on the frontier.

Figure 2.3. Illustration of DEA in input orientation



Linking DEA with the theoretical foundations in the previous section, we can show how the feasible set defined in (2.1) is estimated by DEA. Using the envelopment form we have:

$$\widehat{\Psi}_{DEA} = \left\{ (x, y) \in \mathbb{R}_{+}^{p+q} | y \leq \sum_{i=1}^{n} \lambda_i y_i \, ; \, x \geq \sum_{i=1}^{n} \lambda_i x_i \, ; \, \lambda_i \geq 0, i = 1, \ldots, n \right\} \tag{2.12}$$

The input requirement set and its boundary (which is the estimated frontier under input orientation), would be the same as in (2.3) and (2.4) but replacing Ψ with $\widehat{\Psi}_{DEA}$, $X(y)$ with $\hat{X}(y)$, and $\partial X(y)$ with $\widehat{\partial X}_{DEA}(y_k)$.

2.4.4 The DEA "estimators"

The sample DEA score of a DMU is an estimator of its population value, since it is conditional on the reference set against which it is assessed. Hence, the DEA score is subject to sampling variations and the difference between the sample estimate and the population or "true" value is called model or DEA bias[10]. This bias is in principle positive since the sample DEA score will almost always be higher than the population one (Simar and Wilson, 1998). The latter is attributed to the fact that the probability of all population-efficient DMUs appearing in a finite sample is extremely low[11]. Formally, $\widehat{\Psi}_{DEA} \subseteq \Psi$ and therefore:

$$\widehat{\partial X}_{DEA}(y_i) \subseteq \partial X(y) \Leftrightarrow 1 \leq \theta_k < \hat{\theta}_k \tag{2.13}$$

It becomes apparent that $\hat{\theta}_k$ is an estimator of θ_k which has a distribution attributed to the aforementioned sampling variations.

One topic of interest is the behaviour of the DEA or model bias $\hat{\theta}_k - \theta_k$ with respect to sample size changes. The faster $\hat{\theta}_k - \theta_k$ converges to zero, the higher is said to be its "speed of convergence". Moreover, the consistency[12] of the DEA estimators depends upon their asymptotic convergence, that is $\lim_{n\to\infty}\left(\hat{\theta}_k - \theta_k\right) = 0$, where n is the number of DMUs in the sample.

---

[10] Simar and Wilson (1998) suggest that the unobserved DEA bias could be approximated by bootstrap DEA, a statement that is explained in section 2.6.

[11] It seems worthwhile noting here that the notion of population used by Simar and Wilson (1998) and in this study would be more accurately termed as "super-population". The difference is that the super-population includes theoretically feasible input-output combinations which are not necessarily members of the population and are infinite in number.

[12] Consistency requires that $\hat{\theta}_k$ converges in probability towards θ_k, in that as sample size approaches infinity, the probability $P\left(\left|\hat{\theta}_k - \theta_k\right| < \varepsilon\right) \to 1, \forall\, \varepsilon > 0$ as sample size approaches infinity.

Theoretical studies on the convergence speed of DEA estimators focus on deriving expressions on convergence rates and other asymptotic properties that they possess. These expressions provide a general idea of the effect of sample size on speed of convergence as the latter is expressed as a function of the number of inputs and outputs, the number of DMUs, while it differs depending on the technology assumption used (CRS, VRS or other).

Banker (1993) proves for the VRS, 1 input and 1 output case, that DEA scores of the monotone concave production frontier are asymptotically consistent and they are actually the maximum likelihood estimators of the DEA model. Korostelev et al. (1995) explore the statistical foundations of DEA estimators under VRS and derive theoretical expressions about their speed of convergence for the case of one input and multiple outputs. Their analysis was extended by Kneip et al. (1998) to the general case of multiple inputs and multiple outputs, again under VRS. Recently, Kneip et al. (2008) derive the asymptotic distribution of DEA estimators under VRS for the multiple input and output case. For a further review on this issue the interested reader may refer to Simar and Wilson (2008, 2004, 2000b).

The common conclusion of these studies is that as the dimensions increase (number of inputs and outputs) an exponentially larger data set is required in order to achieve the same accuracy and convergence as with smaller dimensions. Monte Carlo simulations can provide some evidence on the behaviour of the convergence of DEA estimators towards their population values. We will show in our simulations later in this chapter that convergence, apart from the number of inputs and outputs and sample size, is also affected by the assumed data generating process (DGP).

2.5 General information about the bootstrap

The bootstrap, introduced by Efron (1979) and further explored by Efron and Tibshirani (1993)[13], can be used to produce multiple pseudo-samples by resampling with replacement from the empirical distribution of a set of observations. It is an attractive tool in cases where statistical inference is difficult (if not impossible), as the bootstrap distributions can be used to compute quantities of interest, as well as to perform hypothesis testing. The validity of the bootstrap depends on the ability of the process to mimic the data generating process (DGP) of the unobserved population. If we assume that the sample is a "representative" one, then the properties of the population should be reflected in the properties of the sample and therefore the bootstrap should yield meaningful results. In particular, if the moments of the empirical distribution are similar to the moments of the population distribution, the bootstrap will perform well as the bootstrap samples will have the same properties as if they were drawn directly from the population.

[13] The bootstrap is based on a series of properties analyzed in Efron and Tibshirani (1993), the most important of which is that the empirical distribution function should be a good approximation of the actual distribution function of the population.

Bootstrapping within a model framework follows a similar logic. A model uses a structure to compute or estimate of quantities interest. For example, in the regression framework, a model such as OLS is used to estimate the coefficients $(\hat{\beta})$ of the independent variables $(\boldsymbol{x})$ which can be used to compute the expected value of the dependent variable $(\hat{y} = \boldsymbol{x}\hat{\beta},\ \text{or}\ \ E(y|x) = \boldsymbol{x}\hat{\beta}\ )$. The deviations of y from $E(y|x)$ are called residuals $(\hat{\varepsilon} = y - \boldsymbol{x}\hat{\beta})$ and should be normally distributed. Bootstrapping the OLS estimators can be done in two ways: either by bootstrapping pairs of observations (also called "case resampling") or by bootstrapping residuals (also called "fixed resampling" as $\boldsymbol{x}$'s remain unchanged in each iteration). The bootstrap would enable us in this case to extract the distribution of the model's parameters (the betas) and examine, for example, whether they are significantly different from some predetermined value. The source of variability is assumed to be the random distribution of regression residuals and the bootstrap is implemented by reallocating residuals (or deviations from the regression line) among sample observations and regressing again to obtain a new set of parameters[14].

One of the most important issues in bootstrapping models is to identify the source of variability and apply the bootstrap accordingly. For example, if the source of variability seems to be the unconditional distribution of residuals (where $\boldsymbol{x}$ is not correlated with the residuals), it would be preferable to bootstrap residuals. However, if the model's parameters are sensitive towards sampling variations, it would be preferable to

[14] The analysis in this paragraph and terminology used follows Stine (1989) who provides an intuitive and thorough introduction to the bootstrap.

bootstrap pairs (Stine, 1989). Due to the strong assumptions in bootstrapping residuals (residuals have to be uncorrelated with independent variables), this approach is more sensitive to model assumptions compared to bootstrapping pairs; however, they should asymptotically provide similar results (Efron and Tibshirani, 1993).

Another important concept associated with bootstrapping is that of the bootstrap bias and of the model bias. The bootstrap bias is the difference between the bootstrap mean and the model's estimated parameter(s) whereas the model bias is the difference between the estimated parameters and their "true" value or population value. The bootstrap bias occurs (to a large extent) due to the randomness in the resampling process. Therefore increasing the number of bootstrap replications reduces the randomness element in the bootstrap bias and the remaining bias is due to other factors such as sampling variations[15]. The model bias occurs due to sampling variations but it can also be caused by model misspecification or measurement errors. The bootstrap should converge faster if the sampling variations are trivial (i.e. if any randomly selected sample is fairly representative) and if there are no other errors. In the presence of the specification or measurement errors, the bootstrap will not necessarily fail (as it will still reproduce the observable variations of the empirical distribution), but results might not be as meaningful.

---

[15] In fact, other methods such as subsampling or the m out of n bootstrap (either with replacement or not) might be more suitable in cases where the estimated parameters depend on the sample size. However these methods require large samples and tend to work better asymptotically. For more details see Politis et al. (1999) and Bickel et al. (1997).

Hence, if there are no such errors and if the sample is a representative one, then the bootstrap bias can approximate the model bias. More generally and formally, if the estimated data generating process $(\hat{\mathcal{P}})$ is a consistent estimator of the true one $(\mathcal{P})$, then the estimated bias should have similar distribution to that of the true bias:

$$\widehat{bias}|\hat{\mathcal{P}} \sim bias|\mathcal{P} \tag{2.14}$$

This assumption has important implications in the bootstrap world as it is used to construct confidence intervals. Asymptotically this assumption becomes a property as both biases converge to zero since the estimated (model) parameters approximate the true ones. However, the finite validity of this assumption is of interest and practical value and it can be explored with Monte Carlo simulations.

2.6 Bootstrapping DEA efficiency scores

In this section we provide more information about bootstrapping DEA efficiency scores. Bootstrap DEA was first introduced by Simar and Wilson (1998) who used it to extract the sensitivity of DEA efficiency scores towards “sampling variations”. We introduce the logic of applying the bootstrap within the DEA framework, we then explain in more detail the method and we comment on the recent developments on bootstrap DEA and extensions.

2.6.1 Bootstrap DEA: a practical consideration

The principles of bootstrapping within the model framework also apply in DEA. In particular, in DEA the source of variability is the distribution of (in)efficiency scores, while the estimated parameters are the efficiency scores of the DMUs in the sample. Simar and Wilson (1998) introduce bootstrap DEA where efficiency scores are resampled rather than input-output combinations (although the latter is also possible). To this end, one could loosely associate Simar and Wilson's approach to that of fixed resampling in the previous section.

Similar to the residual resampling, under bootstrap DEA one effectively resamples DEA scores and applies DEA repeatedly, keeping outputs fixed (assuming input orientation). This raises, though, an issue which has not been mentioned in the literature. In particular, the random resampling of efficiency scores suggests that any DMU in the sample could achieve any of the observed efficiency scores. Hence, bootstrap DEA implicitly assumes that any bootstrap replication yields pseudo-inputs which are members of the feasible set.

The latter point will become clearer after the mathematical exposition in the next section, but let us first consider an intuitive example. Suppose DEA is applied to a set of DMUs under CRS and input orientation. The sample comprises one "super-star", a few relatively efficient DMUs and quite a few substantially inefficient DMUs. Graphically this is associated with a histogram of efficiency scores with a thin tail towards 1. Applying bootstrap DEA on this dataset means that the efficiency scores are randomly reallocated to each DMU through the resampling process. It is possible in some replications that a poor performer will be allocated with an efficiency score of 1, suggesting that it would have been possible for this DMU to operate efficiently. If the poor performer can indeed drastically reduce its inputs and still produce the same outputs then the bootstrap will yield meaningful results.

Practically, this simply suggests that bootstrap DEA scores will be meaningful as long as the DEA scores suggest input contractions which could have been achieved contemporaneously[16]. On the contrary, if we believe that the suggested input contractions are counterintuitive (if not non-feasible), bootstrap DEA might not be a good idea to use. This is because bootstrap DEA automatically assumes that any DMU could achieve any efficiency score. In such a case one should also explore the reasons why the “super-star” performs so well: is it because of the excellent management practices followed or is it due to access to superior technology which allows the production of outputs with considerably less inputs? We will refer to this case as the “technologically heterogeneous” case and we will investigate its implications for bootstrap DEA in our simulations later in this chapter.

We should clarify at this point that even in the presence of technological heterogeneity, bootstrap DEA will still be consistent. That is, as the sample size approaches infinity the bootstrap will replicate the behaviour of the population. The

[16] By specifying “contemporaneously” we want to make clear that the notion of feasibility relates to the present and not to potential improvements in the future. If this is not the case then it would be counterintuitive to use bootstrap methods as effectively the resampling process would suggest that the improvement in performance would have been feasible. This point will become clearer in our simulations as we include a case which violates this principle.

consistency of bootstrap DEA is well-established in the literature (Kneip et al., 2011, 2008), but it is important for the applied researcher to ensure that it is practically meaningful to apply these methods and avoid counter-intuitive interpretations. It is not within the scope of this study to propose methods to identify technologically heterogeneous DMUs and classify them as outliers; in fact we believe that this should be done on a one-by-one basis using experts' knowledge. We merely suggest that one should be aware of the implications of including such DMUs in the sample for the implementation of bootstrap DEA.

2.6.2 The Simar and Wilson's (1998) bootstrap DEA algorithm

The principle of bootstrap DEA is to generate various reference sets which would produce a distribution of efficiency scores for each DMU in the sample. The first step in implementing the algorithm of Simar and Wilson (1998) is effectively to smooth the empirical distribution of DEA efficiency scores $(\hat{\theta})$; however, the smoothing process is complicated and it might not be clear from the first instance what is actually being smoothed[17]. Then pseudo-efficiency scores (θ^*) are drawn with replacement from the smoothed distribution and, assuming input orientation, a new set of pseudo-inputs (x^*) is obtained by dividing the original efficient input levels $(\hat{\theta}x)$ by θ^*. Finally, the bootstrapped efficiency scores are computed by applying DEA on the original data but

[17] This is done to avoid repeated values showing up in bootstrap loops. More explanations are provided in the next subsection while the smoothing process is analyzed in Appendix I.

using as a reference set the pseudo-inputs and original outputs (x^{*}, y). This procedure is repeated B times and the resulting distribution of bootstrapped DEA scores can be used for statistical inference.

Let us introduce some formality now and assume that in a CRS setup, inputs (x) and outputs (y) are generated by a process $\mathcal{P}$, which depends on the true attainable set and the joint probability density function $f(x, y)$ of inputs and outputs (Simar and Wilson, 2000b):

$\mathcal{P} = \mathcal{P}\big(\Psi, f(x, y)\big)$, where (2.15)

$f(x, y) = f(x|y)f(y)$ (2.16)

It is clear that we can write the joint *pdf* of inputs and outputs as the conditional *pdf* of inputs on outputs, multiplied by the unconditional *pdf* of outputs: this is the case of input orientation. Straightforward interpretation of Simar and Wilson (2000a) implies that in the case of the "homogeneous bootstrap" (as they named bootstrap DEA in their 1998 paper), output is observed with certainty in input orientation, so $f(y) = 1$ and $f(x, y) = f(x|y)$ and the assumed true DGP is:

$\mathcal{P} = \mathcal{P}\big(\Psi, f(x|y)\big)$ (2.17)

Simply, (2.17) tells us that the DGP will produce input-output combinations which belong in the feasible set, using a *pdf* of inputs conditional on outputs but not depending on the distribution of outputs. Since we observe only a sample derived from the underlying population, the DEA attainable set is a subset of the true one and it is defined by the restrictions of the DEA linear program. Thus, the DGP under DEA, $\hat{\mathcal{P}}$ is:

$$\hat{\mathcal{P}} = \hat{\mathcal{P}}\left(\widehat{\Psi}_{DEA}, f_{DEA}(x|y)\right) \tag{2.18}$$

The steps followed in Simar and Wilson (1998) to obtain the bootstrapped efficiency scores and the maths involved are quite straightforward. Again, we assume a CRS frontier technology and we focus on input orientation:

1. Use observed inputs and outputs to estimate DEA efficiency scores

$$\hat{\theta}_i, i = 1,2 \dots n \tag{2.19}$$

2. Use the procedure in Appendix I to smooth the empirical distribution of efficiency scores
3. Generate a sample of pseudo-efficiency scores from the smoothed distribution:

$$\theta_i^*, i = 1,2 \dots n \tag{2.20}$$

4. In each bootstrap replication b, generate a pseudo-sample $\mathcal{X}_b^* = (x_i^*, y_i)_b,\ i = 1,2, \dots n$ where x_i^* is:

$$x_i^* = \frac{\hat{x}^\partial(x_i|y_i)}{\theta_i^*} = \frac{\hat{\theta}_i x_i}{\theta_i^*},\ \ i = 1,2 \dots n \tag{2.21}$$

5. Compute the bootstrapped efficiency scores $\left(\hat{\theta}_k^*\right)$ for a firm k using the initial input-output values (x_k, y_k) and as a reference set $\mathcal{X}_b^*$[18].

$$\hat{\theta}_{kb}^* = min\left\{\theta|y_k \leq \sum_{i=1}^{n} \lambda_i y_i\,;\ \theta x_k \geq \sum_{i=1}^{n} \lambda_i x_i^*\,; \theta > 0; \gamma_i \geq 0\ \forall\ i = 1, \dots, n\right\} \tag{2.22}$$

6. Repeat steps (3)-(5) B times to obtain a distribution of bootstrap estimated efficiency scores $\hat{\theta}_{kb}^*,\ b = 1,2, \dots B$.

[18] The envelopment form is preferred as the linear programming problem involves fewer constraints compared to the multiplier form $(p + q < n + 1)$ and it is therefore faster.

It is important to note that from (2.21) that $x_i^* \geq x_i$, suggesting that the feasible set defined by bootstrap DEA will be a subset of the one defined by DEA, which mimics the fact that the sample DEA feasible set is a subset of the "true" or population one (Simar and Wilson, 1998). This means that the bootstrap DEA frontiers will be always enveloped within the DEA ones and therefore $\hat{\theta}_{kb}^* \geq \hat{\theta}_k$ just as $\hat{\theta}_k \geq \theta_k$.

Now looking at (2.22) we also realise that it is possible for $\hat{\theta}_{kb}^*$ to exceed one as the initial data (x_i, y_i) could lie outside the feasible set, with the latter being defined in each bootstrap replication by $(x_i^*, y_i)_b$ and regardless of (x_i, y_i). In this case bootstrap DEA mimics the fact that drawing randomly DMUs from the population will necessarily leave out some DMUs which would have otherwise been efficient. Hence, $\hat{\theta}_{kb}^*$ exceeding one shows by how much the bootstrap DEA frontier could have been "pushed" to coincide with the initial DEA frontier, just as the DEA frontier should be "pushed" to coincide with the population frontier.

A graphical illustration of what bootstrap DEA does is provided in Figure 2.4, which is a modified version of Figure 4.5 in Simar and Wilson (2008). The figure shows how the true efficiency score, the DEA estimate and the bootstrap DEA scores are computed for DMU $k(x_1^k, x_2^k | y^k)$ in input orientation in a 2-inputs/1-output specification[19]. The unobservable "true" or population frontier, $\partial X_{DEA}(y)$, is depicted by the solid green line, the DEA frontier, $\widehat{\partial X}_{DEA}(y)$, is depicted by the solid black piecewise linear sections, while the bootstrap DEA frontiers, $\widehat{\partial X_{b,DEA}^*}(y)$, $b = 1,2 \ldots B$, are represented by the

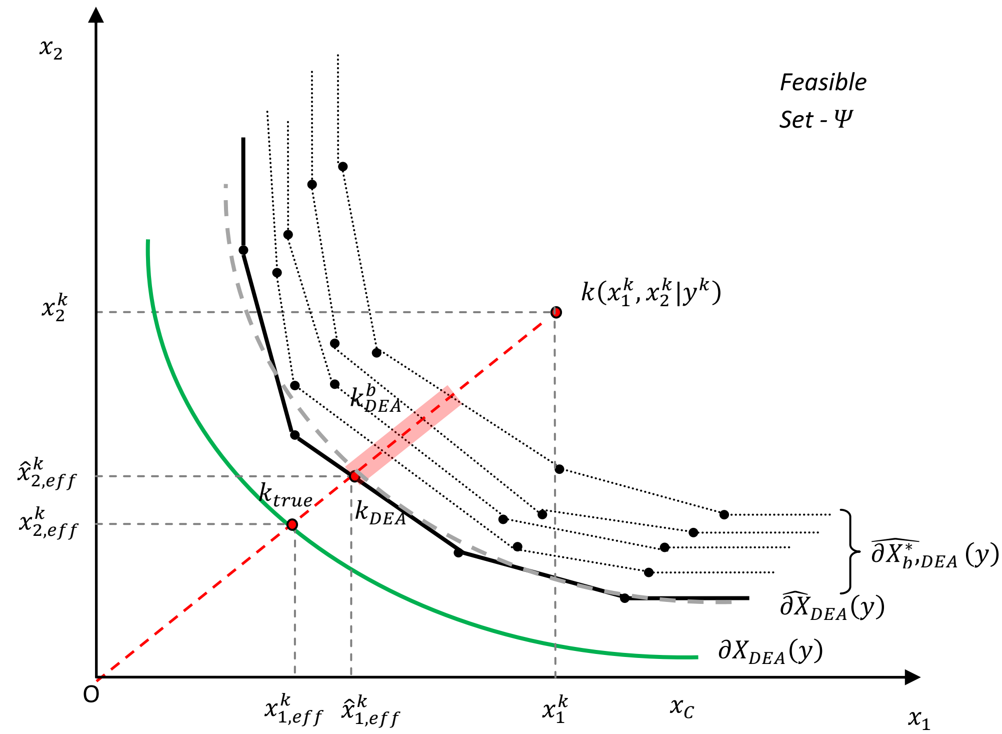
[19] You can either think that all DMUs use a common output or that the axes represent input divided by output.

dashed black piecewise linear sections. We have also included a curved dotted, light-grey line to graphically represent loosely the effects of smoothing the empirical distribution of efficiency scores[20].

Suppose that we want to extract the efficiency distribution of DMU k. The process of bootstrap DEA can be thought of as keeping DMU k, and hence the ray Ok, fixed while generating frontiers through bootstrap DEA. Each bootstrap frontier is associated with a different efficiency score, yielding a range of bootstrapped efficiency scores which is graphically represented by the red-shaded box. The figure also demonstrates that the DEA frontier overestimates the "true" efficiency score and how bootstrap DEA tries to mimic this "overestimation", as previously discussed.

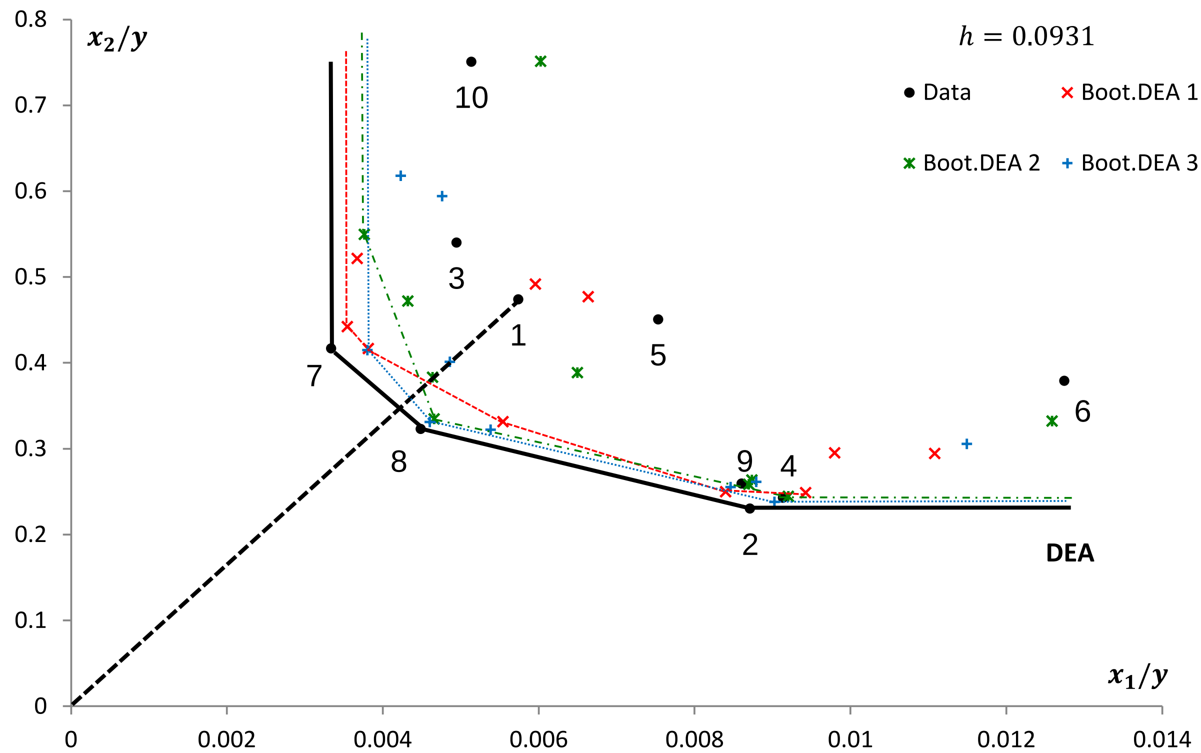
[20] We are not suggesting that the frontier is smoothed; it is the distribution of efficiency scores that is smoothed. However, the richer support provided by the smoothing process yields a continuum of efficiency scores which can be thought of as having an effect on the frontier as well.

Figure 2.4. Illustration of bootstrap DEA in input orientation



Finally, to provide a practical visualisation of how bootstrap DEA works we have reproduced Figure 2.4 in Figure 2.5 using generated data in a 2-input/1-output model. The axes of the figure below are the inputs divided by the outputs so that the frontiers can be interpreted as isoquants. The reported value for h is a smoothing parameter required to smooth the empirical distribution. Regarding DMU 1, its DEA score is 0.7314 while its bootstrap DEA scores are 0.799, 0.7848 and 0.7567 for bootstrap replications 1, 2 and 3, respectively. DMU 4 has a DEA score of 0.9499 while its bootstrap scores are 1.0243, 1.0072 and 0.9836, which is an example of how bootstrap DEA scores can exceed 1. Finally, it is interesting to note that the efficient DMUs (2, 7 and 8) which are associated with a DEA score of one, have bootstrap DEA scores greater than 1 in this example.

Figure 2.5. Graphical illustration of bootstrap DEA using data



2.6.3 Bootstrap DEA: statistical inference and confidence intervals

Let us now consider how $\hat{\theta}_k^* = \left\{\hat{\theta}_{kb}^*,\ b = 1,2, \dots B\right\}$, can be used to construct confidence intervals. The idea is to construct confidence intervals which contain the "true" or population efficiency score of a DMU k. This requires assuming that the bootstrap bias is equal to the DEA or model bias. We will see in this section how this assumption allows for constructing confidence intervals.

The first step is to compute the mean of the bootstrap distribution:

$$\overline{\hat{\theta}_k^*} = \frac{1}{B}\sum_{b=1}^{\mathrm{B}} \hat{\theta}_{kb}^* \tag{2.23}$$

The mean in (2.23) needs to be corrected for bootstrap bias as follows:

$$\widehat{bias}_k = \overline{\hat{\theta}_k^*} - \hat{\theta}_k \tag{2.24}$$

Correcting for bias once, tough, would centre the bootstrap distribution on the DEA score of DMU k. If we denote this shifted distribution with $\hat{\theta}_k^{*c}$, then:

$$\overline{\hat{\theta}_k^{*c}} = \overline{\hat{\theta}_k^* - \widehat{bias}_k} = \overline{\hat{\theta}_k^*} - \left(\overline{\hat{\theta}_k^*} - \hat{\theta}_k\right) = \hat{\theta}_k \tag{2.25}$$

Simar and Wilson (1998) suggest correcting for bootstrap bias twice as it would approximately centre the bootstrap distribution on the population efficiency score. The resulting double-corrected distribution for firm k would be[21]:

$$\tilde{\theta}_k^* = \hat{\theta}_k^* - 2\widehat{bias}_k, \quad b = 1,2,\dots B \tag{2.26}$$

with a mean which is assumed to be approximately equal to the "true" efficiency score:

$$\overline{\tilde{\theta}_k^*} = \overline{\hat{\theta}_k^*} - 2\widehat{bias}_k = \overline{\hat{\theta}_k^*} - 2\left(\overline{\hat{\theta}_k^*} - \hat{\theta}_k\right) = 2\hat{\theta}_k - \overline{\hat{\theta}_k^*} \simeq \theta_k \tag{2.27}$$

Although this assumption is valid asymptotically, it has not been yet confirmed for finite samples, especially for smaller ones which are frequently met in the empirical literature. The accuracy of (2.26) depends on the assumption that the bootstrap bias closely approximates the model (or DEA) bias (2.14):

[21] We need to make a note at this point to avoid confusion with notation. Simar and Wilson (1998) use $\tilde{\theta}_k^*$ to denote the mean of the distribution of $\left\{\tilde{\theta}_{kb}^*, b = 1 \dots B\right\}$ while we use it to denote the set of bootstrap values of $\tilde{\theta}_{kb}^*$. In general, we find more clear to denote with ψ_{kb}^* the b^{th} bootstrap value of ψ attached to DMU k, with $\psi_k^* = \{\psi_{kb}^*,\ b = 1 \dots B\}$ the vector of the bootstrap values for DMU k and with $\bar{\psi}_k^*$ the central moment of ψ_k^*, where ψ can be either $\hat{\theta}$ or $\tilde{\theta}$.

$$\left(\hat{\theta}_k^* - \hat{\theta}_k\right)|\hat{\mathcal{P}} \sim \left(\hat{\theta}_k - \theta_k\right)|\mathcal{P} \tag{2.28}$$

The assumption in (2.28) is asymptotically valid and it allows considering the centre of the distribution of $\tilde{\theta}_k^*$ as the "true" efficiency score (see equation (2.27)). Hence, Simar and Wilson (1998) propose constructing confidence intervals using the $\left({}^{a}/_{2}\right)$% and $\left(1 - {}^{a}/_{2}\right)$% percentiles of this distribution. Hence, the confidence interval that includes the true efficiency score θ_k with a probability $(1-a)$% is:

$$\left(\tilde{\theta}_{k,low}^*\,,\, \tilde{\theta}_{k,up}^*\right) = \left(\tilde{\theta}_k^{*,(a/2)}\,,\, \tilde{\theta}_k^{*,(1-a/2)}\right) \tag{2.29}$$

In a later paper, Simar and Wilson (2000a)[22] propose using the distribution of the bootstrap bias to construct confidence intervals[23]. If we denote with s and $\hat{s}$ the percentiles of the distribution of the DEA bias and of the bootstrap bias, then:

$$1 - a = \Pr\left(s_{a/2} < \hat{\theta}_k - \theta_k < s_{1-a/2}\right) = \Pr\left(\hat{s}_{a/2} < \hat{\theta}_k^* - \hat{\theta}_k < \hat{s}_{1-a/2}\right) \tag{2.30}$$

Implementing the assumption (2.28) here it follows that the endpoints of these distributions are approximately equal or: $s_{a/2} \simeq \hat{s}_{a/2} = \Delta\hat{\theta}_k^{*\,(a/2)}$ and $s_{1-a/2} \simeq \hat{s}_{1-a/2} = \Delta\hat{\theta}_k^{*\,(1-a/2)}$, where $\Delta\hat{\theta}_k^* = \hat{\theta}_k^* - \hat{\theta}_k$. Using this assumption, Simar and Wilson (2000a) propose the following intervals about θ_k:

[22] In fact this approach was first proposed by Simar and Wilson (1999) in the context of bootstrapping Malmquist indices and it was first adopted for the case of bootstrap DEA by Simar and Wilson (2000a).
[23] Simar and Wilson (2000a) state that the basic confidence intervals should be preferred over the intervals constructed under the percentile method of Simar and Wilson (1998) as the bias-corrected bootstrap estimates are associated with excess variation, and in particular that $Var\left(\tilde{\theta}_k^*\right) = 4Var\left(\hat{\theta}_k\right)$.

$$
\begin{aligned}
1 - a &= \Pr\left(\hat{\theta}_k - s_{1-a/2} < \theta_k < \hat{\theta}_k - s_{a/2}\right) \\
&\simeq \Pr\left(\hat{\theta}_k - \hat{s}_{1-a/2} < \theta_k < \hat{\theta}_k - \hat{s}_{a/2}\right) \\
&= \Pr\left(\hat{\theta}_k - \Delta\hat{\theta}_k^{*(1-a/2)} < \theta_k < \hat{\theta}_k - \Delta\hat{\theta}_k^{*(a/2)}\right)
\end{aligned}
\tag{2.31}
$$

That is, they use the endpoints of the distribution of the bootstrap bias to approximate the unobservable endpoints of the distribution of DEA bias. Again, these confidence intervals are asymptotically consistent but it is necessary to establish finite performance before using them.

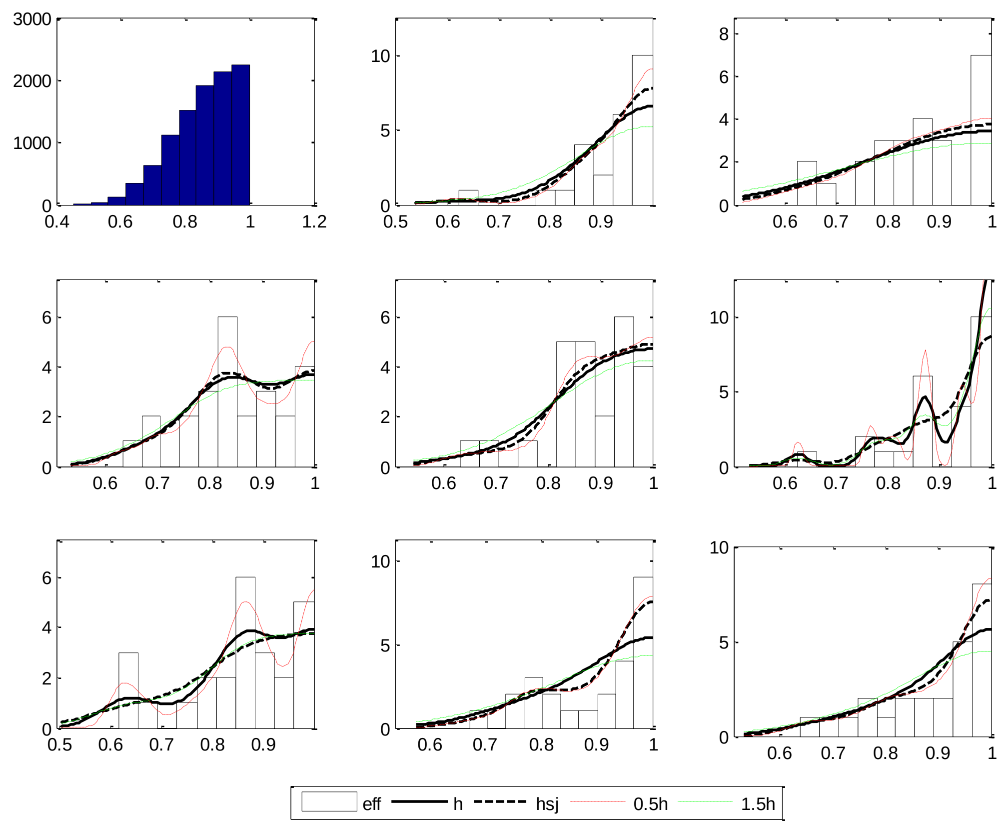
2.6.4 On smoothing the empirical distribution[24]

Simar and Wilson (1998) suggest that the empirical distribution of efficiency scores should be smoothed before bootstrapping. They refer to the standard bootstrap procedure (re-sampling with replacement from the empirical distribution) as the "naïve" bootstrap and they state that it produces inconsistent estimates due to the bounded support of the empirical distribution. The main argument against using the "naïve" bootstrap is that the algorithm produces repeated values (especially in smaller samples), resulting in distributions that cannot be used for statistical inference. Smoothing the empirical distribution, instead, produces bootstrap samples with richer support and therefore bootstrap distributions will be more suitable for statistical inference.

[24] Appendix I elaborates on smoothing techniques and reviews the literature which compares the strengths and weaknesses of some popular approaches. This section assumes previous knowledge of these methods so the interested reader should refer to Appendix I prior to proceeding.

A graphical illustration of smoothing is provided in Figure 2.6 below. On the top left corner we present an assumed population distribution of efficiency scores while the rest subplots present samples of size 25 drawn from the population and on which smoothing has been applied (the various lines)[25]. Ideally, smoothing would estimate a distribution which resembles the population one. It is easy to observe that smoothing sometimes performs well in that respect but sometimes less so.

Figure 2.6. Graphical illustration of smoothing



[25] See footnote 30 for a description of the population. Also note that h corresponds to an estimated bandwidth using the Least Squares Cross Validation (LSCV) method, $0.5h$ and $1.5h$ shows the LSCV-smoothed line with 50% less or more smoothing, while hsj corresponds to a bandwidth that has been estimated using the Sheather and Jones (1991)technique. More information on these methods is provided in Appendix I.

A discussion in support of the smooth bootstrap is given in Simar and Wilson (2004). In particular, they refer to the works of Bickel and Freedman (1981), Swanepoel (1986), Beran and Ducharme (1991), and Efron and Tibshirani (1993) who examine the use of smoothing in general bootstrap applications. In fact, Efron and Tibshirani (1993) demonstrate an example of the failure of the ("non-parametric") bootstrap[26] and state that:

> "*What goes wrong with the non-parametric bootstrap[27]? The difficulty occurs because the empirical distribution function $\hat{F}$ is not a good estimate of the true distribution F in the extreme tail. Either parametric knowledge of F or some smoothing is needed to rectify matters.*" (Efron and Tibshirani, 1993; pp.81)

Indeed, Efron (1979) had already mentioned that, in cases where the empirical distribution function is discrete, it would be probably better to apply smoothing as bootstrapping such a distribution would result into degenerate distributions of repeated values.

Bickel and Freedman (1981) provide further support to the argument above for the case of bootstrapping the mean, under the assumption that the parameterized distribution is a good approximation of the true underlying one. Swanepoel (1986) argues that drawing from an approximated empirical distribution is asymptotically valid. Beran and Ducharme (1991) provide a review of the work thus far on the asymptotics of the bootstrap.

---

[26] This refers to their example experiment where they sampled 50 observations from the uniform distribution, for which the maximum likelihood is the greatest value observed. They compared the performance of the algorithm with drawing with replacement from the 50 observations and another algorithm where they draw with replacement from the uniform distribution on [0, max θ]. They find that the first one (non-parametric) is a poor approximation of the latter (parametric) due to the fact that there is a large probability mass at a level lower than the maximum observed value of the sample.

[27] That is, drawing from a sample rather than from some parametric model or distribution.

Silverman and Young (1987) impose the question of whether smoothing should be employed or not. They emphasize that smoothing is a valuable tool in cases where the empirical distribution is discrete because simple re-sampling would produce samples with peculiar properties. They prove that smoothing will give better results if the approximated function is a linear (affine) transformation of a symmetric distribution but not of a uniform one. They also suggest that future research should empirically explore the appropriateness of smoothing under different assumptions about the distribution of the population[28].

One of the limitations of smoothing approaches is that noise might be introduced in the system when resampling from the smoothed distribution. This is not surprising as smoothing transforms the empirical distribution to one which tries to capture the asymptotic properties of the true distribution. In fact, Simar and Wilson (2002) have mentioned this problem in their paper and have stated in particular that:

> "*The bootstrap procedures… may involve errors in finite samples due to sampling variation in the distance function estimators as well as additional noise introduced by the resampling process itself*" (Simar and Wilson, 2002; pp.124)

And they continue in a footnote on the same page:

> "*In particular, kernel estimators, while consistent, are slow to converge. Resampling from kernel estimates of the density of distance function estimates might be a significant source of noise in the bootstrap process*" (Simar and Wilson, 2002; pp.124; footnote 10)

The mathematics of the consistency of smoothing techniques on bootstrap DEA is a very challenging topic[29]. However, some intuition in support of smoothing can be gained by inspecting Figure 2.7. The figure demonstrates the histograms of the bootstrap

[28] This has motivated our Monte Carlo exercise over the different population assumptions.
[29] The interested reader may refer to Kneip et al. (2011, 2008) who derive theoretical expressions in support of smoothing in bootstrap DEA.

distribution of efficiency scores for a DMU under two smooth bootstraps and the "naïve" bootstrap (last row)[30]. The two smoothing procedures considered are the least squares cross validation method (LSCV) and the "plug-in" method of Sheather and Jones (1991) (SJ). The bootstrap is applied on the same data and for sample sizes of 25 (first column) and 800 (second column) while a rescaled version of the latter is provided in the final column to distinguish among the different cases.

The two smoothing methods in Figure 2.7 have similar distributions for the case of 25 DMUs, while the naïve bootstrap is associated with a discrete degenerate distribution. It is obvious that the naïve bootstrap should not be used for statistical inference as being inconsistent and associated with counter-intuitive confidence intervals. For the case of 800 DMUs, although the smooth bootstrap still produces more variation compared to the naïve bootstrap (last column), the resulting endpoints of the distribution become very narrow when viewed on the same scaling as in that of the smaller sample case (second column). This is in support of the asymptotic convergence of bootstrap DEA (as confidence intervals become narrower).

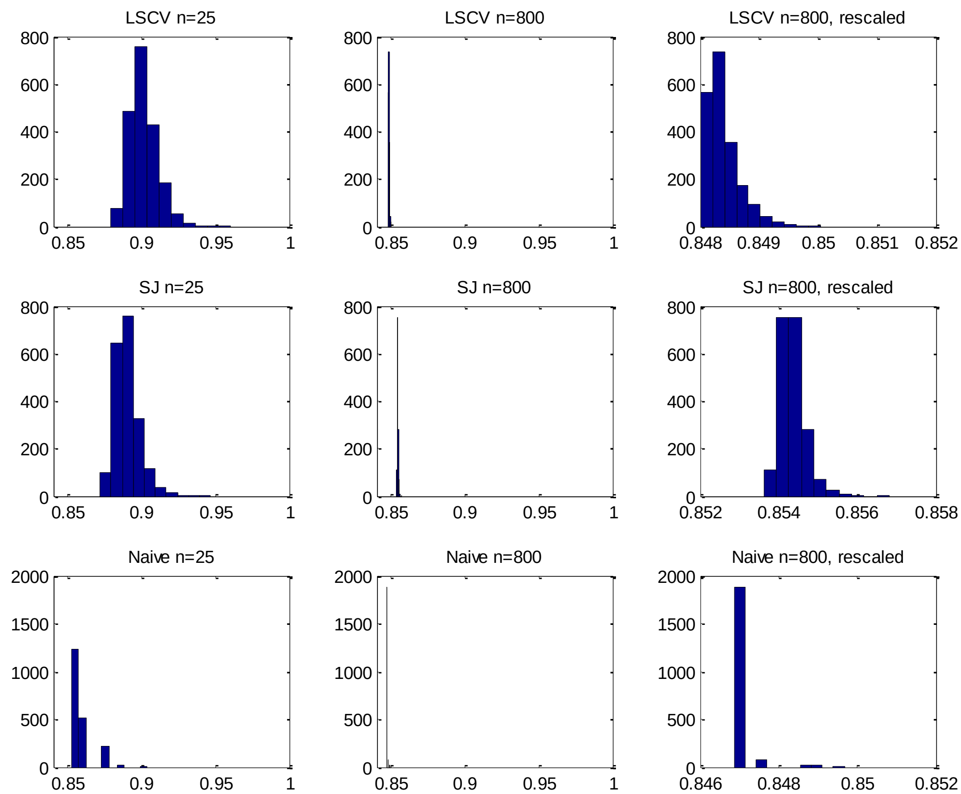
Apart from the insights relevant to smoothing, the example in Figure 2.7 shows that the bootstrap as a process is useful in smaller samples where the researcher has limited knowledge of the population's estimated parameters. However, its use in large samples is limited as the very narrow confidence intervals supress the scope for hypothesis

---

[30] Although it is not important at this stage, these graphs have been produced from a sample of 25 DMUs where a CRS input oriented model is applied on a 1-input/1-output specification, while the bootstrap procedure involves 2000 repetitions. The data have been generated from a process that we name "Standard" in our Monte Carlo simulations that will be presented is section 2.8.

testing since the inferred population parameters are estimated with a very narrow range. This is also evidenced in Simar and Wilson (2004) who report an average 95% confidence interval width of 0.0019 for a sample of 800, which is consistent with our findings.

Figure 2.7. Smooth vs naïve bootstrap: distributions of bootstrapped efficiency scores



It is crucial to explore how these smoothing procedures affect the performance of confidence intervals in finite samples (this will be addressed later in this chapter). There is no clear evidence as to whether LSCV should be preferred to SJ, but from the literature review in Appendix I we would expect LSCV to perform better in smaller samples and SJ better in larger ones[31].

2.6.5 Developments and extensions

Since the introduction of bootstrap DEA there have been various developments and extensions to the algorithm, mainly by Simar and Wilson and co-authors. The most well-known extensions of bootstrap DEA include the bootstrap Malmquist Index (Simar and Wilson, 1999), the heterogeneous bootstrap (Simar and Wilson, 2000a)[32], the tests on returns to scale using bootstrap DEA (Simar and Wilson, 2002) and the two-stage procedure for the regression of efficiency scores on environmental variables (Simar and Wilson, 2007). One assumption/principle that is used in all these studies as well as in Simar and Wilson (1998, 2000a) is that the bootstrap bias is approximately equal to the DEA bias, which is utilised in constructing confidence intervals. Hence, the finite sample performance of bootstrap DEA with respect to this assumption carries important implications for the extensions of the model.

---

[31] We deduce that from the findings in the literature that LSCV performs better when the distribution is degenerate or with multiple peaks, as most likely in small samples, while SJ has a better performance when the empirical distribution has a more clear structure and it is smoother (without peaks), as we would expect to find in large samples.

[32] In an informal discussion with Prof L. Simar, he suggested that the heterogeneous bootstrap might produce very wide confidence intervals and that it is not preferable to the homogeneous bootstrap DEA. In terms of Figure 2.4, the heterogeneous bootstrap DEA would produce a shaded area (bootstrap distribution) that would not lie just on the ray kk_{DEA} but it would it would spread around it at some angle.

The logic in bootstrapping DEA scores has not changed since it was first introduced. The various developments have focused on optimising the smoothing process to increase the finite sample efficiency of bootstrap DEA. One such development is the introduction of a double smoothing process (Kneip et al., 2008) which has been argued to be very complicated and computationally intensive (Kneip et al., 2011). Another alternative is to smooth the empirical distribution about the centre of the bootstrap distribution and use naïve bootstrap for the tails (Kneip et al., 2011). Despite that the latter method is more tractable and efficient, the minimum sample size cannot be small as the naïve bootstrap requires bigger samples to produce adequate tails[33]. In a recent paper, Simar and Wilson (2011) propose subsampling and present evidence from the m/n bootstrap using a data-driven procedure to determine the optimal m. It reduces the computational burden from complicated smoothing procedures and it is more accessible to the practitioner. However their method requires large samples; in fact, their simulations use a minimum size of 100 DMUs while considerably better results are obtained for the alternative sample of 1000 DMUs.

An interesting suggestion is the use of the iterated bootstrap, provided in a short note in Simar and Wilson (2004). The authors suggest iterating the bootstrap (that is, applying bootstrap DEA on each bootstrapped sample) to construct more accurate confidence intervals for the true efficiency score. The authors suggest that this approach would return more accurate confidence intervals by defining better nominal

[33] See the tails in Figure 2.7 (where a sample size of 25 is used) and see tables 2 and 3 in Kneip et al. (2011).

probabilities to perform hypothesis testing and thus extracting more accurate endpoints for the confidence intervals[34]. The major drawback of this process is the very high computational time which would be B_2 times greater than the simple bootstrap DEA, where B_2 is the number of second-stage bootstraps (or iterations) and would normally exceed 1000[35]. Moreover, no Monte Carlo results are provided for this method to evaluate the benefits along with the additional computational costs involved[36].

To our knowledge, the alternative bootstrap DEA procedures are mostly related to optimising the smoothing process or the sampling procedure. Unfortunately, they do not offer a clear-cut solution in applying bootstrap DEA in small samples (at least not with the desirable computational efficiency). Applied researchers use the methods of Simar and Wilson (1998, 2000a) to perform hypothesis testing and it is therefore crucial to establish the finite sample behaviour of these algorithms.

[34] In particular, the double bootstrap would determine a more accurate level of confidence on which Simar and Wilson's (2000a) confidence intervals would be constructed. Hence, instead of using the $\left({}^{a}/_{2}\right)$% and $\left(1 - {}^{a}/_{2}\right)$% percentiles of the bootstrap distribution, iterating the bootstrap would provide a more accurate $\hat{a}$ instead of a.

[35] To demonstrate the magnitude of computational time, the applied researcher would need about 3 hours on an i5 3.6GHz PC (a standard desktop PC) and programmed on Matlab (with parallel computing) to obtain results from the application of the iterated bootstrap on a sample of 30 firms, implementing a CRS 2-inputs/2-outputs specification and using 2000 replications in each stage.

[36] A Monte Carlo experiment with 1000 replications for the specification in the previous footnote would require approximately 125 days to run. Hence, a proper Monte Carlo study with various sample sizes would need several years! Obviously, these times could be reduced significantly by using alternative programming languages (such as C, Fortran or any language that would allow for hyper-programming) and using supercomputers.

2.7 Monte Carlo simulations and previous results on bootstrap DEA

Monte Carlo simulations are commonly used, among other uses, to examine the plausibility of certain assumptions of a model or the performance of confidence intervals. In general, Monte Carlo simulations involve assuming a data generating process that produces an unobservable, "true" population. Then the model, whose performance is being assessed, is applied on random samples (draws) from that population. The model is said to be performing well if (i) the model can replicate on average the moments of the population (mean, standard deviation, skewness and kurtosis), or (ii) if the model can accept (or reject) a pre-defined null hypothesis at a rate that is approximately equal to the nominal probability[37].

The standard approach in bootstrap DEA for performance evaluation is to use coverage probabilities, which count the frequency that the bootstrap confidence intervals include the "true" (population) efficiency score of a "fixed" DMU[38]. If the coverage probabilities converge towards the nominal ones, then this is an indication of good finite sample behaviour. Coverage probabilities are affected by sample size, the dimensions of the linear program (number of inputs and outputs) and by the data generating process (although the last point has not been thoroughly investigated in the literature). More importantly, the convergence of coverage probabilities depends on the finite validity of the assumption that the bootstrap bias is equal to the DEA bias.

---

[37] The nominal probability is the probability used to define the acceptance region for the pre-defined null hypothesis.

[38] A "fixed" DMU is a DMU that is programmed to appear in every Monte Carlo replication.

There are only a few papers which assess the performance of bootstrap DEA, which is not surprising as it is a specialized area. Furthermore, once Monte Carlo results have been published for one bootstrap DEA method, it would be pointless to replicate them. However, as Silverman and Young (1987) suggest, to properly evaluate the performance of a bootstrap procedure it is almost a requirement to use a wide range of population assumptions, especially if smoothing is involved.

In the literature the only well-known Monte Carlo exercises on the performance of the Simar and Wilson's (1998) bootstrap DEA are by Simar and Wilson (2004, 2000a)[39]. Simar and Wilson (2000a) use a one-input/one-output specification under the assumption of output orientation, under both CRS and VRS. They report coverage probabilities for their "enhanced" confidence intervals, which are summarized in Table 2.1 for the CRS case. The first column reports the sample size used in each Monte Carlo repetition, columns 2 to 6 report the coverage probabilities for five different levels of significance, column 7 presents the average width of the 95% confidence intervals, while the last column reports the average size of the difference between the bootstrap bias and the DEA or model bias (the latter is reported as "true" bias in the paper).

Their results suggest that even in smaller samples (such as 25 or 50), the coverage probabilities are quite close to the nominal ones. However, this is not surprising as the average width of the confidence intervals is quite high for smaller samples, which is not

[39] Some results are also provided in Löthgren (1998) who applies a similar exercise to compare the approach of Simar and Wilson (1998) with his. However, this is a working paper and Prof L. Simar expressed his concerns in the EWEPA 2013 conference (Helsinki) that it is flawed in many occasions. Therefore the results of Löthgren (1998) are not discussed here.

a desirable property for applied hypothesis testing. However, the latter is unrelated to the validity of Simar and Wilson's (1998) approach; it may be due to the data generating process chosen[40]. In fact, an indication that both their method and simulations are correct is that confidence intervals become narrower as sample size increases.

The final and perhaps most important point is that the difference between the average bootstrap bias and the average DEA bias is quite substantial for smaller sample sizes. Hence, although the coverage probabilities are very close to the nominal ones in smaller sample sizes, the finite sample performance of Simar and Wilson's bootstrap DEA is affected by the big difference in biases[41]. This implies that samples larger than 200 would be required in this example to combine good coverage probabilities and small differences in bootstrap and DEA biases.

Table 2.1. Simar and Wilson (2000a) Monte Carlo results

	Nominal Coverage Levels					Av. CI width	Av. Boot.bias
n	*0.8*	*0.9*	*0.95*	*0.975*	*0.99*	(95%)	minus DEA bias
10	0.693	0.814	0.886	0.919	0.942	0.911	-0.088
25	0.772	0.883	0.935	0.973	0.983	0.586	-0.075
50	0.784	0.894	0.940	0.970	0.985	0.351	-0.045
100	0.794	0.911	0.946	0.973	0.988	0.187	-0.024
200	0.810	0.899	0.946	0.970	0.994	0.095	-0.012
400	0.807	0.903	0.953	0.977	0.995	0.047	-0.005

Source: Simar and Wilson (2000a), Table 1 and Table 2

Similar evidence is found by Simar and Wilson (2004) who perform Monte Carlo experiments under the assumption of output orientation under both CRS and VRS, in a

[40] For the CRS case they assume $y = xe^{-|v|}$, $v \in N(0,1)$, and $x \in Uniform(1,9)$.
[41] It is reminded that one of the fundamental assumptions for the validity of Simar and Wilson's bootstrap DEA and confidence intervals is that the difference between the two bias is approximately zero.

1-input/1-output setup. In their simulations they compare the coverage probabilities of their confidence intervals (that is, of Simar and Wilson (2000a)) and two “naïve” (non-smooth bootstrap) alternatives: one which draws from the input-output data (case resampling) and one drawing from the empirical distribution of efficiency scores (fixed resampling). Their results for the CRS technology assumption are presented in Table 2.2. The first column reports the sample size while the next three columns present the coverage probabilities for the Simar and Wilson (2000a) method (“SW2000”), the naïve bootstrap with case resampling and the naïve bootstrap with fixed resampling. Columns (5) to (7) report the average confidence interval widths for each of the aforementioned cases while the last four columns report the DEA (or model or “true”) bias and the average bootstrap biases for each procedure.

Their findings suggest that the smooth bootstrap achieves higher coverage than the other two, while comparing the two naïve procedures the coverage probabilities are quite close and there is no clear “superiority” of the one over the other. Confidence intervals become narrower with sample size, while bootstrap and DEA biases become smaller. This is shown in the last block of Table 2.2 where both the model and bootstrap biases converge to zero as sample size increases.

In contrast with their previous Monte Carlo study, the confidence intervals in Simar and Wilson (2004) are substantially narrower. Coverage probabilities seem to converge to the nominal ones when the sample size becomes 800 while they are fairly high for reasonably small samples (25 to 50)[42]. Again, the bootstrap bias adequately approximates the DEA bias for sample sizes greater than 50, while this difference becomes very small when the sample size exceeds 400. Hence, we would deduce that the applied researcher could use bootstrap DEA in smaller samples if he is ready to accept some degree of bias. Finally, we need to note that according to the results in Simar and Wilson (2004), for large samples the average confidence interval width becomes so narrow that they seem to actually converge to a certain point, suggesting that hypothesis testing would reject the null hypothesis of equal efficiency almost every time. That is, any differences observed between DEA scores would automatically be significant.

Table 2.2. Simar and Wilson (2004) Monte Carlo results (95%) for the CRS case

	Coverage Probabilities (95%)			Av. CI Width (95%)			DEA and Bootstrap Biases			
n	*Smooth*	*Case*	*Fixed*	*Smooth*	*Case*	*Fixed*	*DEA*	*Smooth*	*Case*	*Fixed*
10	0.916	0.899	0.899	0.1384	0.2018	0.2018	0.0517	0.0362	0.0324	0.0324
25	0.932	0.894	0.890	0.0551	0.0664	0.0693	0.0203	0.0147	0.0117	0.0121
50	0.920	0.896	0.891	0.0283	0.0320	0.0315	0.0101	0.0076	0.0058	0.0057
100	0.921	0.889	0.891	0.0146	0.0154	0.0157	0.0048	0.0039	0.0028	0.0030
200	0.937	0.879	0.888	0.0076	0.0078	0.0074	0.0024	0.0020	0.0014	0.0014
400	0.936	0.883	0.889	0.0039	0.0037	0.0038	0.0012	0.0010	0.0007	0.0007
800	0.950	0.886	0.871	0.0019	0.0019	0.0019	0.0006	0.0005	0.0004	0.0004
1600	0.957	0.876	0.868	0.0010	0.0009	0.0009	0.0003	0.0003	0.0002	0.0002
3200	0.951	0.897	0.864	0.0005	0.0005	0.0005	0.0002	0.0001	0.0001	0.0001
6400	0.960	0.878	0.868	0.0003	0.0002	0.0002	0.0001	0.0001	0.0000	0.0000

Source: Simar and Wilson (2004), Tables 10.1, 10.3 and 10.6

[42] In fact, Simar and Wilson (2004) state that the deviation of coverage probabilities from their nominal values could be due to: "*sampling variations in the Monte Carlo experiment, and due to the fact that a finite number of bootstrap replications are being used*" (Simar and Wilson, 2004; pp. 285).

The Monte Carlo evidence in Simar and Wilson (2000a, 2004) indicate that bootstrap DEA is associated with substantial sample requirements. Although the issue of minimum sample size is not discussed in their paper, Simar and Wilson (2004) state that:

> *"The results … show that in less favorable situations, even if the bootstrap is consistent, the coverage probabilities could be poorly approximated in finite samples"* (Simar and Wilson, 2004; pp. 292)

Moreover, given that the simulation exercises are based on the smallest possible dimension (1-input/1-output) we deduce that for higher dimensions the requirements should be even larger. Therefore the assumption of similar bootstrap and DEA biases might not have the desirable finite sample performance, carrying important implications for the use of Simar and Wilson's (1998, 2000a) confidence intervals in small samples. In addition, when the sample size becomes large enough, the confidence intervals become so narrow that it would probably reject most null hypotheses (this was also shown in Figure 2.7). Before deducing this implication, it is necessary to establish the behaviour of bootstrap DEA under various data generating processes, smoothing procedures and model dimensions: this is exactly what this simulation exercise is about.

2.8 The Monte Carlo experiments

2.8.1 The experiment outline

The Monte Carlo experiments are performed using samples drawn from four different populations which we name "Standard", "Truncated Normal Low", "Truncated Normal

High" and "Uniform". The motivation for including multiple data generating processes in our exercise stems from Silverman and Young (1987) who suggested than when smoothing is applied, Monte Carlo evidence should be provided under various data generating processes. Moreover, Simar and Wilson (2004) found in their simulations that:

> *"… the structure of the underlying true model plays a crucial role in determining how well the bootstrap will perform in a given applied setting."*
> (Simar and Wilson, 2004; pp.295)

The simulations are performed over 7 different sample sizes (15, 20, 25, 30, 60 and 120) and three different model dimensions (1-input/ 1-output, 2-inputs/1-output and 2-inputs/2-outputs)[43]. Moreover, for the 1-input/1-output dimension we perform one extra exercise by including large samples (25, 50, 100, 200, 400, 800 and 1600), since the computational costs are permissible[44]. Each of the $M = 1000$ repetitions of bootstrap DEA involves $B = 2000$ loops. The experiments are performed with two smooth processes (LSCV and SJ) and one "naïve", under the assumption of constant returns to scale (CRS) and input orientation. All calculations were performed in Matlab, using a straightforward code written by the author, which repeatedly calls an appropriately modified Matlab code for bootstrap DEA written by L. Simar (last updated in November of 2002) while most auxiliary functions (especially for the SJ smoothing process) are

[43] To our know knowledge this is the only simulation study on bootstrap DEA that uses three different dimensions while the 2-input/2-output case has only been included in simulations on bootstrap DEA extensions (Kneip et al., 2011, 2008). For the standard bootstrap DEA the two studies in the literature only use 1-input/1-output. At the moment the computational costs are prohibitive to increase the dimensions and it is left for future research.

[44] We tried to include even larger samples of 3200 and 6400, however due to technological restrictions (memory issues) it was not possible to do so. For future work an advanced computer could be used to overcome these difficulties.

called from the codes of Simar and Zelenyuk (2007)[45]. All main codes, along with line-by-line explanations have been uploaded online are also available upon request by the author[46].

The computational costs in seconds, using a desktop PC Intel i5 3.8MHz processor, are presented in Table 2.3 for each population assumption "Standard", "Truncated Normal Low", "Truncated Normal High" and "Uniform") and each model dimension (1I/1O, 2I/1O and 2I/2O). As expected, computational costs increase with model dimensions. The "naïve" bootstrap is occasionally slightly faster than the smooth bootstrap but not always: this is due to the fact that 5 different PCs were used for the simulations and differences in expected performance can be due to that. The cumulative computational costs were 34.4 days.

Table 2.3. Computational costs in seconds of the Monte Carlo exercise

	Standard			**Trun. Normal Low**			**Trun. Normal High**			**Uniform**		
	1I/1O	*2I/1O*	*2I/2O*	*1I/1O*	*2I/1O*	*2I/2O*	*1I/1O*	*2I/1O*	*2I/2O*	*1I/1O*	*2I/1O*	*2I/2O*
LSCV	44146	97727	172476	44122	71690	103457	44388	72241	105675	44340	80621	125804
SJ	42814	93852	165975	42724	69469	99807	42742	69887	101231	43039	77517	119072
Naïve	42661	98263	167177	42757	70878	100342	42731	75005	106901	42583	81583	125630

An important note about comparing the two smooth bootstraps and the "naïve" is that we take care to use exactly the same samples on which the bootstrap DEA

[45] The codes of Simar and Zelenyuk (2007) are provided from the *Journal of Econometrics Data Archive* and can be downloaded here: http://econ.queensu.ca/jae/2007-v22.7/simar-zelenyuk/. Also note that the paper of Simar and Zelenyuk (2007) is not directly related to the bootstrap DEA of Simar and Wilson (1998) but it is an extension to multiple groups and deriving aggregate efficiency scores. However, there are many auxiliary functions in this paper which are also used in the simple bootstrap and one that is used for the SJ smoothing process and which is slightly adjusted to the univariate case here. In fact, the auxiliary functions used in Simar and Zelenyuk (2007) were the exact ones used in the codes written by L. Simar, however we prefer using the former since they have been officially published in a well-known journal.

[46] Follow the link: https://www.dropbox.com/sh/3btckmd0sqwhqlq/AAAVlFL2cU5DzYUx6sKT7KIDa?dl=0

procedures are run. Hence, the generated samples are common in all cases of smooth and naïve bootstraps. Therefore any potential differences due to the random sampling of the Monte Carlo algorithm have been mitigated and any differences observed are solely due the bootstrap procedures.

The main focus of this exercise is to: (i) examine whether the assumption of equal bootstrap and model biases is plausible in finite samples and (ii) examine whether the bootstrap confidence intervals of Simar and Wilson (1998, 2000a) are associated with coverage probabilities which converge to the nominal ones in finite samples. To evaluate the assumption of equal bootstrap and model (or DEA) biases, which carries important implications for the performance of Simar and Wilson's (1998, 2000a) confidence intervals, we compute the average bootstrap and DEA biases in the Monte Carlo trials and compare them. Although we know that the two biases will converge asymptotically to zero, we are mostly interested in their behaviour in smaller samples as upon this condition depends the performance of Simar and Wilson's (1998, 2000a) confidence intervals and of the extensions of Simar and Wilson's works which make use of this assumption (already discussed in section 2.6.5).

To compute coverage probabilities we follow the common practice of using a "fixed" DMU[47], that is a DMU which is programmed to appear in every Monte Carlo trial. Then coverage is calculated by the frequency that the "true" efficiency score of the "fixed" DMU lies within the bootstrap confidence intervals. It will be discussed later in this

[47] More details on the definition of the "fixed" DMU are provided in section 2.8.4.

chapter that the assumed fixed point returns robust results since it is relatively far from the frontier which would yield higher overage probabilities.

Apart from examining coverage probabilities, we evaluate the behaviour of confidence intervals in two ways: (i) we inspect the convergence behaviour of the average 95% confidence intervals about the "true" efficiency score (along with their width) and (ii) we compute the average moments of the bootstrap distribution of the fixed DMU[48].

2.8.2 The data generating process

The data generating processes (DGP) have been designed to have an economic interpretation, discussed in the next subsection. Since we assume input orientation and since the source of variability is attributed to the deviations of inputs from their efficient levels, the DGPs are designed to generate these deviations. Output is produced in each process by a CRS Cobb Douglas function which uses the efficient input levels of DMUs; the deviation of inputs from their efficient level is the source of inefficiency. The processes of these deviations are presented below for each input $i = 1,2, \dots n$:

Standard: $x_i = {x_i}^{eff} e^{0.2|v|}$ where $v \sim N(0, 1)$

Trunc. Normal Low: $x_i = {x_i}^{eff} e^{0.2\omega}$ where $\omega \sim N^+(0.5, 1)$

Trunc. Normal High: $x_i = {x_i}^{eff} e^{0.8\xi}$ where $\xi \sim N^+(0.5, 1)$

[48] We would like to thank Prof L. Simar for his suggestion to explore the moments of the bootstrap distribution of the fixed DMU.

Uniform: $x_i = x_i^{eff} e^{0.8u}$ where $u \sim Uniform[0, 0.8]$

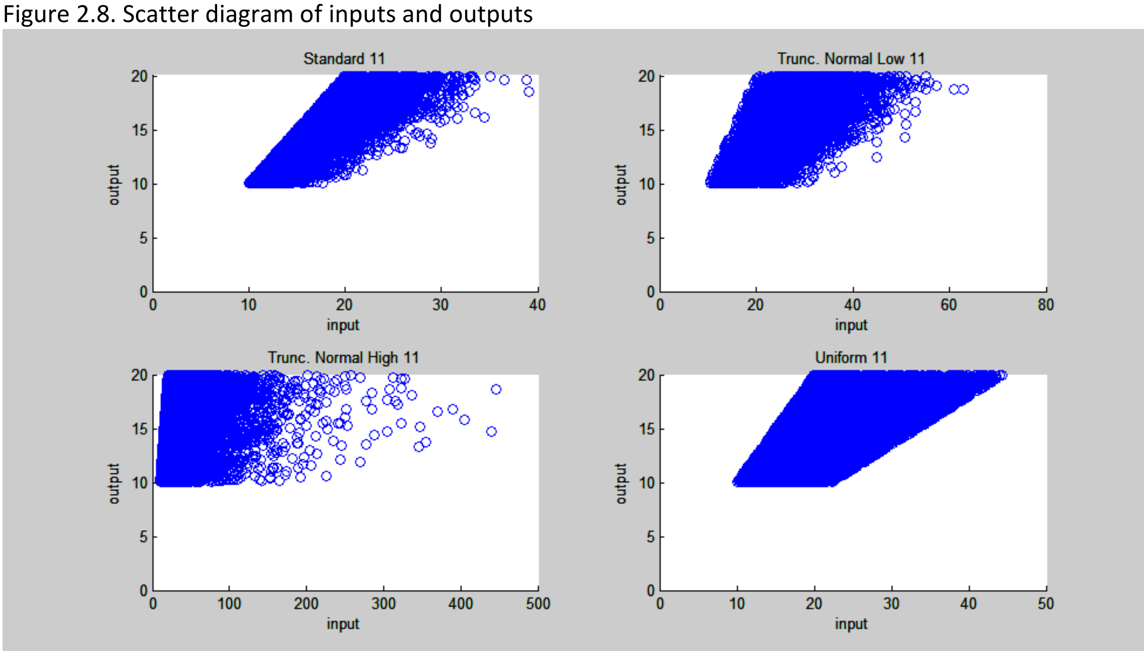
A very important clarification is that in the case of multiple inputs, the random components are common to all inputs. This is due to the definition of the input oriented efficiency: it is the input contraction factor that needs to be applied to *all* inputs of a DMU in order to become efficient. That is, if a DMU has an efficiency score of 0.8, then it will need to use 80% of all its inputs to become input-efficient and the assumption used here reflects this definition.

The efficient inputs in the 1 input and 1 output case are generated from a uniform distribution on the [10,20] interval while output is produced according to the following simple CRS production function: $y = x^{eff} \sim U[10,20]$. Figure 2.8 presents a scatterplot of the generated input-output combinations for the 1-input/1-output case[49]. The resulting scatter plots reflect the expected behaviour: the range of values for the output ranges between 10 and 20 (as it is equal with the efficient input level) while inputs vary according to the assumed distribution of the disturbance. In particular, for the standard case the observations are gathered closer to the frontier, for the truncated normal with low variance the observations are a bit more scattered to the right compared to the standard, in the truncated normal with high variance the observations are substantially more scattered, while in the case of the uniform the observations are equally scattered in the feasible set of values. Regarding the frontiers, they all lie on the 45^o line as

[49] The presentation of the scatterplot is used to address the concerns raised by Pror L. Simar in the EWEPA 2013 conference (Helsinki) that the DGPs used by the author are inconsistent. We would therefore like to thank Prof L. Simar for pointing out potential inconsistencies with previously used DGPs. The DGPs used here are clearly consistent with a well-defined population frontier and behaviour.

expected, although this is not obvious in the two cases with truncation due to the different scaling of the axes.

Figure 2.8. Scatter diagram of inputs and outputs

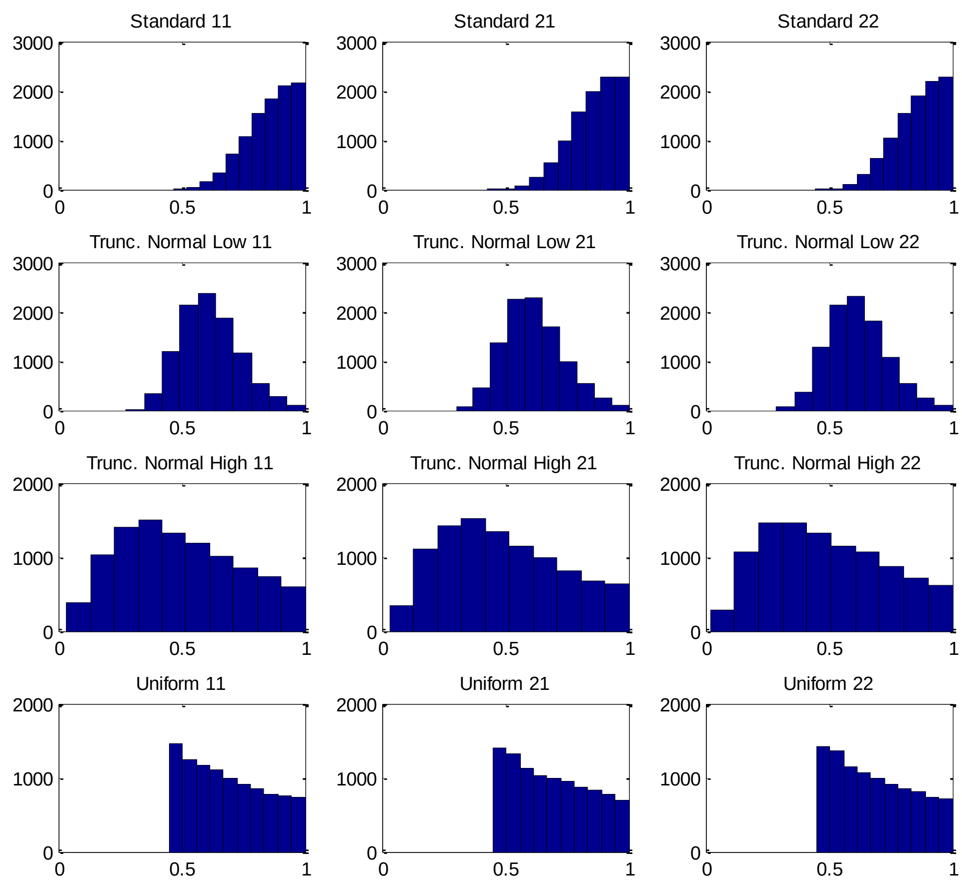


In the case of 2-inputs/1-output, the efficient levels of inputs are uniformly distributed on the [10,20] and [20,30] intervals: $x_1^{eff} \sim Uniform[10, 20]$ and $x_2^{eff} \sim Uniform[20, 30]$. Output is produced using a standard Cobb Douglas CRS production function[50]: $y = \left(x_1^{eff}\right)^{0.5}\left(x_2^{eff}\right)^{0.5}$ Finally, for the case of 2-inputs/2-outputs, the efficient levels of inputs are generated using the same process as in the previous case. Outputs are produced using the following CRS Cobb Douglas functions:

[50] Although the simulations should not be sensitive to the choice of the input elasticities in the production function (as long as they sum up to 1), it would be interesting in the future to examine the robustness of our results under various combinations of these parameters.

$y_1 = \left(x_1^{eff}\right)^{0.5}\left(x_2^{eff}\right)^{0.5}$ and $y_2 = \left(x_1^{eff}\right)^{0.3}\left(x_2^{eff}\right)^{0.7}$. The resulting population distributions for all DGPs and model dimensions are presented in Figure 2.9. The labels above each histogram represent the different combinations of DGP and model dimensions and are self-explanative.

Figure 2.9. Population distributions of efficiency scores for each DGP



2.8.3 The economic interpretation of the DGPs[51]

Each population is constructed to be both consistent with DEA assumptions[52], but also to have an economic interpretation. Hence, we associate the evidence on the performance of bootstrap DEA with certain market conditions which might be useful to the applied researcher. Hence, the user of bootstrap DEA will have more evidence about the finite sample performance of these methods in various market structures.

Regarding the standard case, the actual input levels are created by random positive deviations of inputs from their efficient levels. This is in accordance with input orientated models where $x > x_{eff}$, hence we named this case "Standard". Moreover, the DMUs are homogeneous and produce their outputs using the same CRS technology, which is consistent with the case of perfect competition. In a perfectly competitive industry we would expect all firms to be efficient while inefficiencies should be attributed to randomness, since all firms produce the same output using the same inputs and the same technology. It could be also associated with long-run monopolistic competition, which could be evidenced in non-perfectly competitive industries where well-established and large firms, operating under tight market conditions.

The truncated normal case with low-variance produces histograms of efficiency scores which look like normal distributions. In this case both tails of the distribution are

[51] The author would like to thank Prof M. Tsionas for his time to discuss the association of efficiency distribution and market structure. Prof Mike Tsionas agreed with the opinions expressed in this subsection. In fact, in one of his current works in progress he associated half-normal distributions with perfect competition as we do here.

[52] See section 2.3 and footnote 7. Most importantly, the generated data are convex combinations of a feasible set which exhibits certain technological characteristics (CRS in this case).

both relatively thin, indicating that a small proportion of these firms will operate efficiently (or not). The efficient firms use substantially less inputs than their peers while the inefficient firms use considerably more. In the context of DEA and of production economics this could be attributed to access to different technologies rather than random deviations from the efficient levels (like in the standard case). Hence, efficient firms are expected to have access to superior technology while very inefficient ones probably fail to adopt these technologies (perhaps due to size restrictions, various entry barriers or patents). We therefore associated “Truncated Normal Low” with monopoly. Moreover, it is important to note that applying DEA on such a market would violate the assumption of technological homogeneity and could be therefore associated with a form of model specification error.

The truncated case with high variance produces distributions which look like “flat normal”. The tails are fat, implying that a greater number of efficient firms have access to superior technology compared to the previous case. Moreover, the number of very inefficient firms is relatively high, indicating that inefficiency can be attributed to a reasonable extent to random deviations. Since inefficiency is both due to randomness and technological differences[53], this case is a mixture of the previous two and can be associated with monopolistic competition in the medium-run. That is, the initial patents that some firms used to have are now accessible to other firms, while the entry barriers

[53] In this case there are firms with efficiency score as low as 4% which is due to the high variance introduced. If we wanted to attach an economic intuition behind this behaviour, we could state that the low-performers are firms which failed to catch-up with modern practices that the efficient firms have adopted. These low extremes do not affect the validity of the Monte Carlo exercise as the DGP is valid.

are gradually lifted. Thus, all firms could achieve higher performance using these technologies and are expected to perform efficiently in the near future. Therefore, deviations from the efficient input levels can also be due to random events (apart from inability) which have prevented these firms from being efficient.

Finally, the uniform case cannot necessarily be associated with a specific market structure. We decided to include this case for the sake of completeness in order to evaluate the sensitivity of our results with respect to various assumptions about the DGP. Despite the fact that the DGP does not exhibit technological heterogeneities (as in the previous case), we argue that there is a different type of error; either the DMUs or the input-output variables chosen do not accurately reflect the underlying production process. We therefore suggest that the practitioner should first rethink about the DMUs or the inputs and outputs chosen; however, we provide some results to inform on the expected behaviour of bootstrap DEA in such cases.

2.8.4 Defining the fixed DMU[54]

The Monte Carlo simulations can be used to analyse the behaviour of bootstrap DEA in finite samples. As already explained, the main purpose of bootstrap DEA is to construct confidence intervals about the true efficiency score of a certain DMU of interest

[54] This section serves as a response to the concerns expressed by Prof L. Simar that the fixed point in a previous version was not properly defined. In an informal discussion, Prof L. Simar agreed that the approach that the author had followed was correct but the way presented was unclear and confusing. We have therefore decided to introduce some mathematical sophistication and proofs to show that the fixed point is properly defined and theoretically consistent. The author would like to thank Prof L. Simar for his time and valuable feedback on this issue.

(x_0, y_0)[55]. The Monte Carlo simulations evaluate the ability of bootstrap DEA to produce confidence intervals that actually include the true efficiency score $\theta(x_0, y_0)$, over a number of M trials. The frequency that $\theta(x_0, y_0)$ is included in each of the M constructed confidence intervals (coverage probability) is a popular approach of such an evaluation and we will use it in our analysis. For coverage probabilities to be computed for DMU (x_0, y_0), it has to appear in every Monte Carlo trial and it is therefore termed as the *fixed DMU* or the *fixed point*. Hence, defining the fixed DMU is an important part of the simulation exercise.

An important consideration in defining the fixed point (x_0, y_0) is the position of $\theta(x_0, y_0)$ in relevance to the population distribution of efficiency scores. One case that we could easily exclude is to choose (x_0, y_0) such that $\theta(x_0, y_0) \simeq 1$. In this case we would expect coverage probabilities to be overstated since this DMU would belong in the reference set in (almost) every Monte Carlo sample. A more reasonable choice would be a fixed point in a middle data point[56]; in our case we choose $(x_0, y_0) = (\bar{x}, \bar{y})$ suggesting that $\theta(x_0, y_0) = \theta(\bar{x}, \bar{y})$ would be near $\bar{\theta}$. We could therefore state that in this case we examine the behaviour of bootstrap DEA for a typical DMU, the latter being represented by a DMU that uses average levels of inputs to produce average levels of outputs.

[55] We will denote the fixed point or fixed DMU as (x_0, y_0) and its efficiency score as $\theta(x_0, y_0)$, following a suggestion by Prof L. Simar to avoid confusion.

[56] See for example Simar and Wilson (2004) and Kneip et al. (2008, 2011) where the fixed points lie in the middle of the input and output data.

Finally, one could choose a DMU whose efficiency lies towards the lower tail of the distribution. Considering again Figure 2.5 we deem that as long as that fixed point is not a member of the population reference set (or very close to it), then the performance of bootstrap DEA should not be considerably affected by the exact position of the fixed point. This is because a DMU which is inefficient in (most) Monte Carlo samples, it will also be inefficient with respect to the bootstrap reference sets and therefore the associated coverage probabilities should now be affected by choosing a different fixed point. To make sure that our statement is robust we included a second fixed point which uses one standard deviation of each input extra to produce the same output as the first fixed point $(x_0 + \sigma_x, y_0) = (\bar{x} + \sigma_x, \bar{y})$. The computed coverage probabilities are very close for the two fixed points, providing support to our argument; we therefore only present here the results for the fixed point $(x_0, y_0) = (\bar{x}, \bar{y})$[57]. In terms of Figure 2.5, if we think of DMU 1 as our fixed point then the second fixed point would lie towards the top right corner of the scatterplot, but not (necessarily) on the same ray as that of DMU 1. It would be interesting in the future to examine alternative fixed points that exhibit specialisation in using one of the inputs; they could be thought of as being situated towards the top-left or bottom-right boundaries of the isoquant. However, we would not expect to observe any substantial differences.

The true efficiency scores of the fixed DMU for each data generating process, along with their input and output values are presented in Table 2.4. To support the validity of

[57] The results for the alternative fixed point are available upon request by the author. The differences are so small that could be attributed to randomness.

our approach we will show how the true efficiency score of the fixed DMU can be derived on the basis of production economics while we will also prove that the DEA linear program computes the same efficiency scores as the theoretically derived ones. Without loss of generality we will perform these tasks for the 1-input/1-output case.

Table 2.4. True efficiency score and input/output values of the fixed DMU

	x1	*x2*	*y1*	*y2*	*Efficiency*
Standard 1/1	17.771		15.011		0.8447
Standard 2/1	17.768	29.594	19.373		0.8450
Standard 2/2	17.755	29.567	19.380	21.463	0.8462
Trun. Normal Low 1/1	25.459		15.046		0.5916
Trun. Normal Low 2/1	25.310	42.221	19.312		0.5912
Trun. Normal Low 2/2	25.362	42.240	19.380	21.463	0.5927
Trun. Normal High 1/1	41.903		14.990		0.3578
Trun. Normal High 2/1	42.855	71.414	19.367		0.3502
Trun. Normal High 2/2	43.130	71.731	19.422	21.506	0.3494
Uniform 1/1	22.976		15.011		0.6533
Uniform 2/1	22.944	38.218	19.377		0.6545
Uniform 2/2	23.020	38.353	19.367	21.451	0.6520

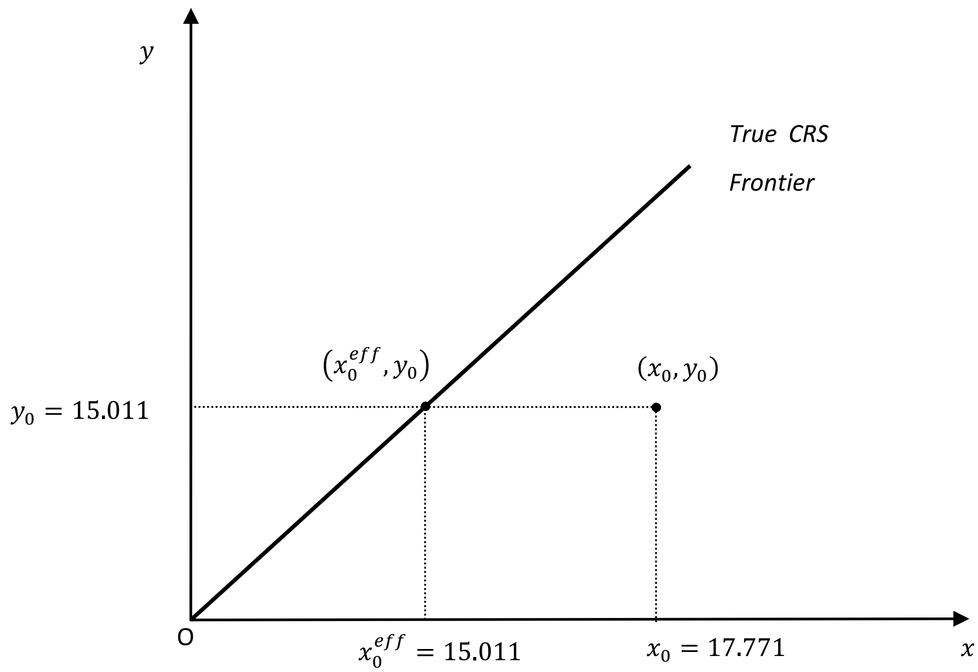
Consider the fixed DMU under the "Standard 1-input/1-output" case and under CRS. It has an input value of 17.771 and an output value of 15.011, while its true efficiency is 0.8447 based on the computations that we will now show. It is reminded that $y = x^{eff} \sim U[10,20]$ and $x_i = {x_i}^{eff} e^{0.2|v|}$, $v \sim N(0,1)$ in our case. In theory, input oriented inefficiency is defined as the horizontal distance of any DMU from the frontier, while the CRS frontier is determined by the ray which has a slope (or tangent) equal to the maximum observed average product (1-input/1-output case). The maximum average product in the population is found by:

$$AP_{max} = \max\frac{y}{x} = \max\frac{x^{eff}}{x^{eff}e^{0.2|v|}} \tag{2.32}$$

There are two equivalent ways to proceed: (i) the maximum average product is associated with efficient inputs and outputs hence $v = 0$, so $AP_{max} = 1$, or (ii) in order to maximize (2.32) and therefore $1/e^{0.2|v|}$, v has to be zero so that $AP_{max} = 1$. Therefore, in all of our 1-input/1-output cases the true frontier is defined by a 45^0 line, as in Figure 2.10 below. Then the efficient input level for the fixed DMU will be $x_0^{eff} = y_0$ and the true efficiency score will be $\theta(x_0, y_0) = x_0^{eff}/x_0 = y_0/x_0$. Hence, for the "Standard 1/1" case, the theoretically-derived, true efficiency score of the fixed DMU is $\theta(x_0, y_0) = 15.011/17.771 = 0.8447$. For the other 1-input/1-output cases the theoretically derived true efficiency is 0.5910 for "Trunc. Normal Low", 0.3577 for "Trunc. Normal High" and 0.6533 for "Uniform". In all cases the theoretical scores are equal to the efficiency scores computed by the application of DEA on the population at a 4 digit precision[58] and therefore applying DEA on the population is a valid means of determining the "true" efficiency score.

[58] As a technical note, any difference between manually-computed and DEA-computed efficiency scores is due to the randomness in generating $v{\sim}N(0, 1)$ and the fact that in the computing world, zero can only be approximated (known as machine epsilon). However, these differences are negligible.

Figure 2.10. Efficiency of the fixed DMU: illustration of the “Standard 1/1” case



We have shown that in our case the manually (or theoretically) derived efficiency scores of the population would be the same if we had applied DEA on the population. We now provide a proof for this statement for the 1-input/1-output case and under CRS[59]. The efficient frontier is defined by DMU(s) c; that is, any DMU c represents an efficient DMU $\theta_c = 1$. We also assume that $x_i = x_i^{eff} e^{u_i},\ u_i \sim iid^+,\ i = 1,2 \ldots N$ and

[59] In a previous version the notion of the “true” efficiency score of the fixed point caused confusion to Prof L. Simar in the EWPA 2014 conference. In particular, the author stated the efficiency scores were the population DEA scores. Prof Simar thought that the author was referring to sample efficiency scores as, according to Prof Simar, when referring to a “DEA score” it is not usually implied the population efficiency score as the latter is e^{-u}. In a private conversation the author explained the procedure followed in detail to Prof Simar and he agreed that the way the population or true efficiency score had been valid was valid but the exposition was confusing. We therefore decided to make clear how the population or “true” efficiency score is defined. Also, proving that applying DEA on the population yields the same efficiency score as its theoretical value (e^{-u}), we establish that our approach is valid.

that output is generated as before by the simple CRS Cobb-Douglas function[60] $y_i = x_i^{eff} = x_i e^{-u_i}$ .The efficiency score of each DMU is:

$$\theta_i = \frac{x_i^{eff}}{x_i} = \frac{x_i^{eff}}{x_i{}^{eff} e^{u_i}} \Rightarrow \theta_i = e^{-u_i}, \quad i = 1,2, \dots N \tag{2.33}$$

We will show that applying DEA on the population to compute θ_k, yields the same solution as in (2.33): $\theta_k = e^{-u_k}$. For this proof we will use both the envelopment and multiplier forms of DEA. In both cases we will need to assume that the frontier comprises a set of C efficient DMUs for which $\theta_c = 1, c = 1,2, \dots C$ and for which $u_c = 0$ and therefore $y_c = x_c^{eff} = x_c$.

Using the multiplier form in (2.10), the efficiency score of DMU k is:

$$\hat{\theta}_k = \max\{\theta = \mu y_k | v x_k = 1; \mu y_i \leq v x_i;\ v, \mu \geq 0;\ \forall\, i = 1,2 \dots, N\} \tag{2.34}$$

By definition $y_i = x_i e^{-u_i}$, while from the first restriction we get $v = 1/x_k$. Hence:

$$\hat{\theta}_k = \max\{\theta = \mu x_k e^{-u_k} | \mu x_i e^{-u_i} \leq x_i / x_k;\ v, \mu \geq 0;\ \forall\, i = 1,2 \dots, N\} \tag{2.35}$$

which reduces to:

$$\hat{\theta}_k = \max\{\theta = \mu x_k e^{-u_k} | \mu x_k \leq e^{u_i};\ v, \mu \geq 0;\ \forall\, i = 1,2 \dots, N\} \tag{2.36}$$

Since $\min(e^{u_i}) = 1$ for $u_i = u_c = 0$, the constraint in (2.36) becomes $\mu x_k \leq 1$. This suggests for the objective function that $\mu x_k e^{-u_k} \leq e^{-u_k}$ and therefore, to maximise θ the constraint needs to be binding so that $\max(\theta) = e^{-u_k}$ for $\mu^* = v^* = 1/x_k$.

Let us now consider the envelopment form in (2.11). Note that $\lambda_i > 0$ only for the efficient DMUs which constitute the set of benchmarks for DMU k (assume there are C_k

[60] It is essentially $y_i = A\left(x_i^{eff}\right)^a$ with $A = 1$ and $a = 1$.

benchmarks). Since $\lambda_i = 0$ for all other DMUs, we can disregard these for now and reformulate the constraints as follows:

$$y_k \leq \sum_{c=1}^{C_k} \lambda_c y_c \quad \text{and} \quad \theta_k x_k \geq \sum_{c=1}^{C_k} \lambda_c x_c \tag{2.37}$$

Note that both constraints need to be binding to minimise θ otherwise $\theta > \sum_{c=1}^{C_k} \lambda_c x_c / x_k$. By definition $y_i = x_i e^{-u_i}$ and $u_c = 0$, so:

$$x_k e^{-u_k} = \sum_{c=1}^{C_k} \lambda_c x_c \quad \text{and} \quad \theta_k x_k = \sum_{c=1}^{C_k} \lambda_c x_c \tag{2.38}$$

We find $x_k e^{-u_k} = \sum_{c=1}^{C_k} \lambda_c x_c = \theta x_k$, and therefore $\theta_k = e^{-u_k}$. Therefore we have proven that applying DEA on the population yields the same technical efficiency score as in the theoretical computation: $\theta_k = e^{-u_k}$.

2.8.5 Performing Monte Carlo simulations and associated issues

The procedure followed in our Monte Carlo simulations is the following:

- Use a data generating process $(\mathcal{P})$ to produce the population data (x, y) according to the specifications in subsection 2.8.2.
- Define the first DMU as the fixed point $(x_0, y_0) = (\bar{x}, \bar{y})$ (simulation assumption)
- Compute the population or true efficiency score of the fixed DMU $\theta(x_0, y_0|\mathcal{P})$ by applying the DEA linear program:

$$\theta(x_0, y_0|\mathcal{P}) = min\left\{\theta\left|y_0 \leq \sum_{i=1}^{N}\lambda_i y_i\,;\ \theta x_0 \geq \sum_{i=1}^{N}\lambda_i x_i\,;\ \theta > 0; \lambda_i \geq 0,\right.\right. \qquad (2.39)$$
$$\left.\forall\ i = 1, \dots, N\right|\mathcal{P}\Bigg\}$$

- Program the fixed DMU to appear as the first observation in every Monte Carlo replication. Hence, its input and output values will always be the same but its sample efficiency scores will be different in each Monte Carlo repetition compared to its population score ("true"). Since each of the M Monte Carlo samples can be considered as generated by a DGP $\hat{\mathcal{P}}_m, m = 1,2, \dots M$ which is an estimate of $\mathcal{P}$, the sample DEA score of the fixed DMU at the m^{th} trial $\hat{\theta}(x_0, y_0|\hat{\mathcal{P}}_m)$ will be:

$$\hat{\theta}(x_0, y_0|\hat{\mathcal{P}}_m) = min\left\{\theta\left|y_0 \leq \sum_{i=1}^{n}\lambda_i y_i\,;\ \theta x_0 \geq \sum_{i=1}^{n}\lambda_i x_i\,;\ \theta > 0; \lambda_i \geq 0,\right.\right. \qquad (2.40)$$
$$\left.\forall\ i = 1, \dots, n\right|\hat{\mathcal{P}}_m\Bigg\}$$

- For each DGP $\hat{\mathcal{P}}_m, m = 1,2, \dots M$, apply bootstrap DEA using the steps (2.19) to (2.22) in section 2.6.2 to generate a distribution of B bootstrapped scores for each $m = 1,2, \dots M$:

$$\hat{\theta}_b^*(x_0, y_0|\hat{\mathcal{P}}_m) = min\left\{\theta\left|y_0 \leq \sum_{i=1}^{n}\lambda_i y_i\,;\ \theta x_0 \geq \sum_{i=1}^{n}\lambda_i x_i^*\,;\ \theta > 0; \lambda_i \geq 0,\quad i\right.\right. \qquad (2.41)$$
$$\left.= 1, \dots, n\right|\hat{\mathcal{P}}_m\Bigg\}$$

- For each $m = 1,2, \dots M$ construct a confidence interval where $\theta(x_0, y_0|\mathcal{P})$ is expected to lie. The Simar and Wilson's (1998) confidence intervals (see Eq. (2.29)) are given by:

$$\theta(x_0, y_0|\mathcal{P}) \in \left(\tilde{\theta}^{*,(a/2)}_{(x_0,y_0|\hat{\mathcal{P}}_m)} \; , \; \tilde{\theta}^{*,(1-a/2)}_{(x_0,y_0|\hat{\mathcal{P}}_m)}\right) \tag{2.42}$$

while Simar and Wilson's (2000a) confidence intervals by (see Eq. (2.31)):

$$\theta(x_0, y_0|\mathcal{P}) \in \left(\hat{\theta}_{(x_0,y_0|\hat{\mathcal{P}}_m)} - \Delta\hat{\theta}^{*\,(1-a/2)}_{(x_0,y_0|\hat{\mathcal{P}}_m)} \; , \; \hat{\theta}_{(x_0,y_0|\hat{\mathcal{P}}_m)} - \Delta\hat{\theta}^{*\,(a/2)}_{(x_0,y_0|\hat{\mathcal{P}}_m)}\right) \tag{2.43}$$

- Use the M confidence intervals constructed by Monte Carlo to compute coverage probabilities as:

$$CP_{SW1998} = \frac{\#\theta(x_0, y_0|\mathcal{P}) \in \left(\tilde{\theta}^{*\,(a/2)}_{(x_0,y_0|\hat{\mathcal{P}}_m)} \; , \; \tilde{\theta}^{*\,(1-a/2)}_{(x_0,y_0|\hat{\mathcal{P}}_m)}\right)}{M}, \quad m = 1,2 \dots M \tag{2.44}$$

for the Simar and Wilson (1998) intervals (2.42) and for the Simar and Wilson's (2000a) confidence intervals (2.43):

$$CP_{SW2000} = \frac{\#\theta(x_0, y_0|\mathcal{P}) \in \left(\hat{\theta}_{(x_0,y_0|\hat{\mathcal{P}}_m)} - \Delta\hat{\theta}^{*\,(1-a/2)}_{(x_0,y_0|\hat{\mathcal{P}}_m)} \; , \; \hat{\theta}_{(x_0,y_0|\hat{\mathcal{P}}_m)} - \Delta\hat{\theta}^{*\,(a/2)}_{(x_0,y_0|\hat{\mathcal{P}}_m)}\right)}{M}, \quad m = 1,2 \dots M \tag{2.45}$$

In performing the simulations we encountered two minor issues that required some light interventions in the codes to help the simulations run, which do not affect the validity of our results. However, they might be of interest to researchers or practitioners.

The first one concerns the Sheather-Jones (1991) smoothing procedure (SJ) which would not yield a solution in a few occasions. The problem is that the differential equation solving process could not converge to a solution after a number of iterations.

The source of the problem was purely data-driven[61] and we therefore decided to substitute in these few cases the smoothing parameter with one derived from the least squares cross validation process (LSCV). We could have alternatively omitted these few cases from our results, but it would require a substantial investment in programming time while the difference in results would be negligible, given that in many occasions the LSCV and SJ smoothing parameters are very close to each other. The number of "SJ discrepancies" is presented in Table 2.5 below, for each combination of data generating process and input-output combination. We observe that in most cases no such discrepancy occurred or less frequently there were 1 or 2 among the 1000 Monte Carlo repetitions. Then there were 5 cases where the number of discrepancies was higher, all of which observed in very small samples (mainly 10 and 15). This suggests that our interventions have not affected results and that perhaps this failure of the SJ smoothing process is limited to very small samples.

[61] The author performed a small experiment on this issue. In particular he used the data of the samples exhibited these discrepancies. By trying different values for the numbers of iterations no result was reached, indicating that the problem was caused most likely by the specific data used. The author did not look further into this issue by trying different numerical approximation methods, but it seems more likely that there is an incompatibility between the specific "problematic" data sets and the SJ method.

Table 2.5. Number of SJ discrepancies

	n = 10	*n* = 15	*n* = 20	*n* = 25	*n* = 30	*n* = 60	*n* = 120
Standard 1-1	9	3	1	2	2	1	0
Standard 2-1	1	7	4	2	0	1	0
Standard 2-2	2	2	2	1	1	0	0
Trun. Normal Low 1-1	1	0	0	0	0	0	0
Trun. Normal Low 2-1	1	1	0	0	0	0	0
Trun. Normal Low 2-2	3	1	0	0	1	0	0
Trun. Normal High 1-1	0	0	0	1	0	0	0
Trun. Normal High 2-1	1	0	0	0	0	0	0
Trun. Normal High 2-2	2	0	0	0	1	0	0
Uniform 1-1	2	0	0	0	0	0	0
Uniform 2-1	0	0	0	0	0	0	0
Uniform 2-2	1	0	0	0	1	0	0

Another minor issue relates to the code modification for the LSCV method when applied in large samples. To avoid "out of memory"[62] problems we changed the precision of real values to "single" from "double", which means that we changed the number format from 15 decimal places to 7. The value of the smoothing parameter is determined by a "grid-search" algorithm that searches for the value of the smoothing parameter that minimizes the value of a function of interest[63]. This procedure considers a range of values for the smoothing parameter from 0 to 1, moving from one value to the next at a certain "step". The reduced precision resulted in a few situations where two consecutive values of the smoothing parameter were associated with the same

[62] The "out of memory" message appears in computing when the available memory of the computer is not adequate to perform an operation. This occurs when the number of elements or the size of a vector exceed some limit which depends on the characteristics of the PC. The usual approach is to reduce the size of the problematic elements by various techniques (such as partition) where possible (not here), to reduce the memory allocation for each element (done here by transforming numbers to have single precision) or to increase the random access memory (RAM) of the computer (not possible at this stage).

[63] This function actually measures the mean integrated squared error (MISE). For more details see equation (I.14) in the Appendix.

minimizing value for the aforementioned function of interest, returning as a solution two smoothing parameters. In these few situations we used the smaller of the two, which is highly unlikely to affect the validity of our results.

2.9 Monte Carlo Results: small samples

The performance of bootstrap DEA and the behaviour of the associated confidence intervals, as already mentioned, is characterized by 4 aspects which will be examined in the following subsections: (i) the equality of bootstrap and DEA biases, (ii) convergence of coverage probabilities to their nominal values, (iii) the behaviour of confidence intervals, and (iv) the distributional aspects of bootstrapped efficiency scores. Subsections 2.9.2 to 2.9.5, thus, present results that correspond to these four aspects. The first subsection, though, tries to address the question of identifying the correct population DGP using sample data which would be useful to practitioners.

2.9.1 Identifying the population DGP from the data

The identification of the underlying population DGP using sample data is not an easy task; especially when the sample distributions are not similar to the population ones. That is, it is not necessary that the distribution of efficiency scores in each sample will always have the same properties as the ones of the underlying population. However, this is a well-known issue in statistical inference and therefore the assumption that the observed sample is a "representative" one is implicit. This means, that both in general statistical applications and in bootstrap DEA, in particular, we hope (and assume) that the observed sample is a good representation of "reality". In this subsection we will examine if such an assumption is plausible in the case of DEA using simple diagnostics, while we will argue that theoretical intuition could be useful in assuming a valid population DGP (as in statistical modelling[64]).

We will first discuss how theoretical intuition can help identifying the underlying population DGP. In subsection 2.8.3 we attached an economic interpretation to each DGP. The first one ("Standard") was argued to be associated with (perfect) competition or monopolistic competition in the long run, whereas the second one ("Trun.Normal Low") was linked to monopoly and technological heterogeneity. It is reasonable to assume that the practitioner knows which of these two cases applies to the sample under examination and therefore infer the correct DGP. This information could be either knowledge of the market under which the DMUs operate, knowledge of the operations of each DMU, or it could be in the form of studies on the industrial organisation or competitive conditions of the market under examination. Especially for the case of technological heterogeneity, this could be easily detected by inspecting data as there should be substantial differences in the proportions of outputs to inputs among DMUs.

[64] It is not within the scope of this subsection to analyse these principles but the interested reader is directed to any introductory textbook in statistics. For example, one principle of sampling is that data should not be collected from certain clusters of the population if the statistical question in hand concerns the whole population. Another example concerns employing distributional assumptions which have some theoretical basis (such as the assumption of normality for financial stock returns).

In addition, careless data selection can also lead to distributions which look similar to those under technological heterogeneity[65]. We would therefore like to highlight the importance of inspecting the sample efficiency distributions and comparing them with what was expected to be observed; if expectations are not realised then the data should be looked at again.

Despite theoretical intuition is clear in these two cases, it is less so in the other two. In particular, it is challenging to identify the exact conditions under which we could detect the third case (“Trun.Normal High”) in sample data. That is, the practitioner cannot easily recognise the conditions under which the sample data can be associated with medium-run monopolistic competition where the market is in transition (it is becoming increasingly competitive). Regarding the last case (“Uniform”) there is no economic interpretation and the DGP is only used for experimental purposes so we do not need to comment on that.

Let us now examine if we could use some simple diagnostics to perform the same task. Since the samples are drawn from a population distribution, comparing the moments of the population and sample distributions could be informative, especially if the latter have unique patterns which could help identifying the underlying DGPs. We argue that this information cannot be found in the measures of central tendency and

[65] For example, mixing commercial banks and state development banks in the same sample and applying DEA using the intermediation approach, would most probably make development banks look much more efficient as the proportions of their deposits compared to loans is much lower compared to those for commercial banks. This is due to the fact that state development banks fund national projects (among others) while they do not (need to) perform commercial deposit operations as their liquidity is injected by the central bank. In the empirical application to the Greek banking sector we will illustrate the implications of such a “malpractice” for DEA, which can be extended to bootstrap DEA.

dispersion. We support this by the fact that one could have generated different DGPs but with similar means and standard deviations. On the other hand, the higher moments (skewness and kurtosis) carry information about the shape of the distribution and it seems reasonable to use these instead. Thus, we will compare the skewness and kurtosis of each population with those of the generated samples[66] and we will try to associate observable patterns to certain DGPs.

Table 2.6 reports the values of skewness and kurtosis for each population and for different sample sizes. We only report here the case of 2-inputs/2-outputs as the dimensions do not affect (and are not relevant to) the identification of the population DGP from the sample[67]; this is because the shape of the distribution is not affected. Finally, we need to underline that the discussion is relevant to the input-oriented efficiency scores under CRS. However, it should be straightforward for the practitioner to perform this simulation exercise (of comparing skewness and kurtosis) for different models.

[66] To be precise, we will use the medians of the samples' skewness and kurtosis values as there are $M = 1000$ samples generated. Hence, the reported sample values for the higher moments can be thought of as the ones of a "typical" sample for each DGP.

[67] In Chapter 4 we propose a method which we call "Moments Bootstrap" and all moments for all cases and dimensions are reported there. However, this is neither necessary nor relevant here.

Table 2.6. Identifying underlying DGP using skewness and kurtosis

	Standard		Trun. Normal Low		Trun. Normal High		Uniform	
Population	**Skew**	**Kurt**	**Skew**	**Kurt**	**Skew**	**Kurt**	**Skew**	**Kurt**
N = 10,000	-0.675	2.893	0.412	3.003	0.284	2.074	0.286	1.885
Samples	**Skew**	**Kurt**	**Skew**	**Kurt**	**Skew**	**Kurt**	**Skew**	**Kurt**
n = 10	-0.494	2.117	0.146	2.125	0.347	1.921	0.250	1.829
n = 15	-0.472	2.185	0.240	2.266	0.355	1.975	0.256	1.855
n = 20	-0.501	2.235	0.268	2.330	0.303	1.947	0.307	1.862
n = 25	-0.536	2.344	0.270	2.407	0.338	2.000	0.304	1.898
n = 30	-0.519	2.370	0.317	2.505	0.321	2.026	0.314	1.887
n = 60	-0.597	2.579	0.356	2.676	0.309	2.033	0.293	1.888
n = 120	-0.650	2.753	0.381	2.802	0.305	2.066	0.291	1.889

The first thing to observe is that in all cases the higher moments of the samples are close to the population ones and they converge as sample size increases. Furthermore, we indeed observe patterns which can help identifying the population DGP using the sample skewness and kurtosis. We have to note, however, that the observed patterns are easier and safer to distinguish for sample sizes above 30 observations. In particular, the “Standard” DGP is associated with negative skewness of about -0.6 and with kurtosis close to 3 (it roughly ranges from 2 to 3). In the case of “Trun.Normal Low” we observe small positive skewness (around 0.4) and similar kurtosis as in the previous case (close to 3 and roughly ranging from 2 to 3). In the case of “Trun.Normal High” we observe small positive skewness (around 0.3) and kurtosis around 2 (that roughly ranges from 1.9 to 2.1). The case of “Uniform” is only presented for reference as it is not likely to be met in practice; we can observe, though, that it distinguishes from the others as it is associated with skewness and kurtosis which are smaller by 0.2 units compared to the “Trun.Normal High” case. Although the difference sounds small, this combination would yield a noticeably flatter distribution.

The discussion of the values in Table 2.6 suggests that if one plotted a histogram of the sample efficiency scores, it would be quite similar to that of the corresponding population. This implies that either by visual inspection of the histograms or by computing skewness and kurtosis, the practitioner should be able to associate the sample data to the true DGP. However, it would be safer to use the suggested diagnostics for sample sizes above 30. We would also like to suggest that in empirical work both theoretical intuition and inspection of histograms and higher moments is employed to reach safer conclusions.

2.9.2 Bootstrap and DEA biases

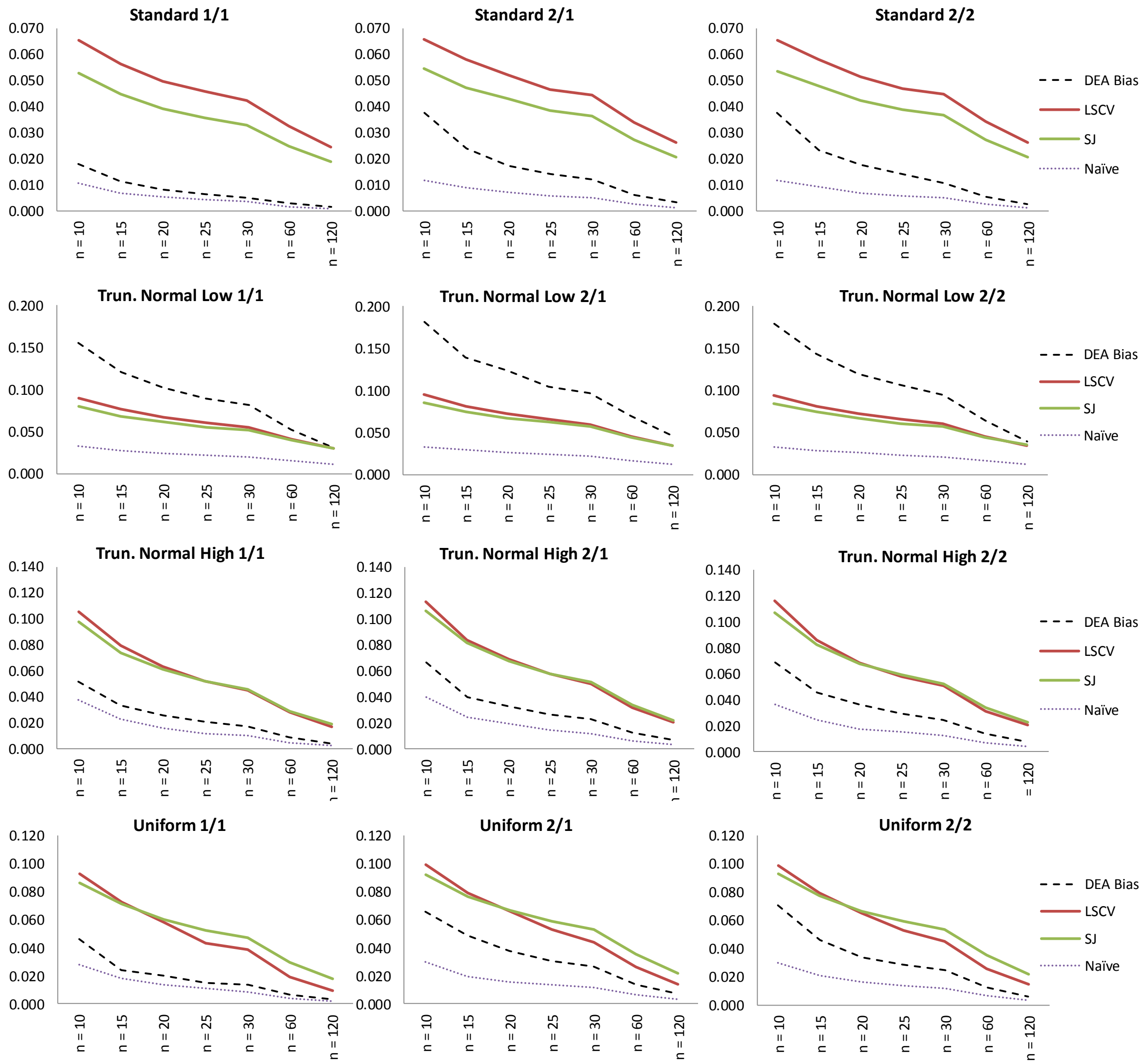
The equality of the bootstrap and DEA biases is examined in this subsection. Figure 2.11 presents these biases for each DGP and for all bootstrap procedures: LSCV (least squares cross-validation), SJ (Sheather-Jones plug-in estimator) and the naïve bootstrap. Each row of Figure 2.11 presents results for the different population assumptions and each column for the three different model dimensions. In each subplot, the DEA bias (or model or “true” bias) is depicted by the black dotted line, the LSCV-smooth bootstrap bias is given by the solid magenta line, the SJ-smooth bootstrap bias by the solid green line while the naïve bootstrap bias is presented by the thin dotted grey line.

The general finding is that for small samples the two biases are not equal, suggesting that the relevant assumption in (2.28) is not plausible for the cases examined. Perhaps larger samples than 120 would be required for this assumption to work, but such a statement should be examined in more depth. However, there is a clear tendency for all biases to converge to zero asymptotically which confirms the consistency of the method: $\left(\hat{\theta}_k^* - \hat{\theta}_k\right)|\hat{\mathcal{P}} \overset{a}{\rightarrow} \left(\hat{\theta}_k - \theta_k\right)|\hat{\mathcal{P}}$. Another way to look at convergence is by considering the ratio of the two biases

$$\left(\hat{\theta}_k^* - \hat{\theta}_k\right)|\hat{\mathcal{P}}/\left(\hat{\theta}_k - \theta_k\right)|\hat{\mathcal{P}} \simeq 1 \tag{2.46}$$

The reason we include this in the discussion is because in some cases the ratios of bootstrap to DEA bias diverge instead of converging to 1; although we expect (2.46) to apply asymptotically. Graphically we observe in some cases that both biases fall and the difference between the two becoming smaller which is in support of the assumption of the equal biases; however, a closer inspection will reveal that (2.46) does not apply. The implication of this is that coverage probabilities fall as sample size increases as we will see in the next subsection; this does not invalidate, though, the consistency of the method but it suggests that its applicability in small samples needs to be well-considered.

Figure 2.11. Bootstrap and DEA biases



Regarding the effect of dimensions we find that biases increase with the number of input and output variables. In some cases the increase is more pronounced and in other cases less so. It is worthwhile noting, though, that dimensionality affects mostly the DEA biases as the effect on bootstrap biases is so small in some cases that one could argue that it is due to randomness. We could state however that as the dimensions increase the biases increase suggesting that larger samples are probably required to make the assumption of equal biases plausible.

Finally, considering the two smoothing methods, we would suggest that in cases such as the “Standard”, which is associated with perfect competition, the SJ is clearly superior to LSCV while in all other cases (monopoly, monopolistic competition, unclear market structure) LSCV performs, in principle, better that SJ. With regards to the naïve bootstrap, it is clear that in all cases the DEA bias is greater than the bootstrap bias, which we will see later that plays an important role in the performance of bootstrap DEA. The very small bootstrap bias is not surprising as the naïve bootstrap resamples from a discrete distribution and therefore the majority of the bootstrapped efficiency scores are equal to the DEA score (which is the main reason why the naïve bootstrap is considered inconsistent). The interesting observation, though, is that smoothing the empirical distribution seems to generate bootstrap biases which are considerably greater than the naïve bootstrap bias and in some cases well–above the DEA bias. This confirms Simar and Wilson (2002) who stated that smoothing the empirical distribution can introduce additional noise in the bootstrap. We certainly do not suggest that the naïve bootstrap should be preferred as its inconsistency has been well-documented in the literature; it seems reasonable, that research should focus on approaches that bring the two biases close to each other[68].

[68] This is the motivation for the “moments bootstrap” that we propose as an alternative to the smooth bootstrap in chapter 4.

Comparing our results with the simulations in Simar and Wilson (2000b, 2004) we find that the behaviour of the bootstrap and DEA biases is similar, although in our case the bootstrap biases fall with a slower pace which is most probably due to the different data generating processes used and to some extent possibly due to the different orientation used[69]. Moreover, in Simar and Wilson's (2004) simulations the ratio of the bootstrap to DEA bias is monotonically converging to one which explains the observed well-behaved coverage probabilities. The examination of every case in isolation does not lead to substantially different conclusions compared to the general ones that we have already mentioned. However, there are some interesting features associated with each DGP which we will now discuss.

The "Standard" cases exhibits the most pronounced absolute differences between bootstrap and DEA biases. Especially if we consider the ratio of the two biases as in (2.46), this increases from a value of 2 for $n = 10$ to about 15 for $n = 120$; and this is observed in all dimensions examined. We will see in the next subsection that this causes coverage probabilities to decline as sample size increases. Apart from attributing these findings to the assumed DGP, we could state that the slower declining bootstrap bias could be due to smoothing (Simar and Wilson, 2002). On the other hand the observed DEA bias is substantially smaller compared to other cases, suggesting that the observed DEA scores are not far from the population ones.

[69] Although it has not been explored in the literature, there is a good chance that input orientation (used here) to be associated with narrower confidence intervals as the support of efficiency scores is $(0,1]$, while in output orientation it is $[1,\infty)$. The author believes that this richer support of output orientation might allow DEA to converge faster and to produce confidence intervals with higher coverage probabilities. The validity of this argument should be explored in the future with further simulations.

The “Trun.Normal Low” case exhibits particular interest because of the fact that it is associated with technological heterogeneity as already mentioned. At a first glance, one might be tempted to conclude that the biases converge as sample size increases, while when $n = 120$ they seem to be very close to each other. Especially since the ratio of the two biases converges monotonically to 1, exhibiting similar behaviour to that in Simar and Wilson (2004). However, looking at the behaviour of the biases in larger samples (see subsection 2.10) we cannot conclude that a sample size of 120 or greater will yield good results as the DEA bias keeps converging fast to zero for $n > 120$ while the bootstrap bias converges slowly (which again might be due to smoothing). We document that the technological heterogeneity introduces a substantial DEA bias which confirms our previous concerns that even applying DEA in such cases might not be a great idea. And given the fact that the DEA bias is considerably underestimated by the bootstrap bias, and underestimated after some point, the use of bootstrap DEA is not suggested in these cases as its performance is hard to evaluate. Regarding dimensionality, it only slightly introduces an increase in the DEA and bootstrap bias.

In the “Trun.Normal High” case, we evidence a similar behaviour as in the “Standard” case with the difference that in the latter case the magnitude of the biases is smaller. The biases increase with dimensions and with regards to the bootstrap biases they are almost identical for both smooth bootstraps. The DEA bias seems to converge faster than the bootstrap biases but only slightly, suggesting that the assumption of equal biases holds better compared to the “Standard” case; yet, we could not consider that the two biases are equal.

The “Uniform” case does not exhibit particular economic interest but simulation-wise it offers well-behaved results compared to the other cases. It seems that the bootstrap bias (especially under LSCV) converges to the DEA bias (and to zero) as sample increases and this improves even more in larger samples. Despite the fact that assumption of equal biases seems more plausible in this case, it still doesn’t hold and this might affect coverage probabilities. Regarding dimensionality, there seems to be a small effect when moving from 2 variables to 3, but the effect is quite smaller when moving from 3 to 4.

To summarize, the assumption of equal DEA and bootstrap biases does not hold in small samples in the cases examined. Other times it fails considerably and other less so; this is to be determined by the associated coverage probabilities examined in the next subsection. Perhaps, larger samples are required or the assumption might only apply asymptotically when both biases are equal to zero. Information on larger samples will provide useful information and will be presented later in this chapter, while it would be interesting in the future to perform the same exercise under alternative DGPs and assumptions on RTS and orientation.

2.9.3 Coverage probabilities

The results on coverage probabilities are presented in this subsection and are summarised in Table 2.7. To conserve space we only report coverage probabilities for Simar and Wilson’s (1998) 95% confidence intervals (SW1998) and for Simar and Wilson’s (2000a) 95% confidence intervals (SW2000), for all DGPs and sample sizes and for the 2-inputs/2-outputs dimension. Results for other levels of significance (20%, 10%, 5% and 1%) and dimensions can be found in Appendix II. Monte Carlo experiments were performed for both LSCV and SJ smooth bootstrap procedures as well as for the naïve bootstrap. The coverage probabilities for the naïve bootstrap are only provided for information and carry no implications for the performance of bootstrap DEA[70]. It is worthwhile noting, though, that they are very similar to the ones reported in Simar and Wilson (2004), which provides support to the fact that our computations are correct.

The overall evaluation of the finite sample performance of bootstrap DEA suggests that Simar and Wilson's (1998 and 2000a) confidence intervals cannot be safely used in small samples. In particular, we do not observe any convergence of coverage probabilities to their nominal values, apart from few cases where coverage probabilities are relatively close to the nominal ones. For example, under "Trun.Normal High", which is associated with monopolistic competition, we find relatively good performance using the SW1998 intervals and for sample sizes of 30 or less. But it would not be convincing to generalise such a result.

[70] It is reminded that the naïve bootstrap produces distributions with peculiar properties and the resulting confidence intervals are inconsistent.

Table 2.7. Coverage of SW1998 and SW2000 95% confidence intervals

	Standard 2/2		T.N. Low 2/2		T.N. High 2/2		Uniform 2/2	
LSCV	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 10	0.743	0.563	0.389	0.517	0.874	0.698	0.755	0.659
n = 15	0.574	0.401	0.385	0.500	0.828	0.621	0.776	0.601
n = 20	0.473	0.325	0.433	0.514	0.819	0.569	0.733	0.581
n = 25	0.421	0.302	0.441	0.511	0.811	0.513	0.745	0.574
n = 30	0.342	0.253	0.446	0.510	0.810	0.511	0.734	0.557
n = 60	0.226	0.151	0.497	0.528	0.690	0.407	0.739	0.494
n = 120	0.148	0.094	0.571	0.576	0.577	0.300	0.756	0.461
SJ	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 10	0.830	0.649	0.363	0.513	0.898	0.712	0.817	0.663
n = 15	0.764	0.498	0.387	0.487	0.920	0.592	0.862	0.605
n = 20	0.670	0.393	0.436	0.496	0.916	0.533	0.833	0.502
n = 25	0.566	0.315	0.434	0.513	0.889	0.486	0.825	0.450
n = 30	0.466	0.227	0.434	0.515	0.873	0.444	0.800	0.432
n = 60	0.165	0.079	0.512	0.525	0.722	0.300	0.593	0.249
n = 120	0.022	0.009	0.589	0.584	0.492	0.158	0.412	0.160
Naïve	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 10	0.346	0.727	0.201	0.441	0.516	0.852	0.459	0.780
n = 15	0.405	0.771	0.215	0.437	0.527	0.847	0.447	0.799
n = 20	0.403	0.763	0.259	0.477	0.501	0.819	0.487	0.820
n = 25	0.430	0.791	0.238	0.481	0.528	0.835	0.515	0.853
n = 30	0.475	0.809	0.257	0.490	0.512	0.839	0.506	0.843
n = 60	0.459	0.809	0.323	0.553	0.507	0.842	0.556	0.864
n = 120	0.424	0.841	0.348	0.615	0.527	0.872	0.555	0.860

In addition to the low coverage, the behaviour of probabilities is not steady in that the reported values may change non-monotonically with sample size. One exception is the "Trun.Normal Low" case where convergence is monotonic for samples up to $n = 120$ (that is, coverage probabilities increase with sample size) and where the behaviour of the bootstrap and DEA biases is similar to that in Simar and Wilson (2004). However, apart from the low coverage probabilities reported, this DGP is associated with technological heterogeneity and it is not a good idea to apply even simple DEA. Therefore we deduce that, based on coverage probabilities and on the particular DGPs examined, it is not safe to apply bootstrap DEA in small samples; at least not as they were proposed by Simar and Wilson (1998, 2000a).

Let us now try to gain a deeper insight about the behaviour of coverage probabilities. One interesting observation is that in some cases they fall as sample size increases. This can be justified by (i) the behaviour of the ratio of the bootstrap bias over DEA bias (2.46), which is observed to increase in all cases except for "Trunc.Normal Low", and (ii) by the fact that as sample size increases the estimated confidence intervals become narrower but targeting at a different efficiency score than the true one due to the persistent mismatch between the bootstrap and DEA biases[71]. The latter is graphically represented in Figure 2.12 and Figure 2.13 in the next subsection.

Among the factors that affect coverage probabilities, the most important ones are the DGP and the choice between the SW1998 and SW2000 confidence intervals. The smoothing process seems to play a role as the LSCV method seems to be associated with higher (in most cases) and more stable coverage probabilities with the exception of the "Standard" case where SJ performs better. Finally, model dimensions, in principle, affect coverage probabilities; though to a small extent. However, they do not always decrease with model dimensions, although in bigger samples we observe this pattern more consistently (see Appendix II).

Regarding the choice between SW1998 and SW2000 intervals we find an interesting pattern: we observe that when the bootstrap bias is greater than the DEA bias, the

[71] We show in the next chapter that both SW1998 and SW2000 include the bias corrected estimate $\tilde{\theta}_k^* = \theta_k$ from (2.26). As they both become narrower with sample size, this suggests that coverage will only be high if the assumption of equal biases (2.28) is satisfied and hence the intervals lie about $\tilde{\theta}_k^* = \theta_k$.

SW1998 intervals perform better, while the opposite is true when the DEA bias is greater than the bootstrap bias. Hence, the SW2000 are associated with higher coverage probabilities under all naïve bootstraps and under the "Trun.Normal Low" case (technological heterogeneity), which are both cases for which we have expressed concerns about their applicability with bootstrap DEA. One might argue that this result is specific to the simulations examined here, but we show in Appendix IV that it can be generalised to a good extent. In particular, we show that the SW2000 intervals perform better than the SW1998 intervals only if the DEA bias is greater than the bootstrap bias[72]. This is confirmed in all of our simulations while it is important to note that in all simulations of Simar and Wilson (2000, 2004) the DEA bias is always greater than the bootstrap bias, explaining the high coverage probabilities reported there.

The results of this subsection have indicated that the coverage probabilities in all cases are not as high as the nominal ones in small samples, providing further support to our suggestion in the previous subsection that bootstrap DEA might not be always applicable in small samples. The factors affecting coverage probabilities are mainly the DGP and the confidence intervals used, while dimensionality or the smoothing technique used were found to be less impactful. Perhaps the most interesting finding, which applies more generally, is that the SW1998 intervals seem to perform better compared to the SW2000 intervals (with the exception of a few cases for which we are

[72] We also require that the bootstrap distribution is positively skewed which is observed in all of our simulations. Obviously under positive skewness and greater bootstrap DEA bias the SW1998 intervals perform better. If there is no skewness (the distribution is symmetrical) then both intervals perform equally well.

concerned about applying bootstrap DEA) and that it should be carefully consider carefully whether the latter should be used.

2.9.4 Bootstrap confidence intervals

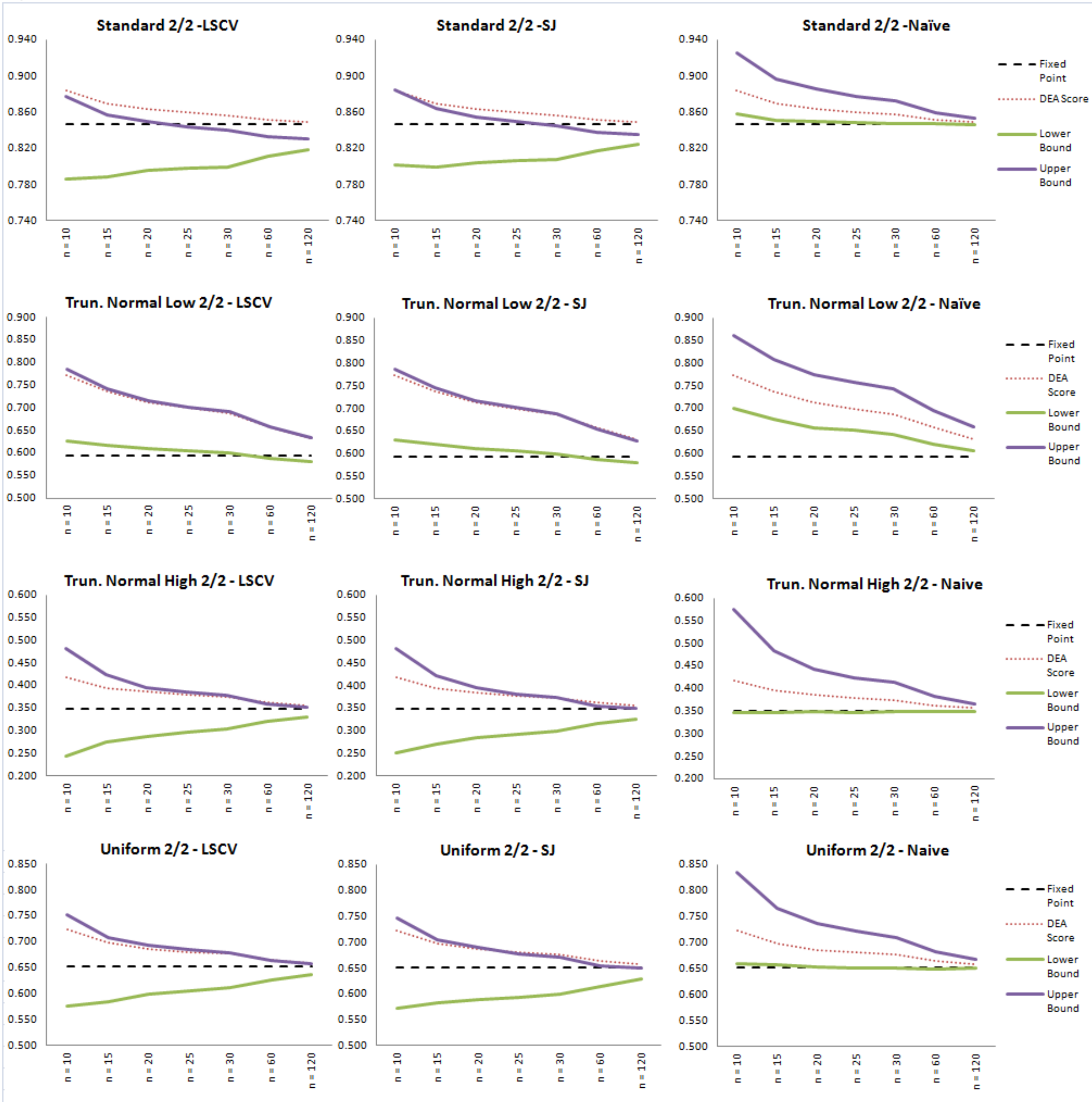
The results on coverage probabilities are further explained in this section which analyses the behaviour of the confidence intervals that correspond to the coverage probabilities presented in Table 2.7. Figure 2.12 and Figure 2.13 plot the average lower (green solid line) and upper (purple solid line) bounds of the 95% SW1998 and SW2000 intervals, respectively, along with the true efficiency score (black dotted line) and average DEA score (magenta dotted line) in the Monte Carlo simulations. The labels on each graph indicate the DGP and smoothing process considered. To conserve space the discussion is based on the 2-inputs/2-outputs cases while results for all cases can be found in Appendix III.

In all cases the intervals exhibit a behaviour which is in accordance with the coverage probabilities in Table 2.7. That is, the highest coverage probabilities correspond to cases where θ_k is better centred by the intervals. Moreover, we observe both in Figure 2.12 and Figure 2.13 that the intervals are wider for small samples and become narrower as the sample size increases. However, in some cases they narrow down towards a different fixed point than θ_k but there seems to be a tendency for this to be corrected asymptotically. In subsection 3.2.3 of the next chapter we show that both intervals include $\overline{\tilde{\theta}_k^*} = \overline{\hat{\theta}_k^*} - 2\widehat{bias}_k$. It seems that as n increases and the confidence intervals become very narrow (targeting $\overline{\tilde{\theta}_k^*}$), the bootstrap intervals will perform better if $\overline{\tilde{\theta}_k^*} = \theta_k$ which can only happen if the bootstrap and DEA biases are equal; otherwise θ_k will be either overestimated or underestimated. This provides more insight into the falling coverage probabilities that we observed in the previous section.

In any case we cannot safely conclude that bootstrap DEA can be applied in small samples as the behaviour of the intervals is not "steady" as it changes with sample size. Regarding other factors that affect the intervals, we observe that width slightly increases with dimensions while the smoothing process has a smaller effect on width (with the exception of the inconsistent naïve bootstrap). Once again, the most important factor that affects the behaviour of the intervals is the assumed DGP yielding either relatively narrow intervals ("Standard" case) or substantially wider ones ("Trun.Normal High" case) or even dislocated ones ("Trun.Normal Low" case).

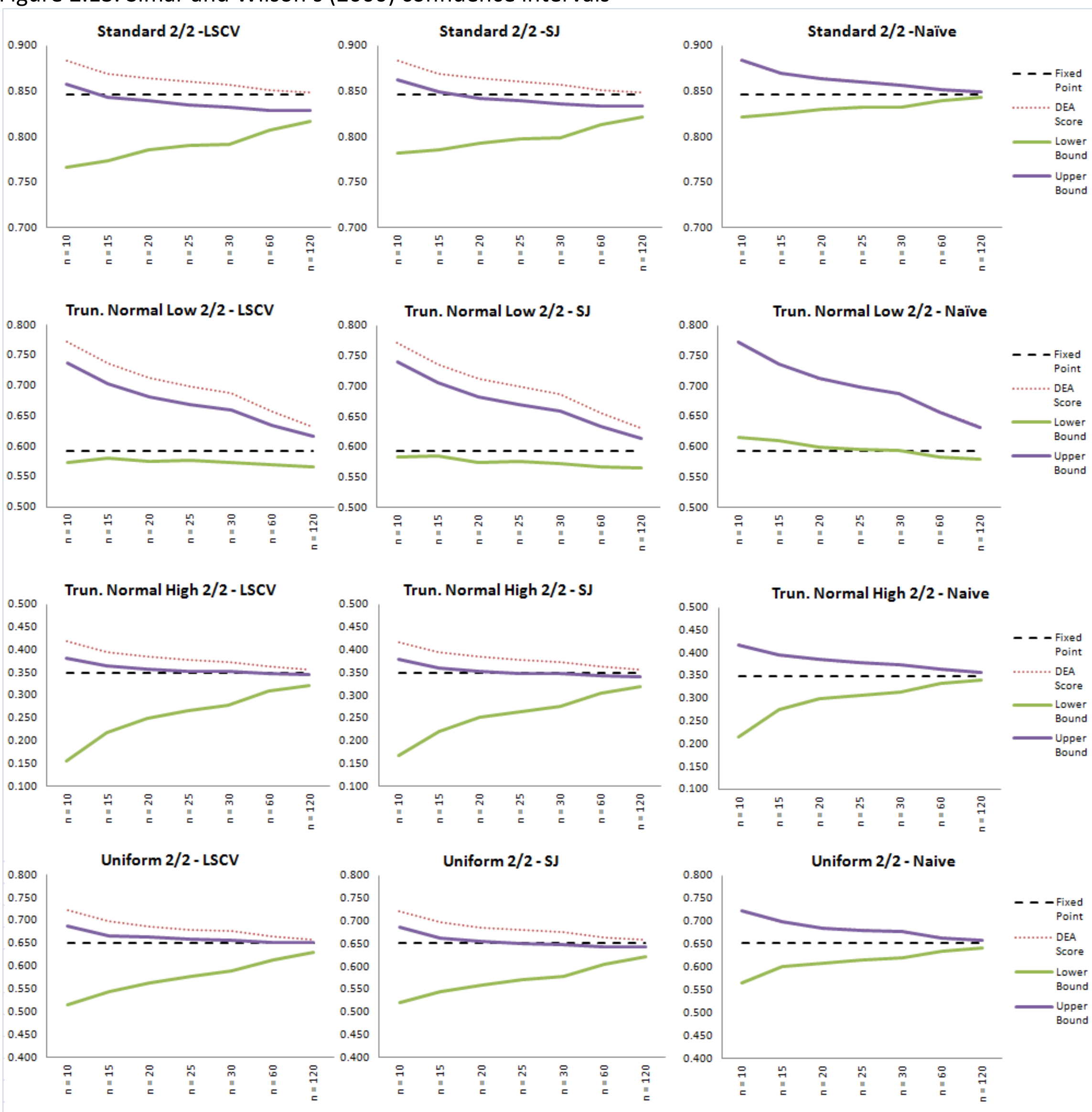
The SW1998 intervals in Figure 2.12 seem to underestimate θ_k in all cases except for the "Trun.Normal Low" case and the naïve bootstraps. This is not surprising as θ_k is underestimated when the bootstrap bias is greater than the DEA bias while it is overestimated in the opposite case. This is more pronounced for the "Standard" case where the DEA bias is very small compared to the other cases while the bootstrap bias is proportionately quite bigger. In all other cases where the DEA bias is greater than the bootstrap bias, the SW1998 intervals overestimate θ_k.

Figure 2.12. Simar and Wilson's (1998) confidence intervals



Similarly, Figure 2.13 provides information for the SW2000 intervals. We observe that when the bootstrap bias is greater than the DEA bias, the intervals underestimate θ_k but to a greater extent compared to the SW1998 intervals. This is in accordance with the discussion in the previous subsection where the respective coverage probabilities where lower. On the other hand, when the DEA bias is greater, θ_k is in principle overestimated to a lesser extent compared to the SW1998 intervals, explaining the higher coverage probabilities.

Figure 2.13. Simar and Wilson's (2000) confidence intervals



Comparing the two figures above, we confirm that the SW2000 intervals will perform better compared to the SW1998 intervals only if the DEA bias is greater than the bootstrap bias. _*This explains the differences in our results with those of Simar and*_

Wilson (2000a, 2004): in our simulations the DEA (or true) bias is always smaller than the smooth bootstrap bias (with the exception of the "Trun.Normal Low" case) whereas in their simulations the opposite is true. Therefore, our results are in accordance with Simar and Wilson (2000a, 2004) and they also indicate a case where SW2000 intervals might underperform[73].

We should also note that, as previously discussed, the bias in the "Trun. Normal Low" case is due to technological heterogeneity which is not desirable. If we accept that large DEA biases are associated with such sample heterogeneity, then the SW2000 intervals have better chances to perform well in cases where DEA might not be a good idea to apply. Especially if we consider the fact that theoretical works have focused on the convergence and consistency of DEA (Kneip et al., 1998; Korostelev et al., 1995)[74], we deduce that small and fast declining DEA biases are desirable and that the opposite

---

[73] This point is intentionally highlighted and underlined as it addresses the most important line of criticism of Prof L. Simar against a paper presented by the author at the EWEPA (2013) conference (Helsinki) with title "The Simar and Wilson's bootstrap DEA: a critique". Prof L. Simar suggested that the observed differences in the results were due to programming mistakes or some misunderstanding of the bootstrap procedure. Our analysis shows that the observed differences are purely due to the DGP used here which generates bootstrap biases that are, in most cases, larger than the DEA biases and for which cases the SW2000 have been shown to underperform. Moreover, bootstrap DEA performs as expected, suggesting that there is no programming mistake. Another interesting fact is that in these cases SW1998 intervals perform better and the author feels that in the simulations in Simar and Wilson (2004) the corresponding SW1998 intervals would perform worse if this exercise had been conducted. Therefore, our results are not in contrast with those of Simar and Wilson (2004) but actually in accordance. Moreover, we indicate cases where either SW1998 or SW2000 intervals might not perform well. It is therefore a case for future research to find DGPs which will balance the ratio of the DEA bias to bootstrap bias with the latter being smaller and explore the conditions in the input/output relations that help generate these conditions; however this does not seem to be a straightforward exercise on a theoretical basis. Most importantly, we need to explore the market structures that would be associated with slightly larger (if not equal) DEA biases compared to bootstrap biases and attach an economic interpretation as we have done in our experiments.

[74] Note that in both papers Prof L. Simar is a co-author which underlines his interest towards the minimization of DEA bias.

should be avoided. This puts serious thoughts on whether the well-established SW2000 confidence intervals should be preferred over the SW1998 ones.

A reasonable question to ask is how we could know upfront whether the DEA bias is greater than the bootstrap bias or not. Our simulations suggest that when the distribution of efficiency scores has a relatively thin tail towards 1 and when values are concentrated symmetrically well below 1 (as in the "Trun.Normal Low" case) then the DEA bias tends to be bigger. It is quite obvious that the DEA bias under "Trun.Normal Low" is greater than the bias in "Trun.Normal High", which in turn is greater than in the "Standard" case. In each of the aforementioned cases the efficiency scores are increasingly concentrated towards 1 and the shape of the distribution transforms to a half-normal one. This suggest that the distribution of the DEA scores can serve as an indication of whether the bootstrap bias is greater than the DEA bias or not (at least under the smooth bootstrap procedures under consideration). We have already discussed in subsection 2.9.1 that skewness and kurtosis can serve as diagnostic tools in identifying the underlying DGP and we can therefore also use them here as an indication of whether the bootstrap bias is greater than the DEA bias or not.

An alternative approach would be to "bootstrap the bootstrap" in the spirit of the iterated bootstrap proposed in Simar and Wilson (2004). We have seen that DEA generates sample distributions which are similar to the population ones. By iterating the bootstrap we would generate samples from the bootstrapped DEA scores and we could then compare the double-bootstrap bias with the single-bootstrap bias. That would mimic the relationship between the bootstrap bias and the DEA bias and we could therefore use the iterated bootstrap as a diagnostic tool. On the downside, and as already discussed, this approach is extremely costly computationally while we would need simulation evidence to explore the validity of our argument. This is a proposed area for future research.

To summarize, in this subsection we have provided a graphical visualisation of the behaviour of the SW1998 and SW2000 intervals, which is complementary to the previous analysis of coverage probabilities. In particular, we confirmed the major findings of the previous subsection and the discussion in Appendix IV which support that the SW2000 intervals might not be a good idea to use. We have also suggested ways to detect upfront the conditions under which we should expect such behaviour. The fact that the inferior performance of the SW2000 intervals is associated with a larger DEA bias compared to bootstrap bias implies that future research in this field should address questions such as: "why would the DEA bias be greater than the bootstrap bias and what are the implications" as well as whether this is something desirable or not.

2.9.5 Bootstrap distributions[75]

So far we have explored the performance of bootstrap DEA on the basis of coverage probabilities and the ability of confidence intervals to capture the true efficiency score θ_k, which is the standard approach. One of the issues, though, that has been ignored in the literature is the behaviour of the moments of the bootstrap distributions of

[75] Once again we would like thank Prof Simar for his suggestion to explore the moments of the fixed point.

efficiency scores[76], which carry information about the location, the variability and shape of those distributions. Ultimately, the examination of bootstrap moments can indicate if the DEA sampling variations are captured adequately by the bootstrap. This information might be relevant for the assessment of the performance of bootstrap DEA from another perspective or for the construction of bootstrap confidence intervals.

The location and variability of the bootstrap distribution of a fixed point is important for two reasons: (i) it shows how close the bootstrap bias is to the DEA bias, while (ii) it indicates how sensitive (or robust) the estimated efficiency scores are towards sampling variations. The standard deviation carries information about the variability of the bootstrap distribution and the width of confidence intervals. If the standard deviation approaches zero (likely in very large samples), then the confidence intervals will be extremely narrow and therefore it would be meaningless to apply bootstrap DEA; the estimated region for θ_k would actually be a point and therefore there would be no need to test hypotheses. Moreover, if the standard deviation in the DEA distributions is very low and the DEA scores are close to their population value, then the observed DEA scores would be good proxies of the population and scores and robust to sampling variations; therefore the application of the bootstrap would not be necessary.

Regarding the shape of the distributions, as already discussed in this section and shown in Appendix IV, if they are positively skewed and leptokurtic then the SW2000 intervals might underperform. This is because under these conditions the SW2000

[76] Please note the difference between the moments of the sample DEA scores and the moments of the bootstrap distribution; the former refer to the DEA scores in the sample while the latter refer to the bootstrapped efficiency scores of the DMU of interest (in our case the "fixed" DMU).

intervals will always lie below the SW1998 ones, and therefore will only perform better when the DEA bias is greater than the bootstrap bias; a condition either associated with technological heterogeneity or observed under the naïve bootstrap. Hence, it is important to know whether it would be safer to avoid using the SW2000 intervals in general.

Another reason why one should look at moments relates to the suggestion by Simar and Wilson (1998) that in the presence of skewness it might be a better idea to use the median when correcting for bootstrap bias and to adopt the bias-corrected intervals of Efron (1982). If there is no skewness the distribution will be symmetric and the bias-corrected intervals will be the same as the simple SW1998 ones. However, the higher the skewness (in absolute terms), the greater the degree of correction of the intervals with the Efron (1982) will be. Hence, it would be useful to know whether skewness is the "rule" or the "exception".

Hence, the examination of bootstrap moments may uncover details about the behaviour of bootstrap DEA that would not be possible to detect with the conventional approach of computing coverage probabilities. We have to note at this point that there are actually no "true" (or population) moments for the "fixed" DMU as it is a fixed observation. Therefore we will use the DEA scores from each of the M Monte-Carlo-generated samples to create a distribution of values for the fixed point and we will deem the moments of this distribution as the true ones. This distribution is due to the sampling variations when randomly drawing observations from the population, which resembles the resampling process of bootstrap DEA. Hence, considering these values as the true ones seems to be valid in principle.

The results of this exercise are summarized for the 2-inputs/2-outputs case in Table 2.8 to conserve space, while the moments for all model dimensions can be found in Appendix V. The moments of the smooth bootstrap procedures approach the DEA ones, while this is not true for the naïve bootstrap which overestimates higher moments. The means of these distributions have been already examined in the analysis of bootstrap biases (section 2.9.2). Regarding, standard deviation we observe that it converges with sample size which is desirable. On the other hand, in larger samples the standard deviation becomes very small (monotonically) and according to the discussion above this limits the relevance of applying bootstrap DEA (or even testing hypotheses) in larger samples (see also section 2.6.4).

Regarding skewness and kurtosis we observe a non-monotonic behaviour, which is not surprising since the distribution of efficiency scores for the fixed point is affected by the randomness in the sampling process. More importantly, we find that in all cases the bootstrap distributions are positively skewed and leptokurtic, providing support to our claims for the superiority of the SW1998 intervals (see previous discussion in this section and in Appendix IV). Moreover, the observed skewness suggests that there might be some benefit from adopting relevant approaches when constructing confidence intervals[77] while it might be better to use one-sided tests when testing

[77] In the next chapter we elaborate on this idea and we propose using the bias corrected and accelerated confidence intervals of Efron (1987) which are an extension of the bias corrected confidence intervals of

hypotheses, especially under the "Trun.Normal High" case where skewness is higher compared to the other cases.

Table 2.8. Moments of bootstrap distribution of the fixed point

	Standard 2/2				Trun. Normal Low 2/2				Trun. Normal High 2/2				Uniform 2/2			
Population	**Eff. Score**				**Eff. Score**				**Eff. Score**				**Eff. Score**			
N = 10,000	0.846				0.593				0.349				0.652			
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.889	0.027	1.043	4.228	0.774	0.083	0.251	2.459	0.440	0.076	1.898	8.496	0.737	0.063	1.323	4.916
n = 15	0.874	0.019	1.151	4.453	0.738	0.067	0.212	2.796	0.409	0.048	1.773	8.456	0.710	0.044	1.846	8.293
n = 20	0.867	0.014	1.146	4.535	0.715	0.062	0.398	2.780	0.394	0.034	1.510	6.450	0.696	0.034	1.662	7.130
n = 25	0.863	0.011	1.115	4.500	0.703	0.054	0.337	2.640	0.386	0.028	1.571	6.208	0.686	0.025	1.557	6.459
n = 30	0.859	0.009	1.111	4.198	0.694	0.052	0.391	2.675	0.381	0.025	1.898	8.371	0.681	0.021	1.247	5.109
n = 60	0.853	0.005	1.365	5.097	0.660	0.038	0.607	3.033	0.366	0.013	1.657	7.525	0.667	0.011	1.364	5.482
n = 120	0.850	0.002	1.532	7.484	0.637	0.026	0.766	3.523	0.358	0.006	1.106	4.315	0.660	0.006	1.509	6.176
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.950	0.023	0.852	4.210	0.860	0.039	1.088	4.723	0.533	0.058	1.836	8.443	0.822	0.042	1.235	5.355
n = 15	0.928	0.018	0.819	4.141	0.815	0.031	1.143	4.810	0.479	0.038	1.770	8.024	0.781	0.031	1.302	5.615
n = 20	0.915	0.014	0.820	4.153	0.781	0.027	1.163	4.824	0.453	0.027	1.646	7.330	0.754	0.025	1.390	6.063
n = 25	0.909	0.012	0.814	4.150	0.763	0.025	1.208	4.913	0.437	0.022	1.581	6.991	0.736	0.021	1.372	5.894
n = 30	0.902	0.010	0.802	4.116	0.746	0.023	1.195	4.879	0.428	0.019	1.548	6.891	0.725	0.018	1.371	5.887
n = 60	0.886	0.006	0.784	4.093	0.701	0.017	1.201	4.830	0.395	0.010	1.364	5.914	0.692	0.010	1.264	5.463
n = 120	0.875	0.003	0.765	4.041	0.669	0.013	1.194	4.817	0.377	0.006	1.253	5.372	0.674	0.006	1.185	5.062
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.938	0.022	1.045	4.684	0.853	0.038	1.211	5.059	0.527	0.057	1.938	9.182	0.818	0.042	1.319	5.635
n = 15	0.918	0.017	0.977	4.435	0.812	0.031	1.205	4.936	0.479	0.038	1.777	8.112	0.778	0.031	1.323	5.704
n = 20	0.906	0.013	0.967	4.409	0.780	0.028	1.192	4.910	0.454	0.027	1.671	7.564	0.755	0.025	1.331	5.767
n = 25	0.900	0.011	0.955	4.390	0.762	0.024	1.220	4.989	0.437	0.022	1.598	7.152	0.741	0.021	1.290	5.597
n = 30	0.894	0.010	0.932	4.345	0.748	0.023	1.216	4.971	0.427	0.019	1.540	6.772	0.731	0.018	1.276	5.502
n = 60	0.880	0.005	0.909	4.285	0.703	0.017	1.178	4.826	0.397	0.010	1.347	5.758	0.700	0.010	1.178	5.058
n = 120	0.870	0.003	0.903	4.267	0.669	0.013	1.152	4.720	0.379	0.006	1.238	5.297	0.681	0.006	1.129	4.818
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.898	0.018	2.267	9.734	0.808	0.045	1.786	5.959	0.467	0.063	2.862	14.018	0.763	0.045	2.348	9.415
n = 15	0.881	0.013	2.133	8.856	0.771	0.038	1.857	6.319	0.427	0.039	2.729	13.664	0.724	0.030	2.507	11.667
n = 20	0.872	0.010	2.285	10.025	0.741	0.034	1.760	6.117	0.408	0.026	2.633	12.789	0.708	0.023	2.525	11.731
n = 25	0.867	0.008	2.188	9.365	0.727	0.030	1.792	6.339	0.397	0.022	2.544	12.126	0.697	0.020	2.403	10.765
n = 30	0.863	0.007	2.248	9.595	0.714	0.027	1.823	6.521	0.391	0.018	2.513	11.605	0.691	0.017	2.311	10.238
n = 60	0.855	0.004	2.321	10.098	0.676	0.021	1.715	6.113	0.371	0.009	2.353	10.600	0.672	0.009	2.217	9.620
n = 120	0.850	0.002	2.613	11.846	0.648	0.015	1.694	6.220	0.361	0.005	2.253	9.982	0.662	0.005	2.262	9.783

Efron (1982) which have been suggested by Simar and Wilson (1998) in cases where the distribution is skewed.

2.10 Monte Carlo Results: large samples

Despite the fact that with bootstrap DEA the interest lies on its applicability in smaller samples, it is important to check its behaviour in larger samples. The examination of larger samples is a standard practice in Monte Carlo simulations and has been also examined by Simar and Wilson (2000b) and Simar and Wilson (2004). One of the reasons for looking at larger samples is to confirm the asymptotic convergence of bootstrap DEA in that both DEA and bootstrap biases approach zero as sample size increases. Moreover, it might be the case that the performance improves in samples larger than 120 as the results thus far have not been encouraging. Finally, examining the behaviour of bootstrap distributions we gain an insight about the meaningfulness of constructing confidence intervals in large samples. To avoid repetition, we will only focus on three issues of interest: bootstrap and DEA biases, coverage probabilities and moments of bootstrap distributions. We examine samples from 25 up to 1600 DMUs, but due to computational limitations we only examine the 1-input/1-output case from each DGP, using an efficient Matlab code developed by the author[78].

The behaviour of bootstrap and DEA biases in larger samples is reported in Table 2.9. The results indicate that in absolute terms both the bootstrap and DEA biases become very small and monotonically approach zero as sample size increases, confirming the consistency of the method. However, in relative terms, the ratio of bootstrap to DEA

[78] The main difference is that DEA scores are computed using straightforward computations instead of solving linear programmes which is only possible for the 1-input/1-output case. This function is available upon request.

bias does not converge to one which suggests that bootstrap DEA will only yield the desired results asymptotically. This difference is more pronounced for the “Standard” case, while it is worthwhile noting that the bootstrap biases under the “Trun.Normal Low” case become larger than the DEA ones from $n = 100$ onwards. Furthermore, the fact that the smooth bootstraps yield larger biases compared to the naïve bootstrap even asymptotically, indicates that smoothing the empirical kernel introduces additional variability which might be responsible for the observed behaviour, as suggested by Simar and Wilson (2002). Overall, the assumption of equal bootstrap and DEA biases seems to be quite generous and can only apply asymptotically.

Table 2.9. Bootstrap and DEA biases: large samples

	Standard				**Trun. Normal Low**			
	DEA Bias	**LSCV**	**SJ**	**Naïve**	**DEA Bias**	**LSCV**	**SJ**	**Naïve**
n = 25	0.006	0.046	0.036	0.004	0.086	0.059	0.055	0.022
n = 50	0.003	0.034	0.026	0.002	0.057	0.043	0.042	0.016
n = 100	0.001	0.027	0.021	0.001	0.031	0.032	0.032	0.011
n = 200	0.001	0.020	0.016	0.000	0.019	0.024	0.024	0.008
n = 400	0.000	0.015	0.012	0.000	0.011	0.017	0.017	0.005
n = 800	0.000	0.012	0.009	0.000	0.006	0.012	0.012	0.003
n = 1600	0.000	0.008	0.007	0.000	0.004	0.008	0.008	0.002
	Trun. Normal High				**Uniform**			
	DEA Bias	**LSCV**	**SJ**	**Naïve**	**DEA Bias**	**LSCV**	**SJ**	**Naïve**
n = 25	0.019	0.053	0.052	0.013	0.015	0.046	0.053	0.010
n = 50	0.007	0.029	0.032	0.006	0.008	0.022	0.033	0.005
n = 100	0.004	0.018	0.021	0.003	0.004	0.012	0.021	0.003
n = 200	0.002	0.012	0.013	0.001	0.002	0.006	0.012	0.001
n = 400	0.001	0.007	0.009	0.001	0.001	0.003	0.008	0.001
n = 800	0.001	0.005	0.006	0.000	0.000	0.001	0.005	0.000
n = 1600	0.000	0.003	0.004	0.000	0.000	0.001	0.003	0.000

The associated coverage probabilities are reported in Table 2.10 below for the 95% intervals while results for other significances are available upon request. Unfortunately, we cannot confirm for any sample size that coverage is adequate under the examined

DGPs and it is therefore not safe to test hypotheses using bootstrap DEA as proposed by Simar and Wilson (1998, 2000a). Coverage probabilities do not always increase, which is attributed to the fact that the ratio of bootstrap to DEA bias might increase with sample size, despite the fact that both biases reduce in absolute terms.

Table 2.10. Coverage of SW1998 and SW2000 95% confidence intervals: large samples

	Standard		Trun.N. Low		Trun.N. High		Uniform	
LSCV	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 25	0.250	0.127	0.481	0.526	0.805	0.427	0.711	0.402
n = 50	0.177	0.083	0.540	0.543	0.720	0.290	0.696	0.355
n = 100	0.118	0.054	0.662	0.592	0.535	0.208	0.699	0.335
n = 200	0.082	0.024	0.697	0.515	0.377	0.134	0.674	0.299
n = 400	0.058	0.020	0.711	0.467	0.250	0.097	0.716	0.318
n = 800	0.020	0.005	0.718	0.350	0.170	0.071	0.739	0.288
n = 1600	0.001	0.000	0.664	0.293	0.149	0.061	0.837	0.170
SJ	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 25	0.319	0.118	0.479	0.496	0.894	0.345	0.761	0.257
n = 50	0.101	0.033	0.555	0.520	0.737	0.178	0.538	0.132
n = 100	0.015	0.002	0.687	0.558	0.416	0.079	0.312	0.056
n = 200	0.002	0.000	0.735	0.533	0.153	0.029	0.126	0.037
n = 400	0.000	0.000	0.752	0.407	0.048	0.004	0.052	0.008
n = 800	0.000	0.000	0.755	0.297	0.008	0.005	0.021	0.004
n = 1600	0.000	0.000	0.637	0.226	0.000	0.000	0.006	0.001
Naïve	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 25	0.610	0.895	0.331	0.599	0.622	0.881	0.582	0.878
n = 50	0.633	0.900	0.350	0.631	0.628	0.879	0.575	0.857
n = 100	0.621	0.897	0.378	0.732	0.628	0.906	0.578	0.865
n = 200	0.613	0.875	0.429	0.742	0.566	0.867	0.618	0.864
n = 400	0.623	0.871	0.472	0.740	0.538	0.841	0.609	0.872
n = 800	0.649	0.895	0.535	0.816	0.530	0.819	0.691	0.879
n = 1600	0.641	0.897	0.544	0.834	0.530	0.816	0.746	0.933

Finally, the results on the moments of the bootstrap distribution are similar to those for smaller samples. The interesting point, though, is that standard deviation becomes negligibly small after a sample size of 200, suggesting that the associated confidence intervals become very narrow; almost point estimates[79]. This implies that it is not meaningful to apply hypothesis testing on large samples as in practice there is almost no

[79] The graphical representation of the intervals can be found in Appendix V.

confidence interval. This argument also carries to extensions of bootstrap DEA such as the second-stage regressions of Simar and Wilson (2007).

Table 2.11. Moments of bootstrap distribution of the fixed point: large samples

	Standard				Trun. Normal Low				Trun. Normal High				Uniform			
Population	**Eff. Score**				**Eff. Score**				**Eff. Score**				**Eff. Score**			
N = 10,000	0.847				0.592				0.349				0.655			
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 25	0.855	0.008	1.724	6.855	0.684	0.056	0.536	2.796	0.377	0.027	1.714	6.466	0.677	0.023	1.826	7.157
n = 50	0.851	0.004	1.886	7.610	0.654	0.038	0.607	2.881	0.361	0.012	1.968	8.250	0.665	0.011	1.917	7.648
n = 100	0.849	0.002	2.279	11.076	0.629	0.026	0.942	3.558	0.355	0.006	2.321	10.589	0.660	0.005	1.710	7.516
n = 200	0.848	0.001	2.089	9.238	0.616	0.017	1.056	4.366	0.352	0.003	1.655	6.330	0.657	0.003	2.399	13.563
n = 400	0.847	0.001	1.999	8.899	0.607	0.012	1.097	4.220	0.350	0.001	1.742	7.361	0.656	0.001	1.797	8.675
n = 800	0.847	0.000	2.038	7.809	0.600	0.006	1.091	3.821	0.350	0.001	1.949	9.330	0.655	0.001	2.139	8.935
n = 1600	0.847	0.000	2.637	13.080	0.597	0.004	1.534	6.166	0.349	0.000	1.554	6.307	0.655	0.000	2.589	11.203
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 25	0.900	0.010	0.929	4.937	0.735	0.025	1.596	6.298	0.420	0.021	2.005	8.985	0.721	0.017	1.866	8.026
n = 50	0.884	0.005	0.903	4.931	0.694	0.019	1.703	6.650	0.389	0.010	2.032	9.190	0.687	0.009	1.964	8.692
n = 100	0.876	0.003	0.795	4.673	0.658	0.014	1.781	7.131	0.372	0.005	1.994	8.975	0.672	0.005	1.963	8.645
n = 200	0.868	0.001	0.742	4.498	0.637	0.010	1.807	7.212	0.363	0.003	1.964	8.627	0.663	0.002	1.969	8.609
n = 400	0.863	0.001	0.670	4.365	0.622	0.006	1.886	7.833	0.358	0.001	1.977	8.762	0.659	0.001	1.960	8.531
n = 800	0.859	0.000	0.622	4.266	0.610	0.004	1.884	7.858	0.354	0.001	1.951	8.597	0.657	0.001	1.966	8.463
n = 1600	0.855	0.000	0.668	4.267	0.604	0.002	1.910	8.067	0.352	0.000	1.944	8.434	0.656	0.000	1.957	8.476
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 25	0.890	0.009	1.332	6.119	0.735	0.025	1.612	6.329	0.421	0.021	2.036	9.390	0.723	0.018	1.834	7.957
n = 50	0.876	0.005	1.262	6.002	0.695	0.019	1.701	6.695	0.389	0.010	2.006	9.065	0.696	0.009	1.875	8.289
n = 100	0.869	0.002	1.123	5.628	0.657	0.014	1.779	7.089	0.374	0.005	1.974	8.699	0.680	0.005	1.888	8.254
n = 200	0.863	0.001	1.039	5.374	0.638	0.010	1.797	7.152	0.365	0.003	1.947	8.553	0.670	0.002	1.894	8.276
n = 400	0.859	0.001	0.941	5.074	0.623	0.006	1.864	7.678	0.359	0.001	1.952	8.619	0.664	0.001	1.913	8.302
n = 800	0.856	0.000	0.826	4.741	0.611	0.004	1.896	7.938	0.355	0.001	1.926	8.437	0.660	0.001	1.923	8.336
n = 1600	0.854	0.000	0.743	4.481	0.604	0.002	1.907	8.028	0.353	0.000	1.923	8.330	0.658	0.000	1.928	8.401
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 25	0.858	0.008	2.519	10.901	0.703	0.036	1.630	5.532	0.386	0.025	2.485	10.320	0.684	0.020	2.532	10.992
n = 50	0.852	0.004	2.502	10.766	0.669	0.027	1.724	5.613	0.364	0.011	2.773	13.084	0.668	0.009	2.751	13.282
n = 100	0.849	0.002	2.616	11.862	0.637	0.019	1.965	6.458	0.357	0.005	2.711	12.473	0.662	0.005	2.492	10.735
n = 200	0.848	0.001	2.737	12.917	0.622	0.014	1.952	6.571	0.353	0.003	2.642	12.261	0.658	0.002	2.404	10.286
n = 400	0.847	0.000	2.606	12.119	0.610	0.008	2.074	7.395	0.351	0.001	2.470	10.455	0.657	0.001	2.382	10.175
n = 800	0.847	0.000	2.623	11.629	0.602	0.006	2.017	6.934	0.350	0.001	2.538	10.912	0.656	0.001	2.437	10.486
n = 1600	0.847	0.000	2.989	14.845	0.598	0.003	2.238	8.104	0.349	0.000	2.311	9.040	0.655	0.000	2.665	11.521

2.11 Conclusions

This chapter has explored the behaviour of bootstrap DEA both theoretically and with Monte Carlo simulations. The theoretical explorations provide a detailed analysis on how bootstrap DEA works, with a view to providing a deep understanding on the logic of the method, its mechanics and the implications of various assumptions. The Monte Carlo simulations assess the performance of bootstrap DEA and of the associated confidence intervals in finite samples, providing suggestions on the applicability of the method.

Since bootstrap DEA involves smoothing the empirical distribution, we followed the suggestion of Silverman and Young (1987) and performed the simulations under various data generating processes. We attached an economic interpretation to each DGP, while we proposed simple diagnostic tools to identify these cases through sample observations. The experiments were performed under the assumption of CRS and input orientation in three model dimensions and sample sizes which spanned from 10 to 120, while two smooth bootstraps and the naïve were considered. Regarding the behaviour of bootstrap DEA with respect to certain factors that affect its performance we find that results are more sensitive with respect to the DGP and sample bias and to a lesser extent due to dimensionality or the smoothing technique used here.

The results of the Monte Carlo simulations indicate that bootstrap DEA cannot be used either in small or large samples safely to construct confidence intervals and test hypotheses; it is however, consistent. This is attributed to the fact that that the assumption of bootstrap and DEA biases was violated in our simulations and seemed quite generous. In particular, although the two biases have a clear tendency to converge asymptotically to zero and despite being very small in larger samples, they are not exactly equal, affecting the associated intervals.

Of equal importance (if not more important) is the finding that the confidence intervals of Simar and Wilson (1998) perform better than those of Simar and Wilson (2000a). In particular, we have shown that under reasonable conditions[80] the SW2000 intervals perform better only if the DEA bias is greater than the bootstrap bias, which corresponds to the case of technological heterogeneity and the naïve bootstrap in our simulations. That is, we argued that the performance of the SW2000 is better in cases where DEA or bootstrap DEA should not be applied, putting serious thoughts on whether the SW2000 intervals should be preferred over the SW1998 ones. The implications become more important if we consider the popular extensions of bootstrap DEA which make use of the SW2000 intervals such as the bootstrap Malmquist index (Simar and Wilson, 1999), tests for RTS (Simar and Wilson, 2002) or the two-stage bootstrap DEA (Simar and Wilson, 2007). This suggests that these extensions might need to be reconsidered.

In our simulations we also examined the behaviour of the moments of the bootstrap distributions, which has been ignored in the literature. We found that confidence intervals may become extremely narrow in large samples, suggesting that it is more

[80] The bootstrap distributions need to be positively skewed and leptokurtic, which is confirmed in our simulations.

meaningful to use bootstrap intervals in smaller samples; therefore research should be focusing on improving the small-sample performance of bootstrap DEA. Moreover, in all cases we found positive skewness suggesting that there might be benefits from employing confidence intervals which account for skewness such as Efron's (1982) bias-corrected intervals, suggested by Simar and Wilson (1998).

Our research comes with some limitations which we aim to address in future research. First, only the CRS technology assumption was considered, although the implications should be transferable to any technology assumption; at least to some extent. Second, we examined only the input oriented case which might yield different results compared to the output oriented case. We believe that this is due to the support of the latter which spans from 1 to infinity, affecting the shape of the population distribution of efficiency and potentially affecting the results of our simulations. Furthermore, despite using 4 different DGPs to perform our experiments, we could still try alternative ones which would exhibit different behaviour with regards to the bootstrap and DEA biases and therefore different results. Finally, it must be noted that some of our suggestions about the appropriate sample size are case-specific[81] they do not necessarily constitute general advice on the exact number of DMUs required to apply these procedures.

The agenda for future research is rich in this field since bootstrap confidence intervals do not achieve yet the desirable finite sample performance. Theoretical work

[81] It is specific with respect to the particular DEA model, technology assumption, orientation, sample size, smoothing method and DGP chosen.

should focus on the conditions that produce appropriately large bootstrap biases compared to the DEA biases and examine the practical implications of accepting larger DEA biases. On the same wavelength, smoothing procedures (or similar) should be proposed which ensure good performance in small samples and not just asymptotically. In addition, future research could focus on confidence interval construction techniques which are based on less generous assumptions and which can establish a desirable performance in small samples. Finally, further work needs to be done towards the direction of designing and performing hypothesis tests, which will be associated with plausible assumptions.

A personal opinion is that the future of bootstrap DEA holds within double-bootstrap procedures such as the iterated bootstrap DEA proposed in a short note by Simar and Wilson (2004). Apart from providing a more accurate approximation of confidence interval endpoints, it could also be used to examine whether the bootstrap bias is smaller or larger than the DEA bias and construct confidence intervals accordingly. Unfortunately, with the current technology it would require an implausibly big computational time on a standard PC in order to obtain results, even for a small sample. It is hoped that with efficient manipulations of the algorithms and with the fast evolution of technology the computational requirements will soon become reasonable.

3 Testing hypotheses with bootstrap DEA

3.1 Introduction

Bootstrap DEA has been mainly used in applied research for confidence interval construction and hypothesis testing. Despite being a well-established technique, there is limited theoretical background as to how hypothesis testing should be performed. It is no surprise that in empirical applications there is no clear description of the hypothesis testing procedure followed and the technicalities involved. One of the possible reasons for the lack of theoretical works on testing hypotheses using bootstrap DEA, might be the lack of detailed evidence for the distributional aspects of bootstrapped efficiency scores. This exercise was performed in the previous chapter under the examination of the moments of the bootstrap distributions, and offered valuable insights with respect to the shape of these distributions.

Perhaps the only theoretical work on testing hypotheses using bootstrap DEA can be found in a relevant book chapter by Simar and Wilson (2008). The authors provide guidance on using their techniques and offer an implementation example for the case of mean efficiency score differences between two groups. In their general rules they suggest what is obvious: the test statistic has to be a function of the data, the critical value should result from the bootstrap distribution while the null hypothesis and the alternative should be clearly stated and be theoretically sensible. Another well-known work is by Simar and Wilson (2002) who propose a test for returns to scale with bootstrap DEA, which, however, cannot be extended to other cases. Also Simar and Wilson (1999) propose a test for efficiency and productivity change over time and where the null hypothesis is explained in an example by the authors (however, it makes use of the SW2000 intervals which are associated with questionable performance as we saw in the previous chapter).

Despite the fact that empirical studies use bootstrap DEA to test for efficiency differences between DMUs or between a DMU and a benchmark (or rank DMUs based on their bias-corrected scores)[82], there is no theoretical paper establishing the methodology for such tests. The prevailing, implied methodology is to construct bootstrap confidence intervals (Simar and Wilson, 1998, 2000a) and examine whether some "fixed point" of interest (a benchmark, a sample mean or a peer DMU/competitor) in included in the confidence region where the respective population value is expected to lie. However, the simulations of the previous chapter have indicated that the finite sample performance of these intervals is not satisfactory, carrying implications for the validity of hypothesis tests. This underperformance is probably linked with the violation of the assumption of equal biases, so it is of interest to explore how bias "asymmetries" may affect confidence intervals and hypothesis testing.

There seems to be a gap in the literature on how hypothesis tests should be performed using bootstrap DEA. More importantly, and to the extent of our knowledge, there is no paper providing recommendations about when one should avoid employing

[82] See for example the illustrative example in Simar and Wilson (1998) or other empirical studies (Gocht and Balcombe, 2006; Gonzales and Miles, 2002; Hawdon, 2003; Magnussen and Nyland, 2008; Sadjadi and Omrani, 2010; Sanhueza et al., 2004; Tsolas, 2011).

bootstrap DEA for hypothesis testing or about issues that could arise when testing hypotheses. This motivated our theoretical explorations in this chapter where we investigate how hypothesis testing should be performed using the SW1998 and SW2000 intervals. In particular, we first consider the assumption of equal biases as valid and explain how the null hypothesis for a hypothesis test should be outlined, while we propose an approach to compute the associated p-values of the tests where possible. Moreover, we show how the presence of unequal bootstrap and DEA biases can affect both confidence intervals and the validity of the hypothesis tests and we examine the possibility of adopting alternative approaches in certain extreme cases. Furthermore we outline some considerations and limitations while some theoretical ideas to overcome these issues are proposed along with future research avenues. We then extend the discussion to the case of testing for returns to scale using bootstrap DEA (Simar and Wilson, 2002) where a similar assumption on bias equality needs to be satisfied. Finally, we propose a test for RTS which does not make use of the equal biases assumption by incorporating the Banker et al. (1996) approach in bootstrap DEA.

Our results indicate that the assumption of equal biases is crucial for the hypothesis tests to be meaningful, while despite the fact that some alternatives might considerably improve the performance of the confidence intervals, they would require large samples to perform well. On a positive note, we argue that when the sample exhibits technological homogeneity (as in the “Standard” case in the previous chapter’s simulations), then it is not necessary to apply bootstrap DEA for sample sizes greater than 120 as the DEA bias becomes very small and the sampling variations negligible;

that is, DEA scores can be considered as robust “estimators” of the population efficiency scores. Finally, we show that the proposed test for RTS is not sensitive to the DGP specification though it would require further simulations to evaluate its performance and sample size requirements.

The remainder of the chapter is structured as follows: section 3.2 discusses the implied hypothesis testing procedures using Simar and Wilson’s (1998, 2000a) confidence intervals when the assumption of equal biases is valid and invalid; section 3.3 discusses some issues that need to be considered when testing hypotheses with bootstrap DEA and proposes lines of action; section 3.4 explores the possibility of adopting alternative approaches in the presence of substantially unequal bias; section 3.5 extends the testing to tests of returns to scale and proposes a bootstrap approach at a theoretical level; finally, section 3.6 concludes the chapter, highlights limitations and suggests areas for future research.

3.2 Simar and Wilson’s intervals and implied tests

Simar and Wilson (1998, 2000a) propose confidence intervals where the true efficiency score of a DMU of interest should lie. It is therefore implied that these intervals could be used for hypothesis testing, despite not explicitly stated in the literature. Two examples of null hypotheses that could be tested for DMU k are $H_0: \theta_k = 1$ (or some other constant) or $H_0: \theta_k = \theta_\nu$ where DMU ν is some other DMU of interest. One special case for the latter type of hypothesis test is $H_0\colon \theta_k^t = \theta_k^{t+1}$ where DMUs k^t and k^{t+1} represent the operations of firm k in two consecutive time periods. This test is relevant in cases where the number of firms per year is very small and where one of the possible solutions is to pool data (Fried et al., 2008, pp.54); and which means that Malmquist-type approaches cannot be applied on a year-by-year basis[83].

This section first explores the mathematics behind the SW1998 and SW2000 intervals, focusing on how they behave when the assumption of equal bootstrap and DEA biases is not satisfied. Then, we explain how the aforementioned hypotheses could be tested if we assumed that the confidence intervals of SW1998 and SW2000 (or similar) performed well. Since the first type $(H_0\colon \theta_k = 1)$ is more straightforward, the discussion will focus on the latter case $(H_0\colon \theta_k = \theta_v)$. The interest lies in the fact that both DMUs k and v are subject to sampling variations and therefore they are both associated with a distribution of efficiency scores. For the purposes of this analysis we will thereafter assume that the "fixed point" v is a DMU that belongs in the same dataset as DMU k and hence they are both associated with the same DGP.

3.2.1 Simar and Wilson's (1998) intervals

The SW1998 confidence intervals have been explained in the previous chapter (2.6.3). The principle is that the distribution of the (double) bias-corrected bootstrapped efficiency scores $\left(\tilde{\theta}_k^*\right)$ is used to construct confidence intervals and therefore test

[83] See for example Siriopoulos and Tziogkidis (2010).

hypotheses. In particular, the $\left({}^{a}/_{2}\right)\%$ and $\left(1-{}^{a}/_{2}\right)\%$ percentiles of this distribution, which we denote as $\hat{q}_{a/2}$ and $\hat{q}_{1-a/2}$, respectively, define a region where the "true" efficiency score of DMU k (θ_k) lies with a probability of $(1-a)$:

$$\Pr\left(\hat{q}_{a/2} < \tilde{\theta}_k^* < \hat{q}_{1-a/2}\right) \simeq \Pr\left(\hat{q}_{a/2} < \theta_k < \hat{q}_{1-a/2}\right) = 1-a \tag{3.1}$$

This results from the assumption of Simar and Wilson (1998) that the bootstrap bias is approximately equal to the DEA bias and therefore the centre of the distribution of $\tilde{\theta}_k^*$ is approximately equal to θ_k, as shown in (2.27).

We have already explained in the previous chapter that if there is an "asymmetry of biases" (the bootstrap bias is either smaller or greater than the DEA bias), then both SW1998 and SW2000 intervals will underperform. We will now show how the SW1998 intervals behave when there is such an "asymmetry of bias". Suppose that the bootstrap bias is $\widehat{bias}_k = \hat{z}$ and the DEA $bias_k = z$ and that $\hat{z} \neq z$. Note that asymptotically $\lim_{n\to N} \hat{z} = \lim_{n\to N} z = 0$ due to consistency. The centre of the bootstrap distribution will be $\overline{\hat{\theta}_k^*} = \hat{\theta}_k + \hat{z}$ while $\hat{\theta}_k = \theta_k + z$, while from (2.27):

$$\overline{\tilde{\theta}_k^*} = \hat{\theta}_k + \hat{z} - 2\hat{z} = \theta_k + z - \hat{z} \neq \theta_k \tag{3.2}$$

The SW1998 intervals assume that $\hat{z} \simeq z$ or $\overline{\tilde{\theta}_k^*} \simeq \theta_k$ and hence:

$$\begin{aligned} 1-a = \Pr\left(\hat{q}_{a/2} < \theta_k < \hat{q}_{1-a/2}\right) &= \Pr\left(\hat{q}_{a/2} < \overline{\tilde{\theta}_k^*} + \hat{z} - z < \hat{q}_{1-a/2}\right) \\ &\simeq \Pr\left(\hat{q}_{a/2} + z - \hat{z} < \theta_k < \hat{q}_{1-a/2} + z - \hat{z}\right) \end{aligned} \tag{3.3}$$

This simply suggests that the estimated SW1998 intervals will lie below their "accurate" [84] position if $\hat{z} > z$ (bias overestimation), and above if $\hat{z} < z$ (bias underestimation). This is reasonable and has been confirmed in our Monte Carlo simulations.

More importantly, if $\hat{z} \gg z$, then the inequality in (3.3) could be violated with the upper bound of the SW1998 intervals lying below θ_k, indicating a failure in interval estimation. Similarly, if $\hat{z} \ll z$, then the lower bound of the intervals could lie above θ_k which is another possibility of failure. The first case is evidenced under the "Standard" case and for sample sizes greater than $n = 25$. The latter seemingly extreme case is in fact observed under the naïve bootstraps and under the DGP associated with technological heterogeneity (for sample sizes up to $n = 60$) in Figure 2.12. The other DGPs examined are not associated with extreme "bias asymmetry" but they still underperform in the way suggested here. In any case, though, the presence of "bias asymmetries" will lead to both Type I and II errors (depending on the null hypothesis tested), reducing the validity of associated hypothesis tests.

3.2.2 Simar and Wilson's (1998) implied tests

Let us now consider how hypothesis testing could be performed using the SW1998 intervals when the assumption of equal biases is satisfied. We will examine the case of

[84] By "accurate" we will mean for the remainder of this chapter the position of the intervals which is associated with the nominal probability of $1 - a$.

testing for efficiency differences between two DMUs, which could be expressed as follows[85]:

$$H_0: \theta_k = \theta_v, \qquad H_1: \theta_k \neq \theta_v \tag{3.4}$$

One might think that it is not possible to perform this test since both θ_k and θ_v are unobservable. However, under the assumption of equal biases we could use their estimated values from (2.27) and express this test as follows:

$$H_0: \theta_k = \overline{\tilde{\theta}_v^*}, \qquad H_1: \theta_k \neq \overline{\tilde{\theta}_v^*} \tag{3.5}$$

Hence, this suggests that we could construct the SW1998 intervals for DMU k and examine whether the value $\overline{\tilde{\theta}_v^*} \simeq \theta_v$ falls within the two endpoints of the intervals for DMU k, or $\overline{\tilde{\theta}_v^*} \in \left(\tilde{\theta}_k^{*,(a/2)}\ ,\ \tilde{\theta}_k^{*,(1-a/2)}\right)$. Moreover, one could compute the following probabilities which could serve as an indication of how “well-included” $\overline{\tilde{\theta}_v^*}$ is within the interval:

$$plow = \frac{\#\left(\tilde{\theta}_{b,k}^* < \overline{\tilde{\theta}_v^*}\right)}{B} \quad \text{and} \quad phigh = \frac{\#\left(\tilde{\theta}_{b,k}^* > \overline{\tilde{\theta}_v^*}\right)}{B}, \quad b = 1,2, \ldots B \tag{3.6}$$

where # indicates “number of times” (technically termed “cardinality”). These probabilities would indicate how often $\overline{\tilde{\theta}_v^*} \simeq \theta_v$ lies in the tails of the bootstrap distribution of $\tilde{\theta}_k^*$. In fact, they could be considered as p-values for one-sided tests[86]; if $plow < a$ we could accept $H_1: \theta_k > \theta_v$, while if $phigh < a$ we could accept $H_1: \theta_k < \theta_v$.

[85] The formulation of the null hypotheses for the one-sided tests is straightforward. In particular the null would be the same but the alternatives would be $H_1: \theta_k < \theta_v$ or $H_1: \theta_k > \theta_v$.

[86] One could also check the extent to which the two distributions overlap by computing the following probability: $prob = \#\left(\hat{q}_{k,{}^a/_2} < \tilde{\theta}_{b,v}^* < \hat{q}_{k,1-{}^a/_2}\right)/B,\ b = 1,2, \ldots B$ as an p-value-alike measure.

3.2.3 Simar and Wilson's (2000a) intervals

We have already discussed how the SW2000 intervals can be constructed (see section 2.6.3); we shortly present the approach here again as this information is of importance. From (2.30) we have:

$$1 - a = \Pr\left(s_{a/2} < \hat{\theta}_k - \theta_k < s_{1-a/2}\right) = \Pr\left(\hat{s}_{a/2} < \hat{\theta}_k^* - \hat{\theta}_k < \hat{s}_{1-a/2}\right) \tag{3.7}$$

Assuming $\left(\hat{\theta}_k^* - \hat{\theta}_k\right) \sim \left(\hat{\theta}_k - \theta_k\right)$ then from (2.31) we have:

$$\begin{aligned} 1 - a &= \Pr\left(\hat{\theta}_k - s_{1-a/2} < \theta_k < \hat{\theta}_k - s_{a/2}\right) \\ &\simeq \Pr\left(\hat{\theta}_k - \hat{s}_{1-a/2} < \theta_k < \hat{\theta}_k - \hat{s}_{a/2}\right) \end{aligned} \tag{3.8}$$

Also note that $\hat{\theta}_{kb}^* - \hat{\theta}_k > 0 \; \forall \; b = 1,2, \dots B$ by definition, as explained in the previous chapter, indicating that the upper bound of the confidence interval will always lie on or below the DEA score (Simar and Wilson, 2008). That is, $\hat{s}_{a/2} > 0$ implying that $\hat{\theta}_k - \hat{s}_{a/2} < \hat{\theta}_k$. Hence, the logic of the intervals is to correct downwards the DEA estimator since it is upwards biased. To evaluate the implications of violating the assumption of equal bootstrap and DEA biases, suppose that $\hat{s}_{a/2} = s_{a/2} + \varepsilon_L$ and $\hat{s}_{1-a/2} = s_{1-a/2} + \varepsilon_U$, where ε_L and ε_U represent the deviations of the lower and upper estimated percentiles from their true values. Hence:

- If the bootstrap bias is equal to the DEA bias ($\hat{z} = z$), then $\varepsilon_L, \varepsilon_U = 0$
- If the bootstrap bias is greater than the DEA bias ($\hat{z} > z$), then $\varepsilon_L, \varepsilon_U > 0$
- If the bootstrap bias is smaller than the DEA bias ($\hat{z} < z$), then $\varepsilon_L, \varepsilon_U < 0$

Then, (3.8) becomes:

$$
\begin{aligned}
1-a &= \Pr\left(\hat{\theta}_k - s_{1-a/2} < \theta_k < \hat{\theta}_k - s_{a/2}\right) \\
&= \Pr\left(\hat{\theta}_k - s_{1-a/2} - \varepsilon_U < \theta_k < \hat{\theta}_k - s_{a/2} - \varepsilon_L\right)
\end{aligned}
\tag{3.9}
$$

The important finding from (3.9) is that if the bootstrap bias is greater than the DEA bias, the estimated intervals will be below their "accurate" position. If the bootstrap bias is smaller than the DEA bias, the estimated intervals will lie above the estimated intervals. Obviously, the larger the difference between the bootstrap and DEA biases is, the further Simar and Wilson's (2000a) intervals will deviate from their "accurate" position and therefore the worse will be their finite sample performance. However, asymptotically both biases will necessarily be zero suggesting that consistency is not violated.

As with the SW1998 intervals, in the case of "extreme bias asymmetry" it would be possible for the SW2000 intervals to completely leave θ_k outside the two endpoints. In particular, if $\hat{z} \gg z$, then $\varepsilon_L, \varepsilon_U$ could be large enough so that $\theta_k > \hat{\theta}_k - s_{a/2} - \varepsilon_L$ in (3.9). Similarly, if $\hat{z} \ll z$, then $\varepsilon_L, \varepsilon_U$ could be small enough so that $\hat{\theta}_k - s_{1-a/2} - \varepsilon_U > \theta_k$. The only case we observe the latter is to some extent under the "Trun.Normal Low" DGP (associated with technological heterogeneity) in the previous chapter (Figure 2.13) and mainly under the naïve bootstraps and in very small samples. Though, the SW2000 intervals seem to be more sensitive towards the first asymmetry ($\hat{z} \gg z$) as evidenced under the "Standard" DGP. In fact, even if the asymmetry is not extreme ($\hat{z} > z$), as with the "Trun.Normal High" and "Uniform" DGPs, the upper bounds of the SW2000 intervals tend to lie below θ_k in small samples. The theoretical explanation for this behaviour is that if $z = \hat{\theta}_k - \theta_k$ converges to zero fast enough, then it is possible for the SW2000 intervals to lie below θ_k since by definition their upper endpoint has to lie on or below $\hat{\theta}_k$ as previously discussed. This indicates a potential weakness of the SW2000 intervals: if the DEA bias converges to zero fast enough (or at least faster than the bootstrap bias) then they will tend to underestimate the true efficiency score. And given that smoothing techniques tend to introduce more variability (Simar and Wilson, 2002) it is possible that this conclusion is not limited to the particular DGPs examined in the previous chapter, suggesting once again that it might be a better idea to use the SW1998 intervals instead.

3.2.4 Simar and Wilson's (2000a) implied tests

We will now outline how hypothesis testing could be performed with the SW2000 intervals if the assumption of equal bootstrap and DEA biases is valid. The test is the same as in (3.4) where $H_0\colon \theta_k = \theta_v \simeq \overline{\tilde{\theta}_v^*}$ due to the equal biases assumption. Hence, if $\overline{\tilde{\theta}_v^*} \in \left(\hat{\theta}_k - s_{1-{a}/{2}},\ \hat{\theta}_k - s_{{a}/{2}}\right)$ we would accept H_0. However, for the sake of completeness we should first show that the SW2000 intervals are designed so that $\theta_k \simeq \overline{\tilde{\theta}_k^*} \in \left(\hat{\theta}_k - s_{1-{a}/{2}}, \hat{\theta}_k - s_{{a}/{2}}\right)$ to ensure that the null is consistent.

Let us now denote the *j*% percentile of the (non-corrected) bootstrap distribution of DMU k $\left(\hat{\theta}_k^*\right)$ as $\hat{p}\left(\hat{\theta}_k^*\right)_{(j)}$. Hence the *j*% percentile of the distribution of $\left(\hat{\theta}_k^* - \hat{\theta}_k\right)$

would be $\hat{s}_{(j)} = \hat{s}\left(\hat{\theta}_k^* - \hat{\theta}_k\right)_{(j)} = \hat{p}\left(\hat{\theta}_k^*\right)_{(j)} - \hat{\theta}_k$, since $\hat{\theta}_k$ is a constant. Using this result and substituting θ_k with $\overline{\tilde{\theta}_k^*} = \overline{\hat{\theta}_k^*} - 2\widehat{bias}_k$ in (3.8), where $\widehat{bias}_k = \overline{\hat{\theta}_k^*} - \hat{\theta}_k$, we have:

$$\left(\hat{\theta}_k - \hat{s}_{1-a/2} < \overline{\tilde{\theta}_k^*} < \hat{\theta}_k - \hat{s}_{a/2}\right) = \left(\hat{\theta}_k - \hat{s}_{1-a/2} < \overline{\hat{\theta}_k^*} - 2\widehat{bias}_k < \hat{\theta}_k - \hat{s}_{a/2}\right)$$
$$= \left(2\hat{\theta}_k - \hat{p}_{1-a/2} < \overline{\hat{\theta}_k^*} - 2\left(\overline{\hat{\theta}_k^*} - \hat{\theta}_k\right) < 2\hat{\theta}_k - \hat{p}_{a/2}\right) \tag{3.10}$$
$$= \left(-\hat{p}_{1-a/2} < -\overline{\hat{\theta}_k^*} < -\hat{p}_{a/2}\right) = \left(\hat{p}_{a/2} < \overline{\hat{\theta}_k^*} < \hat{p}_{1-a/2}\right)$$

The equations in (3.10) simply state that if we substitute θ_k with $\overline{\tilde{\theta}_k^*}$ it would result in a consistent transformation as $\overline{\hat{\theta}_k^*}$ is the centre of the bootstrap distribution and it will always lie within its $a/2$ % and $(1 - a/2)$% percentiles. Thus, $\overline{\tilde{\theta}_k^*}$ will always lie within the lower and upper bound of the SW2000 intervals and therefore the null is valid. Therefore we could state that if $\overline{\tilde{\theta}_v^*} \in \left(\hat{\theta}_k - \hat{s}_{k,1-a/2}, \hat{\theta}_k - \hat{s}_{k,a/2}\right)$ we accept the null hypothesis of equal efficiency between DMUs k and v[87].

3.3 Considerations and limitations

So far we have explained how one could perform hypothesis tests using the SW1998 and SW2000 intervals. In both cases we have demonstrated that the assumption of equal biases should hold otherwise both confidence intervals would have limited coverage while the hypothesis tests would not be consistent. We will now share some

[87] The computation of probabilities as in (3.6) is not straightforward in this case. However, we could use a similar p-value-alike probability as in footnote 85 which could serve as an indication of the overall of the two distributions: $prob = \#\left(\hat{\theta}_k - \hat{s}_{k,1-a/2} < \tilde{\theta}_{b,v}^* < \hat{\theta}_k - \hat{s}_{k,a/2}\right)/B,\ b = 1,2, \dots B$

considerations/observations which we deem of importance and potentially of interest to the potential bootstrap DEA user. In particular, our considerations are with regards to the potential skewness of the bootstrap distributions, on performing same-sample comparisons and on the feasibility of performing cross-sample comparisons.

3.3.1 Dealing with skewness

Simar and Wilson (1998) suggested that if the bootstrap distribution is skewed, it could be preferable to employ Efron's (1982) bias-corrected intervals which apply a median-correction to the percentile intervals. Hence, instead of using the SW1998 intervals $\left(\tilde{\theta}_k^{*,a/2}, \tilde{\theta}_k^{*,1-a/2}\right)$, two endpoints a_1 and a_2 are determined and the following intervals are estimated $\theta \in \left(\tilde{\theta}_k^{*,a_1}, \tilde{\theta}_k^{*,a_2}\right)$, where $a_1 = \Phi\left(2\hat{z}_0 + z^{(a/2)}\right)$, $a_2 = \Phi$, $\hat{z}_0 = \#\Phi^{-1}\left(\tilde{\theta}_k^* < \overline{\tilde{\theta}_k^*}\right)$, and where Φ is the standard normal cumulative density function while $z^{(a/2)}$ is the $a/2$ percentile of the standard normal distribution $\left(\Phi\left(z^{(a/2)}\right) = a/2\right)$. In the same paper, Simar and Wilson (1998) perform an empirical illustration under input orientation using data from Färe et al. (1989) and they use both the SW1998 and SW1998bc intervals (standing for Efron's (1982) bias-corrected intervals). They report small differences between the two intervals which they attribute to the fact that the means of $\tilde{\theta}_k^*$ are close to the medians (in particular in most cases the difference between the two is 0.01 to 0.02). Their results also indicate that the SW1998bc intervals are wider, mainly with respect to the upper bound in their input-oriented model. In particular, the intervals are in most cases wider by 0.015 to 0.03, compared to the SW1998 intervals.

Apart from the seminal paper of Simar and Wilson (1998), the issue of skewness has only been mentioned in subsequent book chapters (Simar and Wilson, 2004; Simar and Wilson, 2008) where the same suggestion of using Efron's (1982) bias correction is given. Subsequent works seem favour the SW2000 intervals but no consideration on the potential effects of skewness is provided[88], apart from the fact that may underperform compared to the SW1998 ones as we explained in the previous chapter. The examination of bootstrap moments in the previous chapter has indicated that the skewness of bootstrap distributions varies with the underlying DGP and it may range from about 0.5 to about 2. The severity of the effect of skewness on confidence intervals could be examined with further Monte Carlo simulations whereby a variety of DGPs associated with a range of skewness values for the bootstrap distributions could be chosen. Then, the effect on coverage probabilities could be monitored and the benefit of employing techniques which account for skewness can be considered but at the same time measuring the potential costs due to potentially wider interval widths.

A development of this approach would be to construct confidence intervals which, apart from providing a median-correction to the intervals, they can also correct for skewness (Efron, 1987). In particular, the bias-corrected and accelerated intervals (BC_a)

---

[88] We could provide a suggestion here of how this could be performed. Denote $\varDelta\hat{\theta}_k^* = \hat{\theta}_k^* - \hat{\theta}_k$ and compute a_1 and a_2 as before, but now $\hat{z}_0 = \#\Phi^{-1}\left(\varDelta\hat{\theta}_k^* < \overline{\varDelta\hat{\theta}_k^*}\right) = \#\Phi^{-1}\left(\hat{\theta}_k^* < \overline{\hat{\theta}_k^*}\right)$. Hence, instead of $\theta_k \in \left(\hat{\theta}_k - \varDelta\hat{\theta}_k^{*,1-\alpha/2}, \hat{\theta}_k - \varDelta\hat{\theta}_k^{*,\alpha/2}\right)$ we have for the SW2000bc intervals: $\theta_k \in \left(\hat{\theta}_k - \varDelta\hat{\theta}_k^{*,a_2}, \hat{\theta}_k - \varDelta\hat{\theta}_k^{*,a_1}\right)$.

of Efron (1987) correct for skewness through the acceleration parameter and are superior to the Efron's (1982) ones. In fact Efron's (1982) intervals comprise a special case of Efron's (1987) intervals where the acceleration parameter is equal to zero. However, the difficulty in implementing this approach is the computation of the acceleration parameter which can be very challenging when the problem in hand is complex (Shao and Tu, 1995) as in the case of DEA. We have included the underlying ideas and the progress of our current work on adapting Efron's (1987) intervals on DEA in Appendix VII for the interested reader.

To the extent of our knowledge there is no work that focuses on the issue of skewness on bootstrap DEA which seems a field for further development. It is within the author's immediate research plans to investigate in-depth the effects of skewness on the performance of bootstrap DEA and to analyse the benefits of implementing the BC_a intervals in the case of DEA.

3.3.2 Same-sample comparisons

We have already discussed in the previous section how one could use the SW1998 and SW2000 intervals to perform hypothesis tests. When testing $H_0: \theta_k = c$ where c is a constant (e.g. $c = 1$), the testing procedure is straightforward and does not present any issues to the extent of our knowledge. When testing, though, $H_0: \theta_k = \theta_v \simeq \overline{\tilde{\theta}_v^*}$, apart from the fact that the assumption of equal biases must hold, one needs to consider that it would be possible to test $H_0: \theta_v = \theta_k \simeq \overline{\tilde{\theta}_k^*}$ as well. Despite the fact that both tests are valid under the assumption of equal biases, it might be possible to receive different outcomes from each test. In particular, one possibility is that different sampling variations for each DMU may lead to bootstrap distributions with wider or narrower confidence intervals. Moreover, skewness can cause the endpoints of the confidence intervals to lie asymmetrically about the centre of the bootstrap distribution which could allow for such an eventuality if the distribution is skewed enough.

One possible solution would be to transform the null hypothesis as follows[89]:

$$H_0: \psi = \frac{\theta_k}{\theta_v} = 1, \qquad H_1: \psi \neq 1 \tag{3.11}$$

To perform this test one could use the bootstrap distribution of the ratios of “the two DMUs and compute the following distribution of ratios:

$$\hat{\psi}_b^* = \frac{\hat{\theta}_{k,b}^*}{\hat{\theta}_{v,b}^*}, \qquad b = 1,2, \dots B \tag{3.12}$$

And then we could perform the usual bias correction to obtain an estimate of ψ:

$$\tilde{\psi}_b^* = \hat{\psi}_b^* - 2\left(\bar{\psi}_b^* - \hat{\psi}\right) = \hat{\psi}_b^* - 2\left(\frac{1}{B}\sum_{b=1}^{B} \hat{\psi}_b^* - \frac{\hat{\theta}_k}{\hat{\theta}_v}\right) \tag{3.13}$$

That is, if $\left(\hat{\psi}^* - \hat{\psi}\right)\big|\hat{\mathcal{P}} \sim \left(\hat{\psi} - \psi\right)\big|\mathcal{P}$, then $E\left(\tilde{\psi}_b^*\right) \simeq \psi$. Then we could use the bootstrap distribution of $\tilde{\psi}_b^*$ to construct confidence intervals for the population value of ψ. The rationale for this hypothesis test is similar to that in Simar and Wilson (1998, 2000a) while it has also been used in Simar and Wilson (1999) for the construction of confidence intervals for the Malmquist index and its components (efficiency change and productivity change).

[89] Alternatively it could be $H_0: \theta_k - \theta_v = 0$ and $H_1: \theta_k - \theta_v \neq 0$. Both tests would yield the same results by definition which the author has also confirmed with simulations.

The proposed confidence intervals could be either constructed using the SW1998 or SW2000 approaches, but we are in favour of the former due to the low performance that the latter exhibited in the previous chapter. Hence, using the SW1998 intervals we would reject (3.11) if $1 \notin \left(\tilde{\psi}_b^{*,(a/2)}, \tilde{\psi}_b^{*,(1-a/2)}\right)$.

If we find that the null is rejected, we could go one step further and test if $H_1: \frac{\theta_k}{\theta_v} > 1$ or $H_1: \frac{\theta_k}{\theta_v} < 1$. And we could compute probabilities as in (3.6) which would help us identify the position of 1 with respect to the distribution of $\tilde{\psi}_b^*$:

$$plow = \frac{\#\left(\tilde{\psi}_b^* < 1\right)}{B} \quad \text{and} \quad phigh = \frac{\#\left(\tilde{\psi}_b^* > 1\right)}{B}, \quad b = 1,2, \dots B \tag{3.14}$$

And as previously, these probabilities could be considered as p-values for one-sided tests; if (3.11) is rejected and $plow < a$ we could accept $H_1: \frac{\theta_k}{\theta_v} > 1$, while if (3.11) is rejected and $phigh < a$ we could accept $H_1: \frac{\theta_k}{\theta_v} < 1$[90].

For future research we propose exploring the power of the proposed test with Monte Carlo simulations which should be carefully designed to represent a "true" H_0. One way would be to include two fixed DMUs, modelled to differ in efficiency to various degrees in various simulations. This would serve as a sensitivity analysis of the

[90] Another possibility would be to test (3.4) as explained previously and accept H_0 if in both cases H_0 is accepted and reject H_0 if it is rejected in at least one test, with reference to (3.5). The logic in this approach would be to reduce the probability of a Type II error, which is the most serious in hypothesis testing and which seems reasonable in the sense that if one of the tests rejects H_0 then there is evidence that the efficiency of the two DMUs is different. However, there is some degree of subjectivity in this approach while the probability of a Type I error is increased due to the trade-off between the two error types.

acceptance/rejection decisions (or of the distribution of the associated p-values where relevant) towards different initial input/output setups[91].

3.3.3 Cross-sample comparisons

It might be the case that the researcher is interested in performing efficiency comparisons between two groups of DMUs. In cases like this it would be more interesting (if not meaningful) to compare, for example, the means of the two samples instead of comparing a DMU from one sample with a DMU from another. Simar and Wilson (2008) outline a hypothesis testing procedure for comparing the means of two groups of DMUs using as an example the "program-follow-through" schools and the "non-program-follow-through" schools in Charnes et al. (1981). They suggest using the ratio of means as a sample statistic and they propose as a p-value the relative frequency that the bootstrap ratio of means is greater than the sample statistic. Kneip et al. (2012) are currently working on the issue of testing differences between sample means, treating the issue from a statistical perspective suggesting that this area of research is under development.

[91] The author has experimented to some extent on this issue by comparing two fixed DMUs under the DGPs described in the previous chapter. In particular, a second fixed DMU was introduced which uses one standard deviation of extra input, hence being more inefficient. The simulations have shown that the test proposed in this subsection would reject the null at a rate close to 100% even in very small samples. However, to arrive at a general conclusion we would need to perform simulations using other input/output combinations for the second fixed point which would make the differences more marginal and hence more sensitive to the required sample size for the test to exhibit a satisfying power. The examination of the power of various hypothesis tests is within the intermediate research plans of the author.

We believe that one of the challenging issues that should be taken into account is the fact that the two samples might be associated with different DGPs. In that case they would exhibit different performance with respect to coverage probabilities and with respect to the plausibility of the equal biases assumption. In fact, from the mathematical formulations in Simar and Wilson (2008) it is implied that the test assumes that the two samples stem from the same feasible set. One suggestion for the researchers who wish to adopt the approach of Simar and Wilson (2008) would be to compare the skewness and kurtosis of the DEA distributions of the two samples which could serve as an indication of whether the underlying DGPs are similar or not.

3.4 Can we “bypass” the issue of unequal biases?

The simulations of the previous chapter have indicated that the assumption of equal bootstrap and DEA biases does not hold well under the chosen DGPs. In fact we evidenced an asymmetry of biases with the two extreme cases being the “Standard” and the “Trunc.Normal Low” case which have been associated with technological homogeneity and heterogeneity, respectively. In the “Standard” case the bootstrap bias is large compared to the DEA bias which fast becomes very small, while in the “Trunc.Normal Low” case the DEA bias is larger than the DEA bias in smaller samples (though it becomes smaller after $n = 120$). In this section we will explore the possibility of adopting alternative approaches towards the direction of confidence interval construction in these “extreme” cases. We highlight that throughout this section we assume that there is an extreme asymmetry of the two biases and therefore the confidence intervals and hypothesis testing approaches discussed thus far would not work.

We will begin with the case of technological heterogeneity which should be diagnosed by a positive skewness about 0.4 and a kurtosis value close to 3. We argue that in this case it is neither worthwhile nor feasible to propose an alternative approach for confidence interval construction or hypothesis testing. It is not worthwhile because, as we have argued in the previous chapter, the fact that the DEA bias is greater than the bootstrap bias suggests issues from the very application of DEA; perhaps a different dataset should be used or the input/output variables should be reconsidered. It is also not feasible as it would require knowledge of the true efficiency score for DMU k, θ_{κ}, which would allow us (perhaps) to inflate the bootstrap bias to make it equal to the DEA bias. Moreover, the argument of non-feasibility is reinforced by the fact that after some sample size the bootstrap bias becomes larger than the DEA bias, which would make any proposed alternative questionable as we cannot be certain about when this turning point should occur. Therefore, in cases where the distribution of efficiency scores resembles the case of technological heterogeneity it is generally advisable not only to avoid bootstrap DEA, but to reconsider the DEA application as well.

Let us now consider the other extreme case where the bootstrap bias is large compared to the DEA bias, with the latter being relatively small and fast converging towards zero. This corresponds to the “Standard” case and we have already shown that we can identify the underlying DGP from the empirical distribution of DEA scores; in particular it should exhibit negative skewness (about -0.65) and kurtosis of about 2.8. Table 3.1 below presents again the DEA and bootstrap bias under the “Standard” DGP and for the 2-input/2-output specification. We have also included the standard deviation of the DEA score of the fixed DMU across the $M = 1{,}000$ generated samples from the population (see also Table 2.8 under “Standard 2/2” and DEA), which serves as an indication of the variability of the DEA scores with respect to sampling variations.

Table 3.1. “Standard” DEA and bootstrap biases

	DEA Bias	**Std**	**LSCV**	**SJ**
n = 10	0.038	0.027	0.065	0.054
n = 15	0.023	0.019	0.058	0.048
n = 20	0.018	0.014	0.051	0.042
n = 25	0.014	0.011	0.047	0.039
n = 30	0.011	0.009	0.045	0.037
n = 60	0.006	0.005	0.034	0.027
n = 120	0.003	0.002	0.026	0.021

We can see clearly that the DEA bias reduces at a faster rate compared to the two bootstrap biases while after $n = 30$ the DEA bias drops below 0.01 while the bootstrap biases are around 0.03. Especially for $n = 120$ the bootstrap bias is about 7 times larger than the DEA bias, indicating that the assumption of equal biases is violated to a considerable extent. Moreover, the sensitivity of the DEA score to sampling variations seems to significantly reduce with sample size. The question now is how should the researcher proceed in this particular case if he still wishes to test hypotheses? In the rest of this section we will explore two potential courses of action and comment on their plausibility.

The first suggestion is really an empirical observation; if the sample is large enough, and especially if $n \geq 120$ in our 2-input/2-output model, the DEA bias could be considered small enough so that $\hat{\theta}_k \simeq \theta_k$. Moreover, given that the standard deviation of the DEA score of the fixed DMU is quite small (below 0.002) across the $M = 1000$ samples, we could argue that for large enough sample sizes the DEA scores become robust to sampling variations while they are approximately equal to their population values. We therefore suggest that when approximately $n \geq 120$, it is not necessary to apply bootstrap DEA for hypothesis testing; observing the DEA scores will be adequate. We would like to remind at this point that the “Standard” DGP is associated with technological homogeneity and perfect competition. Therefore, we could generalise our argument and suggest that if the sample is technologically homogeneous (perhaps derived from a perfectly competitive market) and the sample size is large enough $(n \geq 120)$, then the DEA scores can be considered as good estimates of the population efficiency scores and any observed differences will be significant and robust to sampling variations; that is, we simply suggest applying DEA and avoid using bootstrap DEA.

However, the DEA scores are more sensitive towards sampling variations in smaller samples, evidenced by the higher standard deviation in Table 3.1 above. Hence, although the DEA bias is quite small one might want to consider an alternative approach which involves bootstrapping in order to account for the sampling variations. We could therefore explore the possibility of correcting for bootstrap bias once $\big(\hat{\theta}_k^{*,c} = \hat{\theta}_k^* -$

$\widehat{bias_k})$ instead of twice and construct confidence intervals and test hypotheses following the instructions in the previous section. The idea is that correcting once for bias would centre the bootstrap distribution on the DEA efficiency score, which is close to the population efficiency score (due to the assumed small DEA bias), and at the same time accounting for sampling variations. The assumption and at the same time the limitation of this approach is that we use as a proxy for θ_κ, the mean of $\hat{\theta}_k^{*,c}$ which is equal to $\hat{\theta}_k$.

Before elaborating on theoretical technicalities and the meaningfulness of this approach we will examine if correcting for bootstrap bias once would yield reasonable coverage probabilities. We therefore perform a Monte Carlo exercise where we employ the SW1998 intervals but corrected for bias once (denote them with SW1998c) and using the "Standard" DGP as in the previous chapter which is of interest here. The results for the "Standard" DGP using both the SW1998 and SW2000 intervals and the SW19998c ones are presented in Table 3.2 below. Despite the fact that the proposed intervals perform much better in this special case (especially as sample size increases) compared to the SW1998 and SW2000 ones, the coverage probabilities are still far from their nominal levels[92]. Therefore, bootstrap DEA is not advisable to be used under the "Standard" case with smaller samples, based on the particular simulations.

[92] We also performed an exercise with large samples under the 1-input/1-output specification and we found that the coverage probabilities converge to the nominal ones when $n = 1600$ which supports the consistency of the intervals but which makes clear that practically this approach would not be particularly successful.

Table 3.2. Coverage of SW1998c intervals compared to SW1998 and SW2000 ones

	LSCV			SJ		
Sample	***SW1998***	***SW2000***	***SW1998c***	***SW1998***	***SW2000***	***SW1998c***
n = 10	0.743	0.563	0.426	0.830	0.649	0.358
n = 15	0.574	0.401	0.507	0.764	0.498	0.461
n = 20	0.473	0.325	0.505	0.670	0.393	0.495
n = 25	0.421	0.302	0.550	0.566	0.315	0.518
n = 30	0.342	0.253	0.595	0.466	0.227	0.571
n = 60	0.226	0.151	0.663	0.165	0.079	0.638
n = 120	0.148	0.094	0.715	0.022	0.009	0.668

To summarise, in this section we explored whether any alternative approaches could be followed when the assumption of equal biases is violated to a considerable extent. In the presence of the technological heterogeneity where the DEA bias is considerably larger than the bootstrap bias we argued that it is neither feasible nor worthwhile to propose an alternative approach. In the opposite case, which is associated with perfectly competitive markets, we proposed a solution which performed better but not adequately to be considered as a practically useful approach. Thus, we conclude that if there is substantial bias asymmetry bootstrap DEA should be avoided. On a positive note, we found that in the latter case (perfect competition), and for reasonably large samples $(n \geq 120)$, DEA scores become robust to sampling variations and are approximately equal to their population values, suggesting that any observed efficiency differences can be considered as significant and robust.

3.5 On testing returns to scale

We have already established that bootstrap DEA cannot be safely applied if there is substantial asymmetry in the bootstrap and DEA biases. It is logical to expect that this finding is transferable to other extensions of bootstrap DEA which also make use of this assumption. One such popular extension is that of Simar and Wilson (2002) who test for returns to scale (RTS) using bootstrap DEA, thus accounting for the sensitivity of the characterisation of RTS towards sampling variations. In this section we explain how their method works and indicate where the assumption of equal biases is used and how bias asymmetry could affect the validity of their approach. Finally, we propose an approach for testing RTS in DEA which (i) employs the bootstrap and hence accounts for sampling variations and (ii) it does not make use of the equal biases assumption and it is therefore independent of the performance of bootstrap DEA with respect to coverage probabilities. Despite the fact that the proposed approach is at a theoretical level and requires to be examined through simulations, we believe that it is promising due to the benefits that it is associated with.

3.5.1 Measuring RTS in DEA

Returns to scale are usually tested in the DEA world to provide support on the relevant technology assumption used, unless there is theoretical intuition for using a certain RTS specification. As already explained in the previous chapter, excluding or including the concavity restriction $(\sum_{i=1}^{n} \lambda_i = 1)$, allows for the evaluation of efficiency under constant returns to scale (CRS) or variable returns to scale (VRS), respectively. Returns to scale can be computed using various techniques, depending on the specific model used[93]. A common way of assessing RTS in all models is by computing their scale efficiency, which is computed in DEA by the ratio of CRS over VRS efficiency scores (in input orientation). The idea is that a DMU which exhibits CRS has to operate under the most productive scale size (MPSS). There are two issues, though, with this: (i) DMUs have to be efficient to compute their scale efficiency otherwise their projections on the VRS frontier need to be used, and (ii) although CRS implies MPSS, the opposite might not always be true as the association of economies of scale with RTS requires the assumption of constant factor pricing[94]. Moreover, these tests might be sensitive towards sampling variations and therefore it might be sensible to consider bootstrap approaches.

3.5.2 Simar and Wilson's (2002) approach of testing RTS

The method of Simar and Wilson (2002) uses a bootstrap procedure to test for RTS which takes into account sampling variations and where the distribution of the bootstrap scale efficiency scores is used to perform the test. The attractive feature of their method compared to others in the literature is that the hypothesis or RTS is tested

[93] A comprehensive discussion of returns to scale computation can be found in Banker et al. (2004).
[94] This assumption suggests that the input prices do not change with the scale of operations or that the vector of input prices is common to all DMUs. See for example (Färe and Grosskopf, 1985)

without using assumptions about the distribution of scale efficiency, as opposed to Banker (1996). It could be also argued that it allows for examining the sensitivity of the RTS specification due to sampling variations since the bootstrap is used. On the other hand, it only tests for RTS on a sample of DMUs rather than testing for RTS of a certain DMU. In Simar and Wilson (2002) the null hypothesis is that the production technology in a sample of DMUs exhibits CRS versus the alternative of VRS:

H_0: constant returns to scale

H_1: variable returns to scale

Simar and Wilson (2002) assume output orientation and they use the mean of ratios of CRS over VRS distance functions as their test statistic, given in equation (4.5) in their paper:

$$\hat{S}_{1n}^{crs} = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{D}_n^{crs}(\boldsymbol{x}_i, \boldsymbol{y}_i)}{\widehat{D}_n^{vrs}(\boldsymbol{x}_i, \boldsymbol{y}_i)} = \widehat{\omega}_{obs} \tag{3.15}$$

where n is the number of DMUs in the sample and $\widehat{D}$ denotes the estimated distance function (which is used to calculate efficiency scores in a general non-parametric setup). In input orientation and using efficiency scores instead of distance functions, (3.15) becomes:

$$\hat{S}_{1n}^{crs} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\theta}_n^{crs}(\boldsymbol{x}_i, \boldsymbol{y}_i)}{\hat{\theta}_n^{vrs}(\boldsymbol{x}_i, \boldsymbol{y}_i)} = \widehat{\omega}_{obs} \tag{3.16}$$

In their method, Simar and Wilson (2002) compute both the CRS and VRS efficiency scores in each bootstrap loop, they divide them to compute the bootstrap scale efficiency scores and then they calculate their means in each bootstrap replication, generating a bootstrap distribution of average scale efficiencies ($\widehat{\omega}^*$). The resulting distribution is then used to perform hypothesis tests for RTS, which is outlined in equation (5.11) in Simar and Wilson (2002) and which is shown to be asymptotically consistent:

$$\hat{p} = \Pr(\widehat{\omega}^* \leq \widehat{\omega}_{obs} | H_0, \Phi_n) \tag{3.17}$$

where Φ_n is the observed sample of inputs and outputs of the n DMUs. Hence, if the chosen level of significance is a then the null hypothesis of CRS is rejected if $\hat{p} \leq a$.

The intuition behind this test lies in the fact that $\hat{S}_{1n}^{crs}$ is a ratio of CRS over VRS efficiency scores; the maximum value of this ratio is 1 while the higher it is, the smaller will be the distance between the CRS and VRS frontiers. If we knew the population value ω and we could observe the sampling variations of its estimate $\widehat{\omega}$, then we could examine how sensitive the distance between the CRS and VRS frontiers is towards sampling variations. For example, if we observed quite frequently that $\widehat{\omega} < \omega$, we would deduce that there is a low chance for a random sample generated from the population to be associated with a sample VRS frontier closer to the CRS one, compared to the distance between the population VRS and CRS frontiers. This suggests that this smaller distance is robust to sampling variations and we therefore conclude that the population exhibits CRS. On the other hand, if we would rarely evidence $\widehat{\omega} < \omega$, then we would consider that the population exhibits VRS as in the vast majority of the random samples we would observe a larger distance between the sample CRS and VRS frontiers compared to the distance of the population frontiers.

However, we cannot observe ω and we therefore employ the bootstrap in order to mimic the aforementioned sampling variations and we perform the hypothesis test as outlined above. The assumption is that under the null hypothesis H_0, the bootstrap bias is similarly distributed as the DEA or model bias: $(\widehat{\omega}^* - \widehat{\omega})|H_0 \sim (\widehat{\omega} - \omega)|H_0$. This assumption is similar to the ones used for Simar and Wilson's (1998) bootstrap DEA and given the results of the previous chapter we believe it may not be plausible, at least under certain DGPs. The Monte Carlo evidence provided in Simar and Wilson (2002) suggest that in small samples the computed probabilities do not converge to the nominal ones, although they approach them. In particular, the largest sample examined consists of 60 DMUs and under a 2-inputs/1-output specification the computed probabilities where 0.15 for a nominal probability of 0.05. It is worthwhile, though, to note that Simar and Wilson (2002) consider the computed probabilities as "close enough" to the nominal ones. Certainly, the literature would benefit from a more extensive simulation study on testing for RTS with this approach.

Let us now examine what the violation of the assumption of equal biases implies in this case. Suppose that the bootstrap bias is substantially greater than the DEA bias, which suggests that $\Pr(\widehat{\omega}^* - \widehat{\omega} > 0) > \Pr(\widehat{\omega} - \omega > 0) \Rightarrow \Pr(\widehat{\omega}^* < \widehat{\omega}) < \Pr(\widehat{\omega} < \omega)$. That is, it would be possible to reject a true null which means that the probability of a Type I error is higher. Similarly, if the bootstrap bias was substantially smaller than the DEA bias, the probability of a Type II error would be higher (accept a false null). This supports our previous argument that there is scope for further research on this area with Monte Carlo simulations which report among others the bootstrap and DEA biases.

The advantage of the method of Simar and Wilson (2002) is that it allows testing for RTS for a group of DMUs while employing the bootstrap which accounts for sampling variations. The disadvantage is that it might perform poorly in some cases as its validity depends on the plausibility of the assumption of equal biases. Given our concerns in the previous chapter on the plausibility of the assumption of equal biases it becomes apparent that it would be desirable to use a method that would not depend on this assumption while accounting for sampling variations.

3.5.3 *A proposed approach for testing RTS*

We will now propose an approach for testing RTS which does not depend on the assumption of equal bootstrap and DEA biases, but uses a less restrictive assumption. The approach is at a theoretical stage, requiring Monte Carlo simulations to explore its performance and sample size requirements. The idea is simple and it is based on the definition of RTS by Banker and Thrall (1992), which was later developed by Banker et al. (1996). Here we only discuss the case of testing for RTS under the assumption of a CRS frontier and input orientation.

Banker and Thrall (1992) prove that the RTS of DMU k are defined by the sum of weights $(\sum_{i=1}^{n} \lambda_i)$. In particular,

$$\text{if } \sum_{i=1}^{n} \lambda_i \begin{cases} < 1, \text{ then IRS} \\ = 1, \text{ then CRS} \\ > 1, \text{ then DRS} \end{cases} \tag{3.18}$$

There are two issues here: (i) DMU k must either be efficient or its projection on the frontier should be used, while (ii) we need to reach at the same RTS characterisation for all alternate optima.

Banker et al. (1996) propose a test for RTS which is free of both assumptions: DMUs do not need to be efficient while it is not necessary to examine RTS under all alternate optima. In particular they propose a two-step procedure, the first step of which involves solving the envelopment form of DEA in (2.11):

$$\hat{\theta}_k = \min\left\{\theta | y_k \leq \sum_{i=1}^{n} \lambda_i y_i \,;\ \theta x_k \geq \sum_{i=1}^{n} \lambda_i x_i \,;\ \theta > 0; \lambda_i \geq 0,\ \ \forall\ i = 1, \ldots, n\right\} \tag{3.19}$$

Assuming that the first step has reached a solution for DMU k with $\sum_{i=1}^{n} \lambda_i > 1$, the second step involves solving the following linear program:

$$\begin{aligned} \min\Bigg\{\sum_{i=1}^{n} \hat{\lambda}_i - \varepsilon\left(\sum_{s=1}^{p} \hat{s}_s^- + \sum_{r=1}^{q} \hat{s}_r^+\right) | \ y_k = \sum_{i=1}^{n} \hat{\lambda}_i y_i - \hat{s}^+;\ \hat{\theta}_k^* x_k \\ = \sum_{i=1}^{n} \hat{\lambda}_i x_i + \hat{s}^-;\ \sum_{i=1}^{n} \hat{\lambda}_i \geq 1;\ \ \hat{\lambda}_i, \hat{s}_s^-, \hat{s}_r^+ \geq 0,\ \ \forall\ i = 1, \ldots, n \Bigg\} \end{aligned} \tag{3.20}$$

Here $\hat{s}^-$ is a vector of p input slacks, $\hat{s}^+$ is a vector of q output slacks, while $\hat{\theta}_k^*$ is computed from the first stage and is treated as a constant (Banker et al., 1996). The quantity $\varepsilon > 0$ is a non-Archimedean element which is smaller than any positive real number and which is used to indicate that (3.20) is computed in two phases. In particular, in the first phase $\sum_{i=1}^{n} \hat{\lambda}_i$ is minimised subject to the constraints in (3.20) while in the second phase the sum of slacks $\left(\sum_{s=1}^{p} \hat{s}_s^- + \sum_{r=1}^{q} \hat{s}_r^+\right)$ is maximised subject to the same constraints. If $\sum_{i=1}^{n} \lambda_i < 1$ in (3.19), we solve the same linear program as in (3.20) by changing the objective function appropriately as $\max\left\{\sum_{i=1}^{n} \hat{\lambda}_i + \varepsilon\left(\sum_{s=1}^{p} \hat{s}_s^- + \sum_{r=1}^{q} \hat{s}_r^+\right)\right\}$ while also changing the last constraint to $\sum_{i=1}^{n} \hat{\lambda}_i \leq 1$. The optimised values of the weights on the second stage in these two cases will return values for $\sum_{i=1}^{n} \hat{\lambda}_i$ that will either confirm $\sum_{i=1}^{n} \lambda_i >< 1$ or they will return $\sum_{i=1}^{n} \hat{\lambda}_i = 1$ indicating CRS.

Finally, if the first stage in (3.19) yields $\sum_{i=1}^{n} \lambda_i = 1$, then no further treatment is required and CRS will prevail. This point is also explained in Cooper et al. (2006; pp.139) where it is stated (and shown) that “CRS will prevail at the efficient point” (meaning the projection on the frontier). This suggests that if the sum of weights in the first stage for any DMU is equal to one, then necessarily this DMU exhibits CRS. On the other hand if it exhibits IRS or DRS then the linear program in (3.20) will either confirm this finding or will suggest CRS. It has to be noted, though, that this RTS test by Banker et al. (1996), as with most RTS tests, is sensitive to orientation and this is one of the limitations of this approach.

It has already been established that DEA is subject to sampling variations and therefore the computation of either $\sum_{i=1}^{n} \lambda_i$ or $\sum_{i=1}^{n} \hat{\lambda}_i$ might be affected. Since the bootstrap is an efficient way of simulating the sampling variations, we propose implementing the bootstrap and performing the test of Banker et al. (1996) on each replication[95]. This will yield a distribution for $\sum_{i=1}^{n} \hat{\lambda}_i$ which we could use to test for RTS in DEA while taking into account the sampling variability. The only assumption of our proposed approach is that the observed sample is a representative one and that the sampling variations are adequately simulated by the bootstrap.

[95] The author has already produced a Matlab code for the Banker et al. (1996) test and is in the process of adapting it for bootstrap computations.

The null hypothesis, as with Simar and Wilson (2002), is CRS and the alternative is VRS. When implementing the Banker et al. (1996) approach one could also test for increasing or decreasing returns to scale in the second stage if interested. The important point, though, which stems from Banker et al. (1996), is that at a first stage we could examine whether the sum of weights for any DMU is equal to one or not as this would determine if we should proceed with the second-stage linear program. If the level of significance is a, then we could compute the following probability for DMU k:

$$prob = \frac{\#(\sum_{i=1}^{n} \lambda_i = 1)_{k,b}}{B}, \ \ b = 1,2, \dots B \tag{3.21}$$

and examine if $prob > a$. That is, we could examine how frequently we obtain $\sum_{i=1}^{n} \lambda_i = 1$ for DMU k across the B bootstrap loops and if this exceeds a, then we could accept the null hypothesis of CRS. If not we could proceed with the second stage computations of Banker et al. (1996). However, to establish the performance of the proposed test it would require Monte Carlo simulations with DGPs that simulate the null hypothesis to be examined and which is proposed for future research.

To gain a first insight on the sensitivity of RTS characterisation with respect to sampling variations and to further motivate our test we have performed a simulation exercise. In particular, using the DGPs of the previous chapter, we have computed the medians of the distributions of $\sum_{i=1}^{n} \lambda_i$ for the fixed DMU for both the DEA samples[96] and the bootstrap replications. The computation of the medians serves two purposes: (i) we can examine how well the bootstrap simulates the sampling variations by comparing

[96] This refers to the $M = 1000$ samples generated from the population.

the DEA and bootstrap values and (ii) we can get an indication of the acceptance rates for H_0 as a median of 1 would suggest that a considerable proportion of the bootstrap values has $\sum_{i=1}^{n} \lambda_i = 1$ and hence it is likely that H_0 would not be rejected.

Table 3.3 reports the medians of $\sum_{i=1}^{n} \lambda_i = 1$ for the fixed DMU, for both the DEA samples and the bootstrap resamples. Despite the fact that DMU k is inefficient, we do not need to consider its projections on the frontier according to Banker et al. (1996). Moreover, since the fixed point lies in the centre of the data which is generated from a DGP associated with CRS, it is quite likely for it to exhibit CRS as well and we will therefore consider values close to 1 as a good indication.

Inspecting the results, we first observe that for $n > 30$ the bootstrap values are very close to the DEA ones suggesting that the bootstrap simulates adequately the DEA sampling variations even in small samples. Another interesting observation is that this aspect of performance is independent of the DGP used, even under the "Trun.Normal Low" which is associated with technological heterogeneity and exhibited poor performance in the previous chapter. Finally, we find that values of either 1 or very close to 1 are reported for the fixed DMU, which means that $\sum_{i=1}^{n} \lambda_i = 1$ should be observed a considerable number of times, which is not surprising as all DGPs are associated with CRS. This also means that perhaps it would not be necessary to employ the second stage computations of Banker et al. (1996), though further simulations would be required to confirm this.

Table 3.3. Monte Carlo first insights on proposed RTS test

	Standard			Trunc. Normal Low			Trunc. Normal High			Uniform		
	1I -1O	*2I -1O*	*2I -2O*	*1I -1O*	*2I -1O*	*2I -2O*	*1I -1O*	*2I -1O*	*2I -2O*	*1I -1O*	*2I -1O*	*2I -2O*
DEA												
n = 10	0.9971	0.9992	0.9997	1.0039	0.9991	0.9994	0.9910	0.9993	0.9999	0.9950	0.9991	1.0000
n = 15	0.9997	0.9994	0.9994	0.9984	0.9996	1.0000	0.9837	1.0001	1.0001	0.9833	0.9995	0.9998
n = 20	1.0070	0.9993	0.9995	0.9935	0.9996	1.0002	0.9866	1.0000	0.9993	0.9556	0.9996	0.9997
n = 25	1.0182	0.9993	0.9996	0.9699	0.9994	0.9997	0.9971	1.0000	0.9993	0.9664	0.9995	0.9996
n = 30	1.0041	0.9993	0.9996	0.9626	0.9994	0.9997	1.0164	1.0001	0.9990	0.9970	0.9995	0.9999
n = 60	1.0185	0.9994	0.9995	0.9501	0.9993	0.9993	1.0128	1.0001	0.9992	0.9624	0.9997	0.9998
n = 120	1.0040	0.9994	0.9995	0.9590	0.9994	0.9994	0.9972	1.0001	0.9993	0.9396	0.9997	0.9997
LSCV												
n = 10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 15	1.0000	0.9997	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 20	1.0000	0.9994	0.9996	1.0000	0.9998	1.0000	1.0000	1.0000	0.9998	1.0000	0.9999	0.9999
n = 25	1.0000	0.9993	0.9995	1.0000	0.9995	0.9997	1.0000	1.0000	0.9993	1.0000	0.9998	0.9998
n = 30	1.0000	0.9993	0.9995	1.0000	0.9995	0.9995	1.0000	1.0000	0.9991	1.0000	0.9996	0.9998
n = 60	1.0000	0.9993	0.9994	1.0000	0.9992	0.9994	1.0000	1.0000	0.9990	1.0000	0.9997	0.9997
n = 120	1.0000	0.9993	0.9994	1.0000	0.9993	0.9994	1.0000	1.0000	0.9990	1.0000	0.9996	0.9997
SJ												
n = 10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 15	1.0000	0.9997	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 20	1.0000	0.9994	0.9996	1.0000	0.9999	1.0000	1.0000	1.0000	0.9998	1.0000	1.0000	1.0000
n = 25	1.0000	0.9994	0.9996	1.0000	0.9995	0.9997	1.0000	1.0000	0.9993	1.0000	0.9998	0.9998
n = 30	1.0000	0.9993	0.9995	1.0000	0.9995	0.9995	1.0000	1.0000	0.9991	1.0000	0.9996	0.9998
n = 60	1.0000	0.9993	0.9994	1.0000	0.9992	0.9994	1.0000	1.0000	0.9991	1.0000	0.9997	0.9997
n = 120	1.0000	0.9993	0.9994	1.0000	0.9993	0.9994	1.0000	1.0000	0.9990	1.0000	0.9996	0.9997
Naïve												
n = 10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 15	1.0000	0.9996	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 20	1.0000	0.9994	0.9996	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	1.0000
n = 25	1.0000	0.9993	0.9995	1.0000	0.9995	0.9998	1.0000	1.0000	0.9994	1.0000	0.9997	0.9997
n = 30	1.0000	0.9993	0.9995	1.0000	0.9995	0.9996	1.0000	1.0000	0.9991	1.0000	0.9996	0.9998
n = 60	1.0000	0.9992	0.9994	1.0000	0.9992	0.9994	1.0000	1.0000	0.9991	1.0000	0.9996	0.9997
n = 120	1.0000	0.9992	0.9994	1.0000	0.9992	0.9994	1.0000	1.0000	0.9990	1.0000	0.9996	0.9996

To summarise, in this section we have discussed how tests on RTS could be performed in bootstrap DEA and in particular using the approach of Simar and Wilson (2002). We have explained that in their test, Simar and Wilson (2002) use an assumption of equal bootstrap and DEA biases which is similar to the ones used in bootstrap DEA and we have shown that violation of this assumption may lead to Type I and Type II errors. Given that the results in the previous chapter were not encouraging with respect to the assumption of equal biases we proposed an alternative approach which employs the bootstrap but makes use of the Banker et al. (1996) test. The proposed approach is, in theory, free of the assumption of equal biases while we have provided some evidence that its performance is independent of the underlying DGP. However, it is only limited to a specific DMU while Monte Carlo evidence is required to establish the power of this test and the sample size requirements, which is left for future research.

3.6 Conclusions

The literature on hypothesis testing using bootstrap DEA is underdeveloped despite the interest in empirical applications. More importantly, there are no theoretical works providing guidance about when hypothesis testing with bootstrap DEA should be avoided and what would be the implications of violating fundamental assumptions (such as the equality of bootstrap and DEA biases) on the performance of such tests. In this chapter we attempted to provide guidance as to how hypothesis testing could be performed when the assumption of equal biases is valid and what are the options when it is violated. Moreover, we discussed a few considerations that we deem important when applying these tests and we proposed lines of action accordingly, along with avenues for future research. Finally, we extended the discussion to the case of testing for RTS with bootstrap DEA (Simar and Wilson, 2002) and we proposed an alternative that does not make use of the assumption of equal biases.

Our findings on the theoretical explorations of the SW1998 and SW2000 confidence intervals lend further support to the Monte Carlo evidence of the previous chapter while the inferior performance of the SW2000 intervals compared to the SW1998 is further investigated. More importantly, we show that the associated hypothesis testing procedures require the assumption of equal biases to be valid to avoid Type I and II errors. But even if this assumption is valid one should take into account the potential positive skewness of the bootstrap distributions and the possibility that different DMUs might be associated with different sensitivity towards sampling variations. With regards to these issues, we proposed lines of action which would benefit from simulations to confirm their effectiveness and which is left for future research.

We also explored the possibility of adopting alternative approaches when the bootstrap bias is either small compared to the DEA bias (which corresponds to the case of technological heterogeneity where large DEA biases are observed) or big (which corresponds to the technologically homogeneous or "Standard" case where the DEA biases are small). For the first case we argue that even the DEA model might need to be reconsidered as the presence of large DEA biases is not desirable. For the latter case we proposed an alternative approach which significantly improves coverage probabilities but which cannot be safely used in practice as convergence is only observed in large samples. We therefore conclude that in the presence of substantial biases bootstrap DEA should not be used and the practitioner/researcher should first explore for such asymmetries. One suggestion would be to use the diagnostics of the previous chapter (that is, examine the skewness and kurtosis of the distribution of DEA scores) while the iterated bootstrap of Simar and Wilson (2004) could be relevant in this case, though it is computationally extremely demanding. On the positive side we argued that when the

DEA bias is substantially smaller than the bootstrap bias (as in the “Standard” case, which is associated with technological homogeneity and perfect competition), then for sample sizes greater than 120 the DEA scores are robust to sampling variations and very close to their population value. Hence, in this case it is not necessary to apply bootstrap DEA as the DEA scores can be considered as the “true” ones.

Finally, we show how the conclusions of our discussion are transferable to extensions of bootstrap DEA such as the test for RTS of Simar and Wilson (2002). In particular, we demonstrate how a similar bias asymmetry can lead to Type I and II errors, suggesting that this test should be applied with caution. Moreover, we introduce a test based on the approach of Banker et al. (1996), which also utilises the bootstrap to account for sampling variations but which is free of any equal biases assumption. First insights from simulations suggest that the performance of the test is independent of the underlying DGP. However, a focused simulation study would be required in order to confirm its validity and assess its performance, though the first evidence seems promising. Moreover, we have only discussed the case of testing for RTS for a certain DMU which could be extended in the future to test for RTS in a sample.

Bootstrap DEA is a valuable approach which allows considering for sampling variations in DEA and therefore to perform hypothesis tests. It depends, however, on assumptions which have been challenged in the previous chapter and which carry implications about the performance of hypothesis tests. If the bootstrap bias is equal to the DEA bias then, as previously mentioned, the hypothesis tests discussed in this chapter can be applied. On the other hand, violation of this assumption will lead to inconsistent results. One possibility for future research could be to look at the effects of skewness on confidence intervals and the use of methods such as Efron's (1987) BC_a intervals which might improve coverage probabilities (we proposed an approach to compute the acceleration parameter in Appendix VII). To improve upon the validity of the assumption of equal biases, though, it would require reconsidering the kernel smoothing approaches which introduce additional noise in their effort to smooth out the empirical distribution (Simar and Wilson, 2002). In fact, some developments on bootstrap DEA focus their efforts on this issue but they seem to perform well in large samples. It might be worthwhile looking at alternatives to kernel density estimation, which can still enrich the support of the empirical distribution and at the same time introduce less variability which might cause distortions in the bootstrap biases. This is discussed in the next chapter where a new approach is introduced which performs well in small samples and which can make the SW1998 and SW2000 intervals along with the approaches discussed in this chapter useful in practice.

4 A simple alternative to smoothing

4.1 Introduction

The simulations in Chapter 2 have shown that despite the fact that Simar and Wilson's (1998) bootstrap DEA has nice asymptotic properties, it is less useful in practice due to its low performance in smaller samples. The unsatisfactory performance is attributed to the fact that the bootstrap biases are not equal to the DEA biases in smaller samples and we have shown that both the accuracy of confidence intervals and the validity of hypothesis testing are affected in this case. Considering alternative confidence intervals might go some way towards improving coverage probabilities, however the problem of unequal biases will not be resolved. The other potential is to improve or find an alternative to smoothing as "*kernel estimators are slow to converge*" and they "*might be a significant source of noise in the bootstrap process*" (Simar and Wilson, 2002; pp.124). This chapter proposes an alternative to smoothing which is shown to perform well and therefore allows using the confidence intervals of Simar and Wilson (1998, 2000a) in hypothesis testing as outlined in the previous chapter.

The necessity to employ smoothing in bootstrap DEA stems from the fact that the support of the empirical distribution is not rich and it would result in repeated values and therefore in bootstrap distributions with peculiar properties. This issue is well-established in the works of Simar and Wilson, while it is also referenced in studies not related to DEA. A review of the arguments in favour of the smooth bootstrap has been provided is subsection 2.6.4 of chapter 2, where it was also shown in Figure 2.7 why the naïve bootstrap is a bad idea.

An important body of the literature on bootstrap DEA focuses on more efficient smoothing processes, as already mentioned in subsection 2.6.5. The methods of Kneip et al. (2008) on double smoothing, of Kneip et al. (2011) on using a mixture of smooth and naïve processes and of Simar and Wilson (2011) on subsampling are the most well-known (if not the only) recent developments on this area. However, as the aforementioned papers state or show through simulations, these methods are either too complicated as well as computationally intensive, or require large samples (certainly well above 100 and ideally close to 1000) to perform well. All smoothing processes thus far employ either simple or complicated kernel smoothing techniques, while no alternative approaches have been proposed to the extent of our knowledge.

In this chapter we propose a simple alternative to smoothing which is based on using a Pearson system moment generator to draw values from a pseudo-population instead of the empirical distribution (naïve bootstrap) or some smoothed function of it (smooth bootstrap). The success of the proposed method is based on the idea that, if the DEA samples have moments (mean, standard deviation, skewness and kurtosis) which approach those of the population, we could use those sample moments to generate a pseudo-population of efficiency scores which would enrich the support of the empirical distribution and produce meaningful confidence intervals. Hence, the "moments-bootstrap", as we name it, has the same purpose as the smooth bootstrap but it uses an alternative technique in doing so.

Using Monte Carlo simulations, we show that the implementation of the SW1998 and SW2000 intervals under the “moments-bootstrap” yields better results compared to using the smooth bootstrap. In fact, the combination of SW1998 intervals and of the moments bootstrap exhibits coverage probabilities which converge to the nominal ones for sample sizes of 120 DMUs or more. The success of the proposed method is due to the fact that the resulting bootstrap biases are very similar to the DEA ones which is the fundamental assumption in Simar and Wilson’s works. Moreover, the confidence intervals have similar widths compared to the ones constructed under the smooth bootstrap, which can be either slightly narrower or slightly wider, depending on the DGP.

The remainder of this chapter is structured as follows: section 4.2 provides further evidence in support of using moment generators to enrich the support of the efficiency distribution, section 4.3 briefly analyses the method of moments, section 4.4 provides details about the Pearson system moment generator which is employed here, section 4.5 describes the exact steps in implementing the “moments-bootstrap”, section 4.6 presents Monte Carlo evidence on the performance of the proposed approach, while section 4.7 concludes the chapter.

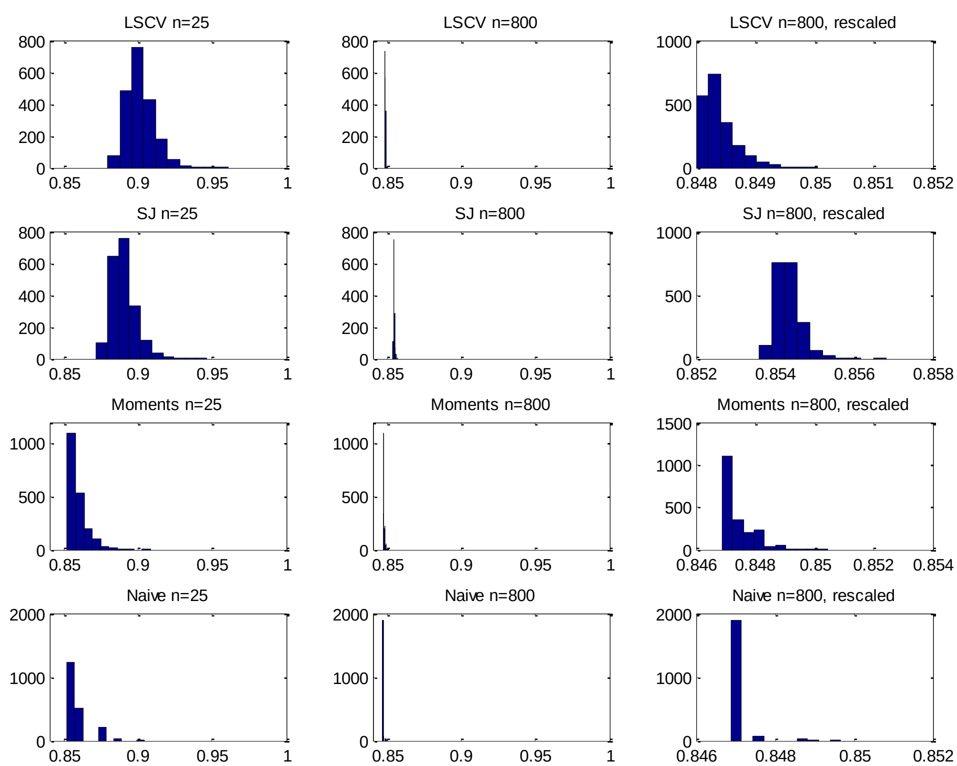
4.2 Why use moments?

The motivation of following this approach stems from the fact that the bootstrap samples mimic the observed samples, which in turn are considered as representative if they have similar properties with the population. Hence, if the sample is a representative one, then the resulting bootstrap distribution will have, in principle, good properties. In that case, the support of the empirical distribution could be consistently enriched by using the sample moments to generate a pseudo-population and apply the bootstrap by drawing values from this pseudo-population.

The resulting bootstrap DEA distribution for a certain DMU should be as rich as that resulting from the smooth bootstrap and therefore the associated confidence intervals will be also meaningful and consistent. To provide an illustration of what the moments-bootstrap does, we have plotted the relevant bootstrap distribution for a certain DMU in Figure 4.1. This is the same example as in Figure 2.7 with the addition of the moments-bootstrap approach. The labels are self-explanative and it is obvious from the figure below that the moments-bootstrap, like the two smooth bootstraps, provides a better support than the naïve bootstrap and is therefore suitable for hypothesis testing. One interesting point to note is that the distribution seems to be peaked close to the sample DEA score and exhibits a tail to the right, suggesting that the moments-bootstrap is perhaps more suitable for one-sided tests.

The advantage of the moments-bootstrap, as it will be explained later, is that it offers the flexibility of choosing an appropriate density function over a selection of distributions as opposed to the kernel density estimation approaches which employ reflection and fit a symmetric distribution with a normal kernel on data. The latter approach has been documented to introduce extra noise in the bootstrap (Simar and Wilson, 2002) which is probably avoided by using the proposed approach given its improved performance. In addition, it would be possible to recognise the corresponding density function and perform further inference using the respective functional forms (though, this is not the focus of this chapter and this is left for future research). Finally, it is computationally less demanding while it can be easily implemented using interpreters such as Matlab or R.

Figure 4.1. Moments-bootstrap and smooth bootstrap histograms



4.3 Method of moments

The foundations of moment-matching mechanisms lie within the method of moments. This method suggests that if the sample is a representative one, then the sample moments can be used to infer those of the population. Using sample moments as estimators of population parameters is a consistent approach. More information can be found in any advanced econometrics book (see for example Greene (2003)) while we will expose here some fundamental information.

Suppose a function of y which is characterized by K parameters, or $f(y|\theta_1, \dots, \theta_K)$. If there are n observations in the sample then the k^{th} sample moment is defined as:

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} y_i^k \tag{4.1}$$

which is associated with the population moment $\mu_k(\theta_1, \dots, \theta_K)$. Hence, we could use the K moment equations $\hat{\mu}_k - \mu_k(\theta_1, \dots, \theta_K),\ \ k = 1, \dots, K$ and solve for $\hat{\theta}_k$ as a function of the sample moments $\hat{\mu}_k$ (Greene, 2003). For example, if $y \sim N(\mu, \sigma^2)$, then $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$ and $\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} {y_i}^2$, so that $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_1)^2 = \hat{\mu}_2 - \hat{\mu}_1^2$, or $\sigma = \sqrt{\hat{\mu}_2 - \hat{\mu}_1^2}$. Hence, the first two sample moments of the Normal distribution can be used to estimate the two population parameters (mean and standard deviation) which will be asymptotically accurate due to consistency.

To make clearer the usefulness of the method of moments, consider the Gamma distribution with a probability density function $f(y) = \frac{1}{\Gamma(a)\beta^a} y^{a-1} e^{-y/b}$, that belongs to the exponential family. It can be shown (Greene, 2003) that $\hat{\mu}_1 = a\beta$ and $\hat{\mu}_2 =$

$a(a+1)\beta^2$. Hence, we could solve for the shape parameter a and the scale parameter β and get: $a = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2}$ and $\beta = \frac{\hat{\mu}_2 - \hat{\mu}_1^2}{\hat{\mu}_1}$.

The purpose of the exposition of the fundamentals of the method of moments in this section is to underline that the moments of each distribution are associated with certain values and relationships that characterize them. The important implication is that each distribution will have a unique combination of moments which cannot be associated with another distribution. Hence, the mean, standard deviation, skewness and kurtosis of a distribution, could be associated with some known distribution and hence with some functional form. Taking also into account the consistency of the method of moments, we infer that sample moments could be potentially used to identify the underlying population distribution, provided that the sample is a representative one.

4.4 Pearson system random number generator

The Pearson system moment generator is a random number generator that draws values from one of the distribution types that belong in the family of Pearson's distributions. The Pearson family includes most types of standard distributions which are most commonly used in the econometrics literature. The 8 types included cover a wide range of potential distributions that could be attached to most empirical distributions and it therefore seems suitable to be used in bootstrap DEA.

The decision of attaching a type from the Pearson system to the empirical distribution depends on the first four moments of the sample under consideration (mean, standard deviation, skewness and kurtosis). The methods involved are mathematically advanced and it is beyond the scope of the thesis to provide a detailed account of them all. The interested reader may refer to the book by Johnson et al. (1994) for further information on distributions and their moments. Here we will provide a summary of the various types of distributions that belong in the Pearson system as well as a description of how random values can be generated from the Pearson system.

The Pearson system includes probability density functions that satisfy a differential equation which has the following form[97]:

$$\frac{1}{p}\frac{dp}{dx} = -\frac{x+a}{c_0 + c_1 x + c_2 x^2} \tag{4.2}$$

The shape of the distribution depends on the parameters a, c_0, c_1 and c_2 while the roots of the equation:

$$c_0 + c_1 x + c_2 x^2 = 0 \tag{4.3}$$

define the solution in (4.2) and therefore the distribution-type of the Pearson system.

Suppose that $c_1 = c_2 = 0$; the solution to (4.2) would be:

$$p(x) = K \exp\left[-\frac{(x+a)^2}{2c_0}\right] \tag{4.4}$$

where K is the integrating constant and has to be $K = \sqrt{2\pi c_0}$ in order to satisfy $\int_{-\infty}^{\infty} p(x)dx = 1$. Hence $p(x) = \sqrt{2\pi c_0} \exp\left[-\frac{(x+a)^2}{2c_0}\right]$ is the resulting probability

[97] See Johnson et al. (1994), section 4.1, pp. 15 for further details. The exposition of the material here largely follows that book.

distribution with expected value a and standard deviation c_0. This is the Normal distribution and it is known as Type 0 in the Pearson system.

Type I corresponds to the case where $a_1 < 0 < a_2$ are the roots of (4.3) so that $c_0 + c_1 x + c_2 x^2 = -c_2(x - a_1)(x - a_2)$. It can be shown that this corresponds to the Beta distribution with the following solution:

$$p(x) = K(x - a_1)^{m_1}(x - a_2)^{m_2}, \quad m_1 = \frac{a + a_1}{c_2(a_2 - a_1)} \text{ and } m_2 = \frac{a + a_2}{c_2(a_2 - a_1)} \tag{4.5}$$

If $m_1 = m_2$, then this gives rise to a Symmetric Beta distribution and corresponds to Type II of the Pearson system.

Type III is the case where $c_2 = 0$ and $c_1, c_2 \neq 0$ which has the following solution:

$$p(x) = K(c_0 + c_1 x)^m \exp\left(\frac{-x}{c_1}\right), \qquad m = c_1^{-1}(c_0 c_1^{-1} - a) \tag{4.6}$$

This is the case of Gamma distribution.

Type IV does not belong to some standard distribution density as (4.3) is assumed to have no real roots. The solution to (4.3) is extremely complicated and it is usually computed by numerical approximations while various papers have tried to come up with an accessible functional form. In all cases the solution is of the form $p(x) = g(a, c_0, c_1, c_2)$ and it involves imaginary numbers[98].

Type V of the Pearson system corresponds to the case where (4.2) is a perfect square, or $c_1^2 = 4c_0 c_2$. The solution to (4.3) now becomes:

---

[98] Johnson et al (1994) provide the following functional form: $p(x) = K[C_0 + c_2(x + C_1)^2]^{-1/2c_2} \exp\left(-\frac{a - C_1}{\sqrt{c_2 C_0}} \tan^{-1} \frac{x + C_1}{\sqrt{C_0/c_2}}\right)$
where $C_0 = c_0 - \frac{1}{4}c_1^2/c_2$ and $C_1 = \frac{1}{2}c_1/c_2$

$$p(x) = K(x + C_1)^{-1/c_2} \exp\left[\frac{a - C_1}{c_2(x + C_1)}\right], \qquad C_1 = \frac{c_1}{2c_2} \tag{4.7}$$

which is the general form of the Inverse Gamma distribution.

Type VI is associated with the case where the roots of (4.2) are all real and have the same sign. The solution is exactly the same as the one in (4.5) and an important distribution that belongs in this family is the F-distribution.

Finally, Type VII distribution corresponds to the case where $c_1 = a = 0$ and $c_0, c_2 > 0$. Now the solution to (4.3) becomes:

$$p(x) = K(c_0 + c_2 x^2)^{-(2c_2)^{-1}} \tag{4.8}$$

A well-known distribution that belongs in this family is the t-distribution with $c_2^{-1} - 1$ degrees of freedom.

The values and restrictions on $a, c_0,\ c_1$ and c_2 make possible the distinction among the 8 different types of the Pearson System (including the normal one). It can be shown that the parameters of (4.3) can be associated with the moments of the distribution and analytical results can be obtained. In particular, the solution to the parameters of interest satisfies the following system (Johnson et al., 1994):

$$\begin{aligned} c_0 &= (4\beta_2 - 3\beta_1)(10\beta_2 - 12\beta_1 - 18)^{-1} \\ c_1 &= \alpha = \sqrt{\beta_1}(\beta_2 + 3)(10\beta_2 - 12\beta_1 - 18)^{-1} \\ c_2 &= (2\beta_2 - 3\beta_1 - 6)(10\beta_2 - 12\beta_1 - 18)^{-1} \\ \beta_1 &= (skewness)^2 \quad \text{and} \quad \beta_2 = kurtosis \end{aligned} \tag{4.9}$$

Depending on the combination of values that these parameters take and on the value that $\kappa = \frac{1}{4} c_1^2 (c_0 c_2)^{-1}$ takes, the distribution is characterized as belonging to one the types of the Pearson system[99]. Afterwards, random values can be drawn from the respective distribution, taking into account the mean, standard deviation, skewness and kurtosis of the sample.

As already mentioned, in practice this is very straightforward to apply. Compilers such as Matlab (or R) can perform this task with only one command line. In particular, the Matlab function (which is used here) is:

```
PEARSRND(MU,SIGMA,SKEW,KURT,M,N)
```

which returns an M by N matrix of values drawn from the Pearson system of distributions with mean “MU”, standard deviation “STD”, skewness “SKEW” and kurtosis “KURT”. Hence, the only step required by the user is to compute the respective sample statistics and feed them into the Matlab function.

4.5 The moments-bootstrap DEA

The moments-bootstrap, as we call it, follows the same steps as the bootstrap DEA of Simar and Wilson (1998), with the only exception being that the Pearson system random number generator is used instead of smoothing. In particular, we replace steps 2 and 3 in subsection 2.6.2 (see (2.20)) with the following two steps:

---

[99] The sufficient criteria for the characterization are: Type 0: $c_1 = 0, \beta_2 = 3$; Type I: $k < 0$; Type II: $\beta_1 = 0, \beta_2 < 3$; Type III: $2\beta_2 - 3\beta_1 - 6 = 0$; Type IV: $0 < \kappa < 1$; Type V: $\kappa = 1$; Type VI: $\kappa > 1$; Type VII: $\beta_1 = 0, \beta_2 > 3$.

- Use the moments of the empirical distribution of $\hat{\theta}_i, i = 1,2 \dots n$ to generate a pseudo-population of efficiency scores $\hat{\theta}_j^M, j = 1,2 \dots N$, so that $\hat{\theta}_j^M \in (0,1]$.
- Randomly draw n values of pseudo-efficiency scores from $\hat{\theta}_j^M$:

$$\theta_i^*, i = 1,2 \dots n \tag{4.10}$$

Hence, we choose a value for N which has to be large enough to generate a smooth pseudo-population distribution. We use $N = 5000$ in our simulations.

One of the limitations of the proposed approach, is that the distribution of the generated pseudo-population has to be truncated so that the generated pseudo-efficiency scores lie between 0 at 1, to avoid theoretical inconsistencies. This is expected to have a small impact on results as the Pearson system would generate distributions that recognize such limitations, especially as sample size increases. However, there is a chance for some generated values on the right tail to "misbehave". In these cases we delete these values and we ask the generator to replace them with others that satisfy our restrictions. This limitation does not restrict the validity of the results[100]; however, future research could examine alternatives to truncation.

[100] To examine the extent to which the results might be affected by the truncation, we compared the moments of the truncated pseudo-population and the moments of the non-truncated one. We find that the median absolute differences (MAD) of these moments becomes very small and certainly too small to be considered as capable of changing the characterisation of the Pearson distribution type. Appendix VIII includes more information about this exercise and presents the relevant results.

4.6 Monte Carlo evidence

We provide Monte Carlo evidence on the performance of the moments-bootstrap. The Monte Carlo exercise is exactly the same as the one performed in chapter 2 and the interested reader may refer to section 2.8 for a recollection of the data generating processes (DGPs) used. The evaluation of coverage probabilities is performed on the basis of the SW1998 and SW2000 intervals to evaluate the enhancement in coverage. We first compare the population, sample and bootstrap moments to assess the plausibility of this method. We then compare the bootstrap bias generated from the moments-bootstrap with that of the other approaches and we compute coverage probabilities and examine the behaviour of confidence intervals.

4.6.1 Population, sample and bootstrap moments

The performance of this approach is based on the assumption that the sample moments are close enough to the population ones. Hence, the moments-bootstrap will return distributions with moments similar to the sample ones, by construction, which are expected to be similar to the population moments, by implication.

A clarification required here is that we do not refer to the moments of the fixed point but to the moments of the distribution of efficiency scores. The bootstrap draws values from the empirical distribution of efficiency scores and it is therefore reasonable to state that if the moments of this distribution are close to the population ones, then the bootstrap results will be meaningful. Considering this point from a different perspective, smoothing procedures discussed in the previous chapters aim at capturing the asymptotic properties of the underlying population distribution. Therefore, our idea of comparing sample and population moments of the efficiency distributions and using the Pearson generator to produce “pseudo-population” values does not lack theoretical or intuitive basis.

Table 4.1 presents the mean, standard deviation, skewness and kurtosis (which we loosely refer to as the first 4 moments) of the population, the sample and the bootstrap. We present findings for the 2-inputs/2-outputs case, to conserve space, while more detailed evidence can be found in Appendix IX. The labels are self-explanative and the results are provided for the population, the sample (DEA), the two smooth bootstraps (LSCV and SJ), the herein introduced moments-bootstrap (moments) and the naïve bootstrap. We need to note at this point that, regarding the bootstrap moments, we actually present the centre (median) of the distribution of the respective moments as an indication of representative behaviour of the Monte Carlo simulations.

Comparing the population moments with the sample ones, we find that in all cases DEA performs well as it approaches the population statistics quite fast. An interesting finding is that in the case of technological heterogeneity (“Trun. Normal Low”), apart from a substantial overestimation of the population mean, the higher moments are substantially underestimated in smaller samples. This suggests that, apart from the issues reported in the previous chapters, in such cases hypothesis testing might not be a safe choice overall.

Comparing the moments of the smooth bootstraps with those of the moments-bootstrap we find that the behaviour is quite similar, with the exception of the mean. The mean under the moments-bootstrap is always closer to the DEA and population means compared to the smooth bootstraps with the exception of the DGP associated with technological heterogeneity. Given the randomness in the Monte Carlo resampling, we cannot consider these differences as substantial and we therefore conclude the that moments-bootstrap produces bootstrap samples which have at least similar properties and behaviour with that of the smooth bootstraps. However, the moments-bootstrap samples are located closer to the true ones and this difference is more evident in smaller samples. This might suggest that the proposed approach is more appropriate to be used in small samples as it will have similar shape to the ones related to the smooth bootstraps but will be displaced towards the population centre.

Table 4.1. Population, sample and bootstrap moments

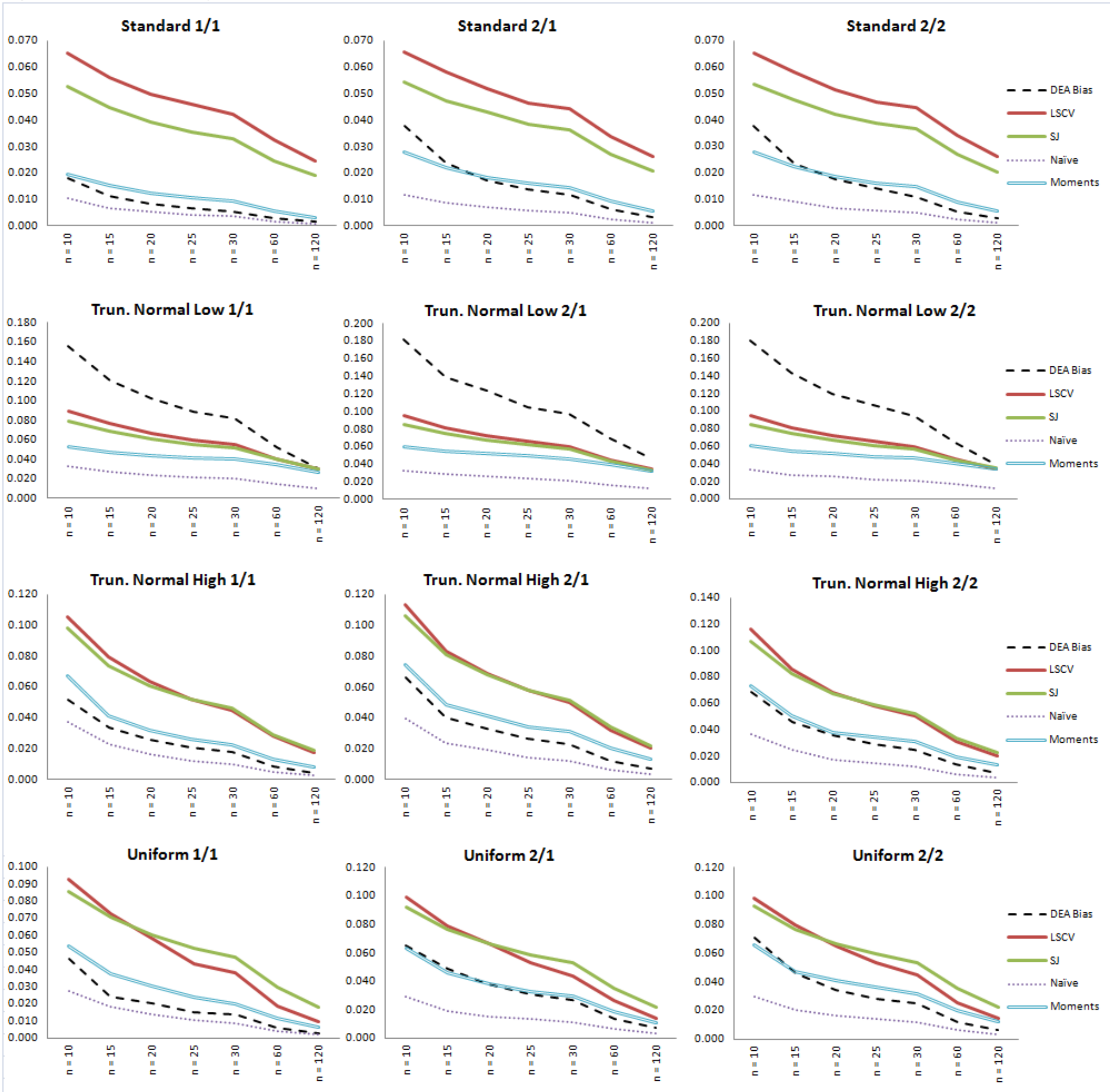
	Standard 2/2				**Trun. Normal Low 2/2**				**Trun. Normal High 2/2**				**Uniform 2/2**			
Population	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
N = 10,000	0.859	0.097	-0.675	2.893	0.617	0.121	0.412	3.003	0.493	0.241	0.284	2.074	0.688	0.158	0.286	1.885
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.891	0.097	-0.494	2.117	0.786	0.139	0.146	2.125	0.555	0.278	0.347	1.921	0.756	0.163	0.250	1.829
n = 15	0.882	0.098	-0.472	2.185	0.759	0.138	0.240	2.266	0.538	0.270	0.355	1.975	0.741	0.164	0.256	1.855
n = 20	0.880	0.098	-0.501	2.235	0.741	0.139	0.268	2.330	0.534	0.267	0.303	1.947	0.729	0.164	0.307	1.862
n = 25	0.876	0.099	-0.536	2.344	0.730	0.138	0.270	2.407	0.526	0.263	0.338	2.000	0.724	0.163	0.304	1.898
n = 30	0.873	0.099	-0.519	2.370	0.720	0.138	0.317	2.505	0.524	0.258	0.321	2.026	0.717	0.164	0.314	1.887
n = 60	0.869	0.098	-0.597	2.579	0.688	0.134	0.356	2.676	0.515	0.255	0.309	2.033	0.707	0.162	0.293	1.888
n = 120	0.865	0.098	-0.650	2.753	0.667	0.131	0.381	2.802	0.504	0.248	0.305	2.066	0.700	0.161	0.291	1.889
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.959	0.108	-0.476	2.470	0.876	0.161	0.318	2.601	0.676	0.348	0.493	2.181	0.845	0.189	0.410	2.062
n = 15	0.943	0.107	-0.450	2.383	0.834	0.157	0.367	2.643	0.630	0.322	0.443	2.181	0.814	0.185	0.351	1.990
n = 20	0.933	0.105	-0.473	2.395	0.808	0.156	0.388	2.667	0.613	0.310	0.370	2.085	0.789	0.180	0.376	1.966
n = 25	0.927	0.106	-0.521	2.479	0.791	0.154	0.368	2.704	0.596	0.301	0.392	2.118	0.774	0.177	0.359	1.996
n = 30	0.921	0.105	-0.509	2.488	0.778	0.152	0.405	2.791	0.584	0.293	0.368	2.131	0.761	0.176	0.365	1.966
n = 60	0.905	0.103	-0.590	2.645	0.735	0.144	0.409	2.877	0.555	0.276	0.332	2.082	0.736	0.169	0.320	1.928
n = 120	0.893	0.101	-0.645	2.786	0.703	0.139	0.410	2.918	0.531	0.263	0.316	2.090	0.717	0.165	0.303	1.909
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.944	0.107	-0.468	2.476	0.870	0.161	0.323	2.602	0.674	0.348	0.495	2.190	0.847	0.190	0.417	2.068
n = 15	0.931	0.106	-0.449	2.381	0.832	0.157	0.369	2.648	0.637	0.325	0.444	2.184	0.819	0.185	0.352	1.991
n = 20	0.924	0.105	-0.469	2.399	0.808	0.156	0.389	2.672	0.619	0.312	0.372	2.088	0.797	0.182	0.377	1.970
n = 25	0.917	0.105	-0.519	2.478	0.792	0.154	0.371	2.699	0.601	0.303	0.393	2.116	0.782	0.179	0.362	1.996
n = 30	0.913	0.105	-0.507	2.488	0.778	0.152	0.406	2.796	0.591	0.294	0.368	2.131	0.772	0.178	0.365	1.969
n = 60	0.898	0.103	-0.590	2.645	0.735	0.144	0.408	2.877	0.560	0.279	0.332	2.082	0.745	0.171	0.321	1.929
n = 120	0.887	0.101	-0.645	2.786	0.705	0.139	0.410	2.918	0.535	0.264	0.316	2.090	0.725	0.167	0.303	1.909
Moments	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.919	0.105	-0.455	2.504	0.843	0.159	0.335	2.633	0.629	0.325	0.513	2.174	0.815	0.183	0.430	2.107
n = 15	0.906	0.103	-0.443	2.397	0.813	0.154	0.382	2.663	0.597	0.305	0.469	2.142	0.787	0.179	0.362	2.007
n = 20	0.900	0.103	-0.465	2.409	0.793	0.154	0.397	2.694	0.583	0.294	0.390	2.068	0.769	0.176	0.382	1.988
n = 25	0.895	0.103	-0.517	2.485	0.780	0.152	0.381	2.714	0.570	0.287	0.404	2.108	0.758	0.174	0.366	2.006
n = 30	0.891	0.102	-0.505	2.487	0.768	0.150	0.417	2.814	0.563	0.281	0.383	2.117	0.750	0.173	0.367	1.981
n = 60	0.880	0.101	-0.587	2.642	0.731	0.144	0.413	2.898	0.541	0.269	0.337	2.082	0.730	0.168	0.322	1.932
n = 120	0.873	0.099	-0.642	2.785	0.703	0.139	0.416	2.931	0.523	0.258	0.318	2.093	0.715	0.164	0.304	1.912
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.904	0.102	-0.483	2.491	0.809	0.152	0.324	2.587	0.577	0.299	0.487	2.176	0.778	0.175	0.411	2.062
n = 15	0.893	0.101	-0.461	2.385	0.783	0.149	0.374	2.642	0.560	0.284	0.443	2.168	0.760	0.171	0.345	1.990
n = 20	0.889	0.101	-0.481	2.411	0.762	0.148	0.390	2.687	0.551	0.278	0.365	2.078	0.743	0.170	0.375	1.961
n = 25	0.884	0.101	-0.526	2.485	0.751	0.146	0.377	2.707	0.542	0.274	0.386	2.107	0.736	0.168	0.360	1.992
n = 30	0.880	0.101	-0.511	2.488	0.740	0.144	0.409	2.790	0.537	0.267	0.365	2.125	0.730	0.168	0.365	1.965
n = 60	0.873	0.099	-0.590	2.650	0.705	0.138	0.406	2.880	0.523	0.260	0.330	2.081	0.715	0.164	0.317	1.927
n = 120	0.868	0.099	-0.647	2.788	0.680	0.135	0.413	2.918	0.510	0.252	0.316	2.089	0.705	0.162	0.302	1.907

4.6.2 Bootstrap and DEA biases

We now turn to the comparison of the bootstrap and DEA biases which is important for the finite performance of Simar and Wilson's approaches. We remind that the SW1998 and SW2000 intervals are based on the assumption that the DEA (or model) and bootstrap biases are equal. Here, we provide Monte Carlo evidence about the behaviour of the moments-bootstrap compared to the smooth and naïve bootstraps.

Figure 4.2 below presents the bootstrap and DEA biases associated with the “fixed DMU”. The fixed DMU is defined exactly as in chapter 2 while the figure below is exactly the same as Figure 2.11 with the addition of the bias of the moments-bootstrap (blue double line). In all cases, except under “Trun. Normal Low” which is associated with technological heterogeneity, the bootstrap bias associated with the moments-bootstrap is very close to the DEA bias (black dotted line). This suggests that the moments-bootstrap satisfies the assumption of Simar and Wilson (1998, 2000a) of equal bootstrap and DEA biases to a greater extent compared to the two smooth bootstraps (and of course the naïve). We would therefore expect that the coverage probabilities for the respective confidence intervals of Simar and Wilson will be higher if the moments-bootstrap is employed instead of the smooth bootstraps. This also suggests that we can make use of the hypothesis testing approaches discussed in the previous chapter more safely.

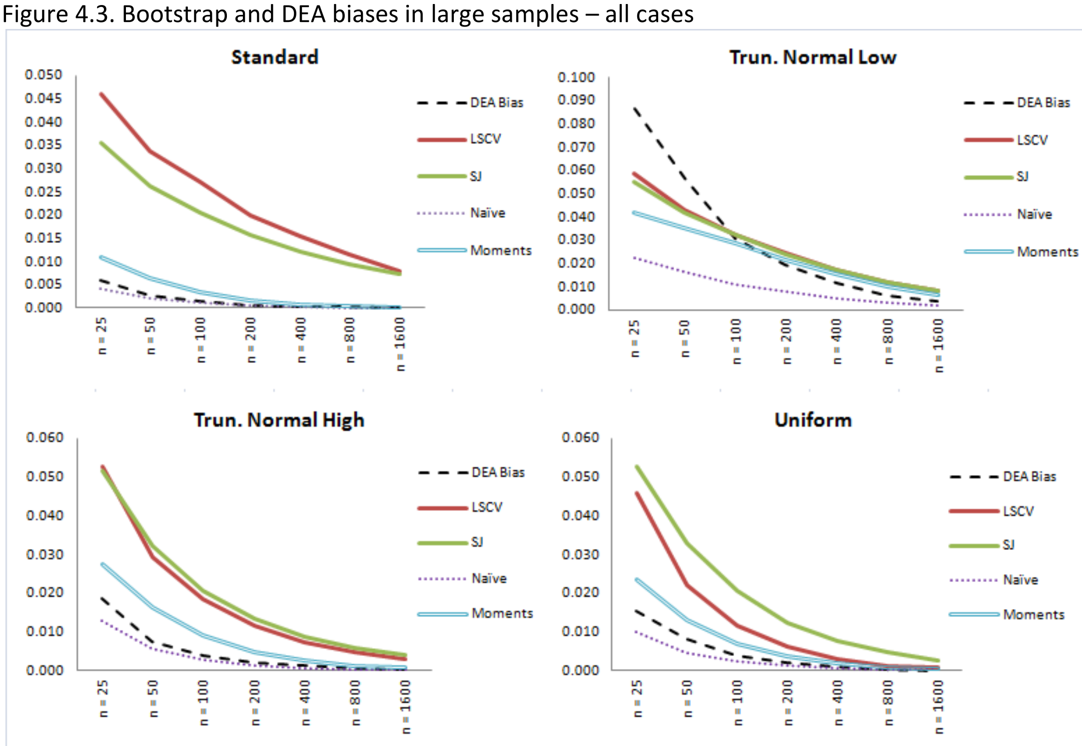
Figure 4.2. Bootstrap and DEA biases – All cases



To confirm that the moments bootstrap generates bootstrap and DEA biases which converge asymptotically we also examined the behaviour of biases in large samples but only for the 1-input/1-output case (due to computational limitations). The results are presented in Figure 4.3 where it is obvious that the good behaviour of the moments bootstrap is preserved asymptotically, providing further evidence that Simar and Wilson's fundamental assumption of equal biases works under the moments bootstrap.

As already mentioned in chapter 2 the case of technological heterogeneity ("Trun.Normal Low") requires special attention as the convergence is considerably slower.

Figure 4.3. Bootstrap and DEA biases in large samples – all cases



4.6.3 Coverage probabilities - Small samples

We now present results on coverage which is a performance indicator of the proposed method. Table 4.2 replicates the information of Table 2.7 on the LSCV and SJ smooth bootstraps for comparison and reports the coverage probabilities for the moments-bootstrap on the last section (we present the 2-input/2-output case here but results for all dimensions can be found in Appendix X).

Table 4.2. Coverage probabilities of 95% intervals – moments-bootstrap

	Standard 2/2		**T.N. Low 2/2**		**T.N. High 2/2**		**Uniform 2/2**	
LSCV	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 10	0.743	0.563	0.389	0.517	0.874	0.698	0.755	0.659
n = 15	0.574	0.401	0.385	0.500	0.828	0.621	0.776	0.601
n = 20	0.473	0.325	0.433	0.514	0.819	0.569	0.733	0.581
n = 25	0.421	0.302	0.441	0.511	0.811	0.513	0.745	0.574
n = 30	0.342	0.253	0.446	0.510	0.810	0.511	0.734	0.557
n = 60	0.226	0.151	0.497	0.528	0.690	0.407	0.739	0.494
n = 120	0.148	0.094	0.571	0.576	0.577	0.300	0.756	0.461
SJ	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 10	0.830	0.649	0.363	0.513	0.898	0.712	0.817	0.663
n = 15	0.764	0.498	0.387	0.487	0.920	0.592	0.862	0.605
n = 20	0.670	0.393	0.436	0.496	0.916	0.533	0.833	0.502
n = 25	0.566	0.315	0.434	0.513	0.889	0.486	0.825	0.450
n = 30	0.466	0.227	0.434	0.515	0.873	0.444	0.800	0.432
n = 60	0.165	0.079	0.512	0.525	0.722	0.300	0.593	0.249
n = 120	0.022	0.009	0.589	0.584	0.492	0.158	0.412	0.160
Moments	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***	***SW1998***	***SW2000***
n = 10	0.637	0.806	0.337	0.487	0.782	0.909	0.702	0.855
n = 15	0.727	0.823	0.358	0.492	0.813	0.916	0.753	0.864
n = 20	0.747	0.825	0.417	0.533	0.800	0.913	0.809	0.878
n = 25	0.779	0.824	0.438	0.534	0.818	0.895	0.840	0.884
n = 30	0.823	0.842	0.466	0.562	0.836	0.901	0.847	0.887
n = 60	0.866	0.814	0.574	0.640	0.885	0.886	0.906	0.860
n = 120	0.929	0.817	0.674	0.702	0.960	0.880	0.930	0.838

The results indicate that the moments-bootstrap is better behaved and associated with higher coverage probabilities. In particular for samples sizes greater than 25 the coverage probabilities under the moments bootstrap exceed the respective ones under the two smooth bootstraps considered. More importantly, for sample sizes equal or greater than 120 the coverage probabilities converge to their nominal levels in all cases and under the SW1998 intervals, except under the case of technological heterogeneity where convergence is slow. Comparing the two confidence intervals we find that the SW1998 intervals perform much better than the SW2000 ones as the latter do not achieve convergence. Finally, it is worthwhile mentioning that the probabilities exhibit almost monotonic convergence which is desirable as it suggests that their performance stabilises as sample size increases. We therefore conclude that it is safe to use the SW1998 intervals in samples sizes of 120 or more, and to apply the hypothesis testing approaches discussed in the previous chapter.

The correction that we achieved by using the moments-bootstrap indicates that there is scope for further research towards the direction of smoothing-alike processes. Enriching the support of the efficiency distribution seems critical for the finite sample performance of bootstrap DEA. Future research should focus on engineering accessible and computationally efficient processes that perform well on small samples. The more recent approaches of Kneip et al. (2011) and Simar and Wilson (2011) seem to enhance to some extent previous approaches; however, they are computationally intensive while they seem to work better in larger samples, as already mentioned.

4.6.4 Confidence intervals

To examine the behaviour of confidence intervals, we have plotted the average 95% SW1998 intervals in Figure 4.4 and the SW2000 ones in Figure 4.5. The plots in the figures below further support the good behaviour of the moments-bootstrap, especially for the SW1998 case. The Simar and Wilson's intervals almost centre the true efficiency score (or "fixed point") in all cases except for the "Trun. Normal Low" (as expected), which suggests that the good performance cannot be attributed to chance. The observed behaviour is well justified by the theoretical explorations and the simulations of the previous two chapters while the good performance is due to the fact that the assumption of equal bootstrap and DEA biases is realised in smaller samples. It also becomes apparent that the SW2000 intervals perform slightly worse than the SW1998 and they will always lie below the SW1998 ones, as already explained previously, suggesting that their inferior performance is probably due to the fact their upper bound tends to underestimate the true efficiency score.

Figure 4.4. Confidence intervals of Simar and Wilson (1998) – Moments-bootstrap

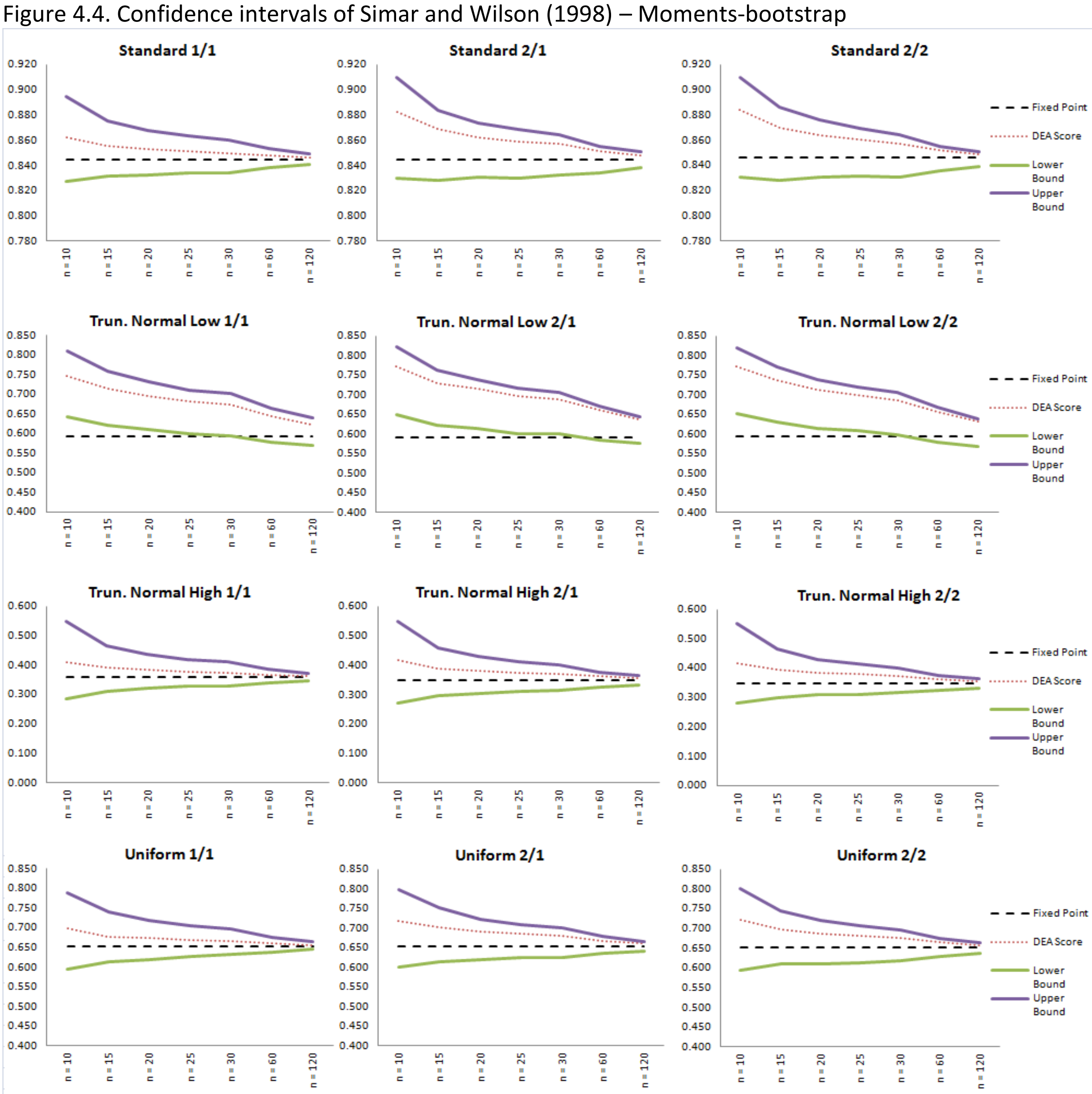
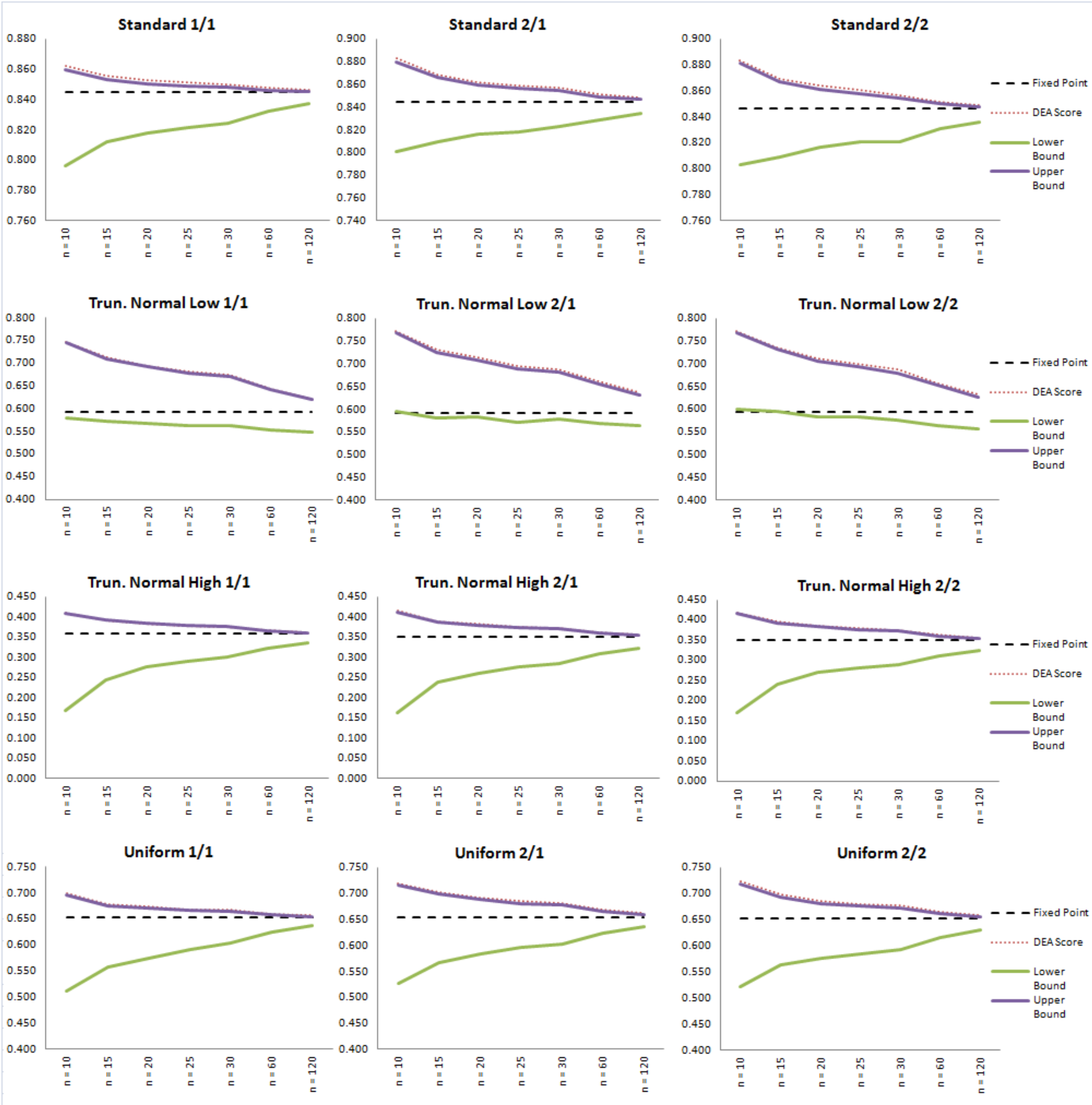


Figure 4.5. Confidence intervals of Simar and Wilson (2000a) – Moments-bootstrap



Having established the good performance of the SW1998 and SW2000 intervals under the moments bootstrap, the next step is to compare the confidence interval widths under the various approaches. We therefore computed the average widths of the 95% SW1998 confidence intervals (which are the best performing) under the moments bootstrap to the respective ones under the LSCV and SJ smooth bootstraps (see subsection 2.9.4). The results are presented in Table 4.3 and the labels are self-explanative. We observe that the moments bootstrap yields narrower SW1998 intervals under the "Standard" DGP, with the exception of $n = 120$ where the intervals are marginally wider, while in all other cases the moments bootstrap yields slightly wider intervals[101]. The differences in widths become smaller with sample size and could be considered unimportant for $n = 120$ (or more) which is the suggested sample size to be used with the moments bootstrap. In fact, any differences are limited to the third decimal place, with the exception of technological heterogeneity where the differences are larger. We therefore conclude that the SW1998 (and SW2000) intervals under the proposed alternative approach to smoothing are much more accurate while having similar widths when compared to the ones under the smooth bootstraps.

[101] It is worthwhile mentioning that the "Standard" DGP which exhibits narrower intervals is associated with technological homogeneity and perfect competition. Hence in this case the intervals are both narrower and more accurate.

Table 4.3. SW1998 average 95% confidence interval widths

	Standard			Trun. Normal Low			Trun. Normal High			Uniform		
LSCV	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***
n = 10	0.085	0.092	0.092	0.161	0.161	0.158	0.235	0.232	0.237	0.173	0.173	0.177
n = 15	0.058	0.067	0.068	0.125	0.128	0.125	0.150	0.147	0.150	0.116	0.121	0.124
n = 20	0.046	0.054	0.055	0.109	0.111	0.106	0.104	0.114	0.108	0.093	0.096	0.096
n = 25	0.037	0.045	0.045	0.094	0.100	0.096	0.081	0.088	0.089	0.070	0.078	0.080
n = 30	0.031	0.039	0.040	0.090	0.087	0.092	0.066	0.074	0.073	0.057	0.067	0.068
n = 60	0.017	0.022	0.022	0.067	0.064	0.067	0.034	0.039	0.039	0.029	0.036	0.038
n = 120	0.009	0.012	0.012	0.050	0.049	0.052	0.018	0.022	0.023	0.014	0.020	0.021
SJ	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***
n = 10	0.081	0.085	0.083	0.151	0.157	0.157	0.233	0.237	0.231	0.170	0.170	0.175
n = 15	0.056	0.063	0.064	0.123	0.124	0.124	0.145	0.147	0.151	0.123	0.127	0.123
n = 20	0.044	0.050	0.050	0.108	0.108	0.105	0.106	0.113	0.110	0.094	0.096	0.100
n = 25	0.035	0.043	0.043	0.096	0.096	0.096	0.081	0.089	0.089	0.075	0.081	0.083
n = 30	0.029	0.037	0.038	0.086	0.088	0.088	0.067	0.075	0.075	0.060	0.069	0.071
n = 60	0.015	0.021	0.021	0.066	0.063	0.067	0.034	0.039	0.040	0.031	0.038	0.040
n = 120	0.008	0.012	0.011	0.048	0.047	0.049	0.018	0.022	0.023	0.015	0.021	0.022
Moments	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***	***1I-1O***	***2I-1O***	***2I-2O***
n = 10	0.067	0.079	0.079	0.167	0.174	0.167	0.264	0.277	0.268	0.192	0.197	0.209
n = 15	0.044	0.055	0.057	0.137	0.142	0.139	0.158	0.160	0.164	0.126	0.137	0.135
n = 20	0.035	0.044	0.045	0.123	0.123	0.125	0.116	0.128	0.120	0.098	0.104	0.110
n = 25	0.030	0.038	0.037	0.113	0.117	0.109	0.093	0.101	0.102	0.079	0.086	0.093
n = 30	0.026	0.032	0.034	0.109	0.104	0.107	0.080	0.087	0.085	0.064	0.077	0.080
n = 60	0.015	0.021	0.019	0.089	0.086	0.088	0.044	0.051	0.050	0.036	0.043	0.046
n = 120	0.009	0.013	0.012	0.071	0.068	0.071	0.026	0.032	0.031	0.019	0.025	0.026

4.7 Conclusions

This chapter has suggested an alternative approach to smoothing, which performs the same task of enriching the empirical distribution of efficiency scores. Based on the observation/assumption that the samples are representative, in that the sample moments are similar to the population ones, we propose using the Pearson system random number generator to produce pseudo-populations of efficiency scores to draw from when bootstrapping. We have named this method the “moments-bootstrap”.

Comparing the population and sample moments we find that there is scope for such an approach as the latter are quite close to the former even in very small samples. Results from the Monte Carlo simulations have indicated that the moments-bootstrap can satisfy the assumption of equal bootstrap and DEA biases (Simar and Wilson, 1998) to a noticeably greater extent compared to the other two smooth bootstraps. Consequently, the coverage probabilities for Simar and Wilson's intervals under the moments-bootstrap are substantially improved; especially, for the SW1998 ones which exhibit coverage probabilities close to their nominal values for sample sizes equal or greater than 120. The only exception is the DGP associated with monopoly and technological heterogeneity where, despite the fact that coverage is improved compared to the smooth bootstraps, the associated coverage probabilities are still far from their nominal values.

The improvement in coverage probabilities comes at no additional cost as the confidence interval widths are comparable to those produced under the two smooth bootstraps. Moreover, as sample size increases, the differences in widths become very small. In particular, we find that under the DGP which is associated with perfect competition (and technological homogeneity) the SW1998 intervals generated under the moments bootstrap are narrower compared to the ones generated under the LSCV or SJ smooth bootstraps. For the other DGPs we find that the moments bootstrap generates slightly wider intervals, but the difference is too small to be considered as a limitation of this approach over the smooth bootstrap; in fact for sample sizes as large as 120 or more, the differences range from 0.004 to 0.008 (with the exception of the technologically heterogeneous case where differences are larger).

We therefore conclude that using the moments bootstrap makes the assumption of equal biases plausible in small samples and to its extent the theoretical works of Simar and Wilson implementable in practice. Using the SW1998 intervals under the proposed alternative to smoothing, allows performing hypothesis testing in samples of 120 DMUs or more following the suggestions in the previous chapter. We believe that this finding carries implications for the previously mentioned extensions of bootstrap DEA (bootstrap Malmquist DEA, tests of returns to scale and two-stage regressions) the validity of which was questioned due to observed bias asymmetries under the smooth bootstraps. Future research could focus on implementing the moments bootstrap into these approaches and on comparing their performance through Monte Carlo simulations.

The limitation of this approach is that the generated pseudo-populations are truncated; however, we have shown that this is not adequate to affect the validity of our results as the resulting truncated pseudo-population would only exhibit small differences compared to a non-truncated one, especially in larger samples. Future research could focus on alternative approaches for this issue, but also taking care not to increase the confidence interval widths as we suspect that the refection method (used in the smooth bootstrap) does. Another suggestion for future research would be the consideration of alternative approaches which would increase coverage probabilities in even smaller samples while preserving or even reducing the width of the associated confidence intervals. The author currently experiments with a “smooth-moments-bootstrap” which involves smoothing the pseudo-population generated from the moments-bootstrap, while in the future research agenda Bayesian methods such as the HPDI (highest probability density interval) could be also considered.

5 Suggested guidelines on applying bootstrap DEA

The previous chapters have discussed various aspects of bootstrap DEA, both in terms of technique as well as in terms of application. Some weaknesses were identified and some suggestions to move forward were proposed. In this short chapter we summarise these recommendations in “manual-style” guidelines for the application of bootstrap DEA. The exhibition of technical material and use of terminology is minimised in order to provide straightforward guidance to the interested practitioner.

5.1 Assumptions

There are three assumptions in bootstrap DEA: (i) the bootstrap bias is equal to the DEA bias, (ii) the sample is representative (in that the observed distribution of DEA scores reflects the distribution of the underlying population), and (iii) efficiency scores reflect practically feasible input reductions or output expansions. The last one is due to the fact that the bootstrap resamples efficiency scores randomly, suggesting that any firm could be assigned with any efficiency score in the sample.

5.2 Applying bootstrap DEA

The simulations have shown that the samples should ideally consist of about 120 firms or more. We believe that smaller samples might exhibit good performance but we would definitely not recommend using less than 60 firms in any case. To apply the bootstrap on DEA we propose the following steps:

i. Identify the underlying population

ii. Enrich he empirical distribution

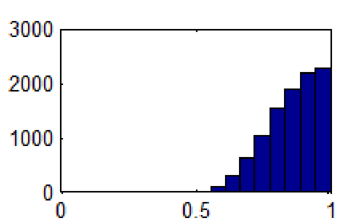
iii. Apply the bootstrap

5.2.1 Step 1: Identify the underlying population

It is important to identify the underlying population as it may affect how we proceed. To perform this task it is suggested inspecting the histogram of the empirical DEA scores and the associated skewness and kurtosis. We discuss 4 cases.

Case 1: Technological Homogeneity

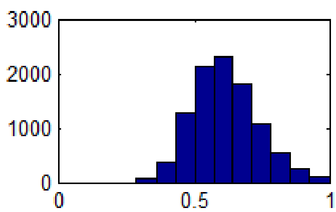
This case corresponds to setups where the firms exhibit technological similarities among them and it could be associated with (almost) perfectly competitive markets. The



underlying population has a half-normal distribution and it can be identified in the sample by a negative skewness of about -0.65 and kurtosis of about 2.8. Under this case the efficiency scores are less sensitive to sampling variations and they tend to be close to their population value, especially as sample size increases. For large enough samples sizes (certainly larger than 120 firms and considerably more if many inputs and outputs are used) the application of simple DEA would be adequate as bootstrap DEA would not add much in practice.

Case 2: Technological Heterogeneity

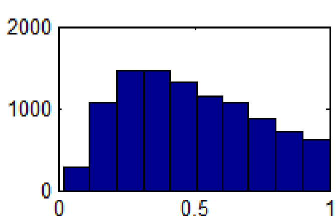
In this case some firms have access to superior technology that other firms do not. This is a form of barrier and could be associated with a monopolistic market (or some form



of oligopoly). The underlying population has a bell-shaped distribution with a thin tail towards 1 and it can be recognised from sample skewness of about 0.4 and kurtosis approaching 3. Bootstrap DEA cannot be applied in this case because apart from violating assumption (iii) above, it would be valid only asymptotically which is practically infeasible. In fact, due to the high and persistent DEA bias we express our concerns on even applying DEA. We recommend reconsidering the inputs and outputs used as well as the firms included in the dataset in case any outliers can be detected.

Case 3: Technological “Variability”

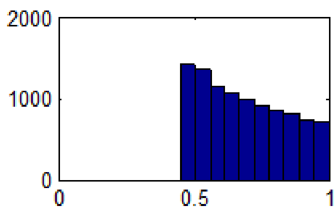
This case represents a “changing” market and it is a mixture of the previous two cases. Intuitively, the firms gradually gain access to superior technology and we therefore



consider this case as a form of monopolistic competition in the medium-run. This case can be identified by skewness close to 0.3 and kurtosis that slightly exceeds 2. The population efficiency scores tend to be sensitive to sampling variations and the use of bootstrap can be very useful even in larger samples.

Case 4: Technological Randomness

This case exhibits an almost random selection of efficiency scores which implies that it cannot be associated with a specific market structure. We would not expect this case to



appear frequently in practice, and if it did it would be a good idea to reconsider the data chosen and input-output specification. It can be identified by a flat, almost uniform distribution of efficiency scores which have skewness slightly below 0.3 and kurtosis below 2. The efficiency scores are sensitive to sampling variations and there is scope to apply bootstrap DEA.

5.2.2 Step 2: Enrich the empirical distribution

It has been established in the literature that the discrete nature of the DEA scores may lead to inconsistencies if the “naïve” bootstrap is applies. In particular, the resulting bootstrap distributions will consist of repeated values and will possibly have peculiar properties. It is therefore necessary to enrich the empirical distribution to deal with this issue. The most popular way is to employ kernel density estimation techniques which, however, introduce additional noise and require very big samples to perform well.

Recent developments which are based on these techniques are sited to perform better but they still require samples much bigger than 100 and ideally close to 1000 firms.

An alternative approach would be to employ the “moments-bootstrap” which uses the sample moments to enrich the support of the empirical distribution by producing pseudo-populations with similar properties. Simulations have shown that this approach performs very well for samples with about 120 firms (or more). The assumption of bias equality, which is the fundamental assumption for Simar and Wilson’s (1998) bootstrap DEA and for its popular extensions, is well-satisfied under the moments bootstrap.

5.2.3 Step 3: Apply the bootstrap

Having established that it is suitable to apply the bootstrap to the sample in hand we are ready to generate bootstrap DEA scores. The procedure followed is the same as in Simar and Wilson (1998) but we recommend using the moments-bootstrap instead of the smooth bootstrap. The resulting distribution of bootstrapped efficiency scores for each firm can be used to construct confidence intervals and test hypotheses as well as to provide more accurate estimates of the population efficiency scores.

5.3 Testing hypotheses

The interested reader should consult chapter 3 which is devoted on testing hypotheses with bootstrap DEA for more details. Here we only describe briefly the steps that could be followed.

5.3.1 Step 1: Define the null

It is important to clearly state what is being tested as this will determine the way to proceed. The tests can be either one-sided or two sided and can take the form of same-sample or cross-sample comparisons. In the first case one could test, among others, if a firm achieves a certain efficiency score or if two firms have similar efficiency. The second test can be particularly useful in cases of pooled panel data where the interest is on testing for efficiency change for a firm over time and where the implementation of the bootstrap Malmquist might not be feasible due to sample size issues.

Cross sample comparisons are also possible where one could test, for example, the equality of the means between two samples (see also Simar and Wilson, (2008)). We recommend care to be taken in this case as the two samples might be associated with different underlying populations, which could affect the validity of the results. Comparing the skewness and kurtosis of the two samples could be useful.

Extensions of bootstrap DEA can be also used to test hypotheses. For example one could test for productivity change using the bootstrap Malmquist index or test for returns to scale using the approach in Simar and Wilson (2002). The two-stage approaches in Simar and Wilson (2007) can be used to test the significance of the impact of environmental factors on efficiency.

5.3.2 Step 2: Define the test statistic

The test statistic determines how the hypothesis test is carried out. In the simple case of testing if a firm has a specific efficiency score or if it has the same efficiency compared to another firm, the test statistic is actually a constant. The latter case can be transformed into a test involving the ratio of efficiency scores in which case the test statistic is this ratio and which will be computed in all bootstrap replications. Another example of a test that requires the careful definition of an appropriate test statistic is that of Simar and Wilson (2002) on testing for returns to scale. In that case, the computed statistic is the average scale efficiency of the sample and it computed in every bootstrap loop. If one wants to construct their own test it is recommended to consider carefully how they define the test statistic.

5.3.3 Step 3: Confidence intervals and p-values

The two most popular methods of constructing confidence intervals is the percentile method used in Simar and Wilson (1998) and the basic bootstrap confidence intervals used in Simar and Wilson (2000a). The theoretical explorations and simulations here have shown that the percentile method provides more accurate intervals and requires fewer observations. Moreover, the SW2000 intervals have been argued to perform well in cases which are not associated with good DEA practice such as in the case of technological heterogeneity. We therefore recommend using the percentile method.

In the presence of high skewness it might be worthwhile considering extensions of the percentile method such as the bias-corrected intervals of Efron (1982), proposed by Simar and Wilson (1998). Another popular extension which is argued to cope better with skewness is the bias-corrected and accelerated intervals of Efron (1987); however, it is still under development and experimentation by the author. The downside of these methods is that they are associated with wider intervals.

Finally, the bootstrap distribution of efficiency scores can be used to compute p-values for any test. One simply needs to compute the number of times that the bootstrap test satisfies the null hypothesis and divide it with the number of bootstrap loops.

5.3.4 Step 4: Accept or reject the null

The null hypothesis can be rejected if either the hypothesised value in the null (the critical value) lies outside the confidence intervals or if the computed p-values are less than the level of significance. In the special case of comparing two firms with each other, it might be worthwhile performing the test twice (using the two different bootstrap distributions for each firm) to check if they reach a common decision. If not, we recommend following the instructions in section 3.3.

6 An illustrative example: the Greek banking case

The previous chapters have investigated the plausibility of certain assumptions of bootstrap DEA in small samples and have shown through simulations that alternative methods to smoothing may perform better towards this direction. The proposed "moments bootstrap" seems to be a promising avenue for bootstrap DEA as under this approach the assumption of equal bootstrap and DEA biases is plausible in small samples while the associated coverage probabilities seem to converge reasonably fast (we proposed a minimum of 120 observations). In this chapter we provide an empirical illustration of the methods examined using as an example the Greek banking sector reforms of the late 80s. This is a subject of topical interest due to the ongoing Greek debt crisis and the expected closer supervision of Greek banks under the umbrella of the recently established Single Supervision Mechanism (SSM).

6.1 Introduction

Since the early stages of the EMU, European banking integration has received criticism. For example, Dermine (2002, 2006) points to the inadequacy of home country supervision and that a pan-European framework would need to finance the costs of a potential bailout, concluding that a common regulatory framework should be created. It is arguable that such arguments have proven to be correct, especially after the subprime crisis in 2007 and the ensuing banking crisis culminating in the EU sovereign debt crisis, which has affected severely the Greek economy. Most Greek banks became technically insolvent by 2012 and the source of liquidity of many Greek banks has been the ELA funds from the Bank of Greece. The 53.5% “haircut” of Greek debt in 2012, which was mostly held by Greek banks, has further worsened the parlous state of the balance sheets, while the writing off of bonds, combined with the significant increase in non-performing loans has eaten the sector’s equity. Greek banks had to undergo a substantial recapitalization process to meet the requirements of the supervisory framework, which has recently become stricter.

To avoid the contagion of the banking crisis to other countries in distress, the creation of a European Support Mechanism (ESM) was proposed from which EU banks could borrow. However, this required the establishment of a Single Supervisory Mechanism (SSM) which would ideally supervise all EU banking institutions and grand access to ESM funds, and which, in fact, resumed duties on the 4th of November 2014. Although prudential regulation is deemed to favour depositors and the economy in the long run, it is not clear whether this would be the case for Greece whose financial sector is already in a transitional process. It is therefore important to investigate how the potential imposition of further controls may affect the performance of Greek banks, using as a reference the Greek banking (de)regulation process of the late 80s and early 90s.

This is achieved by monitoring the effects of each step of the (de)regulation process on bank efficiency and productivity and by analysing their behaviour after the imposition of prudential controls. The step-by-step analysis of the deregulation process as well as the long-run post-event analysis comprises an empirical contribution in the literature of banking regulation. The explorations are utilized by the implementation of the moments bootstrap DEA (introduced in Chapter 4) on a pooled sample of observations, which allows the computation of bootstrapped Global Malmquist indices and the application of the hypothesis testing procedures discussed in Chapter 3. Throughout the analysis we show how the suggested guidelines can be followed in this case and apart from the policy implications extracted, we results across the various approaches, both qualitatively and quantitatively. Our findings confirm theory in that after the provision of commercial freedoms the productivity of Greek banks increases, whereas after the imposition of further controls productivity tends to decrease. We arrive at the same qualitative finding with all approaches reviewed, although we observe that under the moments bootstrap the rejection rate of our null hypotheses is smaller and the p-values slightly different.

The rest of the chapter is structured as follows: section 6.2 provides a contextual background of the Greek banking sector; section 6.3 reviews the relevant literature; section 6.4 describes the data and method used; section 6.5 presents and discusses the empirical results of the study, while section 6.6 concludes the study and provides directions for future research.

6.2 Contextual background

The Greek banking sector until the end of the 80's was heavily regulated and was characterized by high concentration rates relative to the other European countries. It operated under conditions of monopolistic competition (Hondroyiannis et al., 1999) with existing, though declining, economies of scale (Apergis and Rezitis, 2004; Karafolas and Mantakas, 1994).

The Singe Market Act, of 1986, provided the imperative for the Greek banking sector to modernize and become more competitive by 1993. The necessary reforms were implemented over a 5 year period according to a plan outlined in the "Committee for the Restructuring and Modernization of the Banking System" introduced in 1987. Among others, the deregulation process involved[102] (i) the liberalization of interest rates, (ii) the removal of minimum reserve requirements, (iii) the abolition of compulsory purchases of governmental promissory notes and bonds, (iv) the abolition of compulsory financing of public companies and SMEs by commercial banks, and (v) the removal of restrictions on capital mobility among EU state members.

The last few commercial freedoms (de-specialization of special credit institutions) along with the complete liberalization of capital mobility and branching within EU were established by the Second Banking Directive of 1988 and were effective as of 1993. However, they were followed by the imposition of prudential controls in 1993

[102] A detailed analysis of the Greek deregulation process is provided by Gortsos (2002) and Voridis et al. (2003).

(definition of capital for regulatory purposes, minimum 8% of capital adequacy ratio, introduction of accounting standards), in order to harmonize the Greek banking sector with those of other European countries.

Macroeconomic policy was geared towards the requirements of the Maastricht Treaty while competition in the banking sector was intensified as the liberalization attracted more banks into the industry. The macroeconomic outlook of Greece improved after 1995, followed by a bull run on the Athens Stock Exchange market. Moreover, end of 90s sees vivid M&A activity, especially during 1998 and 1999, while the universal banking model is gradually adopted.

The accession of Greece in the Eurozone was a changing point for Greek banks which expanded into new markets (mainly the Balkans, Turkey and Eastern European countries) and offered a wider range of financial products and services. The access to substantially cheaper funds in the European interbank market reduced the cost of borrowing and boosted the credit expansion in Greece.

However, since the outbreak of the Greek debt crisis in 2009, Greek banks have become technically insolvent, especially after the 53.5% debt haircut of March 2012. In fact, the total equity of all commercial banks (according to their annual financial statements) fell to a negative 461.1 million Euros during that year, forcing some banks to shut down and others to merge. Greek banks had to recapitalize in order to meet the appropriate regulatory standards and to gain access to the ESM funds, implying also that they would need to enter the Single Supervisory Mechanism[103] which furthers the pressure due to the stricter supervision.

The basic features of the Greek banking sector during the period of study (1987-1999) and extending until the end of 2011 (making 1999 the midpoint) are depicted in Figure 6.1, below, while fundamental ratios and economic indicators are summarized in Table 6.1. Inspecting Figure 6.1, we observe that deregulation increased banking competition which is evident in the steady reduction of concentration[104] (auxiliary axis) from 1987 to 1999. Indeed, deregulation lifted the entry barriers and relaxed the conditions for the provision of financial intermediation services, therefore increasing the number of domestic commercial banks as well as the branches of foreign banks. Concentration increased again in 2000 due to the M&A wave in Greece while it returned to the 1998 levels after the accession to the EMU, with the latter motivating new entries. After 2010, concentration increased due to the Greek debt crisis as banks merged in order to meet the regulatory requirements and to survive through the crisis.

Size is a key success factor for Greek banks as implied by the high concentration. Big banks can manage to operate under tight margins by exploiting their economies of

[103] The SSM was first announced in 2012 (Ecofin meeting, 15th September 2012, Cyprus) with an initial plan to be implemented by the beginning of 2013. However, after a long debate among EU members on its rules and implementation, it was finally agreed in the Ecofin council of 13th December 2012 (Brussels) that the legal framework of SSM should be ready within 2013 and to be implemented by March 2014. The ECB assumed the supervisory tasks in the framework of the SSM on the 4th of November 2014 with 120 "significant credit institutions" included in the regulators' list. Under this arrangement 4 Greek banks will be directly supervised from the ECB along with their subsidiaries: Alpha Bank (including the recently acquired Emporiki Bank), Eurobank, National Bank of Greece and Piraeus Bank (including the recently acquired General Bank). More information can be found here: https://www.ecb.europa.eu/ssm/html/index.en.html

[104] Concentration is measured here by the contribution of the assets of the 5 largest banks. Chortareas et al. (2008) find that concentration in the Greek banking industry is well above the European average.

scale; a strategy that cannot be easily followed by small banks. Indeed, Greek banks seem to follow the structure-conduct-performance (SCP) paradigm (Rezitis, 2010), whereby banks use their size to gain market power and increase their profitability and efficiency. At the same time, the inflexibility of the labour market (Ayadi, 2008) is an impending factor in Greece in terms of adjusting variable costs, implying that overgrown banks (that is, banks which exhibit diseconomies of scale) are expected to be more cost-inefficient.

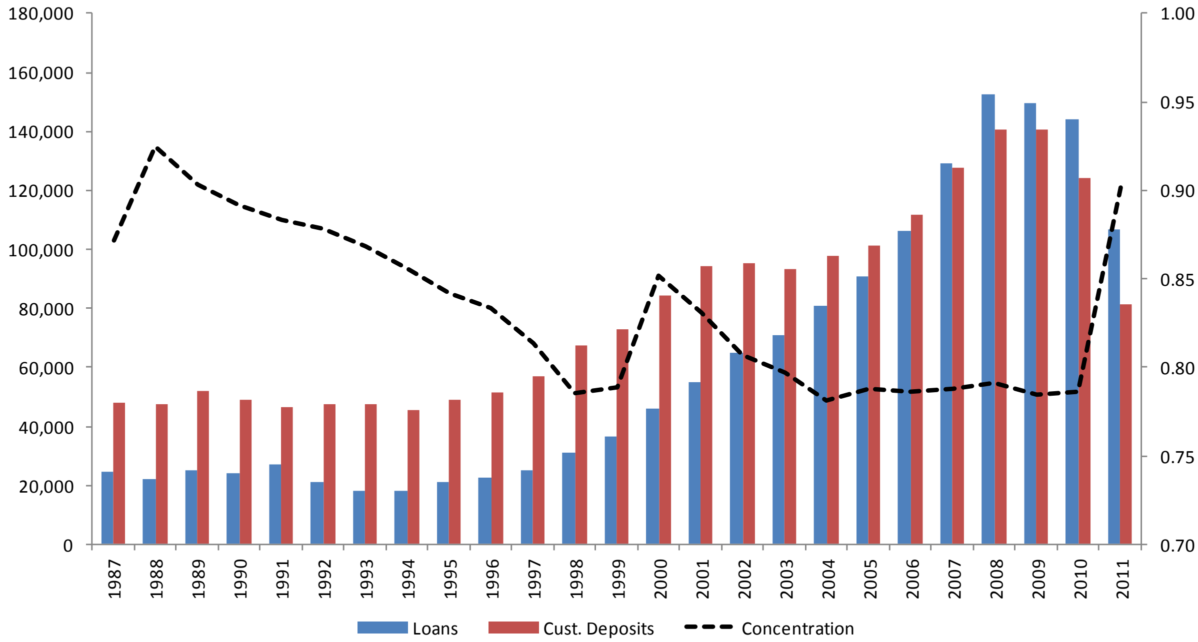
Regarding customer loans (less provisions) and deposits, it is interesting to note that most of the credit expansion in Greece took place after the accession in the EMU, as interest rates on loans, especially mortgages, where historically low. In addition, Greek banks increased their interbank borrowing activity[105] in order to satisfy the increasing demand for loans, explaining the loan-to-deposits ratio which exceeds one in 2007. However, due to the recent Greek debt crisis the value of loans less provisions has substantially decreased, after a considerable proportion of loans being characterized as bad debt and due to the noticeable contraction of credit. Similarly, deposits have also experienced a sharp decline as depositors have become nervous about the safety of their deposits and have moved their deposits out of the country[106].

---

[105] Data on interbank borrowing can be obtained online from *Bank of Greece,* under *"Monetary and Banking Statistics"*

[106] This fact and the unstable political environment received the attention of the press as they contributed in a hasted deposit flight. See for example:
http://www.theguardian.com/world/2012/may/16/greeks-withdraw-3bn-10-days and
http://www.reuters.com/article/2012/06/29/us-ecb-greece-deposits-idUSBRE85S0I720120629

Figure 6.1. Greek banking sector fundamentals



** Values in constant 1995 prices*

Banks operated in an enhancing economic environment until the breakout of the Greek debt crisis, as documented in the last two columns of Table 6.1. The structure of the Greek banking sector seems to change after 2000 as all ratios in the first four columns exhibit a steady increase, especially during the first years after 2000. In particular, the size of the banking sector relative to the size of the Greek economy grows, while the proportions of assets per employee and of loans to deposits increase steadily. This indicates that Greek banks have changed their conduct of business after the accession to the EU suggesting a different "technology" of transformation of their inputs into outputs. This may be relevant to the observation of Molyneux (2009) that reaction of European banks to M&As before and after post-2000 is different and this may be associated with the different way in which banks seem to operate. On the other hand, prior to 2000, the aforementioned ratios only mildly fluctuate, despite the sector reforms; the only exception is the ratio of equity to liabilities which exhibits an increase in the period 1997-1999 due to the bullish exchange market in Greece. This provides further support to our decision to cut-off the sample prior the accession of Greece to the EU.

Regarding the profitability of Greek banks, indicated by the financial ratios of returns to assets (ROA) and net interest margin (NIM) we do not observe a particular pattern. The ROE becomes negative but increases again until 1999, while the highest value of the ratio is observed afterwards. This may suggest that the sector reforms had an initial negative impact on the profitability of Greek banks but it was later improved. Regarding NIM, we can observe that its lowest values are observed during periods of high competition or distress, which is not surprising (Matthews and Thompson, 2014).

Table 6.1. Greek banking sector fundamentals

	Assets/ GDP	*Assets/ Employee*	*Loans/ Deposits*	*Eq./ Liabilities*	*ROA (%)*	*NIM (%)*	*Inflation (%)*	*Real GDP Growth (%)*
1987	0.76	1.41	0.52	2.28	0.46	2.20	16.40	-2.30
1988	0.68	1.33	0.46	3.16	0.37	2.44	13.50	4.30
1989	0.70	1.38	0.48	3.15	-0.51	2.26	13.70	3.80
1990	0.68	1.36	0.49	4.02	-0.08	2.81	20.40	0.00
1991	0.64	1.30	0.58	4.85	0.36	3.45	19.50	3.10
1992	0.67	1.38	0.44	5.22	0.71	3.45	15.90	0.70
1993	0.70	1.39	0.39	4.92	0.70	1.99	14.40	-1.60
1994	0.67	1.30	0.40	5.26	0.99	1.91	10.90	2.00
1995	0.68	1.28	0.43	5.00	0.94	2.34	8.90	2.10
1996	0.70	1.32	0.44	4.82	0.62	2.23	8.20	2.40
1997	0.75	1.46	0.44	5.36	0.76	2.50	5.50	3.60
1998	0.82	1.57	0.46	6.14	0.98	2.48	4.80	3.40
1999	0.91	1.93	0.50	10.81	2.68	2.57	2.60	3.40
2000	1.01	2.00	0.55	9.09	1.73	2.65	3.10	4.50
2001	1.05	2.18	0.58	8.33	1.32	2.78	3.40	4.20
2002	1.05	2.20	0.68	7.41	0.68	2.69	3.60	3.40
2003	1.01	2.22	0.76	7.22	0.87	2.93	3.60	6.00
2004	1.02	2.37	0.83	6.82	0.69	2.95	2.90	4.40
2005	1.17	2.80	0.90	6.33	1.02	3.08	3.60	2.30
2006	1.27	3.15	0.95	7.20	1.03	2.98	3.20	4.50
2007	1.50	3.64	1.01	7.09	1.08	2.73	2.90	4.30
2008	1.76	4.21	1.09	4.89	0.32	2.38	4.20	1.30
2009	1.87	4.38	1.06	6.82	0.02	2.04	1.20	-2.30
2010	1.86	4.36	1.16	6.55	-0.50	2.06	4.70	-4.20
2011	1.46	3.63	1.31	-0.16	-11.18	2.04	3.30	-6.90

6.3 Literature Review

Bank regulation can be either systemic (financial system stability), prudential (consumer protection) or on the conduct of business (Casu et al., 2006). Proponents of systemic regulation support that bank runs can be prevented with the introduction of deposit insurance schemes, the provision of liquidity assistance to financial institutions in distress by central banks (such as emergency liquidity assistance funds or the lender-of-last-resort function) or restrictions imposed on withdrawals (Baltensperger and Dermine, 1987; Diamond and Dybvig, 1983). Prudential controls concern the monitoring of the soundness of financial institutions, the imposition of minimum capital adequacy and reserve requirements as well as the disclosure of information. This monitoring is undertaken by regulatory agencies, hence benefiting consumers who do not have the resources or incentives to perform this task[107]. Finally, regulations on the conduct of business mainly involve authorizing (or not) banks to undertake certain activities (securities trading, investment banking, insurance) as well as maintaining an ethos in banking activities and services provided.

On the other hand, regulation (mainly in its prudential form) induces moral hazard as banks have incentives to take up more risk (Diamond and Dybvig, 1986) while it is associated with high costs for both banks and the society (Goodhart, 1988). In fact Goodhart (1988) reports that regulation costs include, among others, capital and labour costs, social costs arising from the Pareto-inefficient allocation of resources, costs from potentially lower competition (especially for peripheral, non-intermediation services, also offered by bank conglomerates) as well as potential costs from hindering financial innovation[108].

---

[107] This concept is known as the "representation hypothesis" and it was introduced by Dewatripont and Tirole (1994).
[108] For an overview of theories of banking deregulation the interested reader may consult any standard textbook on banking, while a more detailed account of those theories is provided in Hall (1989) and Dewatripont and Tirole (1994).

Deregulation allows the redistribution of inputs allocated on (or restrained by) supervision and compliance to more productive purposes, by lifting certain restrictions and providing commercial freedoms to banks. In theory, it aims at a more efficient allocation of resources and is therefore expected to increase efficiency while the benefits to society include reduced intermediation costs, higher quality and wider range of products and services provided. Deregulation is also used to increase banking sector competitiveness (as happened in Europe during the early 90s in the view of the Single Market) which has a more aggressive character and it is therefore uncertain whether it will lead to efficiency improvements or not (Berger and Humphrey, 1997). On the other hand, deregulation is usually followed by reregulation (Matthews and Thompson, 2014) in order to limit the commercial power given to banks and avoid moral hazard (Dewatripont and Tirole, 1994), this explains the term "(de)regulation" used here. It is therefore possible that the benefits of deregulation will be eliminated by the imposition of prudential controls[109].

Theory suggests that more regulation tends to hinder total factor productivity (TFP) growth. Crafts (2006) reviews the relevant theories and concludes that if regulation reduces the net returns to investment and innovation (through tough regulation controls or high costs of supervision and compliance), then it is expected to have a negative impact on TFP growth. In banking, the effects of (de)regulation on efficiency and productivity depend on the purpose of the reforms (more efficient resource

[109] A nice review on bank regulation and consumer protection (on the lines of both prudential and conduct of business controls) is provided by Benston (2000).

allocation or higher competitiveness), while other factors should be taken into account, such as the economic conditions, monetary policy as well as the timing and process of implementing the reforms.

It is almost certain that (de)regulation affects efficiency and productivity since it involves a reconsideration of the input/output mix used in the banking production process; however, its exact effects may differ across countries and context of reforms. Indeed, Berger and Humphrey (1997) review 130 studies over 22 countries and find that there is no consensus on the effects of (de)regulation on bank efficiency and productivity. They attribute the observed differences to the variety of models, methodologies and approaches followed as well as to the specific characteristics of the various cases examined.

Recent international studies examine the effect of the “state of regulation” (power of regulator, type of regulation, bank activity), bank-specific characteristics and macroeconomic environment on bank efficiency and productivity. Pasiouras (2008) uses a variant of the intermediation approach on a sample of 715 banks from 95 countries to examine the effects of the aforementioned factors on banks’ technical efficiency. He finds that, after using various model specifications, the third pillar of Basel II (“market discipline”, which relates mainly to financial information disclosure) always appears significant, while the significance of the other two pillars (“capitalization” and “internal capital adequacy assessment process”) is sensitive towards model specification.

Pasiouras et al. (2009) extend the study of Pasiouras (2008) and perform a similar analysis for cost and profit efficiency using stochastic frontier analysis. The variables they used, which relate to the regulatory environment, are sourced from the same database as in Pasiouras (2008)[110], but their sample is different and includes 615 banks from 74 countries (selected on the basis of data availability). They find that cost and profit efficiency are positively affected by the second and third pillars of Basel II while capital requirements (first pillar) tend to increase cost efficiency and decrease profit efficiency. On the other hand, restricting bank activities tends to decrease cost efficiency but increase profit efficiency.

Delis et al. (2011) explore the linkage between regulation and productivity from a dataset of 22 transition countries[111]. They find that only market discipline (related to the third pillar of Basel II) and restrictions of bank activities (other non-traditional operations) have a positive impact while the other two pillars gain significance after crises. They attribute the non-significant dependence of the other two Basel pillars to the characteristics of banking systems in transition countries, such as overcapitalization and law enforcement.

The previous studies, although of great importance, do not provide country-specific results due to data limitations on the sophisticated list of regulatory variables constructed by Barth et al. (2001). Studies which focus on certain countries can provide

[110] Barth et al. (2001) have created a very interesting database on regulatory conditions for each country which is available on-line from the World Bank. It is constructed using responses from banking institutions around the world and by aggregating answers per country into a single measure. Since then, the database has been updated in non-regular time intervals and its completeness is subject to the banks' responsiveness.

[111] It is important to note that in their model specification they use one period's lag for regulatory variables on the basis that it needs time for regulatory changes to affect productivity. We also adopt this view in this study as it seems to be a reasonable assumption for Greek banks according to Siriopoulos and Tziogkidis (2010).

a deeper insight about the effects of (de)regulation on bank efficiency at a national level.

Extensive US studies find negative effects on productivity during and after the deregulation of the 80’s (Humphrey and Pulley, 1997; Humphrey, 1990; Wheelock and Wilson, 1999). However, after a 4 year period of continuous adjustment (input reduction and adjustment of output prices), US banks seem to recover and improve their profitability, driven by the enhancing business environment (Humphrey and Pulley, 1997).

Bank deregulation studies in Asia report mixed results. Kumbhakar and Sarkar (2003) examine the effects of deregulation on Indian banks during the pre- and post-deregulation period (1985-1996). They find that productivity increased, however regulatory distortions persisted in the post-deregulation period, especially for public banks, in the form of distortions in input prices (mainly due to over-employment). Positive effects are also documented by Isik and Hassan (2003a) who examine the deregulation process in Turkey during the 80s. On the other hand Chen et al. (2005), who examine the technical and cost efficiency of Chinese banks in the pre and post-deregulation period of 1995, document a decline on the average levels of technical and allocative efficiency, especially after the outset of the Asian financial crisis.

European studies seem to document an increase in productivity after the (de)regulation period of late 80s to early 90s which was implemented by most European Community members in the view of the Single European Market. Altunbas et al. (2001) and Altunbas et al. (1999) use a large sample of banks from 15 EU countries and find that during 1989 to 1997 banks exhibited technical progress which led to cost savings, benefiting mostly large banks. Casu et al. (2004) examine the productivity change in France, Germany, Italy, Spain and UK over the period 1994 to 2000. They use both parametric and non-parametric techniques and find that EU banks in the post deregulation period have increased their productivity on average (with the exception of the first and the last year). More recently, Chortareas et al. (2013), after examining the influence of financial freedoms that the commercial banks of 27 EU countries have enjoyed during 2001-2009, they document a positive effect on productivity. Moreover, Chortareas et al. (2012) in a similar study using data from 22 EU countries for the period 2000-2008, confirm that governmental interventions on private banks' policies and the monitoring of their practices has had a negative effect on efficiency. On the contrary, regulations concerning capital quality tend to have a positive effect on efficiency; however, these effects are mainly evidenced for large banks operating in countries with developed and low-concentrated financial systems.

Country-specific studies for the EU can be found in the literature, although the recent focus is on cross country exercises. For example, Kumbhakar et al. (2001) examine the effects of deregulation on Spanish savings banks and document an increase in productivity but a decline in technical efficiency, whereas profit efficiency first declines and then increases. Also, Berg et al. (1992) find that Norwegian banks experienced technical regress prior to deregulation but technical progress afterwards.

The literature on Greek banking also reports mixed results, depending on the period examined and the approach followed, as documented in the review of Chortareas et al.

(2008). The majority of studies focus on the post-deregulation period, and specifically during 1993-1998. In particular, these studies find that productivity increases with the exception of the first year (Tsionas et al., 2003), that private banks are more technically efficient than public banks (Noulas, 2001) and that large banks are substantially more cost inefficient than small ones (Christopoulos et al., 2002). However, there seems to be room for substantial improvement in cost efficiency for all banks (Christopoulos and Tsionas, 2001; Kamberoglou et al., 2004).

To the extent of our knowledge, the only Greek banking studies which cover the full period of deregulation (that is, from 1987 onwards) are by Apergis and Rezitis (2004) and Rezitis (2006), who use a dataset of 6 banks over the period 1982 to 1997. Although the sample used is the same, the two studies report different effects on productivity, potentially attributed to the different methods and variables (or approach) used or even due to the small number of observations.

The literature on the effects of bank (de)regulation, although vast, seems to be focusing only on the overall or average effects of (de)regulation; the effects of each step of the deregulation process are neglected, which is a gap in the literature that we wish to address. In Greece, there is evidence that significant, destabilizing events have a negative impact on banks' technical efficiency the year after the event, followed by a period of "recovery" which may last from 2 to 4 years (Siriopoulos and Tziogkidis, 2010)[112]. Lagged effects are also assumed by Delis et al. (2011) in his European study on regulation, or by Orea (2002) who examined the M&As of Spanish savings banks. Apart from the fact that it takes time to implement regulations from their date of announcement, in the presence of strong trade unions or labour laws (as in Greece) the potentials for cost reductions or better allocation of resources are not necessarily exploited in the short run and it might lead to decrease in efficiency (Ayadi, 2008).

Deregulation seems to be associated with efficiency and productivity improvements whereas the imposition of prudential controls seems to have opposite results in the short run. The two gaps that we identified in the literature is that no study follows a step-by-step approach to analyse the effects of (de)regulation, while we found no Greek banking studies which cover the full period of deregulation and reregulation. We therefore aim to contribute towards this direction with our empirical exercise.

6.4 Data and Method

6.4.1 Choice of study period

For the purposes of this illustrative example we use Greek commercial banks which operated during the period 1987 to 1999. Due to the fact that the number of Greek

[112] This concept could be related to neoclassical theories of the firm where capital needs one period in order to become productive (termed as "time to build") or in the theories of management change where the effects of a "bad" event appear with a lag (Elrod and Tippett, 2002).

banks in most years is too small (10 to 13 while the maximum is 18) even to apply simple DEA, we pool observations. Hence, the operations of a bank in a certain year are considered as a separate DMU. This is explained in more detail in subsection 6.4.4.3.

The study period covers the (de)regulation era of 1987 to 1994, while it includes another 5 years to explore the existence of longer term benefits from the sector reforms. In mid-1999 the Athens Stock Exchange experienced a crisis (due to a "bubble" burst) while at the same time Greece was working towards entering the European Union, with the Euro being adopted from the beginning of 2001. Hence, we consider that the effects of deregulation could not be identified beyond 1999.

During the deregulation period (1987 to 1994), apart from the reforms, no other event has been observed with the exception of a scandal in 1987 (see subsection 6.4.3) and two privatisations[113]. From 1994 to 1999 we observe 5 M&A events[114] (out of which 4 occurring during 1998-1999) and one partial privatisation[115] (in 1998). Considering that our database is "quite clean" of other major events (at least until 1997) we could argue that deregulation and the fiscal or monetary policies of that time were perhaps the most influential factors to affected bank efficiency and productivity of Greek banks.

[113] Piraeus bank is privatised in 1991 and Bank of Athens in 1993.

[114] In 1995 Emporiki Bank acquires 51% Metrolife (40% through Emporiki Bank and about 11% through one of its subsidiaries); in 1998 EFG Eurobank acquires Cretabank (99.8% of shares) and Bank of Athens; in 1998 Egnatia Bank acquire 51% of shares of Bank of Central Greece from its parental company, "Agricultural Bank" but with an agreed price that was at a 56% discount compared to its market value; in 1998 National Bank merges through absorption with the National Mortgage Bank; in 1998 Piraeus Bank acquires a 37% controlling stake of Macedonia Thrace's shares from the National Bank of Greece and it also acquires the branches of Chase Manhatan and Credit Lyonnais Hellas.

[115] In 1998 General Bank is securitized and partially privatized; though its full privatization and acquisition by Societe General occurred in 2004.

Therefore, we assume that, during the study period, changes to banks' inputs and outputs are a response to the changing regulatory environment, which can be translated into changes of their productivity. In fact, the author has examined the annual reports of each bank for that period (including those of special credit institutions)[116] and the focus is on the sector reforms and the macroeconomic environment in the view of the Single Market. Therefore attributing any substantial efficiency changes to the sector reforms seems reasonable.

6.4.2 Data and variables[117]

To construct the dataset we used a combination of the Bankscope database along with archived and published financial statements of banks (in order to verify Bankscope and include missing entries)[118]. The archived financial statements were obtained from the library of the Bank of Greece (banks' annual reports, Banker's Almanac, Athens Stock Exchange annual catalogue of listed firms), the libraries of banks which maintain historical archives (Agricultural Bank of Greece, Alpha Bank, National Bank of Greece), from the finance divisions of the respective financial institutions, or from the Hellenic

[116] The annual reports are available at the library of the Bank of Greece or in the historical archives of the National Bank of Greece and Alpha Bank.

[117] The author would like to thank the employees at the library of the Bank of Greece for their support on locating entries in the library, on finding missing entries from alternative resources as well as on helping the author with various auxiliary, time-consuming tasks.

[118] To verify the Bankscope database accuracy we compared it with the published accounts on the basis of total assets and earnings before tax on an annual basis; if a difference was detected we reviewed all Bankscope figures accordingly. The procedure of data collection and building up the database was very time consuming (it lasted more than 9 months) as it required several visits at various locations in Athens (also detained by the restricted opening hours of the libraries to the public) as well as typing accounting entries into the computer (they were only available in hardcopy form).

Printing Office[119]. After inputting the data into a “processable” file, we converted data from Drachmas to Euros, using the fixed rate of 1 DRC=340.75 EUR for ease of exposition, while all values were converted to 1995 constant prices using the GDP deflator.

For the purposes of the illustrative example we collected data for both commercial and other financial institutions in order to exhibit the effects of technological heterogeneity on DEA and, to its extent, on bootstrap DEA. The analysis of empirical results, though, is based on commercial banks only, with the exception of a few outliers, the exclusion of which we justify and discuss in subsection 6.4.4.2.

The final list of commercial banks used for the illustrative exercise is provided in Table 6.2. In each year, “YES” denotes that the bank was included in the sample, “N/A” indicates that there were no available data (also shaded in dark tan), while “NO” indicates that the bank was excluded from the sample (also shaded in light orange). We have also included in each year an artificial DMU which we have named “Average Bank”, in order to capture the average behaviour of the Greek banking sector. The inputs and outputs of the “Average Bank” are the average values of the inputs and outputs of all DMUs during a certain year[120]. Hence, the efficiency scores of these artificial DMUs are always less than 1 and their inclusion does not affect the shape or position of the frontier and therefore the efficiency scores of other banks. Also, we have included a

[119] This is the official printing office of the Greek state which, apart from publishing the Greek Government’s Gazette and Presidential Decrees, it maintains an archive of published documents. In most cases, the published financial accounts of Greek banks since 1994 are available on line (www.et.gr – in Greek).
[120] In our view, this provides a better estimate of the average efficiency score per year rather than just calculating the average efficiency score of DMUs after applying DEA, as usually done in the literature.

second artificial bank which uses the weighted averages (weighted each year by total assets) of the variables to examine the extent to which the market is driven by large banks[121]. Overall, the sample comprises 216 DMUs out of which 26 correspond to the aforementioned artificial observations.

Table 6.2. Banks included in the sample

	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	Total
Agricultural Bank of Greece	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES	*8*
Alpha Bank AE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	*13*
Bank of Athens	N/A	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES			*10*
Bank of Attica SA	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	*13*
Bank of Central Greece	YES	N/A	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES		*11*
Bank of Crete - Cretabank	YES	N/A	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES		*11*
Egnatia Bank SA							YES	YES	YES	YES	YES	YES	YES	*7*
Emporiki Bank of Greece SA	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	*13*
Ergobank SA	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	*13*
Eurobank Ergasias (EFG) SA											YES	YES	YES	*3*
General Bank of Greece SA	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	*13*
Interbank									YES	YES				*2*
Ionian and Popular Bank of Greece	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES		*12*
Laiki Bank (Hellas) SA							YES	YES	YES	YES	YES	YES	YES	*7*
Macedonia Thrace Bank SA	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	*13*
National Bank of Greece SA	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	*13*
Piraeus Bank SA	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	*13*
T Bank S.A							YES	YES	YES	YES	YES	YES	YES	*7*
Xiosbank					YES	YES	YES	YES	YES	YES	YES	YES		*8*
Total	*11*	*10*	*12*	*12*	*13*	*14*	*17*	*17*	*18*	*18*	*18*	*17*	*13*	**190**

To measure bank efficiency we use the well-established intermediation approach (Sealey and Lindley, 1977) which deems banks as financial intermediaries that transform their resources (usually related to capital, labour and certain liabilities) into banking

[121] The author would like to thank Prof John Nankervis† for his kind suggestion at a presentation of the author with title “Did (de)regulation deteriorate the performance of Greek banks?” at Essex Business School, in October 2012. Prof Nankervis† had suggested that the addition of the weighted-average DMU would not affect the computation of efficiency of the other DMUs in the dataset and would provide a measure that takes into account the high concentration of the Greek banking sector, hence acting as a “representative large bank”. Moreover, I would like to thank Prof Nankervis† for being encouraging on my work on the theoretical explorations on bootstrap DEA.

outputs (usually related to earning assets). In particular, we use fixed assets, personnel expenses and customer deposits as inputs and net loans (loans minus provisions for bad debts) and other securities[122] as outputs. We should note that we have excluded from our analysis the interbank activity (that is, deposits and loans to other financial institutions) as we want to focus on the customer orientation of banks. Furthermore, we have not included off-balance sheet items due to data unavailability and due to the fact that these items became more important in more recent years. Finally, due to lack of data we are only able to compute technical efficiency and not cost efficiency, which would concern the effects of deregulation on the cost structures of financial institutions (Berger and Humphrey, 1997). But since we are using monetary values in an input oriented model, we have incorporated the concept of cost minimization in our analysis to some extent.

Table 6.3 presents the annual averages of the input and output variables used in the final sample of commercial banks; effectively this is the data for the average bank. Although the values may seem to vary at a first glance, when considering the ratios of outputs over inputs these variations become quite less noticeable. This means that on average, banks have not changed substantially the way they transform the particular inputs of the intermediation approach into outputs.

[122] Bankscope defines other securities as the sum of investments of banks to associates through equity and other securities, which in turn includes bonds, equity derivatives and any other type of securities. Also, we have to note that some studies use other assets (=total assets – fixed assets - loans) instead of other securities. We diverge from this in order to assess financial institutions in terms of their earning assets (assets that are used to produce earnings) while also excluding loans and advances to banks as well as deposits to banks, hence focusing on the customer orientation of intermediaries.

Table 6.3. Averaged of input/output variables per year

	Fixed Assets	*Personnel Expenses*	*Customer Deposits*	*Loans*	*Other Securities*
Average_1987	59.22	65.20	3,808.89	1,588.80	1,393.86
Average_1988	99.84	74.90	4,069.99	1,458.81	1,534.44
Average_1989	79.73	67.49	3,685.07	1,419.01	1,331.52
Average_1990	69.93	66.01	3,470.56	1,327.02	1,329.68
Average_1991	61.71	59.51	3,084.19	1,550.69	786.47
Average_1992	71.33	73.63	3,385.42	1,486.19	1,295.66
Average_1993	43.79	56.52	2,765.71	1,053.51	1,295.45
Average_1994	43.64	59.54	2,653.63	1,056.82	1,142.40
Average_1995	41.55	61.26	2,695.14	1,143.47	1,104.05
Average_1996	43.71	66.29	2,837.50	1,240.62	1,099.11
Average_1997	43.56	68.85	3,136.09	1,367.10	1,338.48
Average_1998	48.12	76.63	3,917.96	1,790.71	1,463.93
Average_1999	67.14	104.00	5,526.23	2,746.27	2,237.03

** Values in million Euros and in 1995 constant prices*

Table 6.4 presents some descriptive statistics for the input and output variables (lower part) and their Spearman's rank correlations[123] along with an indication of significance[124] (upper part). The descriptive statistics suggest that banks range from very small to quite big. Also the correlations indicate a strong, positive and significant association between all variables, which is not of concern (in the sense of multicollinearity) in DEA modelling due to its non-parametric nature[125]. High correlation

[123] It is preferred to the Pearson correlation in cases where the variables might not be linearly related to each other. In our case it would be normal to expect some non-linear input-output relationships and therefore the Spearman correlation seems to be a safer choice. At the same time, the log-transformations reveal that these relationships are monotonic hence Spearman's rho is a valid measure of correlation in our case.

[124] The test statistic is $t = \rho\sqrt{\frac{n-2}{1-\rho^2}}$, where ρ is the Spearman correlation coefficient and n is the sample size. It follows approximately a t-distribution with $t-2$ degrees of freedom.

[125] However, in general the discrimination power can be affected in the presence of high correlation (Charnes et al., 1994). Low discrimination refers to the situation where DEA is favourable only towards a certain group of DMUs that exhibit similar characteristics. The more homogeneous the sample is, the less worrying this issue is. The scatterplots in Figure 6.2 show that our sample is quite homogeneous as there

implies that there is consistency with regards to the input and output variables used, in the sense that they are associated with a certain banking production process. In fact, it would be surprising if the correlation coefficients were low. Moreover, the high significant correlations indicate that the input/output proportions under the intermediation approach have remained almost fixed, explaining the observed technological homogeneity across time periods.

Table 6.4. Correlations and descriptive statistics of input/output variables

	Fixed Assets	*Deposits*	*Pers. Expenses*	*Loans*	*Securities*
Fixed Assets	1				
Deposits	0.943**	1			
Pers. Expenses	0.943**	0.962**	1		
Loans	0.945**	0.987**	0.963**	1	
Securities	0.915**	0.955**	0.939**	0.932**	1
Mean	56.54	3381.94	68.61	1450.12	1316.50
Median	22.86	872.68	25.35	457.36	253.19
St.Deviation	83.64	5775.90	99.13	2178.62	2453.70
Minimum	1.03	56.74	1.34	11.97	2.97
Maximum	548.11	26321.56	450.64	11645.83	11875.60

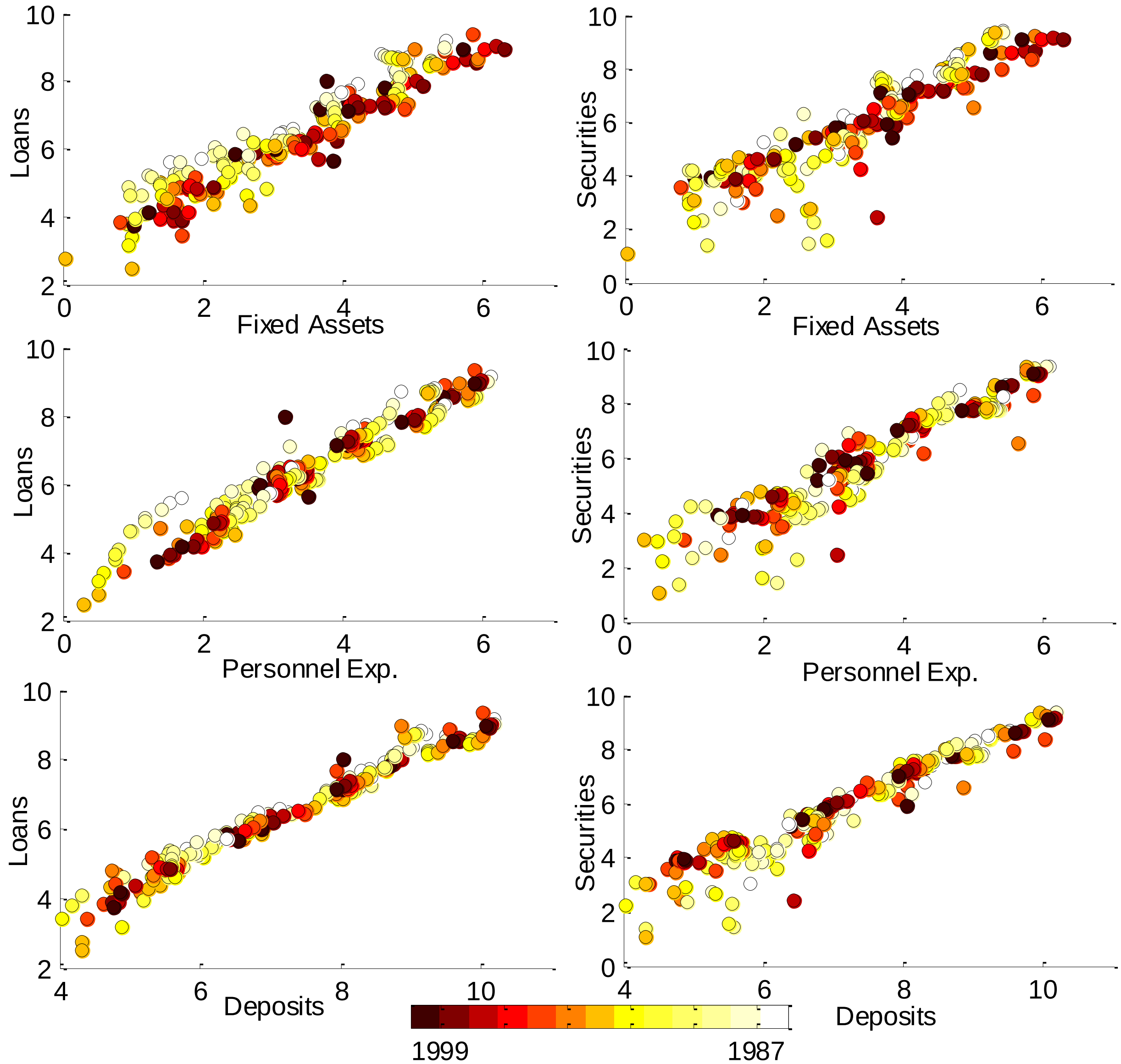
*** Signifcant at the 0.01 level*

To provide a graphical illustration of the input-output relationships in our sample, we have produced relevant scatterplots in Figure 6.2. In particular, the horizontal axes in each "line" correspond to the three inputs used and the vertical axes in each "column" correspond to the two outputs used. The values are expressed in natural logarithms and

is only one homogeneous cluster when considering input/output combinations and therefore the issue of reduced discriminatory power is not of concern in our case.

therefore any movement in the plot can be considered as a percentage change. Finally, the different colours of the filings represent a different year with the darkest ones corresponding to 1999.

Figure 6.2. Inputs/outputs (in logs) per year



The information included in this depiction is quite interesting. One observation is that there is no specific clustering of banks per year of operation. For example, if we believed that in later years banks had access to superior technology, we would expect to observe the majority of dark-coloured dots lying on the north-western part of the cluster while banks operating in early years should lie on the south-eastern region. The mix of colours can be therefore perceived as an indication that the technology of transforming the inputs of the intermediation approach into outputs did not change over the time period; at least not in proportional terms.

Another interesting observation is that the sample is quite homogeneous with almost all observations lying on a dense cluster that approximately forms a straight line. There are only a few scattered observations in the lower part of the scatterplots but we could not state that we observe a "break" in the cluster or another one forming. A simple regression analysis would reveal that the slopes of those "lines" are quite close to 1, which means that a proportional increase in inputs would lead, in principle, to a proportional increase in outputs (given that the axes are expressed in logs). This observation provides further support to our CRS assumption.

6.4.3 An account of the sector reforms examined

In this subsection we will present the sector reforms announced and implemented in each year and we will explain how we expect them to affect efficiency and productivity. A detailed account of the sector reforms and the actions of monetary policy during this period is provided by Voridis et al. (2003), while a more general overview can be found in the annual reports of the National Bank of Greece and in Gortsos (2002).

1987

The sector reforms are announced and include a long list of actions aiming at the modernisation and competitiveness of the Greek banking sector. Given that Greek banks were tightly bound by governmental controls and given the inflexibility of the Greek labour market, the positive effects of such an announcement are not expected to be immediately realised (Ayadi, 2008). This is also supported empirically by Berger and Humphrey (1997) in cases where deregulation has an aggressive character. The first financial freedoms appear in 1987; interest rate controls on loans and deposits are liberalized (to a large extent) and the reserve requirement of 19.5% for large industrial firms' loans (accompanied by a low interest rate floor of 12.5%) is abolished.

Apart from the managerial shock, the minimum reserve for loans and bonds of public sector companies increased to 10.5% (from 3.5%), the minimum deposits with the Bank of Greece increased to 7.5% (from 6%) with a lower interest rate of 14% (from 15.5% during a period with an inflation rate of 16.4%) while the minimum reserve requirements on holding Greek state promissory notes increased by 1% (to 38%); that is, more controls seem to be imposed in the first year of the (de)regulation process.

However, the positive news of deregulation for the banking industry are shadowed by one of the greatest scandals in the history of Greek banking: the "Koskotas scandal". Koskotas was a banker who owned the majority of shares (around 60%) of Cretabank and who was favoured by the ruling party at the time (PASOK) by directing public companies' deposits and assigning their financing to his bank. Koskotas, with the support of certain politicians, was involved in illegal activities using Cretabank's funds. Among others, Koskotas tried to acquire the Bank of Central Greece through Cretabank in 1987, but no clearance was given for the takeover (Dobratz and Whitfield, 1992; Featherstone, 1994, 1990). This resulted in a temporary shock in the Greek banking market and mistrust in state-owned banks which must have had an impact on the productivity of the sector negatively.

1988 and 1989

During 1988 we observe the first substantial set of commercial freedoms to banks. The most important of them include the lifting of restrictions on financing certain sectors of the economy, the abolishment of the 21% interest rate ceiling on loans, as well as the removal of selective credit controls[126]. Moreover, banks are allowed to determine freely loan rates and contract terms with certain industries.

In 1989 the liberalization process is continued. In particular, selective controls are completely removed, interest rates and other contract terms for most types of loans are freely determined, while interest rates on demand and sight deposits are liberalized. Furthermore, some measures aim at increasing competition: housing loan borrowers

[126] Effective as of 31 Dec 1988 (and completed in 1989), Greek commercial banks are no more obliged to allocate their portfolio of loans to certain sectors of the economy according to a predetermined percentage on outstanding loans, introduced in 1966 in order to weather the banks' reluctance to finance certain industries. Voridis et al. (2003) report the following percentages: 9.6% on domestic trade, 9.6% on import trade, 10.8% on export trade and 26.6% on manufacturing (that is, 56.6% in total).

are allowed to use financing from more than one financial institutions and special credit institutions are allowed to finance various sectors of the economy at freely determined rates and contract terms. Finally, the Second Banking Directive in 1989 (although effective from 1993) gave a fresh perspective to banks' expansionary strategies as it permitted the establishment of branches to other European countries without the further permission of the host country authorities. We expect that efficient banks should exploit this opportunity to expand their outputs or contract their previously "reserved" inputs, leading to an increase in productivity during both years.

1990

The climate is reversed in 1990 as inflation jumps to a period high of 20.4%, perhaps due to the oil crisis[127] as the growth rate of money supply was stable. The newly established government promotes a restructuring plan for the economy, including the liberalization of the private sector and the privatization of various public sector companies (including the Agricultural Bank of Greece and Piraeus Bank, though both completed in the following year). At the same time authorities focus their efforts on catching up with the forthcoming Maastricht Treaty's requirements by increasing taxation (in order to reduce the substantial deficit of 19.4%) and by adopting policies to decrease inflation. One of the fiscal measures which is relevant to Greek banks is the announcement of the imposition of 10% tax on interest income in 1991. We believe that the effect of high

[127] After the Iraqi invasion in Kuwait the price of the barrel increased from $17 to $36 in August.

inflation, the moderation efforts and the pending imposition of tax on interest income, have all affected bank efficiency negatively[128].

1991-1992

During the next two years, the government policies succeed in decreasing inflation and in increasing real GDP. The deregulation process enters one of its most important steps as the obligation of banks to invest 40% of their deposits in Greek promissory notes, Greek government bonds or bonds of public sector enterprises reduces to 30% in 1991 and 15% in 1992, allowing banks to use an important fraction of their funds more productively. Moreover, the minimum requirements on low-interest loans to SMEs is gradually lifted in 1992 and abolished by mid-1993. At the same time, the operations of commercial banks and other credit institutions are completely liberalized, allowing banks to expand their operations. We would expect to evidence an increase in banks' productivity during this period.

1993

In 1993 the Greek government decides to adopt Basel I standards and imposes a minimum liquidity ratio of 8% while capital is explicitly defined for regulatory purposes. At the same time Greek accounting standards (GAS) and international accounting standards (IAS) are introduced. The idea of imposing these prudential controls was to

[128] The only important deregulation step was the reduction of the minimum percentage of banks' deposits that should be directed to the financing of Greek enterprises from 10.5% to 6%, which cannot be considered as adequate to offset the negative climate.

harmonize Greek banks with the European ones in the view of the Single Market; hence, supervision became tighter. Although a few more commercial freedoms were given in 1993[129] and competition was further intensified[130], we believe that the impact from the introduction of new regulations would have been quite powerful in terms of restricting a recently liberalised sector, potentially causing productivity to decline.

The reason we expect this behaviour is that, apart from the evidence in the literature (Chortareas et al., 2012; Matthews and Thompson, 2014; Tsionas et al., 2003), compliance with Basel regulation requires allocating substantial resources for this purpose and the reconsideration of banks' portfolio of securities and other assets.

1994 and after

After 1994 the business environment of Greece is gradually improving and by the end of the study period Greece is very close to the Maastricht Treaty requirements. Until 1997, before Greek banks start engaging in M&As, we would expect banks to settle after the volatile period of reforms and reconsider their allocation of resources to increase their technical efficiency. From 1997 to 1999 we observe a further decrease in interest rates and greater improvement of the macroeconomic indicators, which we view as an

[129] In particular, capital movements of medium and long-term funds within EU where completely liberalized in 1993 while for short-term funds liberalisation came in June of 1994. Second, the obligation of commercial banks to hold a certain fraction of their deposits in Greek government bonds and promissory notes is completely abolished by May 1993 (from 40% in 1991 and 15% in 1992). Moreover, banks are no more required to channel funds to SMEs, however they are obliged to refinance the loans of these enterprises, corresponding to 6.5% of deposits in 1993 (Voridis et al., 2003). Finally, the interest rate floor on saving deposits (which comprise about 2/3 of total deposits in 1993) is completely liberalized, which had be proven to be binding (Voridis et al., 2003).

[130] The co-operative bank notion is legally introduced in the Greek banking sector while other financial intermediaries can offer a broad range of products and services that commercial banks traditionally offered.

opportunity for banks to grow. The relation to the sector reforms is that the imposition of prudential controls might have contributed towards building up confidence to depositors and investors, which has been reinforced by the improving business environment which peaks in 1999. The sign of the efficiency change should depend upon whether the increase in inputs is proportionately greater than the increase in outputs or not, while it may have been affected to some extent by the M&A wave of that period. However, we would expect to evidence an increase in efficiency from 1998 to 1999 due to the bullish stock exchange which should have increased the value of securities.

6.4.4 Method and Implementation

To compute technical efficiency we use the input oriented model in (2.11), with the orientation being justified by the fact that banks have more control over their inputs rather than outputs (Cook et al., 2014). Regarding the assumption of the CRS technology, apart from the evidence we provided in the previous subsection, there is a number of reasons for supporting this choice, which are explained in subsection 6.4.4.1. We then explain the procedure of choosing the banks to be included in the final sample in subsection 6.4.4.2, following the suggestions of the suggested guidelines in section 5.2. Finally, in subsection 6.4.4.3 we explain how we apply bootstrap DEA and how we extend the test of significant efficiency change from section 3.3.2 to the case of testing for productivity change using the Global Malmquist index of Pastor and Lovell (2005).

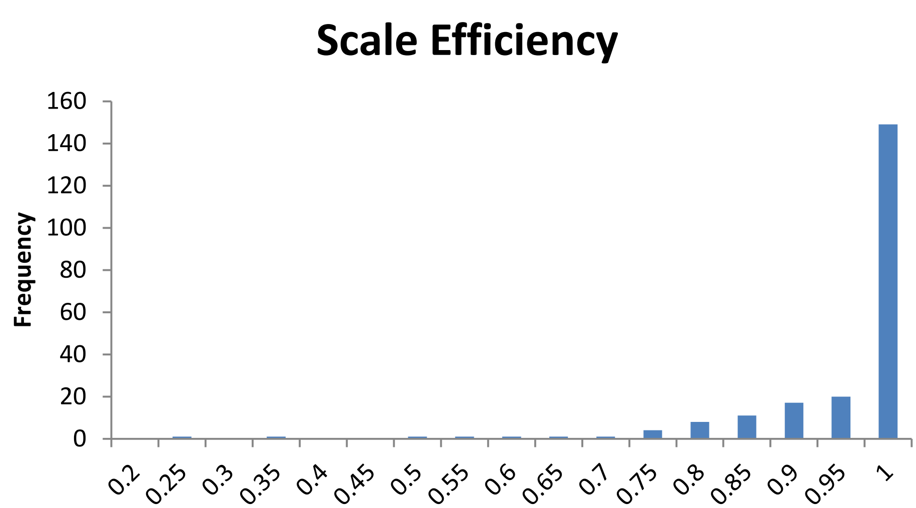
6.4.4.1 Returns to scale

For the purposes of the illustrative example we will adopt the assumption of constant returns to scale (CRS), given that the simulations have also been performed under the same assumption and hence we would like our illustrations to be consistent with the theoretical part of the thesis. Apart from serving the purposes of an empirical illustration and apart from the previous analysis using Figure 6.2, the CRS assumption can be considered appropriate in our case for a number of reasons.

One such reason is that, under CRS, the efficient banks are associated with minimum long-run average costs and have exploited any economies of scale, which can be considered as one of the desirable effects of deregulation. Given also that we are using an input-oriented model and given that inputs are expressed in monetary terms, it could be thought that we are assessing the extent to which banks operate under the minimum costs with reference to the whole study period. Hence, it could be considered that CRS is consistent with the intentions of the policymakers who, through deregulation, may want to encourage banks to appropriately adjust their scale of operations and input mixes. On the other hand, applying VRS would assess some banks (usually relatively small and big ones) with respect to a convex frontier under the justification that it would not be technologically feasible for them to operate under the same input/output proportions as the CRS-efficient banks. However reasonable this may seem for a certain point in time, our reference set comprises 13 years which is adequate time for banks to expand or contract their operations and therefore VRS might not even be appropriate for our purposes in this case.

Another line of argument that provides support to the CRS assumption is that the median scale efficiency is quite high (0.989) suggesting that half of the DMUs in the sample are associated with a scale efficiency between almost 0.99 and 1. Since scale efficiency is the ratio of the CRS over the VRS technical efficiency scores, the high value of the median suggests that the two frontiers are quite close to each other and therefore CRS is a reasonable assumption. There are only a few cases where SE is quite small and we therefore find useful to provide a histogram with the distribution of scale efficiency scores in Figure 6.3.

Figure 6.3. Distribution of scale efficiencies



Moreover, given that the underlying DGP is technologically homogeneous to a considerable extent (this will be further discussed in the following subsection), we could state that the observed scale efficiency scores are a good approximation of the population ones since the computed technical efficiencies are quite robust due to this homogeneity (see subsection 5.2.1 of the suggested guidelines). That is, although the technical efficiencies are subject to sampling variations, we would expect the DEA scores to be relatively close to their population values and therefore the distribution of the sample scale efficiencies to be similar to the population distribution. Assuming that the sample distribution is a representative one, the few low scale efficiencies may correspond to a few isolated cases of banks who failed to catch up with the changes and adjust their size accordingly.

As a final note, there seems to be a non-conclusive debate in the literature on the assumption of returns to scale in DEA. The early literature provides evidence in support of CRS in the form of flat, U-shaped cost curves (Berger et al., 1993). Later studies seem to turn their attention to unexploited scale economies evident by small banks and provide arguments which are in support of VRS (Berger and Mester, 1997). Matthews and Thompson (2014) conclude that the potential for scale economies is left open in the literature. Thus, the assumption of CRS finds support on one stream of the literature, while it also seems to be reasonable in our case as well.

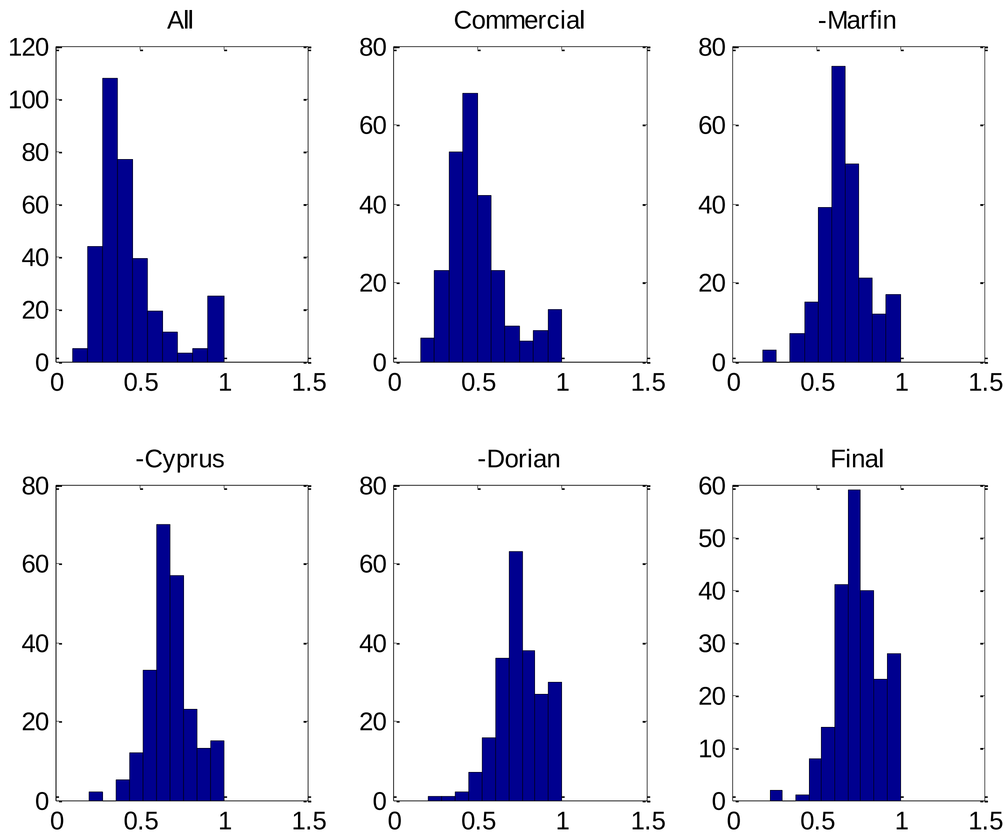
6.4.4.2 The effect of technological heterogeneity

This section discusses the methodological approach followed to decide whether the sample is appropriate to apply bootstrap DEA and this relates to the suggestions in section 5.2 of the suggested guidelines. In particular, we had suggested that technological homogeneity is desirable, which, in our case, translates into homogeneity across both DMUs and time periods (the latter already discussed in the previous subsection).

To exhibit the effect of technological heterogeneity we present histograms of the efficiency distribution by including all financial intermediaries[131] that operated during the study period and then we exclude non-commercial banks. Then we remove commercial banks (one at a time) which we consider as outliers and we observe how the distribution of efficiency scores gradually changes. In particular, we observe that the distribution shifts from a symmetric one with a relatively thin tail to the right, towards an almost half-symmetric distribution with a concentration of values towards 1. In terms of the discussions in chapter 2, we move from a sample associated with technological heterogeneity, where the application of bootstrap DEA is not permissible, towards a more technologically homogeneous sample where bootstrap DEA performs well if the sample is large enough (we have suggested 120 DMUs or more).

[131] Regarding other financial intermediaries, the sample includes 3 investment banks (Aegean Baltic Bank, Euromerchant Bank, Investment Bank of Greece) 1 savings bank (Hellenic Postbank), 2 development banks (Hellenic Industrial Development Bank, National Investment Bank for Industrial Development), 3 mortgage banks and building societies (Deposits Loans and Consignations Fund, National Housing Bank, National Mortgage Bank) and two cooperative/industry-specific banks (Pancretan Bank, Traders' Bank).

Figure 6.4. The effect of deleting outliers on the distribution of technical efficiency scores



It is interesting, though, to explain why removing certain DMUs has this effect on the distribution of efficiency scores. To begin with, consider non-commercial banks: their operations are quite different and could exhibit a high ratio of loans to deposits or a high proportion of financial assets compared to what a typical commercial bank would exhibit. For example, the two development banks included in the “All” sample used to receive their liquidity from the Bank of Greece, hence deposits were very low and at the same time their loans were very high, financing major public projects. Including these two banks in the sample would introduce technological heterogeneity as they operated under a much higher output/input ratio which was not feasible for commercial banks.

Moreover, under the intermediation approach these two banks would always appear efficient while they would set counter-intuitive efficient input targets for inefficient banks.

Regarding the commercial banks removed, the rationale is similar. For example Marfin Bank, being a former investment bank as well as the Greek subsidiary of a Cypriot conglomerate, had limited commercial banking activities while its business plan was different. Regarding Cyprus Bank, it was excluded from the sample as it reflects the operations of the Greek branches of the Cypriot Cyprus Bank and hence the reporting standards or the business model are different compared to the rest of the sample. The next exclusion, Dorian Bank, although officially classified as a commercial bank, it focused its operations on large enterprises, maritime financing as well as private banking and became an investment bank when it merged in 1999 with Telesis Finance (creating Telesis Investment Bank). Finally, the removal of the operations of the Agricultural Bank of Greece until 1991 is justified by the fact that it was a non-for-profit governmental organization and it only became an SA after 1991, expanding its activities to commercial banking and extending its potential clientele outside the agricultural sector.

These banks use a different “technology” compared to commercial banks and this “technology” can be expressed in terms of their conduct of business or business plan, which would imply a different input/output scheme. When these banks are included in the sample, they appear as efficient, distorting the frontier and leading to “unfair” evaluations for the other commercial banks. This lends support to our suggestion in subsection 2.8.3 that symmetric distributions with a thin tail towards 1 suggest technological heterogeneity and it might not be a good idea to even apply DEA as the resulting input contractions (or output expansions in output orientation) would not be feasible. Prior to applying DEA, it should be ensured that all DMUs are members of the same feasible set; hence, our suggestion could be perceived as an exploratory data analysis approach which would inform the data selection process.

Having justified the data selection process we now move to performing the diagnostic checks proposed in the suggested guidelines in section 5.2. In particular we compute the first four moments of the various DEA samples which will be used to associate the sample distribution with the underlying population, which carries implications for the applicability of bootstrap DEA. Table 6.5 below presents this information and it is obvious that the data selection process has significantly increased the mean efficiency and has reduced its variability. At the same time the median converges to the mean while the shape of the distribution becomes less skewed and less peaked. Comparing these results with Table 2.6 we could say that the initial sample ('All') corresponds to the "Trun.Normal Low" case which is associated with technological heterogeneity. In particular, both skewness and kurtosis are quite high, which is the characteristic that stands out in this DGP, while the corresponding histograms are very similar to each other.

On the other hand, the final sample has a kurtosis relatively close to 3 and at the same time negative skewness, which is a combination that we only meet under the "Standard" case which corresponds to technological homogeneity. However, skewness is smaller in absolute terms while the histogram could be characterised as a mixture of the “Standard” and “Trun.Normal High” DGPs. Given that the latter distribution was designed as a mixture of technological homogeneity and heterogeneity, it is reasonable to state that the final sample reflects also such a mixture but with more technological homogeneity compared to that under the Trun.Normal High” DGP. This is also supported by the larger concentration of efficiency scores towards 1.

Table 6.5. Diagnostics to identify the underlying DGP

	Mean	Median	St.Dev.	Skewness	Kurtosis
All	0.441	0.375	0.201	1.495	4.671
Commercial	0.501	0.464	0.180	1.138	4.070
- Marfin	0.666	0.660	0.143	0.107	3.928
- Cyprus	0.684	0.673	0.131	0.089	4.068
- Dorian	0.754	0.746	0.137	-0.345	3.883
Final	0.753	0.747	0.133	-0.335	3.954

The implications of our diagnostic analysis are important for the further examination of the final sample. First, the fact that the sample has a considerable technological homogeneity suggests that the “technology” of transformation of inputs into outputs under the intermediation approach has not changed dramatically during the period of study. If the frontier had shifted out substantially due to technological developments we should have observed a distribution that is associated with technological heterogeneity with the more recent banking operations defining the frontier and the older observations lying on the left tail of the distribution; this is not the case in our sample as the efficient DMUs are scattered across the study period[132]. We are not suggesting that the various "technological" advances in Greek banking were not important; we are just arguing that the results from the pooled sample are credible and can be used for further analysis.

Another important implication, and taking into account the suggested guidelines in chapter 5, is that applying the moments bootstrap DEA on the final sample will yield consistent results and accurate confidence intervals. The relatively high technological homogeneity and the fact that the sample includes 216 observations means that even the DEA scores will be quite accurate; however, they are subject to sampling variations which can be adequately captured by the moments bootstrap. Hence, the discussion on hypothesis testing in chapter 3 is relevant and the therein suggestions can be applied in our case.

6.4.4.3 Implementation

The illustration of the approaches discussed in the previous chapters proceeds in two steps. We first apply bootstrap DEA to compute and compare confidence intervals of interest and then we proceed with examining the effects of sector reforms on banks' efficiency, using the hypothesis testing approaches discussed in chapter 3.

As already mentioned, efficiency is estimated by a CRS, input-oriented DEA model. To gauge the sensitivity of the efficiency scores towards sampling variations we apply

[132] See also Figure 6.2.

bootstrap DEA (we use 2000 repetitions) under all smoothing alternatives considered in the previous chapters: the LSCV bootstrap, the SJ bootstrap, the Moments bootstrap and the Naïve bootstrap. The logic of the bootstrap algorithm has already been explained in section 2.6.2 while section 4.5 describes how this algorithm can be adapted for the case of the moments bootstrap[133]. The resulting bootstrap distributions are used to compute the bias-corrected estimates of the "true" efficiency scores and to construct the percentile confidence intervals of Simar and Wilson (1998). The intervals of Simar and Wilson (2000a) are excluded from the analysis due to the inferior performance evidenced in our simulations[134].

Then we analyse the effects of the sector reforms on the efficiency and productivity of Greek banks. Due to the small number of observations per year we had to pool the dataset in order to satisfy the minimum size requirements for applying bootstrap DEA which is more than 120 observations under the moments bootstrap (though quite higher for the other smoothing approaches). Pooling the sample is an acceptable approach (Fried et al., 2008, pp.54) and it has been followed in DEA empirical studies in Greek banking (Halkos and Salamouris, 2004; Siriopoulos and Tziogkidis, 2010).Then the ratios of those "global" technical efficiency scores for each bank and between adjacent periods are in fact the Global Malmquist indices of Pastor and Lovell (2005) as explained later in this subsection.

---

[133] Information about the LSCV and SJ smoothing approaches is provided in Appendix I.

[134] However, results on this approach are available upon request by the author. Results on naïve are only presented to compare smoothing versus non-smoothing methods.

The bootstrap in this case randomly redistributes efficiency scores of DMUs across all time periods and the resulting bootstrapped values will be members of the same feasible set by construction. The lower the technical heterogeneity across time periods the narrower the confidence intervals will be. In our case we observe that the efficient DMUs are scattered across the study period (this is has been already discussed; see Figure 6.2 and Appendix XII), suggesting that banks have the capacity to operate efficiently in any year; at least with respect to the particular inputs and outputs.

The only popular alternative approach which has been used with bootstrap DEA is the Bootstrap Malmquist index of Simar and Wilson (1999). In our case this approach would not be suitable as the number of DMUs is too small (in some cases even to apply simple DEA). In addition, despite the fact that the manual of Prof Paul Wilson's *FEAR* package states that the Bootstrap Malmquist is fully compatible with unbalanced panels, there are issues of "information loss" in this case. In fact, we demonstrate in Appendix XI the potential problems arising in this case, using the derivations and definitions in Simar and Wilson (1999).

Let us now provide more details on the approach we follow to examine the effects of sector reforms. As already explained, we include all observations under the same frontier, which is also termed as "global" frontier (Pastor and Lovell, 2005). Define the contemporaneous technology (or feasible set) in period t as:

$$\Psi^t = \left\{(x^t, y^t) \in \mathbb{R}_+^{p+q} | x^t \text{ can produce } y^t\right\}, \quad t = 1,2, \dots T \tag{6.1}$$

The global technology is defined as the convex hull of the contemporaneous technologies (Pastor and Lovell, 2005):

$$\Psi^G = \text{conv}\left\{\Psi^1 \bigcup \ldots \bigcup \Psi^T\right\} \tag{6.2}$$

The input-oriented, CRS DEA score of DMU k that operates in period t and benchmarked against the N observations of the global frontier is:

$$\hat{\theta}_k^G(x_k^t, y_k^t) = \min\left\{\theta \,\middle|\, y_k^t \le \sum_{i=1}^{N} \lambda_i y_i\,;\ \theta x_k^t \ge \sum_{i=1}^{N} \lambda_i x_i\,;\ \theta > 0; \lambda_i \ge 0,\ \ \forall\ i = 1, \ldots, N\right\} \tag{6.3}$$

And the linear program above can be also used to compute $\hat{\theta}_k^G(x_k^{t+1}, y_k^{t+1})$. Since $\hat{\theta}_k^G(x_k^t, y_k^t)$ and $\hat{\theta}_k^G(x_k^{t+1}, y_k^{t+1})$ are two different DMUs which are assessed under the same frontier, despite being the same firm k, we can follow the guidelines in 3.3.2 to test for their "efficiency differences". In particular, in the context of subsection 3.3.2 we define $\theta_k = \theta_k^G(x_k^t, y_k^t)$ and $\theta_v = \theta_k^G(x_k^{t+1}, y_k^{t+1})$ and thus the ratio θ_k/θ_v now becomes $\theta_k^G(x_k^t, y_k^t)/\theta_k^G(x_k^{t+1}, y_k^{t+1})$. This ratio is in fact the Global Malmquist index introduced by Pastor and Lovell (2005)[135]:

$$M^G\left(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}\right) = \frac{\theta_k^G(x_k^t, y_k^t)}{\theta_k^G(x_k^{t+1}, y_k^{t+1})} \tag{6.4}$$

If $M^G < 1$ then the productivity of DMU k increased between periods t and $t+1$, while if $M^G > 1$ then the productivity of DMU k decreased, whereas $M^G = 1$ indicates no change in productivity.

Pastor and Lovell (2005) argue that the Global Malmquist Index has four benefits over the simple Malmquist index of Caves et al. (1982). The most important one is that, unlike the standard Malmquist index, it is circular, in that:

---

[135] See equation (2) of their paper. The Global Malmquist index is introduced in terms of distance functions which are the inverse of the technical efficiency scores used in our presentation.

$$M^G(x_k^t, y_k^t, x_k^{t+\tau}, y_k^{t+\tau})$$
$$= M^G\left(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}\right) \times \ldots \times M^G\left(x_k^{t+\tau-1}, y_k^{t+\tau-1}, x_k^{t+\tau}, y_k^{t+\tau}\right) \quad (6.5)$$

Second it provides a single measure (and does not depend upon the time direction) without requiring the computation of geometric means of adjacent time periods. Third, the frontier shift element is with respect to the whole period of study and not relevant to two adjacent time periods. Finally, it can be decomposed to the usual elements which are all immune to linear programming infeasibilities[136].

The proposed test of efficiency differences in subsection 3.3.2 can be easily adapted in this context. We just need to observe that the ratio in (3.11) in this case is the Global Malmquist index in our case and the null hypothesis now becomes:

$$H_0: M^G\left(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}\right) = 1, \qquad H_1: M^G\left(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}\right) \neq 1 \quad (6.6)$$

The bootstrap distribution of efficiency ratios in (3.12) is therefore a bootstrap distribution of Global Malmquist indices:

$$\widehat{M}_b^{G,*} = \frac{\hat{\theta}_k^G(x_k^t, y_k^t)_b^*}{\hat{\theta}_k^G(x_k^{t+1}, y_k^{t+1})_b^*}, \qquad b = 1,2, \ldots B \quad (6.7)$$

[136] Pastor and Lovell (2005) show that the Global Malmquist can be decomposed into efficiency change and technical change as follows: $M^G(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}) = \frac{\theta_k(x_k^t, y_k^t)}{\theta_k(x_k^{t+1}, y_k^{t+1})} \times \left\{\frac{\theta_k(x_k^{t+1}, y_k^{t+1})}{\theta_k^G(x_k^{t+1}, y_k^{t+1})} \cdot \frac{\theta_k^G(x_k^t, y_k^t)}{\theta_k(x_k^t, y_k^t)}\right\}$, where $\theta_k(x_k^t, y_k^t)$ and $\theta_k(x_k^{t+1}, y_k^{t+1})$ are the usual efficiency scores for DMU k at times t and $t+1$, respectively. The first element is the efficiency change component and the latter is the technical change component. It can be also decomposed into its scale efficiency component (Ray and Desli, 1997) which should be a simple extension of the previous decomposition (Pastor and Lovell, 2005). However, these decompositions are not considered here due to the small sample size as in 7 out of 13 years the number of DMUs does not even satisfy the well-known (and in fact challenged) “rule of thumb” for simple DEA which would require at least 15 DMUs in our case; the requirements for bootstrap DEA are obviously much higher as our previous simulations have shown. We also note that another possibility would be to use the Global Frontier shifts of Asmild and Tam (2007), the combination of which with bootstrap DEA is proposes for future research.

Assuming that $\left(\widehat{M}_b^{G,*} - \widehat{M}^G\right)\big|\widehat{\mathcal{P}} \sim \left(\widehat{M}^G - M^G\right)\big|\mathcal{P}$, we can bias-correct the bootstrap distribution above as in (3.13) and use its percentiles to test the hypothesis in (6.6); if $1 \notin \left(\widetilde{M}_b^{G,*,(a/2)}, \widetilde{M}_b^{G,*,(1-a/2)}\right)$ we can accept the alternative hypothesis that productivity has changed from t to $t+1$. And if the null is rejected we proceed with examining the direction of productivity change by testing the two possible alternatives:

$$H_1: M^G\left(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}\right) > 1, \text{ or } H_1: M^G\left(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}\right) < 1 \tag{6.8}$$

The following p-values can be used to test (6.8):

$$plow = \frac{\#\left(\widetilde{M}_b^{G,*} < 1\right)}{B} \quad \text{and} \quad phigh = \frac{\#\left(\widetilde{M}_b^{G,*} > 1\right)}{B}, \quad b = 1,2,\dots B \tag{6.9}$$

If (6.6) is rejected and $plow < a$ we could accept the alternative $H_1: M^G(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}) > 1$ which indicates a decline in productivity, while if (6.6) is rejected and $phigh < a$ we could accept the alternative $H_1: M^G(x_k^t, y_k^t, x_k^{t+1}, y_k^{t+1}) < 1$ which indicates an increase in productivity.

6.5 Empirical Results

This section presents the empirical results of the illustrative example. We first present general results which look at the global efficiency scores of the DMUs examined, the relationship between size and efficiency as well as at the distribution of efficiency scores relative to the inputs and outputs used (subsection 6.5.1). We then look at the shapes of the bootstrap distributions of global efficiency scores and the associated confidence intervals (subsection 6.5.2). Then the hypothesis testing procedures described above is implemented and we look into the numerical results obtained from the various approaches (subsection 6.5.3). Finally, we empirically analyse the effects of the sector reforms using the results obtained from the moments bootstrap which has been argued to perform well in small samples (subsection 6.5.4).

6.5.1 General results

To provide an informative summary of the global efficiency scores, we present in Table 6.6 the averages and standard deviations for the listed size percentiles and for the whole sample. In particular, the size groupings are according to each bank's total assets in each year and in constant 1995 values (for example the top 10% comprises a combination of National Bank and Alpha bank in certain years). We note that the average efficiency (median is 0.747, close to mean) over the period of study is similar to the one documented in the international and Greek literature (Berger and Humphrey, 1997; Chortareas et al., 2008).

There are strong indications in Table 6.6 that larger banks tend to be more efficient as there is a quite monotonic decrease in efficiency as size decreases. At the same time standard deviations are small enough to suggest that the size-efficiency relationships are not due to chance. This implies that during the study period the SCP paradigm seems to be valid[137].

Table 6.6. DEA scores by size percentile

Percentile	*Average Eff.*	*Stand.Dev.*
Big 10%	0.868	0.131
10%-20%	0.863	0.096
20%-30%	0.820	0.109
30%-40%	0.753	0.138
40%-50%	0.704	0.056
50%-60%	0.678	0.076
60%-70%	0.703	0.153
70%-80%	0.713	0.094
80%-90%	0.735	0.144
Small 10%	0.676	0.180
Total	**0.753**	**0.133**

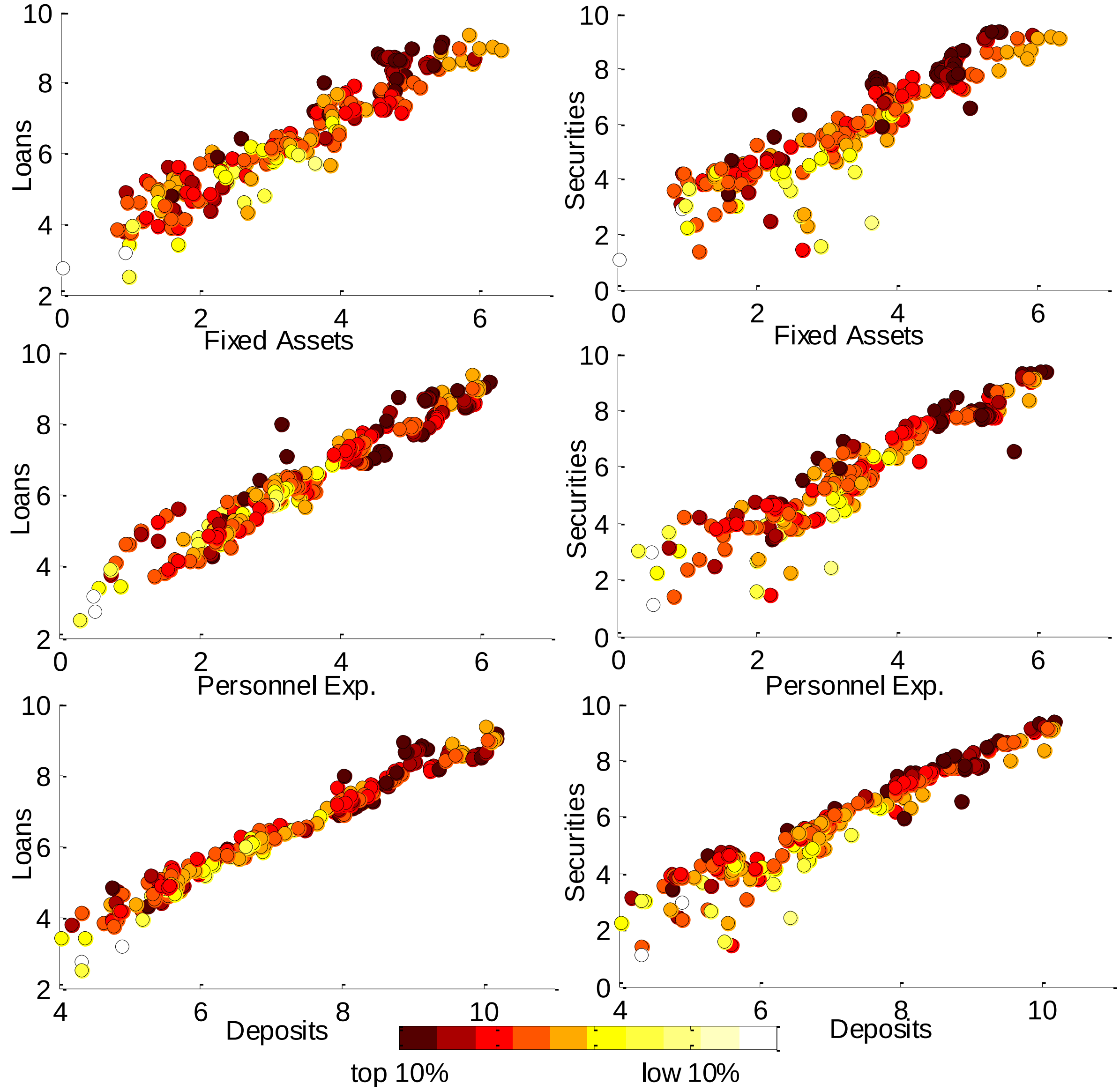
The positive relationship between size and efficiency is also observed in Figure 6.5, which depicts the same scatterplots as in Figure 6.2 but this time the colour mapping corresponds to the efficiency scores observed in the sample; the higher the efficiency score of a DMU the darker the dot filling. It is obvious that the most efficient DMUs lie on the north-eastern part of the scatterplots, confirming that banks which use more inputs and outputs are more efficient. There are a few exceptions of very efficient and inefficient DMUs scattered across the graphs; however, as we move outwards from the origin, the fillings are in principle darker.

[137] The author would like to note that he has also examined the efficiency behaviour of Greek banks after 2000 and this pattern is no longer observed as some small banks appear as efficient. One possible explanation for this change is the adoption of “technologies” or financial innovations by small banks which allowed them to perform operations that previously only large banks could afford to undertake. This is in accordance with the findings in the elaborate review study of Amel et al. (2004).

Another interesting observation which combines Table 6.6 and Figure 6.5 is that most of the variability in efficiency scores is observed for the smallest banks. Indeed, the smallest banks are associated with higher standard deviations according to Table 6.6, while their positions in the lower end of Figure 6.5 seem to be slightly more scattered. This corresponds to the left tail of the empirical distribution of efficiency scores which could potentially affect bootstrap DEA results by introducing additional variability across DMUs when resampling. In our case though, this is not of concern as the dataset is quite homogenous[138]. However, we would recommend practitioners who want to apply bootstrap DEA to be careful when dealing with substantial variability in the lower end of the efficiency distribution. This is, though, something that could be looked at in a future paper.

[138] It is interesting to note that the author has experimented with the linkage of the scattered, inefficient observations and bootstrap DEA. In particular, we observe that by removing the very inefficient and scattered DMUs the resulting confidence intervals become narrower.

Figure 6.5. Inputs/outputs (in logs) and efficiency distribution



6.5.2 Bootstrap distributions and confidence intervals

In this subsection we present results regarding the distributional aspects of bootstrapped efficiency scores along with the associated confidence intervals. We only discuss the bias-corrected distribution and the associated percentile confidence intervals of Simar and Wilson (1998). The "basic intervals" of Simar and Wilson (2000a) are not discussed as the simulations have suggested that convergence is considerably slower and therefore they are not suitable for small samples. The focus is rather on the choice of the smoothing method which has been shown to affect performance.

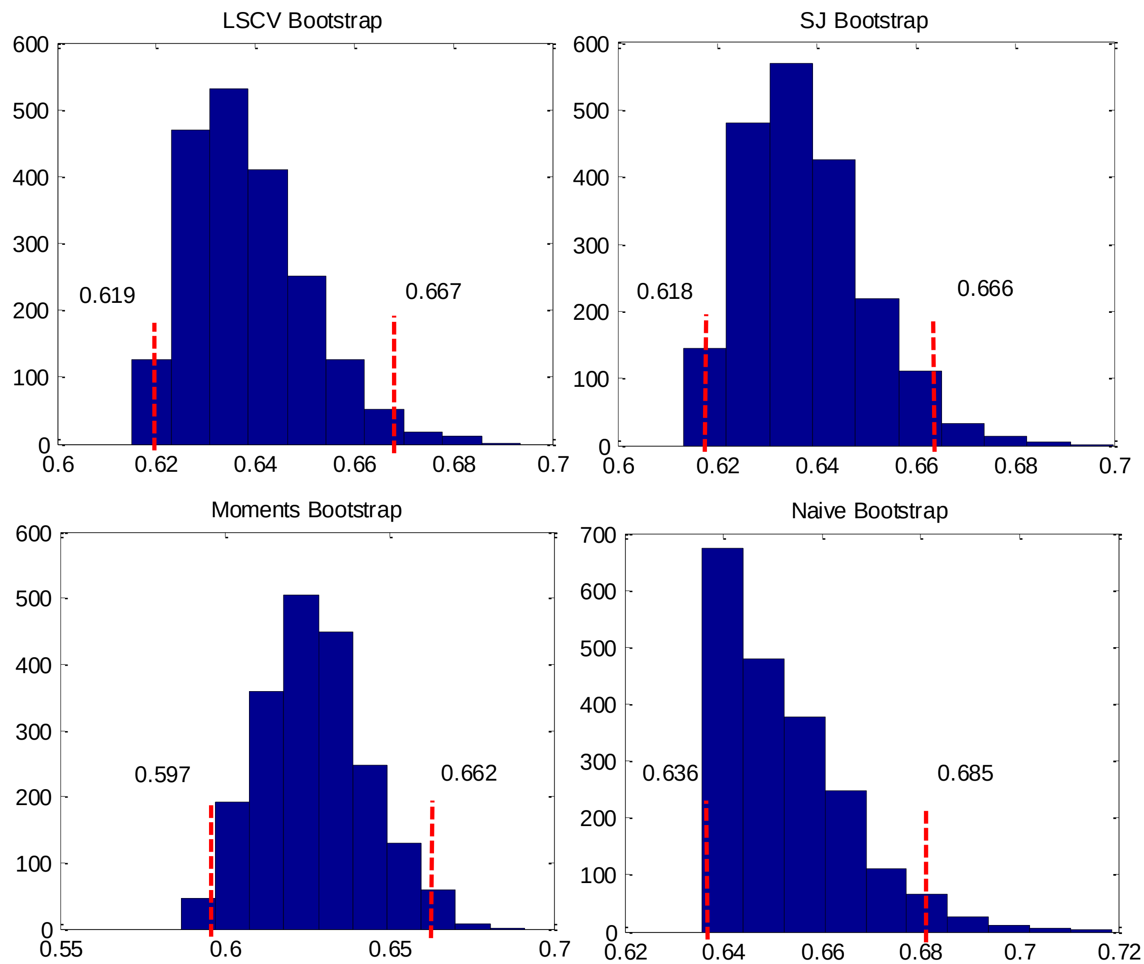
Table 6.7 below presents the average moments of the bias-corrected bootstrap distributions along with the average 95% SW1998 widths. More analytic results for each bank per year of operation can be found in Appendix XII where we also present results for the bias-corrected and accelerated confidence intervals of Efron (1987), the adoption of which for bootstrap DEA was discussed in Appendix VII.

Table 6.7. Bootstrap distribution moments and widths of 95% intervals

	Mean	*St.Dev*	*Skew*	*Kurt*	*SW98 Width*
LSCV	0.699	0.023	0.632	3.651	0.090
SJ	0.699	0.024	0.615	3.542	0.091
Moments	0.686	0.028	0.297	2.984	0.107
Naïve	0.713	0.025	1.083	6.031	0.092

In terms of distributional aspects, we observe that the moments bootstrap is on average less skewed and leptokurtic compared to the other two smoothing methods while the SW1998 widths are slightly wider (by 0.016 units). The average standard deviation of the distributions indicates that there is sampling variability that justifies the application of bootstrap DEA. To provide a better insight regarding the shape of the distributions in each case, we have plotted, as an example, the histograms of the bootstrapped efficiency scores for the DMU "Average Bank 1991" where we have also indicated with red dotted lines the 95% SW1998 confidence intervals. This DMU was selected by chance but it can be also deemed as the middle of the reforms period 1987-1994.

Figure 6.6. Bootstrap distributions for Average Bank in 1991



The descriptive statistics and SW1998 confidence intervals for the histograms above are provided in Table 6.8 below. By inspecting Figure 6.6 and the table below we observe that the moments bootstrap, being more symmetrical, has well-defined tails on either side of the bootstrap distribution. On the other hand, the inconsistent “Naïve” bootstrap is skewed to the right and quite leptokurtic and therefore has a significant mass of bootstrap values to the lower end of the distribution and a thin tail to the right. The two smooth bootstraps (LSCV and SJ) have almost the same performance and are substantially less skewed and leptokurtic compared to the Naïve bootstrap but more skewed compared to the moments bootstrap. This may be one of the contributing factors for the improved performance of the moments bootstrap in our simulations but it also suggests that it may be more meaningful to account for skewness under these smoothing techniques when constructing confidence intervals.

Table 6.8. Details for distribution of Average Bank in 1991

	Mean	*St.Dev*	*Skew*	*Kurt*	*SW98 Low*	*SW98 High*
LSCV	0.638	0.012	0.817	3.796	0.619	0.667
SJ	0.638	0.012	0.817	3.796	0.619	0.667
Moments	0.627	0.017	0.290	2.904	0.597	0.662
Naïve	0.652	0.014	1.087	4.323	0.636	0.685

Regarding the confidence interval widths in either Table 6.7 or the table above, it should not be perceived that the Moments bootstrap underperforms as the simulations in the previous chapters have suggested otherwise. For example, we have already mentioned that, when Simar and Wilson (1998) used the bias-corrected intervals of Efron (1982) to account for skewness, the SW1998bc intervals were wider by 0.015 to 0.03 units compared to the SW1998 ones (see subsection 3.3.1). This suggests that the costs of slightly widely confidence intervals seem to be small compared to the potential benefits. In fact, since the moments bootstrap is associated with higher and converging coverage probabilities for sample sizes as large as ours, any differences in the behaviour of the other smoothing techniques could be interpreted as deviations from the "benchmark" (the moments bootstrap); however, with caution due to the infinite possible underlying DGPs. In our case, it could be argued, for example, that the slightly wider confidence intervals may allow to successfully capture the underlying population efficiency scores, which the other two smooth bootstraps seemed to miss out in the simulations of the previous chapters.

We could go one step further with the previous example of the Average Bank in 1991 and translate the efficiency scores and confidence intervals in terms of its input values. In particular, Table 6.9 below reports the actual input values for the Average Bank in 1991, its DEA-efficient input levels and below it reports the target values computed by the bias-corrected bootstrap distributions. For example, Average Bank in 1991 could have produced the same outputs by using €41.24 million worth of fixed assets according to DEA (along with the required reductions in the other inputs). However, a better estimate of the input level that would make the average bank efficient would be around €38.70 million, focusing on the moments bootstrap. There is a chance of 95% that this "ideal" input level ranges between €36.87 and €40.85 million, which excludes by far the observed value of €61.71 million. Given that the Average Bank in 1991 represents average operations, we could also state, for example, that Greek banks would have operated efficiently if they used on average €23.01 million less of their fixed assets (apart from deposits and personnel expenses) and still produce the same level of output. The same rationale could be applied to the other inputs and smoothing methods.

Table 6.9. Target input levels for Average Bank 1991

	Fixed Assets			Deposits			Pers. Expenses		
Actual Value	61.71			3084.19			59.51		
DEA Efficient	41.24			2061.50			39.78		
		Range			*Range*			*Range*	
LSCV	39.40	*38.22*	*41.16*	1969.20	*1910.49*	*2057.05*	38.00	*36.87*	*39.69*
SJ	39.40	*38.22*	*41.16*	1969.20	*1910.49*	*2057.05*	38.00	*36.87*	*39.69*
Moments	38.70	*36.87*	*40.85*	1934.10	*1842.63*	*2041.85*	37.32	*35.56*	*39.40*
Naïve	40.23	*39.23*	*42.27*	2011.05	*1960.60*	*2112.59*	38.81	*37.83*	*40.77*

* *Values in millions of Euros and in 1995 prices*

6.5.3 Hypothesis testing results

We now present the hypothesis testing results which will be used in the next subsection to analyse the effects of the sector reforms on the efficiency of Greek banks. To perform this task we have computed the ratios of the bias-corrected efficiency scores for the operations of each bank between adjacent periods; a bootstrap version of the Global Malmquist index. We remind that values of the index below 1 indicate productivity increase and values greater than 1 indicate productivity decline.

The results for the Average bank and for the three smoothing alternatives are presented in Table 6.10. In particular, for each of the LSCV, SJ or Moments bootstraps, we present the bias-corrected Bootstrap Global Malmquist index means (*Gl.Mal.BC*) along with an indication of significance. In particular, we test the hypothesis of no change in productivity as in (6.6) and if rejected we test for the direction of productivity change as in (6.8) using the p-values in (6.9), which are represented here by *Prob<1* and *Prob>1*. Hence, *"*"* and *"**"* indicate that the increase or decrease in productivity was significant at the 0.05 or 0.01 level of significance, respectively. The results for all banks along with the 95% intervals used to test (6.6) can be found in Appendix XIII.

Table 6.10. Hypothesis testing results for the Average Bank

	LSCV Bootstrap			**SJ Bootstrap**			**Moments Bootstrap**		
	Gl.Mal.BC	*Prob<1*	*Prob>1*	*Gl.Mal.BC*	*Prob<1*	*Prob>1*	*Gl.Mal.BC*	*Prob<1*	*Prob>1*
1987-1988	1.019	*0.279*	*0.721*	1.019	*0.304*	*0.696*	1.015	*0.361*	*0.640*
1988-1989	0.992	*0.752*	*0.249*	0.993	*0.727*	*0.273*	0.991	*0.721*	*0.279*
1989-1990	0.972**	*1.000*	*0.001*	0.971**	*1.000*	*0.000*	0.973**	*0.998*	*0.002*
1990-1991	1.051*	*0.014*	*0.987*	1.053*	*0.015*	*0.985*	1.051	*0.048*	*0.952*
1991-1992	0.891**	*1.000*	*0.000*	0.889**	*1.000*	*0.000*	0.886**	*1.000*	*0.000*
1992-1993	0.918**	*1.000*	*0.000*	0.918**	*1.000*	*0.000*	0.926**	*0.998*	*0.002*
1993-1994	1.046**	*0.000*	*1.000*	1.047**	*0.000*	*1.000*	1.044**	*0.000*	*1.000*
1994-1995	1.016	*0.105*	*0.895*	1.016	*0.107*	*0.893*	1.017	*0.145*	*0.855*
1995-1996	1.026**	*0.000*	*1.000*	1.026**	*0.000*	*1.000*	1.025**	*0.001*	*1.000*
1996-1997	0.936**	*1.000*	*0.000*	0.936**	*1.000*	*0.000*	0.938**	*1.000*	*0.000*
1997-1998	1.119**	*0.000*	*1.000*	1.121**	*0.000*	*1.000*	1.122**	*0.000*	*1.000*
1998-1999	0.927**	*1.000*	*0.000*	0.927**	*1.000*	*0.000*	0.927**	*1.000*	*0.000*

The results in Table 6.10 suggest that the direction of productivity change is the same across the different approaches but the significance levels can be different as the position of the bootstrap distributions relative to 1 can be different. That is, observing the p-values we find that the moments bootstrap tends to "include" 1 to a greater extent compared to the other two smoothing methods as the p-values under the moments bootstrap towards the tails are smaller. In fact, we observe that in one case (1990-1991) the moments bootstrap does not reject the null hypothesis of no change in productivity while the other two approaches do. This may be attributed to the slightly wider or less skewed intervals generated under the moments bootstrap but it may also suggest that the other two procedures are more likely to reject a null hypothesis. And given the good convergence rates for the moments bootstrap in our simulations, this might be an indication of Type I errors for the two smoothing alternatives.

Looking in Appendix XIII we find that the null hypothesis is rejected 136 times under the LSCV bootstrap (and at a 5% level), 137 times under the SJ bootstrap and 131 times under the Moments bootstrap. In these cases we find only one case where the moments bootstrap rejects the null while the LSCV bootstrap fails in doing so[139], while we find 6 cases where the LSCV and SJ bootstraps both reject the null when the moments bootstrap does not[140]. So the behaviour of the two smoothing alternatives is very similar while under the Moments bootstrap we observe about 6% less rejections[141]. The author believes that these differences could have been more pronounced if the sample was less homogeneous with respect to the input-output relations (see for example Figure 6.5). Moreover, the nature of the particular tests contributes towards this direction as, although the magnitudes of the estimated Global productivity change indices are different, their location within the associated bootstrap distributions seems to be analogous across the different approaches.

[139] This is Xiosbank (97-98) which is rejected by SJ and Moments bootstrap but not under LSCV.

[140] These are: Alpha Bank (93-94), Cretabank (96-97), Emporiki Bank (95-96), General Bank (89-90), Xiosbank (94-95) and Average Bank (90-91).

[141] We note here that if we were using the alternative approach suggested in footnote 89 in subsection 3.3.2, we would end up with 133 rejections of the null for LSCV and SJ (corresponding to the same cases) and with 112 for the Moments bootstrap, all of which being in common with the other two smoothing approaches. That is, the Moments bootstrap rejects the null for the same cases as with the LSCV and SJ bootstrap but does not reject the null in 21 cases. We are not presenting analytical results for this approach as it is associated with more limitations compared to the approach we use here, which have been discussed in chapter 3.

A summary of the productivity changes on an annual basis is provided in Table 6.11. In particular, the table below indicates the direction of the productivity change for the average bank along with a summary of the movements of all other banks (excluding the average and weighted average banks). We find that the change in productivity under all bootstrap approaches is the same, though their magnitude and level of significance might differ. The total number of commercial banks that exhibited an increase or decrease in productivity is presented for each year while we also report how many of those changes were significant under each smoothing alternative in the last three columns. One interesting observation is that the differences in hypothesis testing decisions are scattered across 6 out of 12 time periods which suggests that they should not be disregarded; however, it is true that in terms of policy implications for the whole sector the conclusions are not affected considerably if only the direction of productivity change is considered (and not its magnitude).

Table 6.11. Summary of hypothesis testing results for sample

	Av. Bank	*Increase*	*Decrease*	*LSCV Sig*	*SJ Sig*	*Mom. Sig*
1987-1988	↓	6	4	6	6	6
1988-1989	↑	4	6	8	8	8
1989-1990	↑	10	2	11	11	10
1990-1991	↓	4	8	8	8	8
1991-1992	↑	7	6	9	9	9
1992-1993	↑	8	6	9	9	9
1993-1994	↓	4	13	15	15	14
1994-1995	↓	10	7	11	11	10
1995-1996	↓	9	9	17	17	16
1996-1997	↑	14	3	13	13	12
1997-1998	↓	3	14	5	6	6
1998-1999	↑	8	5	8	8	8
Total		*87*	*83*	*120*	*121*	*116*

We also observe that in three cases the productivity change of the average bank may not be in accordance with that of most banks during a certain year. In particular, during the first two years and during 1994-1995 the behaviour of the average bank and of the sample seem to be different. However, this can be explained by the fact that the behaviour of the average bank is largely driven by the largest banks in the sample. We consider this as reasonable since the Greek banking sector is highly concentrated and therefore the average bank is a good representation of the behaviour of the Greek banking sector as it captures these influences[142].

6.5.4 Examining the effects of sector reforms

The results in Table 6.10 have provided evidence that the provision of commercial freedoms results in higher productivity levels in the following year. On the other hand, in the view of regulation tightening or aggressive economic reforms, banks experience a decline in productivity on average. In this subsection we will use the hypothesis testing results to analyse the effects of sector reforms on Greek banks' efficiency on a year-by-year basis as mentioned in the literature review. We will use the results from the Moments bootstrap since the simulations have suggested that it performs better than the other two smoothing alternatives in terms of providing numerically more accurate results.

[142] The behaviour of the Weighted Average bank is very similar but it attaches even more weight to larger banks.

To aid our analysis we will use Figure 6.7 and Figure 6.8 which demonstrate the (log) input-output scatterplots for the “Average Bank” and its trajectory over time, respectively. In particular, Figure 6.7 indicates the position of the Average bank (red fillings) with respect to the other banks in the sample while Figure 6.8 presents the path of the Average Bank over time with the labels indicating the bias-corrected efficiency scores and the year (1 through 13 correspond to 1987 through 1999)[143]. Moreover, the black solid lines in the latter figure indicate a significant increase or decrease in productivity (under the Moments bootstrap and a 0.05 level of significance), while the grey dashed lines suggest no significant change in efficiency. The individual scatterplots and trajectories for each bank are presented in Appendix XIV.

For the rest of this subsection, when we talk about productivity change and firms that exhibited significant productivity decline or increase we will be referring to Table 6.10 and Table 6.11, respectively. The discussion will be also using Figure 6.8 for illustrative purposes.

The first observation is that during the first 7 years of the study period, where all sector reforms took place, the inputs and outputs of the Greek banking sector were quite volatile. On the contrary, the later years appear more tranquil with smoother movements and with a growing trend during 1997 to 1999. The behaviour of the post-reforms period is more straightforward to be explained and through a quick overview

[143] We could have used the respective bias-corrected bootstrap Global Malmquist index instead; however, we found that their use is confusing in terms of presentation in this case. Besides, it can be confirmed that the direction of productivity change is the same as that of the change in the bias-corrected global efficiency scores by comparing Appendices XIII and XIV. This should not be confused with efficiency change as it is (a) based on the bootstrap bias-corrected efficiency scores and (b) with respect to the global frontier.

we could say that the sector experienced growth in both inputs and outputs. However, the analysis of the years during the sector reforms requires a step-by-step analysis in order to extract useful conclusions.

Figure 6.7. Average Bank input-output scatterplots

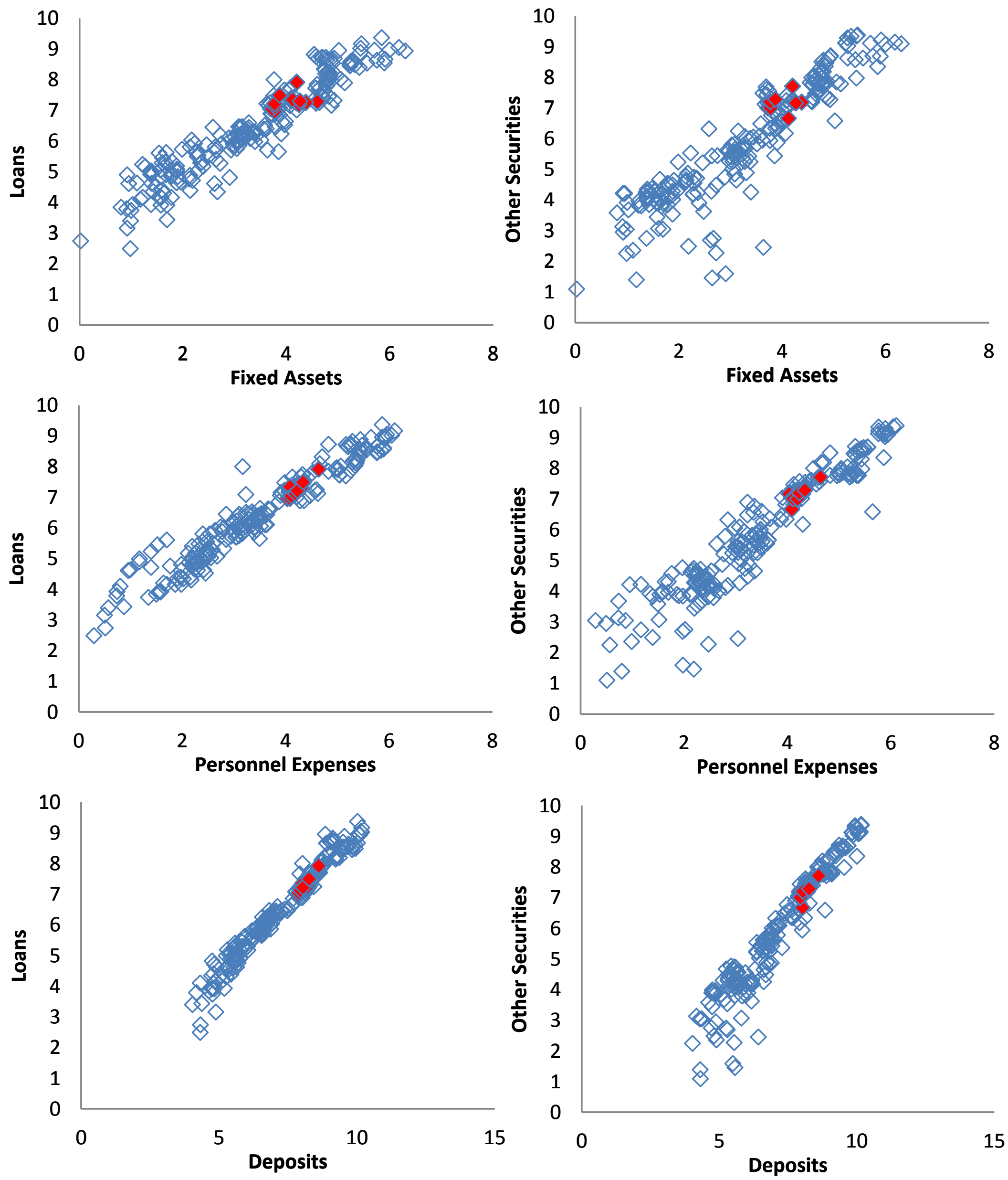
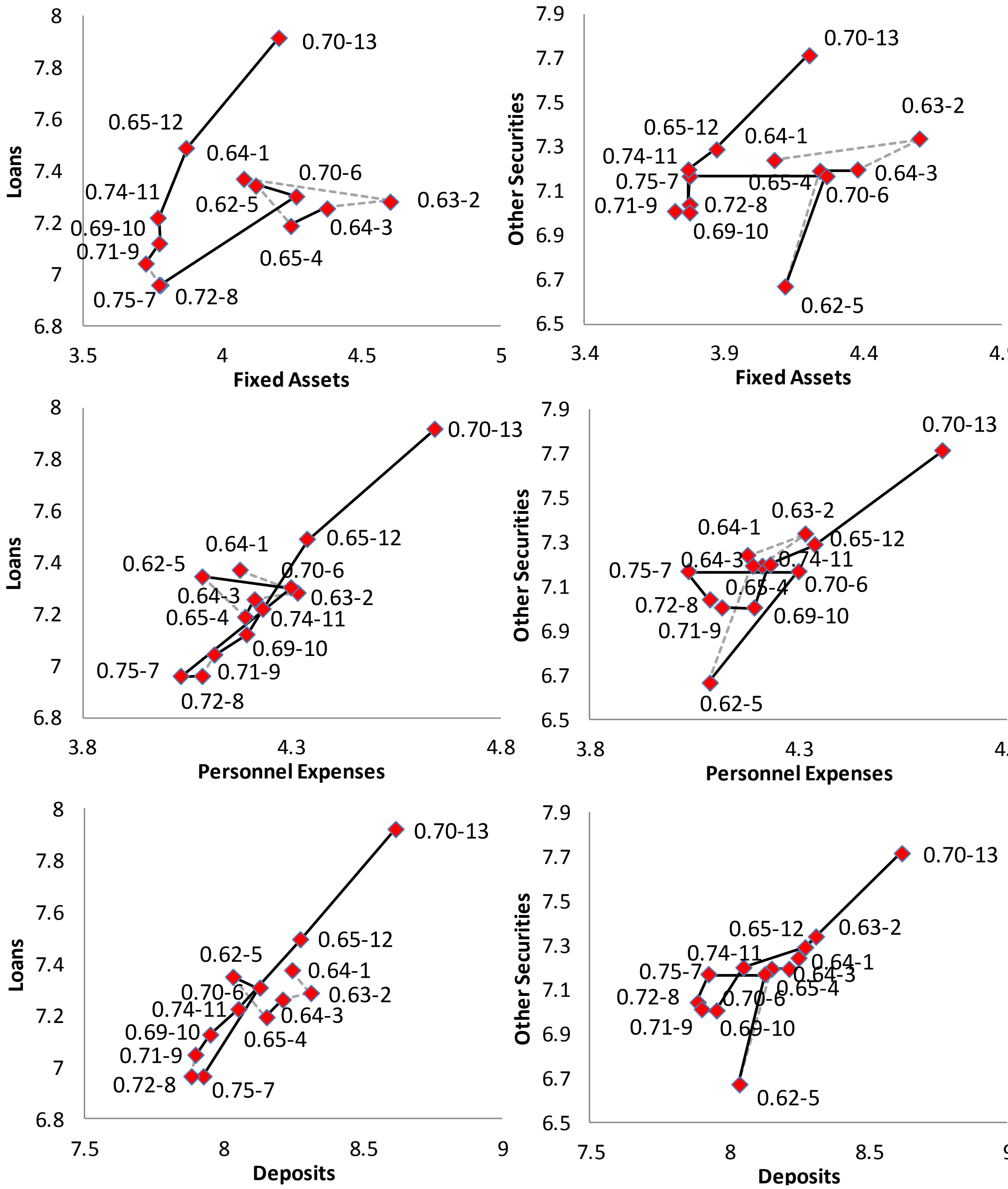


Figure 6.8. Average Bank efficiency trajectory



Considering the whole period of reforms (1987-1994) we would have concluded that the (de)regulation of the Greek banking sector had mixed results if we had followed the standard approach in the literature of comparing years 1987 and 1994 (see for example Gilbert and P. W. Wilson, 1998; Isik and Hassan, 2003a; Chen et al., 2005; Kumbhakar and Sarkar, 2003). In particular, looking in Appendix XIII we find that the change in global bias-corrected efficiency[144] from 1987 to 1994 for 6 banks was positive while for 6 banks it was negative[145]. Given that the positive change was driven by large banks we could state that the overall change was mostly on the positive side as captured by the change in productivity of the Average Bank.

Taking into account, though, each step of the (de)regulation process and of the relevant policy interventions, the results are different. Following Orea (2002), Siriopoulos and Tziogkidis (2010) and Delis et al. (2011) we will assume that the effects of each step of the (de)regulation process appear with one period's lag, which can be also supported by the inflexible Greek labour market (Ayadi, 2008).

The commercial freedoms given to Greek banks were limited during the first year of the reforms; in fact, a few extra controls were imposed, while the scandal of 1987 destabilised the Greek banking sector. We would therefore expect that some banks benefited from the provision of commercial freedoms but we would expect a decline in the productivity levels for the banks which experienced a "managerial shock" in the view of the sector reforms or which were affected by the scandal. Indeed, we observe that the estimated productivity for the Average Bank declines, though insignificantly.

[144] We find easier to inspect the start and end-period bias-corrected efficiency scores with respect to the global frontier. Alternatively we could have computed the bootstrapped ratios of bias-corrected efficiency scores (Global Malmquist indices) between the two periods and compute a bias-corrected Global Malmquist index from the resulting distribution (as we did in Appendix XIII for each bank and for adjacent time periods). We would like to avoid a potential "information overload" and hence we did not present these computations here; however, the author can provide this information upon request.

[145] Increase is documented for Attica Bank, Bank of Central Greece, Emporiki Bank, Ionian and Popular Bank, National Bank and Piraeus Bank. Decline is documented for Alpha Bank, Bank of Athens, Cretabank, Ergobank, General Bank and Macedonia-Thrace Bank.

The average decline was driven by the substantial decrease in productivity of Cretabank (the bank which was primarily affected by the scandal[146]) and the decrease in productivity of National Bank (the biggest bank which was also affected to a small extent by the scandal). The change that stands out is that both fixed assets and personnel expenses increased substantially, especially for large banks; even those that experienced an increase in productivity. One possible explanation is that those banks believed that they should expand their network to exploit the forthcoming commercial freedoms which was registered as a decline in efficiency under the intermediation approach. The other explanation could be related to the scandal as a substantial amount of resources shifted away from Cretabank and the Bank of Central Greece and probably directed to other Greek banks or even abroad.

The initial decline in productivity in 1988 was followed by an increase in 1989 and 1990, which may be attributed to the commercial freedoms given to commercial banks during that period (discussed in the previous section). The increase in productivity from 1988 to 1989 was mainly driven by large banks, despite the fact that the majority of banks exhibited lower productivity. The pattern observed for most banks was an increase in loans, irrespective if their productivity eventually improved or declined. From 1989 to 1990, with the exception of two banks which experienced a small decline in productivity, all other banks recorded a significant increase, mainly due to the decrease in inputs.

[146] Unfortunately we do not have data for the Bank of Central Greece in 1988 which was also involved in the scandal (though not directly). There appears to be a productivity increase from 1987 to 1989 but we cannot be sure about its direction in the first year.

The period from 1990 to 1991 involved more obvious effects; the jump in inflation, combined with the pending imposition of tax on interest on deposits should cause productivity to drop by considerably decreasing the value of securities (the potential decrease in deposits should be proportionately smaller). This expectation is supported by the fact that inflation-indexed bonds had not been established in Greece yet (Garcia and Rixtel, 2007), while inflation would divert investors from securities to real estate according to theory (Fama and Schwert, 1977). Indeed, during that period 8 out 12 banks experienced a decline in productivity due to a substantial decrease in securities, with 6 of these cases being significant.

The following year we observe that the moderation efforts were successful as inflation declined to some extent and at the same time the real GDP growth achieved a 3.1% rate. The tax on interest is also implemented, which contributed to the decrease in deposits as depositors sought alternative options for their money which can probably explain the observed increase in securities. At the same time the deregulation process almost completes during 1991 and 1992, allowing banks to reallocate their inputs more productively. Indeed, we evidence a substantial increase in securities for most banks while 7 out of 13 banks experience an increase in productivity, with 6 cases being significant.

By the end of 1992 the deregulation process has almost been completed with the final important commercial freedoms provided to banks. We observe that banks experience a decrease in all inputs and loans, with the former being greater. On the other hand, securities remained almost the same in most cases, though higher for bigger banks. The author did not find a justification for the observed behaviour in the commercial banks' reports. One possibility is that the final wave of deregulation allowed banks to reallocate their inputs in more productive sources; in this case probably securities. The effect on productivity was overall positive and significant as documented by the results for the Average and for 8 out of 14 commercial banks during that period, including the biggest banks in the sample (in 5 cases the increase was significant).

The deregulation wave was followed by reregulation in 1993, which is a common pattern in the literature (Matthews and Thompson, 2014). The imposition of prudential controls after the introduction of the Basel rules on capital definition and liquidity along with the introduction of financial accounting standards suggested that banks would need to use more inputs and produce their outputs under stricter supervision. Since Basel I focused mainly on credit risk and the risk-weighting of assets for regulatory purposes (to compute the necessary capital ratios) we would expect that banks would reconsider their securities. In fact we observe a decrease in securities along with a small increase in personnel expenses (perhaps, to some extent, due to the higher resourcing requirements for compliance), leading to a decrease in productivity for 13 out of 17 banks, with the decline being significant in 10 cases.

In the first three years following the sector reforms, we document a substantial improvement of the macroeconomic environment with inflation dropping down to 5.5% and the real GDP growth rates averaging 2.7%. The good environment is also reflected in the productivity of Greek banks which, on average, experienced an overall increase in productivity. The results for the Average bank document two small decreases in productivity, followed by a substantial increase. Comparing the bias-corrected efficiency scores between 1994 and 1997 we observe that 12 banks improved their performance during that period whereas 6 documented a fall[147]. The prevailing change in inputs and outputs that explains this behaviour is the relative increase in loans compared to other inputs. Securities exhibit a small decline, fixed assets and deposits are not very volatile while personnel expenses seem to increase (perhaps to serve the extra demand for loans). This overall increase in productivity, supported by the improving conditions of the environment is a pattern also documented in Humphrey and Pulley (1997) for the case of the US deregulation during the 80s.

During 1997-1999 the macroeconomic environment keeps improving, while the bullish market of the Athens stock exchange reaches its peak in late 1999. Inspecting the results for the Average Bank we deduce that during the last two years, banks experienced an increase in both inputs and outputs, which can be attributed to the decreasing interest rates and the possibilities opening up in the view of joining the EU in the near future. The Greek banking sector enters an M&A wave during that period and it is followed by a bullish stock market in 1999. Hence the effects of the deregulation process may be mixed with the effects of the aforementioned events. It seems that in some cases banks increased their inputs (mainly deposits) proportionately more compared to their outputs which is captured as a decline in productivity. In 1998 to

[147] These were: Agricultural Bank, Attica Bank, Bank of Central Greece, Ionian and Popular bank and Emporiki Bank. We note that these banks were acquired in the future by other Greek banks.

1999 we observe, though, that outputs increase by more compared to inputs which might be due to the improving business conditions, combined with the bullish market.

Overall, we observe that the Greek banking sector follows the theoretical pattern that bank productivity increases after deregulation and tends to decrease after the imposition of controls (Matthews and Thompson, 2014). After about 3 years from the end of the (de)regulation period (by 1997), Greek banks seem to recover on average, supported by the good market conditions which is a pattern also evidenced by Humphrey and Pulley (1997) for the US deregulation of the early 80's. Moreover, we observe that commercial banks experience in most cases a decline in productivity during 1993-1994 and in 1994-1995, while productivity increased in most cases over the next two years which is a pattern similar to that in Tsionas et al. (2003).

6.6 Conclusions

In this chapter we provided an illustrative example of the methods discussed earlier under the scope of the Greek banking sector reforms era. In particular, we examined the effect of each step of the deregulation and reregulation process on the efficiency and productivity of Greek banks by applying bootstrap DEA on a pooled sample of commercial banks. In particular we used for our analysis the moments bootstrap which was shown in the previous chapters to perform well in small samples and we compared the results with other smoothing techniques. Quantitatively, the magnitudes of the estimated productivity changes and the associated confidence intervals are different to some extent. However, the qualitative result is the same; the productivity of Greek banks tends to increase after the provision of commercial freedoms whereas the imposition of controls seems to have the opposite effect.

Throughout our analysis we followed the suggested guidelines in Chapter 5. The first important action was to highlight issues of technological heterogeneity which led to the exclusion of certain banks from the sample which did not exhibit the usual characteristics of commercial banks. The resulting sample is technologically homogeneous both with respect to the cross section and the panel, which is supported by the fact that efficient DMUs are scattered across time periods and that they form one homogenous cluster when we looking at the different input/output scatterplots

This homogeneity allows for the implementation of bootstrap DEA under the moments bootstrap and the hypothesis testing procedures described in Chapter 3. In particular, the resulting sample has an almost half-normal distribution of efficiency scores which was shown in the previous simulations to be associated with good performance for samples of size 120 or more. The diagnostics have also confirmed that the underlying DGP exhibits technological homogeneity and it is a mixture of what we called in Chapter 2 “Standard” and “Trun.Normal High”.

The results on the technical efficiency of Greek banks suggest that size is a key success factor as large banks tend to be more efficient compared to smaller ones; a pattern which holds well across the whole study period. This lends support to the SCP paradigm for Greek banking; at least for the period examined. At the same time, Greek banks seem to have exploited their economies of scale as the median scale efficiency is 0.989 with the few exceptions of mainly smaller banks lying on the lower tail. These findings do not seem to be in accordance with those of Christopoulos et al. (2002) who find that big banks are more cost-inefficient. However, this difference can be justified by the different sample and method used, but it certainly makes us consider that the patterns that we observe for the particular period might not be the same for all time periods.

The sensitivity analysis of DEA scores through bootstrap DEA reveals that the moments bootstrap has produced more symmetrical bootstrap distributions compared to the other approaches (at least in the case examined here). In fact, the naïve bootstrap exhibits high skewness to the right and kurtosis and its distribution resembles a peaked half-normal distribution with the tail to the right. The other two smoothing approaches (LSCV and SJ) are less skewed compared to the naïve bootstrap but more skewed compared to the moments-bootstrap, making the smooth bootstraps to look like a mixture between the naïve bootstrap and the moments bootstrap. If we accept that the moments bootstrap is the "benchmark" due to its good behaviour in the simulations, then this extra asymmetry may be associated with the inferior performance of the two smoothing alternatives. We also find that the associated confidence intervals for the moments bootstrap are slightly wider in this example, which could be another reason for its superior performance and we would not perceive it as a weakness.

Then we wanted to examine the effects of sector reforms on the Greek banking sector by implementing the relevant test in subsection 3.3.2 and we therefore computed the ratios of the bootstrapped efficiency scores for adjacent periods and for each bank. And given that the efficiency scores were computed using a global frontier, the resulting ratios can be considered as Pastor and Lovell's (2005) Global Malmquist indices. The hypothesis testing results under all approaches provide the same qualitative result: deregulation improves productivity while reregulation deters it (Casu et al., 2004; Pariouras et al. 2009, Matthews and Thompson, 2014). However, under the moments bootstrap we find that the null is not rejected in 6 cases whereas there was another case where the null was rejected under the moments bootstrap but not under the LSCV bootstrap. And if we accept the moments bootstrap as the benchmark, this could be perceived as an indication of Type I error for the other two bootstraps. It is interesting to note at this point that the naïve bootstrap has provided the same hypothesis testing outcomes as the other two smoothing approaches.

Looking further into the productivity changes over time we observe that the changes were driven by big banks, which is probably due to the fact that the Greek banking sector is highly concentrated. To some extent this lends support to the studies of Altunbas et al. (1999, 2001) across 15 countries and during 1989-1997. The lagged response also confirms the relevant suggestion by Orea (2002), Siriopoulos and Tziogkidis (2010) and Delis et al. (2011). The analysis of the banks' input-output trajectories was not necessarily conclusive but the strongest patterns seem to be the expansions and contractions in securities, as well as the simultaneous increases or decreases of all inputs and outputs. In fact, during the period 1994-1997 we observe an expansion in banks' activities and an overall increase in productivity, on average, which was supported by the enhancing business environment in Greece. This finding is similar to the one of Humphrey and Pulley (1997) on the post-deregulation period of US banks during the 80s.

There is a clear message from this study which we suggest to be taken into account by regulators and policy makers. The imposition of prudential controls on Greek banks will probably reduce the productivity of the already unstable Greek banking sector. When the Basel regulations where implemented in Greece we observed a decrease in securities, as banks had to reconsider their portfolios along with a small increase in personnel expenses, perhaps due to the higher resourcing requirements for compliance. In the view of the closer supervision under the ESM and to the stricter capital requirements, as well as combined with the bad business environment in Greece and the increase in “red” loans, we expect a negative impact on Greek banks in the short run. The author believes that authorities should make sure that the imposition of new controls will not come at extra costs for the banks, especially the ones in distress, in order to ensure a smoother transition towards ECB supervision.

The limitation of this study is the fact that due to sample size restrictions it was not possible to decompose the Global Malmquist index to its components. Hence, it is not clear whether these changes in productivity were due to changes in efficiency or technology. One possibility would be to use a mixture of the Global Malmquist approach and window analysis, however it is not clear if this would solve more problems than create and it is left for future research. We also note that an informal analysis of post 2000 data (not discussed here) shows a change in the patterns observed which might suggest that an approach to account for global frontier shifts, such as Asmild and Tam (2007) would be relevant. In terms of future methodological research, the author believes that there is scope for development of a bootstrap approach on the Global Malmquist index and its decompositions which poses the challenge of using an appropriate smoothing approach.

7 Thesis Conclusions

The thesis has explored the performance of Simar and Wilson's (1998) bootstrap DEA through Monte Carlo simulations and has proposed a modification which makes it applicable in small samples. It has also suggested guidelines for the implementation of bootstrap DEA and hypothesis testing and it has performed an empirical illustration on the Greek banking case. The theoretical explorations have highlighted the importance of the assumption of equal bootstrap and DEA biases for the accuracy of the constructed confidence intervals and, to its extent, for hypothesis testing. Our simulations have indicated that kernel density estimation techniques, used in the seminal paper of Simar and Wilson (1998) and in other developments or extensions, might indeed introduce additional noise (Simar and Wilson, 2002) and contribute towards the violation of the aforementioned assumption. The proposed alternative to smoothing performs better in our simulation towards this respect, justifying the higher coverage probabilities observed. The empirical application indicates that these differences might be reflected in slightly different confidence intervals and shapes of the bootstrap distributions; though the overall qualitative result seems to be the same across all methods. In the sections that follow we discuss the main findings of the thesis, we highlight its limitations and we propose avenues for future research.

7.1 Thesis summary and discussion

The theoretical explorations of the thesis concerned the analysis of the deterministic efficiency measurement technique DEA and its extension, bootstrap DEA, which allows for statistical inference. Bootstrap DEA, proposed by Simar and Wilson (1998), has been shown to be a consistent technique which uses the empirical distribution of DEA scores to generate bootstrap distributions of efficiency scores for each DMU. These distributions can be then used to construct confidence intervals which are supposed to cover the population efficiency score of the DMUs in the sample. The coverage probabilities of these confidence intervals seem to depend heavily on the extent to which the fundamental assumption of equal bootstrap and DEA biases is valid; this is shown both theoretically and through simulations.

The Monte Carlo simulations, which are the most extensive compared to others in the literature, use 4 data generating processes along with 2 different smoothing techniques and cover a range of sample sizes. The results indicate that, although bootstrap DEA is consistent and has nice asymptotic properties, it cannot be safely used with small samples due to the violation of the equal biases assumption. In our simulations we observe an interesting pattern for the confidence intervals of Simar and Wilson (1998, 2000a) which is also explained theoretically; the basic confidence intervals (Simar and Wilson, 2000a) only perform better than the percentile ones (Simar and Wilson, 1998) when the DEA bias is considerably greater than the bootstrap bias. In the Monte Carlo exercises this case was associated with populations that exhibit technological heterogeneity for which we argue that even the application of simple DEA might not be a good idea. We therefore propose that the bootstrap DEA confidence intervals based on the basic interval method (Simar and Wilson, 2000a) should not be preferred over the ones based on the percentile method (Simar and Wilson, 1998), or at least the use of the primer should be carefully considered. This finding carries implications for the later extensions of bootstrap DEA which make use of the basic confidence intervals, such as the bootstrap Malmquist index (Simar and Wilson, 1999), the tests on returns to scale (Simar and Wilson, 2002) or the two-stage bootstrap DEA (Simar and Wilson, 2007).

The investigation of the moments of the bootstrap DEA distributions of the "fixed DMU", which has been disregarded in the literature, has also offered some interesting insights on the behaviour of bootstrap DEA. Firstly, we find that these moments are similar to the moments of the distribution of the DEA scores of the fixed point generated by the various Monte Carlo samples. This could be considered as evidence that bootstrap DEA has the capacity to "mimic" the sampling variations of DEA scores as claimed by Simar and Wilson (1998), providing support to the validity of their method. Secondly, we observe that the greater the technological homogeneity of the population, the faster the standard deviation of the bootstrap distribution will be converging to zero. Given the fast declining bootstrap and DEA biases in these cases, this suggests that for large enough samples and "homogeneous" enough samples, the DEA scores are robust and approximately equal to their population value. Hence, the application of simple DEA is adequate in these cases as the resulting confidence intervals become very narrow. Finally, according to our simulations the bootstrap distributions appear as positively skewed and relatively leptokurtic (on average), which may be relevant when constructing confidence intervals. For example, in the presence of high skewness, as suggested by Simar and Wilson (1998), researchers may want to consider alternative confidence interval construction techniques, such as the bias-corrected intervals of Efron (1982), or they could consider the bias-corrected and accelerated confidence intervals of Efron (1987) proposed in Appendix VII.

The non-satisfactory small sample performance was further investigated in Chapter 3 and its implications for hypothesis testing were explained. In particular, we find that, apart from the low coverage probabilities, the violation of the equal biases assumption can be translated into Type I and II errors when testing hypotheses. Exploring alternatives in the presence of bias asymmetries (mainly on the basis of alternative confidence intervals) resulted in solutions which, although seemed to improve coverage probabilities, they did not exhibit converge to the nominal ones. Moreover, we showed how these asymmetries can affect the popular extension of testing for returns to scale with bootstrap DEA (Simar and Wilson, 2002) and we indicated a possible alternative that could be further looked into in the future.

On the other hand, we argue that when there are no bias asymmetries, bootstrap DEA could work well in small samples. In fact, we indicate how a range of hypothesis tests could be implemented and how p-values could be computed. This further motivated our search towards finding an approach that would make this assumption work and that could be used in practice with small samples. Given that the reduction of the bootstrap and DEA bias asymmetries is not related to confidence interval construction we decided to look into an alternative option; reconsidering the empirical distribution smoothing approach.

The reconsideration of kernel density estimation techniques was also motivated by the comment in Simar and Wilson (2002) that such approaches usually introduce additional variability. The considerably larger bootstrap biases compared to DEA ones in (relatively) technologically homogeneous processes can be deemed as evidence in support of this comment of Simar and Wilson (2002). The alternative approach proposed here uses the moments of the empirical distribution of DEA scores to generate pseudo-populations of efficiency scores from which draws can be performed for bootstrap DEA. Simulations have shown that the moments-bootstrap, as we named it, is associated with considerably lower asymmetry of bootstrap and DEA biases compared to the smooth bootstraps, resulting in coverage probabilities that converge to the nominal ones for samples of 120 observations and under a 2-input/2-output setup.

The theoretical explorations were summarised in a few suggested guidelines on the application of bootstrap DEA and its implementation on hypothesis testing. We emphasised the assumptions used in bootstrap DEA and the need to use the simple proposed diagnostics to identify the underlying data generating process, as it has implications for the performance and even the applicability of bootstrap DEA. On the same note we proposed the investigation of the technological homogeneity of the DMUs included in the sample and, where possible, to exclude from the analysis DMUs that seem to use different processes or have access to different technology. Once a “satisfactory” degree of technological homogeneity has been achieved, we suggested using the moments bootstrap along with the percentile method for constructing confidence intervals (Simar and Wilson, 1998) with samples of 120 observations or more, while we summarised the steps that could be followed for hypothesis testing.

In order to provide an empirical illustration of the theoretical findings we used data from Greek banks during 1987 to 1999, a period which is characterised by a long deregulation process followed by reregulation towards the standards of the Basel I accord. The choice of the data period is influenced by the fact that after 1999 the Greek banking sector is affected by a range of other events (stock exchange crisis, M&As, privatisations and the accession of Greece to the EU), making the long-run effects of the sector reforms no longer discernible. In fact Molyneux (2009) observes that after 2000, the EU banks exhibited different reaction to certain events, which is also confirmed for the Greek case by informal explorations by the author. Apart from the empirical and data contributions of this study, it is also of topical interest due to the current outlook of the Greek economy and the tightening of supervision through the European Supervisory Mechanism (ESM).

The methodological challenge in this application lies within the very small sample size for each year and the fact that the data panel is highly unbalanced. To overcome this issue we decided to use a global frontier approach and therefore the ratios of global efficiency scores for a certain DMU in adjacent periods is the Global Malmquist productivity index of Pastor and Lovell (2005). Then the implementation of the previously discussed hypothesis testing approaches in this case means that we can test for significant changes in productivity change as well as for the direction of that change.

When looking into the quantitative results we observe some differences among the approaches considered. In particular, when testing for productivity change and its direction, we find that the associated bootstrap distributions are different and therefore the associated confidence intervals and p-values are different, to some extent. To be precise, in our case the moments bootstrap is associated with more symmetrical distributions while the other two smooth alternatives and the naïve bootstrap have more skewed and leptokurtic distributions. This extra "symmetricity" could be considered as an explanation for the improved performance of the moments bootstrap in the previous Monte Carlo simulations, something that could be further investigated in the future.

The qualitative results, though, seem to be almost the same across the different approaches. In particular, in all cases we conclude that the provision of commercial freedoms increases the productivity of Greek banks the next year [lagged effects also in Orea (2002), Siriopoulos and Tziogkidis (2010) and Delis et al. (2011)] while the imposition of prudential controls has the opposite effect, which is in accordance with theory and evidence (Casu et al., 2004; Pariouras et al. 2009, Matthews and Thompson, 2014). We also find that these changes where driven by the larger banks, which is in accordance with the European studies of Altunbas et al. (1999, 2001), while the application of simple DEA indicates that larger banks across all time periods tend to be more technically efficient than smaller ones. This suggests that the overall performance of the highly concentrated Greek banking sector was driven by large banks, which seemed to be the leaders of the change.

In the longer term, Greek banks seemed to have experienced an overall improvement in productivity and on average over the next 3-4 years. This was supported by the enhancing economic outlook of Greece, a pattern also observed in Humphrey and Pulley (1997) for the US deregulation of the early 80s.

7.2 Policy implications

There is a useful policy implication from this exercise regarding the ongoing Greek debt crisis which has severely affected the Greek banking sector. The four biggest Greek banks entered the Single Supervisory Mechanism on the 4th of November, 2014. This recent change requires the direct supervision of these banks (and their subsidiaries) by the European Central Bank, tightening the prudential monitoring of those institutions. During the early 90s, when prudential controls were imposed on Greek banks, they had a negative productivity impact for the next 1-2 years, but resulted in an overall (and on average) increase in productivity over a 3-4 year horizon. The productivity increase was supported by the good business environment at that time, which is a pattern that was also observed by Humphrey and Pulley (1997) for the US banking case. This finding may suggest that changes in banking regulation, even if they are considered as "improvements", will not necessarily lead to higher efficiency and productivity in the short run, especially if the environment is "hostile".

Given that the prospects of the Greek economy are not promising and that the big Greek banks only marginally passed the recent stress tests, any further regulations might have a long-lasting negative impact on the productivity of Greek banks. Taking also into account the current rumours for a further haircut of the Greek debt, the overall impact on the Greek banking sector would be hard to manage. We therefore suggest that the imposition of any further controls to be gradual and that any potential changes in regulation to be announced well in advance to give time for banks to adjust their operations accordingly. For the same reasons, we believe that the fact that the entrance of the Greek banks into the SSM came after the end of their recapitalisation process, was a good move by the policymakers; either this was intentional or not.

7.3 Limitations and future directions

The findings from our theoretical and empirical analysis come along with some limitations that have been explained in the previous chapters. Accordingly, suggestions for future research have been also proposed to address these issues as well as to suggest alternative avenues that could be considered in the future. Here we discuss what we consider to be the most important ones.

One of the limitations of the study relates to the fact that the simulations, despite being the most extensive so far in the literature, they are not exhaustive. Despite the fact that we observe common patterns arising, it would require further simulations to allow us to generalise the conclusions derived from our observations. The Monte Carlo simulations involved 4 data generating processes, 7 different sample sizes ranging from 10 to 120, 3 alternative approaches to smoothing (including the moments bootstrap) as well as the assumption of CRS and input orientation. Given the suggestions in the literature that when smoothing is involved a number of alternative simulation setups should be considered (Silverman and Young, 1987), we propose extending the simulations to account for as many possibilities as possible. To this end, future research could look into output orientation and VRS as well as alternative DGPs, since the resulting shapes of the bootstrap distributions might be different.

The additional simulations could also look at the extent to which we can generalise our suggestion that the percentile intervals of Simar and Wilson (1998) should be preferred over the “basic” ones of Simar and Wilson (2000a). Our simulations have shown that the latter perform better (yet not adequately well) only in the presence of technological heterogeneity, where even simple DEA might not be a good idea to use. It is interesting to examine whether the same result will be reached with alternative simulations as, apart from allowing us to generalise this finding, it would also pose questions on the performance of extensions of bootstrap DEA that make use of the latter intervals.

Another fact, which is only partially a limitation, is that, on average, bootstrap distributions are associated with positive skewness, despite the fact that it tends to decrease with sample size. Given the fact that skewness has received some attention in the literature, it may be the case that confidence intervals that account for skewness might be more suitable in the presence of high skewness. For example the Efron's (1982) bias-corrected intervals, proposed by Simar and Wilson (1998), might be appropriate or the Efron's (1987) bias-corrected and accelerated intervals, proposed in Appendix VII in this thesis, might be relevant. In any case, it seems that there is research potential on the issue of the effect of skewness on the performance of bootstrap DEA and simulations could reveal the extent to which alternative confidence intervals would perform better. On the same logic, the effect of kurtosis could also be investigated and ideally linked to certain types of data generating processes.

Given the importance of the potential underperformance of bootstrap DEA and the importance of the unequal bootstrap and DEA biases towards this direction, it seems reasonable to propose the further investigation of the causes of such asymmetries. Our simulations have suggested clearly that the higher the DEA bias the greater will be the degree of technological heterogeneity, identified visually by histograms with a thin tail towards 1. However, we could not necessarily identify what causes the bootstrap bias to be greater than the DEA bias or vice versa. We suspect that the variability in the DGP or the smoothing processes used might be associated with this issue. In any case, a focused study on the causes of bias asymmetries and their identification from sample data would be useful. The author believes that the iterated bootstrap might be promising towards this direction but the extremely high computational costs make it almost impossible to assess its performance through Monte Carlo simulations (they would currently require several months to run, if not years).

The alternative approach to smoothing, the moments bootstrap, is also associated with some limitations that need to be further examined. In particular, our simulations suggested that the associated confidence intervals are slightly narrower, on average, when the underlying DGP is associated with technological homogeneity. On the other hand, the introduction of technological heterogeneity seems to make the intervals wider, though to a small extent. Despite that fact that narrower confidence intervals seem more “attractive”, the high coverage probabilities under the moments bootstrap seem to suggest otherwise; besides, as sample size increase the differences in widths seem to become very small. It would be therefore reasonable to propose for future research an in-depth investigation of the relationship between certain DGPs and the bootstrap distributions, which would explain why the resulting intervals are narrower or wider.

The final limitation, which we would like to point out here, concerns our empirical illustration. The small number of observations per year posed a methodological challenge that we tried to mitigate with the consideration of a global frontier. The resulting ratios for the implementation of the required hypothesis tests (discussed in Chapter 3) were actually the Global Malmquist indices of productivity change of Pastor and Lovell (2005). Due to sample size issues we could not decompose the indices to efficiency change and technical change, while, even if sample size was not an issue, such decomposition would pose other methodological challenges. In particular, it would require an approach to maintain the correlation structure of the local efficiency scores (the usual DEA scores) between adjacent periods, since they are required for these decompositions. Simar and Wilson (1999) dealt with this issue when they proposed bootstrapping the Malmquist index, by introducing a bivariate kernel from which draws could be performed for two adjacent periods, taking into account the correlations between the two samples. Apart from the fact that this method might not be relevant in our case, we have shown in Appendix XI that the approach of Simar and Wilson (1999) is not fully compatible with unbalanced panels, despite the fact that the *FEAR* software manual suggests otherwise. In particular, although results can be obtained, the smoothing process disregards the non-common elements which may have serious implications for the shape of the empirical distribution and hence for smoothing.

Future research could also consider the implementation of the bootstrap on Asmild and Tam's (2007) approach of Global Frontier shifts. This approach might be relevant in our case since the accession of Greece to the EU seems to be a structural break for the operations of Greek banks, as was the case for other European sectors (Molyneaux, 2009). Hence the extension of our dataset after 1999 with the implementation of the bootstrap on the approach of Asmild and Tam (2007) seems to be an interesting extension.

As a final note, the author would like to point out that there seems to be a lot of room for future research on bootstrap DEA. Our explorations have indicated that our understanding of how these methods work could be expanded by additional simulations. Theoretical explanations would also be useful to indicate “when can be used what” and create a manual for bootstrap DEA with general applicability. Therefore, future research on bootstrap DEA could invest some efforts towards further improving the practical understanding of the existent approaches before moving to new ones.

I. Appendix I: Smoothing methods

A. Kernel density estimation

The purpose of density estimation is the determination of a functional form that mimics the empirical distribution of data. In particular, it uses the sample distribution to estimate the kernel of the density function which best approximates the asymptotic characteristics of the underlying population. Silverman (1986) provides a solid review of density estimation methods.

The simplest method of "estimation" is the inspection of histograms, however it lacks precision. To construct a histogram, data are sorted and plotted by using a certain "bin-width" h, which determines the width of the histogram intervals. Therefore, each histogram bar includes the frequency of observations that belong in a certain interval. More formally:

$$\hat{f}_h(t) = \frac{1}{nh}\{\text{number of } \hat{\theta}_i\text{'s that belong in the same bin as t}\} \tag{I.1}$$

where n is the number of observations, $\hat{\theta}_i$ is the estimated DEA efficiency score while the subscript h in $\hat{f}_h$ denotes that the estimated density depends on the bin-width.

The most popular alternative is kernel density estimation, which uses a kernel estimator from a popular distribution (usually a symmetric one) along with an appropriate smoothing parameter (or bandwidth or window width) h which determines the closeness of the estimated density to data. The estimated kernel is determined by:

$$\hat{f}_h(t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-\hat{\theta}_i}{h}\right) \tag{I.2}$$

where $K(\cdot)$ is the kernel estimator used. Obviously, $\hat{f}(t)$ is a probability density with the same continuity and differentiability properties with those of the kernel estimator used (Silverman, 1986). If the kernel estimator is a standard normal one, we have from (I.2):

$$\hat{f}_h(t) = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\hat{\theta}_i}{h}\right)^2} \tag{I.3}$$

However, in the case of efficiency score distributions the distribution is bounded at 1, which is incompatible with this approach.

Simar and Wilson (1998) propose, for the case of DEA, to use a standard normal density and to reflect the efficiency scores of inefficient DMUs about 1 which creates a compatible symmetric distribution. Hence, the kernel of efficiency scores between zero and one will be the mirror image of the kernel of reflected scores (between 1 and 2). In particular, if t_i is a random variable which is defined on the $(0,1)$ interval and t_i^R is its reflected value on the $(1,2)$ interval, then due to symmetricity we have:

$$P\left(\hat{\theta}_i < t_i < 1\right) = P\left(1 < 2 - t_i^R < 2 - \hat{\theta}_i\right), \qquad t_i^R = 2 - t_i \tag{I.4}$$

Equation (I.4) states the obvious: that the probabilities are symmetric about one. The same is valid for the tails, that is:

$$P\left(0 < t_i < \hat{\theta}_i\right) = P\left(2 - \hat{\theta}_i < t_i < 2\right) \tag{I.5}$$

and by standardizing we have:

$$P\left(-\frac{\hat{\theta}_i}{h} < \frac{t_i - \hat{\theta}_i}{h} < 0\right) = P\left(0 < \frac{t_i - 2 + \hat{\theta}_i}{h} < \frac{\hat{\theta}_i}{h}\right) \tag{I.6}$$

The result in (I.6) implies that the aggregated kernel on the (0,2) interval can be defined by the following expression (Silverman and Young, 1987; Simar and Wilson, 1998):

$$\hat{g}_h(t) = \frac{1}{2nh}\sum_{i=1}^{n}\left[K\left(\frac{t-\hat{\theta}_i}{h}\right) + K\left(\frac{t-2+\hat{\theta}_i}{h}\right)\right]$$
$$= \frac{1}{2nh}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-\hat{\theta}_i}{h}\right)^2} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-2+\hat{\theta}_i}{h}\right)^2}\right] \tag{I.7}$$

This is simply the average of the kernels implied by (I.6), which is therefore equivalent to the following expression (Simar and Wilson, 1998):

$$\hat{l}_h(t) = \begin{cases} 2\hat{g}_h(t) = \dfrac{1}{nh}\displaystyle\sum_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-\hat{\theta}_i}{h}\right)^2} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-2+\hat{\theta}_i}{h}\right)^2}\right], & \text{if } t \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{I.8}$$

Hence if $t \leq 1$ we attach a double weight on the density since it has the same probability for $t \geq 1$ resulting from symmetricity and we attach a zero density in the latter case. Thus, the reflected density is reflected back to the $(0,1)$ interval.

To illustrate how this can be implemented in the case of bootstrap DEA, suppose that t_i in (I.6) is determined by the following process:

$$t_i = \beta_i^* + h\varepsilon_i^* = t_i \leq 1, \quad i = 1,2, \ldots n \tag{I.9}$$

where β_i^* is a random resample from the empirical distribution of efficiency scores and ε_i^* is a standard normal error. Using (I.4) and (I.9) we have for the reflected values:

$$t_i^R = 2 - t_i = 2 - \beta_i^* - h\varepsilon_i^* \tag{I.10}$$

Obviously, the expected value of (I.9) and (I.10) is the DEA score or the reflected DEA score (since β_i^* is their random resample) and the standard deviation is equal to h.

Hence, the kernels of the standardised values of t_i correspond to the ones in (I.7), that is:

$$t_i \sim \hat{g}_{1,h}(t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_1 - \hat{\theta}_i}{h}\right) = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t_1-\hat{\theta}_i}{h}\right)^2} \tag{I.11}$$

and

$$t_i^R = 2 - t_i \sim g_{2,h}(t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_2 - 2 + \hat{\theta}_i}{h}\right) = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t_2-2+\hat{\theta}_i}{h}\right)^2} \tag{I.12}$$

Thus, averaging over (I.11) and (I.12) we get (I.7).

From (I.9)and (I.10) we can define the following sequence:

$$\tilde{t}_i^* = \begin{cases} \beta_i^* + h\varepsilon_i^*, & \text{if } \beta_i^* + h\varepsilon_i^* \leq 1 \\ 2 - \beta_i^* - h\varepsilon_i^*, & \text{otherwise} \end{cases} \tag{I.13}$$

The distribution of the sequence $\tilde{t}_i^*$ is distributed as $\hat{l}_h(t)$ in (I.7) as $\tilde{t}_i^* \leq 1$ (Simar and Wilson, 1998) and it can be used to translate the reflected resample of DEA scores into a smoothed resample of non-reflected scores.

To summarize, we have shown how Simar and Wilson (1998) use the empirical distribution of efficiency scores to estimate the kernel in (I.8) which will be used to produce the bootstrap sample of pseudo-efficiency scores. We still need to determine how the smoothing parameter, h is defined.

B. Choice of the smoothing parameter

The estimated density is sensitive towards the choice of the smoothing parameter, h. Low values of h give rise to spurious and under-smoothed estimated densities, hence gaining in precision but lacking in structure (or variability). On the other hand high values cause over-smoothing which leads to the exclusion of potentially interesting variation; therefore, the gain in terms of capturing the asymptotic feature of the empirical distribution is offset by the higher estimation bias for the observed sample. It is obvious that there is a trade-off between bias and variance in density estimation hence bandwidth selection (that is, the choice of h) should take this into account.

The appropriate choice of the smoothing parameter is a long debated topic in the literature which is divided in two main streams: cross validation (or first generation methods) and "plug-in" (or second generation methods). Their common goal is to minimize a measure of distance of the estimated and the true density, usually being the mean integrated square error (MISE):

$$MISE(\hat{l}_h) = \int MSE\,(\hat{l}_h)dt = \int E\{\hat{l}_h(t) - l(t)\}^2 dt \tag{I.14}$$

which, according to Silverman (1986) can be proven to be:

$$MISE(\hat{l}_h) = \int \{E[\hat{l}_h(t)] - l(t)\}^2 dt + \int var[\hat{l}_h(t)]dt \tag{I.15}$$

where

$$E[\hat{l}_h(t)] = \int \hat{l}_h(t) f(t) dt \tag{I.16}$$

and

$$n \cdot var\left[\hat{l}_h(t)\right] = E\left[\hat{l}_h(t)^2\right] - E\left[\hat{l}_h(t)\right]^2 = \int \left[\hat{l}_h(t)\right]^2 f(t)dt - \left[\int \hat{l}_h(t)f(t)dt\right]^2 \qquad \text{(I.17)}$$

Density estimators are assessed in terms of their asymptotic convergence, which is done using the asymptotic MISE[148] (or AMISE). The problem is that MISE and AMISE cannot be estimated directly as the probability density function $f(t)$ in the expectations term (I.16) is not observed. Hence, different approaches are followed in the literature to perform this task.

First generation methods include "rules of thumb", least squares cross validation (LSCV), likelihood cross validation (LCV) and biased cross-validation (BCV). Among these methods the best performing one is LSCV, which is obvious in the simulations in Park and Marron (1990), Jones et al. (1996) and Loader (1999). The general idea behind LSCV, introduced by Rudemo (1982) and Bowman (1984), is to minimize the integrated squared error (ISE) with respect to the smoothing parameter, which should also be the minimizing value for MISE:

$$ISE(h) = \int \left(\hat{l}_h(t) - l(t)\right)^2 dt = \int \hat{l}_h^2(t)dt - 2\int \hat{l}_h(t)l(t)dt + \int l^2(t)dt \qquad \text{(I.18)}$$

where the second term is estimated using "leave-one-out" cross validation.

Second generation methods include, among others, "plug-in" methods, which seem to be quite popular. "Plug-in" methods involve expressing the error of the estimated density in terms of the unknown density and approximating it using Taylor series expansions (Loader, 1999). In particular, both the MISE and the optimal h depend on the

[148] This is a quite complicated issue and describing the details of AMISE or providing further details about MISE is not within the scope of this study. For an introduction on these concepts, the interested reader can look at chapter 3 in Silverman (1986).

integral of the second derivative of the unknown density. Then a pilot (kernel) estimate of the second derivative is used where a certain relationship between the estimated bandwidth and the pilot bandwidth is assumed. And the resulting estimated error approximates MISE. Many "plug-in" approaches have been proposed but the benchmark seems to be these of Park and Marron (1990) and Sheather and Jones (1991), the latter estimator known as *SJPI*, standing for Sheather-Jones Plug-In.

Second generation methods provide an optimum trade-off between error and variance, in contrast to LSCV which focuses in approximating MISE at the cost of excess variance. Models like SJPI introduce much less variance while they achieve a much faster rate of asymptotic convergence. However, they produce meaningful results only when the density to be estimated is already smooth enough. In the opposite case, the estimated density is not a good representation of the actual one (actually it is oversmoothed) and approaches based on "plug-in" techniques, like SJPI, should not be used; LSCV would provide by far more consistent results.

In fact, Park and Marron (1990), when they introduced their popular "plug-in" method, compared simulation results from using their method against different methods and different data sets and state that:

"The main result is that, under strong enough smoothness assumptions on the underlying density, the plug-in bandwidth will dominate in the limit. Nevertheless, there is some trade-off for this, which is caused by the fact that for small amounts of smoothness least squares cross-validation is the most effective"

In one of the simulation exercises they find that LSCV provides unreasonable answers, which is justified by the small-scale clustering in the data, implying that if there is distinct clustering then it may be preferred to use LSCV.

In a comparison study, Jones et al. (1996) review bandwidth selection methods and argue that although LSCV provides the best centring in terms of the distribution of the smoothing parameter, it is associated with excess variability, hence with undersmoothing. Also, they argue that the asymptotic rate of convergence is very slow, in that it would require an enormous amount of data to ensure asymptotic convergence. The authors conclude that in the case of smooth densities new generation methods perform better; however, if there is substantial variability in the density it is implied that LSCV performs better, although it is not clearly stated in the paper.

Loader (1999), argues that plug-in approaches are subject to criticism of arbitrary selection of pilot estimators and that they introduce too much smoothing when dealing with complex problems. However, second generation methods can capture the main trend (in the sense of capturing the asymptotic distribution) and introduce significantly less noise in the kernel estimation. Loader (1999) also tries to address the criticism of excess variability and under-smoothness on the classical methods. In particular Loader states that:

> *"We argue that variability of cross validation is not a problem but a symptom of the difficulty of bandwidth selection. Less variable bandwidth selectors display this difficulty in another way: consistently oversmoothing when presented with problems with small and difficult to detect features."* (Loader, 1999; pp. 417)

Loader (1999) also performs comparisons based on simulations among a set of bandwidth selection methods: Akaike information criterion (AIC), LCV, BCV, LSCV and SJPI. In two distinct examples they highlight the superiority of LSCV and SJPI. In particular, in the first exercise LSCV fails to clearly capture the bi-modal nature of the data set while SJPI achieves the best performance. However, in the second exercise, where the density to be estimated is a multi-modal claw density, SJPI completely fails to capture the behaviour of the density, while LSCV achieves a very good approximation. Loader concludes that there is no distinct superiority between first and second generation methods of bandwidth selection; rather, that each has its advantages and disadvantages and the method used should be carefully chosen, depending on the nature of the data set.

To sum up, when dealing with "hard-to estimate" densities in the sense that data do not follow a smooth distribution, LSCV provides much better results although it introduces excess variability. However, when the density to be estimated is smooth enough, SJPI provides better asymptotic results and LSCV does not provide enough smoothness.

C. Obtaining smoothed bootstrap samples

This section provides more detail on the procedure followed by Simar and Wilson (1998) to obtain a set of smooth pseudo-efficiency scores $\theta_i^*, i = 1,2 \dots n$, as mentioned in Step 2, in section 2.6.

The first step is to use the empirical distribution to determine the smoothing parameter by also assuming a functional form for the kernel estimator, which is the one in (I.8). They apply an appropriate technique to approximate the smoothing parameter and then they correct it for sample size using the following expression:

$$h = h\left(\frac{m}{n}\right)^{1/5} \tag{I.19}$$

Then, they use the sequence in (I.13) to transform the bootstrap resample according to the stochastic properties defined by the estimated kernel. Finally, they correct for variance and they obtained the set of smoothed pseudo-efficiency scores using:

$$\theta_i^* = \bar{\beta}^* + \frac{1}{\sqrt{1 + h^2/\hat{\sigma}_\theta^2}}\left(\tilde{t}_i^* - \bar{\beta}^*\right) \tag{I.20}$$

where $\bar{\beta}^*$ is the average of the re-sampled (with replacement) DEA scores and $\hat{\sigma}_\theta$ is the standard deviation of the DEA scores. The vector θ_i^* is the one that we wish to obtain in (2.20).

II. Appendix II: Coverage probabilities

Table II.1. Coverage of Simar and Wilson's (1998) confidence intervals: "Standard" case

	Standard 1/1				Standard 2/1				Standard 2/2			
Cov. LSCV	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.261	0.409	0.575	0.916	0.425	0.587	0.743	0.896	0.425	0.587	0.743	0.896
n = 15	0.214	0.310	0.409	0.719	0.284	0.420	0.574	0.854	0.284	0.420	0.574	0.854
n = 20	0.178	0.248	0.319	0.541	0.253	0.363	0.473	0.761	0.253	0.363	0.473	0.761
n = 25	0.137	0.191	0.243	0.398	0.239	0.341	0.421	0.657	0.239	0.341	0.421	0.657
n = 30	0.140	0.199	0.254	0.361	0.201	0.269	0.342	0.538	0.201	0.269	0.342	0.538
n = 60	0.091	0.131	0.167	0.234	0.128	0.176	0.226	0.313	0.128	0.176	0.226	0.313
n = 120	0.085	0.119	0.138	0.192	0.082	0.114	0.148	0.208	0.082	0.114	0.148	0.208
Cov. SJ	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.330	0.577	0.814	0.922	0.499	0.695	0.817	0.865	0.522	0.718	0.830	0.857
n = 15	0.227	0.417	0.603	0.962	0.438	0.597	0.771	0.927	0.416	0.582	0.764	0.929
n = 20	0.159	0.295	0.448	0.877	0.336	0.507	0.679	0.933	0.322	0.477	0.670	0.919
n = 25	0.121	0.219	0.341	0.753	0.248	0.397	0.574	0.892	0.256	0.394	0.566	0.891
n = 30	0.079	0.163	0.257	0.595	0.220	0.340	0.472	0.837	0.189	0.322	0.466	0.809
n = 60	0.022	0.051	0.083	0.212	0.075	0.118	0.188	0.390	0.059	0.104	0.165	0.374
n = 120	0.002	0.009	0.017	0.058	0.008	0.020	0.033	0.093	0.009	0.013	0.022	0.084
Cov. Naïve	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.551	0.570	0.570	0.570	0.332	0.332	0.332	0.332	0.346	0.346	0.346	0.346
n = 15	0.586	0.588	0.590	0.590	0.378	0.378	0.379	0.379	0.404	0.405	0.405	0.405
n = 20	0.585	0.585	0.588	0.588	0.386	0.386	0.386	0.386	0.402	0.403	0.403	0.403
n = 25	0.576	0.576	0.579	0.579	0.434	0.434	0.434	0.434	0.430	0.430	0.430	0.430
n = 30	0.570	0.570	0.572	0.572	0.415	0.416	0.417	0.417	0.475	0.475	0.475	0.475
n = 60	0.571	0.571	0.573	0.573	0.435	0.437	0.437	0.437	0.457	0.459	0.459	0.459
n = 120	0.557	0.561	0.561	0.561	0.409	0.410	0.410	0.410	0.418	0.424	0.424	0.424

Table II.2. Coverage of Simar and Wilson's (2000) confidence intervals: "Standard" case

	Standard 1/1				Standard 2/1				Standard 2/2			
Cov. LSCV	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.235	0.277	0.331	0.445	0.383	0.491	0.563	0.690	0.383	0.491	0.563	0.690
n = 15	0.175	0.210	0.242	0.301	0.260	0.333	0.401	0.537	0.260	0.333	0.401	0.537
n = 20	0.128	0.159	0.187	0.233	0.225	0.283	0.325	0.420	0.225	0.283	0.325	0.420
n = 25	0.107	0.125	0.145	0.183	0.218	0.271	0.302	0.372	0.218	0.271	0.302	0.372
n = 30	0.105	0.120	0.135	0.158	0.175	0.216	0.253	0.313	0.175	0.216	0.253	0.313
n = 60	0.062	0.072	0.080	0.086	0.113	0.140	0.151	0.174	0.113	0.140	0.151	0.174
n = 120	0.064	0.074	0.075	0.081	0.074	0.089	0.094	0.107	0.074	0.089	0.094	0.107
Cov. SJ	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.260	0.317	0.382	0.487	0.438	0.549	0.636	0.761	0.492	0.578	0.649	0.788
n = 15	0.179	0.220	0.255	0.325	0.363	0.461	0.519	0.627	0.366	0.451	0.498	0.609
n = 20	0.125	0.151	0.176	0.228	0.281	0.355	0.404	0.490	0.293	0.341	0.393	0.486
n = 25	0.091	0.112	0.135	0.165	0.221	0.274	0.305	0.384	0.216	0.270	0.315	0.391
n = 30	0.056	0.075	0.086	0.111	0.183	0.228	0.270	0.321	0.157	0.202	0.227	0.302
n = 60	0.014	0.020	0.023	0.028	0.063	0.076	0.090	0.108	0.045	0.062	0.079	0.099
n = 120	0.001	0.002	0.003	0.003	0.006	0.007	0.009	0.013	0.005	0.007	0.009	0.010
Cov. Naïve	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.742	0.779	0.871	0.933	0.487	0.623	0.723	0.866	0.480	0.612	0.727	0.853
n = 15	0.763	0.767	0.874	0.953	0.517	0.656	0.778	0.904	0.556	0.669	0.771	0.893
n = 20	0.741	0.741	0.867	0.941	0.551	0.690	0.782	0.910	0.531	0.660	0.763	0.909
n = 25	0.755	0.757	0.874	0.954	0.542	0.684	0.776	0.904	0.574	0.701	0.791	0.910
n = 30	0.761	0.765	0.885	0.955	0.553	0.698	0.790	0.912	0.606	0.724	0.809	0.926
n = 60	0.745	0.774	0.873	0.952	0.594	0.729	0.827	0.922	0.589	0.715	0.809	0.916
n = 120	0.709	0.750	0.858	0.953	0.541	0.683	0.786	0.930	0.579	0.727	0.841	0.938

Table II.3. Coverage of Simar and Wilson's (1998) confidence intervals: "Trun. Normal Low" case

	Trun. Normal Low 1/1				Trun. Normal Low 2/1				Trun. Normal Low 2/2			
Cov. LSCV	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.370	0.409	0.435	0.467	0.303	0.349	0.367	0.410	0.327	0.376	0.389	0.421
n = 15	0.390	0.432	0.450	0.478	0.337	0.386	0.413	0.438	0.318	0.370	0.385	0.420
n = 20	0.382	0.433	0.451	0.468	0.346	0.400	0.429	0.453	0.358	0.412	0.433	0.448
n = 25	0.378	0.439	0.459	0.480	0.348	0.398	0.425	0.455	0.350	0.411	0.441	0.458
n = 30	0.385	0.445	0.468	0.487	0.338	0.402	0.430	0.449	0.330	0.417	0.446	0.480
n = 60	0.409	0.520	0.548	0.556	0.358	0.431	0.459	0.491	0.377	0.459	0.497	0.536
n = 120	0.387	0.536	0.612	0.624	0.395	0.482	0.518	0.546	0.429	0.526	0.571	0.609
Cov. SJ	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.354	0.381	0.397	0.414	0.302	0.327	0.348	0.374	0.319	0.353	0.363	0.387
n = 15	0.371	0.416	0.427	0.445	0.335	0.389	0.407	0.430	0.319	0.369	0.387	0.412
n = 20	0.379	0.433	0.449	0.462	0.340	0.416	0.433	0.460	0.353	0.417	0.436	0.460
n = 25	0.397	0.461	0.477	0.495	0.362	0.419	0.445	0.459	0.341	0.406	0.434	0.455
n = 30	0.384	0.458	0.465	0.480	0.351	0.416	0.439	0.451	0.339	0.412	0.434	0.475
n = 60	0.407	0.546	0.565	0.573	0.372	0.438	0.479	0.493	0.377	0.479	0.512	0.540
n = 120	0.377	0.581	0.636	0.645	0.411	0.503	0.552	0.569	0.450	0.547	0.589	0.634
Cov. Naïve	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.249	0.249	0.249	0.249	0.199	0.199	0.199	0.199	0.201	0.201	0.201	0.201
n = 15	0.275	0.275	0.275	0.275	0.222	0.222	0.222	0.222	0.215	0.215	0.215	0.215
n = 20	0.261	0.261	0.261	0.261	0.232	0.234	0.234	0.234	0.259	0.259	0.259	0.259
n = 25	0.278	0.278	0.278	0.278	0.242	0.244	0.244	0.244	0.237	0.238	0.238	0.238
n = 30	0.296	0.296	0.296	0.296	0.246	0.246	0.246	0.246	0.253	0.257	0.257	0.257
n = 60	0.343	0.343	0.344	0.344	0.263	0.266	0.267	0.267	0.317	0.320	0.323	0.323
n = 120	0.431	0.431	0.431	0.431	0.271	0.277	0.278	0.278	0.337	0.345	0.348	0.349

Table II.4. Coverage of Simar and Wilson's (2000) confidence intervals: "Trun. Normal Low" case

	Trun. Normal Low 1/1				Trun. Normal Low 2/1				Trun. Normal Low 2/2			
Cov. LSCV	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.379	0.471	0.554	0.684	0.339	0.442	0.530	0.663	0.360	0.451	0.517	0.639
n = 15	0.381	0.477	0.545	0.682	0.362	0.459	0.537	0.668	0.324	0.419	0.500	0.656
n = 20	0.383	0.466	0.541	0.657	0.362	0.444	0.498	0.631	0.356	0.432	0.514	0.649
n = 25	0.366	0.451	0.536	0.673	0.360	0.457	0.520	0.663	0.361	0.443	0.511	0.650
n = 30	0.385	0.473	0.546	0.654	0.350	0.429	0.513	0.665	0.354	0.453	0.510	0.626
n = 60	0.367	0.456	0.522	0.636	0.354	0.455	0.534	0.663	0.367	0.454	0.528	0.649
n = 120	0.342	0.417	0.466	0.565	0.396	0.495	0.572	0.699	0.415	0.500	0.576	0.676
Cov. SJ	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.343	0.435	0.516	0.652	0.314	0.412	0.512	0.634	0.326	0.421	0.513	0.614
n = 15	0.360	0.460	0.527	0.655	0.336	0.457	0.527	0.658	0.315	0.407	0.487	0.637
n = 20	0.378	0.466	0.540	0.653	0.359	0.433	0.484	0.622	0.339	0.430	0.496	0.632
n = 25	0.377	0.458	0.523	0.659	0.367	0.468	0.532	0.667	0.334	0.433	0.513	0.642
n = 30	0.368	0.457	0.526	0.633	0.347	0.442	0.519	0.650	0.339	0.449	0.515	0.634
n = 60	0.355	0.452	0.507	0.609	0.369	0.459	0.524	0.661	0.356	0.443	0.525	0.649
n = 120	0.333	0.402	0.446	0.536	0.405	0.491	0.573	0.705	0.422	0.517	0.584	0.692
Cov. Naïve	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.395	0.428	0.529	0.635	0.309	0.363	0.449	0.575	0.321	0.373	0.441	0.551
n = 15	0.453	0.453	0.570	0.659	0.328	0.403	0.488	0.614	0.310	0.370	0.437	0.601
n = 20	0.477	0.478	0.579	0.677	0.317	0.413	0.476	0.604	0.337	0.424	0.477	0.598
n = 25	0.470	0.475	0.604	0.713	0.327	0.424	0.499	0.639	0.324	0.413	0.481	0.622
n = 30	0.481	0.482	0.584	0.703	0.318	0.411	0.479	0.629	0.322	0.420	0.490	0.615
n = 60	0.520	0.545	0.663	0.778	0.338	0.423	0.499	0.655	0.390	0.480	0.553	0.679
n = 120	0.570	0.602	0.683	0.808	0.349	0.460	0.552	0.684	0.422	0.531	0.615	0.737

Table II.5. Coverage of Simar and Wilson's (1998) confidence intervals: "Trun. Normal High" case

	Trun. Normal High 1/1				**Trun. Normal High 2/1**				**Trun. Normal High 2/2**			
Cov. LSCV	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.636	0.816	0.883	0.905	0.646	0.817	0.868	0.885	0.641	0.790	0.874	0.897
n = 15	0.528	0.743	0.837	0.911	0.570	0.767	0.840	0.912	0.555	0.736	0.828	0.890
n = 20	0.469	0.698	0.813	0.897	0.504	0.727	0.826	0.907	0.511	0.714	0.819	0.900
n = 25	0.443	0.668	0.778	0.887	0.447	0.682	0.792	0.906	0.469	0.693	0.811	0.898
n = 30	0.405	0.615	0.773	0.887	0.424	0.652	0.798	0.893	0.457	0.684	0.810	0.907
n = 60	0.284	0.471	0.634	0.887	0.349	0.531	0.718	0.920	0.354	0.538	0.690	0.894
n = 120	0.221	0.351	0.514	0.808	0.236	0.395	0.565	0.857	0.268	0.417	0.577	0.858
Cov. SJ	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.688	0.863	0.900	0.918	0.659	0.847	0.879	0.894	0.685	0.869	0.898	0.912
n = 15	0.545	0.812	0.910	0.936	0.552	0.823	0.918	0.933	0.583	0.819	0.920	0.937
n = 20	0.465	0.782	0.889	0.938	0.518	0.800	0.910	0.954	0.516	0.784	0.916	0.953
n = 25	0.390	0.713	0.868	0.946	0.432	0.728	0.887	0.961	0.474	0.752	0.889	0.957
n = 30	0.327	0.646	0.852	0.951	0.380	0.696	0.871	0.952	0.430	0.682	0.873	0.950
n = 60	0.192	0.378	0.666	0.955	0.243	0.456	0.734	0.966	0.289	0.484	0.722	0.945
n = 120	0.094	0.194	0.379	0.875	0.128	0.273	0.472	0.902	0.151	0.277	0.492	0.916
Cov. Naïve	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.582	0.600	0.603	0.603	0.552	0.562	0.562	0.562	0.509	0.516	0.516	0.516
n = 15	0.601	0.602	0.604	0.604	0.549	0.550	0.550	0.550	0.526	0.527	0.527	0.527
n = 20	0.570	0.570	0.573	0.573	0.575	0.575	0.575	0.575	0.498	0.501	0.501	0.501
n = 25	0.556	0.556	0.559	0.559	0.554	0.556	0.556	0.556	0.526	0.528	0.528	0.528
n = 30	0.561	0.561	0.565	0.565	0.537	0.538	0.538	0.538	0.509	0.512	0.512	0.512
n = 60	0.560	0.560	0.565	0.565	0.535	0.539	0.539	0.539	0.504	0.507	0.507	0.507
n = 120	0.588	0.589	0.591	0.592	0.497	0.500	0.501	0.501	0.524	0.527	0.527	0.527

Table II.6. Coverage of Simar and Wilson's (2000) confidence intervals: "Trun. Normal High" case

	Trun. Normal High 1/1				**Trun. Normal High 2/1**				**Trun. Normal High 2/2**			
Cov. LSCV	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.497	0.564	0.623	0.707	0.546	0.631	0.699	0.788	0.547	0.629	0.698	0.801
n = 15	0.411	0.467	0.520	0.575	0.445	0.552	0.622	0.689	0.477	0.556	0.621	0.706
n = 20	0.356	0.418	0.469	0.524	0.400	0.492	0.544	0.621	0.438	0.506	0.569	0.646
n = 25	0.324	0.385	0.427	0.485	0.376	0.439	0.486	0.556	0.388	0.463	0.513	0.608
n = 30	0.308	0.345	0.377	0.428	0.346	0.415	0.470	0.545	0.380	0.460	0.511	0.585
n = 60	0.216	0.255	0.276	0.308	0.287	0.347	0.387	0.435	0.315	0.373	0.407	0.470
n = 120	0.154	0.183	0.193	0.214	0.191	0.239	0.265	0.308	0.221	0.271	0.300	0.331
Cov. SJ	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.484	0.574	0.617	0.703	0.504	0.597	0.674	0.762	0.560	0.648	0.712	0.814
n = 15	0.355	0.417	0.467	0.548	0.415	0.499	0.551	0.645	0.462	0.535	0.592	0.685
n = 20	0.315	0.376	0.430	0.483	0.394	0.465	0.514	0.595	0.404	0.480	0.533	0.626
n = 25	0.264	0.309	0.345	0.419	0.315	0.386	0.432	0.523	0.353	0.427	0.486	0.575
n = 30	0.224	0.264	0.293	0.355	0.300	0.363	0.405	0.473	0.332	0.391	0.444	0.515
n = 60	0.141	0.169	0.184	0.207	0.191	0.223	0.261	0.301	0.230	0.264	0.300	0.357
n = 120	0.069	0.082	0.092	0.100	0.090	0.105	0.132	0.172	0.120	0.140	0.158	0.189
Cov. Naïve	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.753	0.804	0.906	0.961	0.693	0.759	0.873	0.957	0.653	0.728	0.852	0.927
n = 15	0.767	0.771	0.894	0.949	0.656	0.770	0.863	0.944	0.648	0.745	0.847	0.941
n = 20	0.734	0.735	0.874	0.945	0.674	0.769	0.865	0.949	0.623	0.740	0.819	0.933
n = 25	0.736	0.738	0.876	0.947	0.675	0.788	0.853	0.958	0.645	0.752	0.835	0.934
n = 30	0.728	0.740	0.864	0.938	0.655	0.780	0.850	0.955	0.616	0.757	0.839	0.942
n = 60	0.704	0.728	0.853	0.949	0.646	0.783	0.860	0.940	0.620	0.760	0.842	0.930
n = 120	0.744	0.787	0.861	0.955	0.616	0.759	0.838	0.945	0.661	0.782	0.872	0.956

Table II.7. Coverage of Simar and Wilson's (1998) confidence intervals: "Uniform" case

	Uniform 1/1				**Uniform 2/1**				**Uniform 2/2**			
Cov. LSCV	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.512	0.685	0.772	0.869	0.543	0.697	0.789	0.837	0.542	0.689	0.755	0.809
n = 15	0.434	0.641	0.735	0.863	0.497	0.641	0.738	0.835	0.504	0.671	0.776	0.869
n = 20	0.435	0.613	0.718	0.837	0.467	0.626	0.727	0.846	0.486	0.621	0.733	0.854
n = 25	0.427	0.604	0.721	0.833	0.492	0.631	0.742	0.849	0.490	0.650	0.745	0.866
n = 30	0.396	0.559	0.691	0.837	0.451	0.595	0.708	0.823	0.465	0.629	0.734	0.851
n = 60	0.371	0.552	0.684	0.852	0.453	0.598	0.698	0.851	0.431	0.611	0.739	0.888
n = 120	0.382	0.571	0.700	0.866	0.447	0.657	0.775	0.899	0.411	0.600	0.756	0.911
Cov. SJ	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.519	0.770	0.860	0.883	0.302	0.327	0.348	0.374	0.584	0.761	0.817	0.838
n = 15	0.400	0.699	0.834	0.935	0.335	0.389	0.407	0.430	0.542	0.758	0.862	0.901
n = 20	0.348	0.646	0.828	0.956	0.340	0.416	0.433	0.460	0.447	0.674	0.833	0.923
n = 25	0.270	0.547	0.786	0.948	0.362	0.419	0.445	0.459	0.415	0.669	0.825	0.948
n = 30	0.235	0.476	0.731	0.938	0.351	0.416	0.439	0.451	0.382	0.613	0.800	0.952
n = 60	0.106	0.232	0.487	0.905	0.372	0.438	0.479	0.493	0.223	0.397	0.593	0.932
n = 120	0.050	0.105	0.227	0.720	0.411	0.503	0.552	0.569	0.146	0.251	0.412	0.864
Cov. Naïve	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.562	0.576	0.577	0.577	0.456	0.457	0.457	0.457	0.457	0.459	0.459	0.459
n = 15	0.608	0.610	0.612	0.612	0.436	0.439	0.439	0.439	0.446	0.447	0.447	0.447
n = 20	0.612	0.612	0.617	0.617	0.463	0.463	0.463	0.463	0.483	0.487	0.487	0.487
n = 25	0.602	0.602	0.607	0.607	0.466	0.468	0.468	0.468	0.515	0.515	0.515	0.515
n = 30	0.570	0.570	0.572	0.572	0.450	0.454	0.454	0.454	0.502	0.505	0.506	0.506
n = 60	0.609	0.612	0.614	0.614	0.477	0.480	0.480	0.480	0.555	0.556	0.556	0.556
n = 120	0.606	0.609	0.611	0.611	0.430	0.431	0.431	0.431	0.552	0.555	0.555	0.555

Table II.8. Coverage of Simar and Wilson's (2000) confidence intervals: "Uniform" case

	Uniform 1/1				**Uniform 2/1**				**Uniform 2/2**			
Cov. LSCV	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.427	0.496	0.545	0.637	0.509	0.609	0.677	0.791	0.486	0.593	0.659	0.777
n = 15	0.343	0.399	0.447	0.510	0.451	0.544	0.610	0.718	0.445	0.528	0.601	0.712
n = 20	0.335	0.390	0.428	0.490	0.414	0.499	0.551	0.657	0.440	0.529	0.581	0.656
n = 25	0.322	0.371	0.421	0.490	0.433	0.510	0.570	0.656	0.445	0.520	0.574	0.638
n = 30	0.311	0.351	0.381	0.424	0.419	0.511	0.563	0.632	0.413	0.503	0.557	0.631
n = 60	0.262	0.307	0.332	0.370	0.404	0.470	0.523	0.593	0.357	0.442	0.494	0.561
n = 120	0.256	0.285	0.311	0.340	0.391	0.459	0.509	0.583	0.343	0.420	0.461	0.521
Cov. SJ	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.379	0.456	0.509	0.595	0.314	0.412	0.512	0.634	0.492	0.588	0.663	0.775
n = 15	0.261	0.314	0.355	0.428	0.336	0.457	0.527	0.658	0.432	0.532	0.605	0.727
n = 20	0.222	0.266	0.301	0.367	0.359	0.433	0.484	0.622	0.391	0.454	0.502	0.602
n = 25	0.203	0.241	0.269	0.303	0.367	0.468	0.532	0.667	0.325	0.401	0.450	0.536
n = 30	0.153	0.198	0.222	0.264	0.347	0.442	0.519	0.650	0.320	0.367	0.432	0.508
n = 60	0.074	0.085	0.098	0.111	0.369	0.459	0.524	0.661	0.189	0.226	0.249	0.306
n = 120	0.030	0.040	0.044	0.050	0.405	0.491	0.573	0.705	0.110	0.143	0.160	0.177
Cov. Naïve	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.736	0.779	0.865	0.925	0.594	0.700	0.809	0.909	0.586	0.670	0.780	0.895
n = 15	0.794	0.799	0.910	0.961	0.560	0.696	0.800	0.911	0.568	0.705	0.799	0.905
n = 20	0.761	0.762	0.899	0.958	0.571	0.711	0.797	0.931	0.604	0.723	0.820	0.917
n = 25	0.757	0.758	0.880	0.952	0.587	0.719	0.812	0.928	0.635	0.767	0.853	0.949
n = 30	0.743	0.752	0.879	0.950	0.571	0.698	0.795	0.931	0.628	0.765	0.843	0.952
n = 60	0.769	0.793	0.884	0.965	0.597	0.723	0.813	0.922	0.675	0.786	0.864	0.955
n = 120	0.756	0.804	0.875	0.958	0.562	0.735	0.842	0.946	0.658	0.800	0.860	0.949

III. Appendix III: Confidence intervals

Figure III.1. Simar and Wilson (1998) confidence intervals – LSCV smoothing

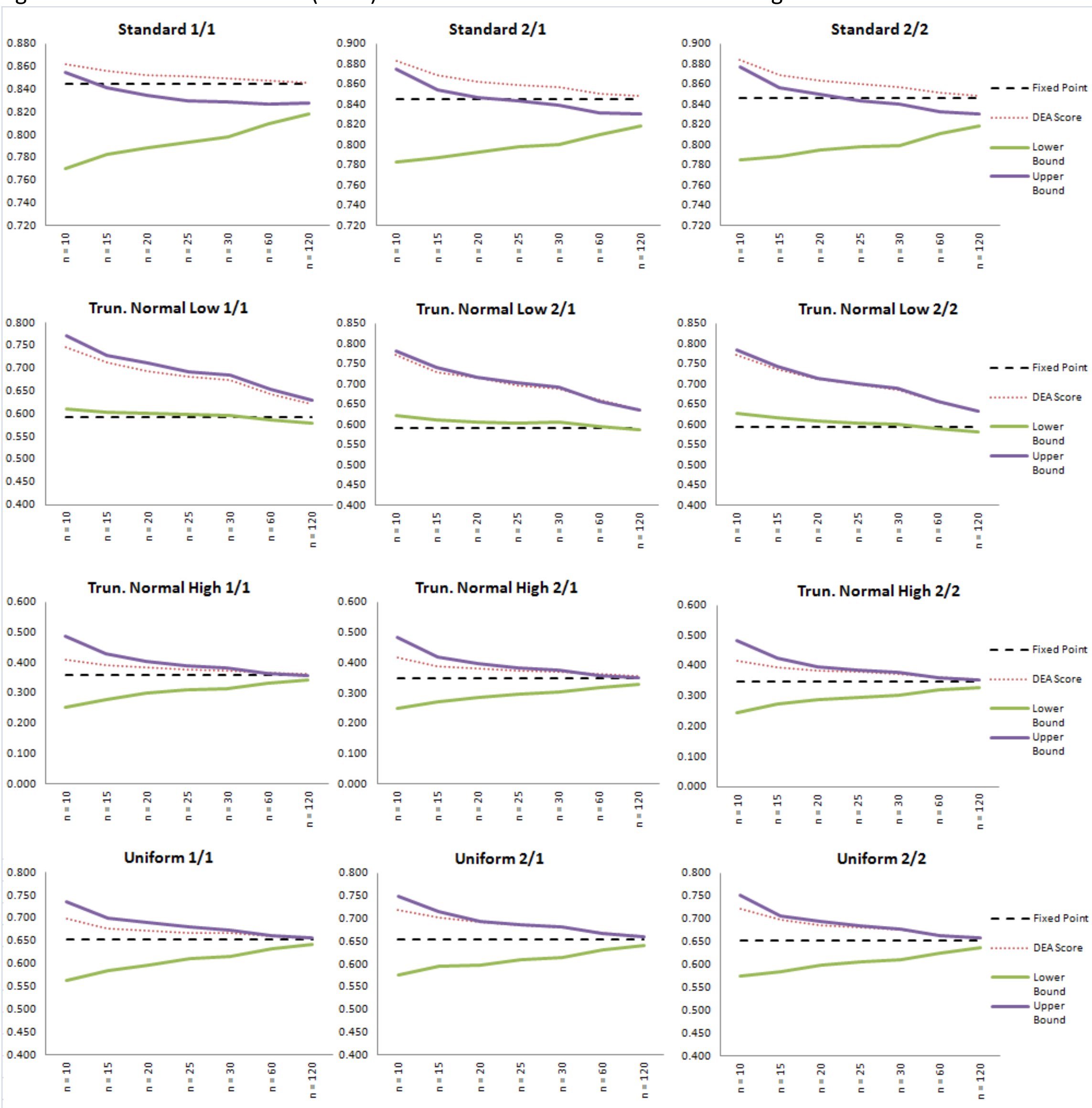


Figure III.2. Simar and Wilson (2000) confidence intervals – LSCV smoothing

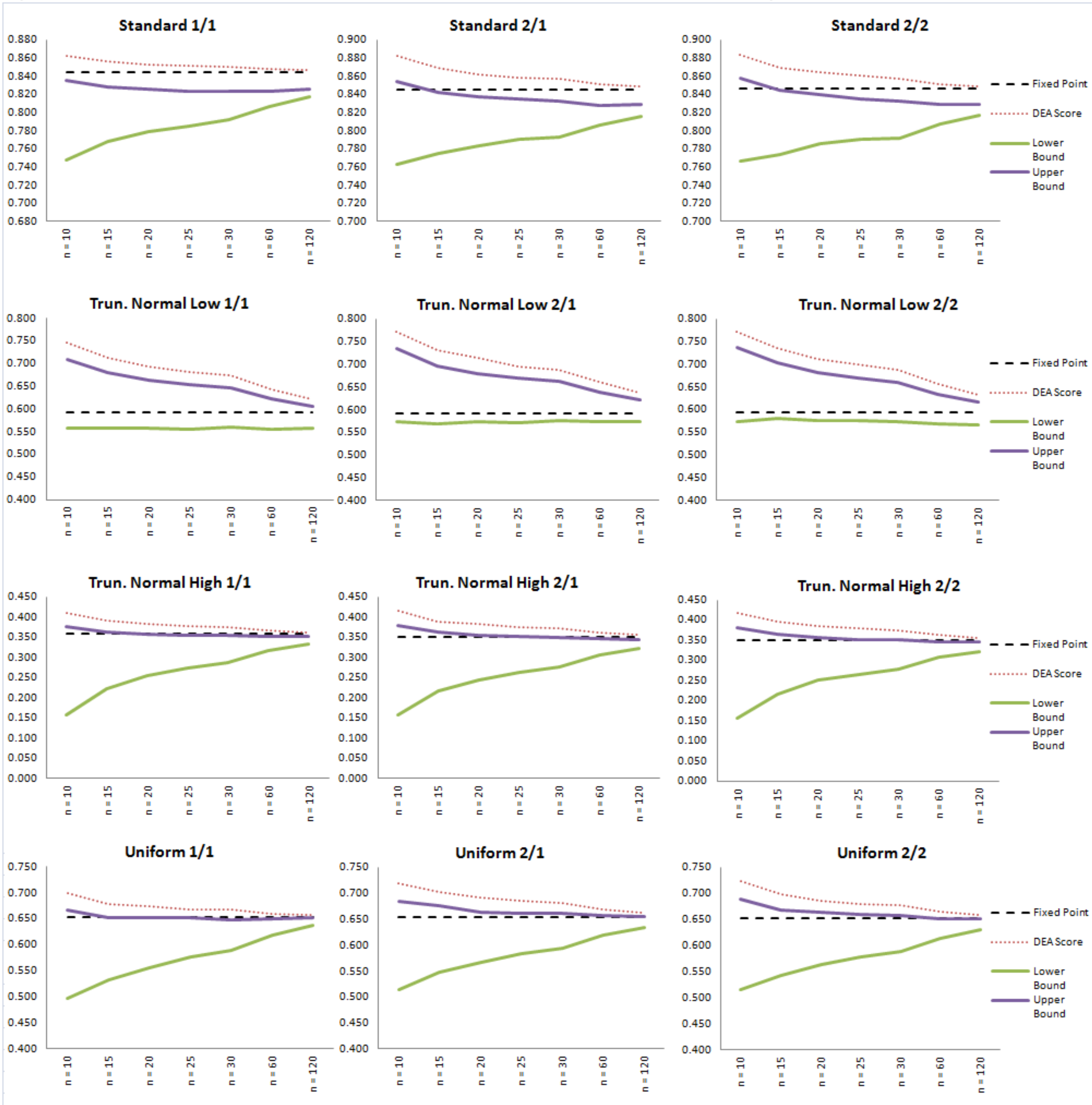


Figure III.3. Simar and Wilson (1998) confidence intervals – SJ smoothing

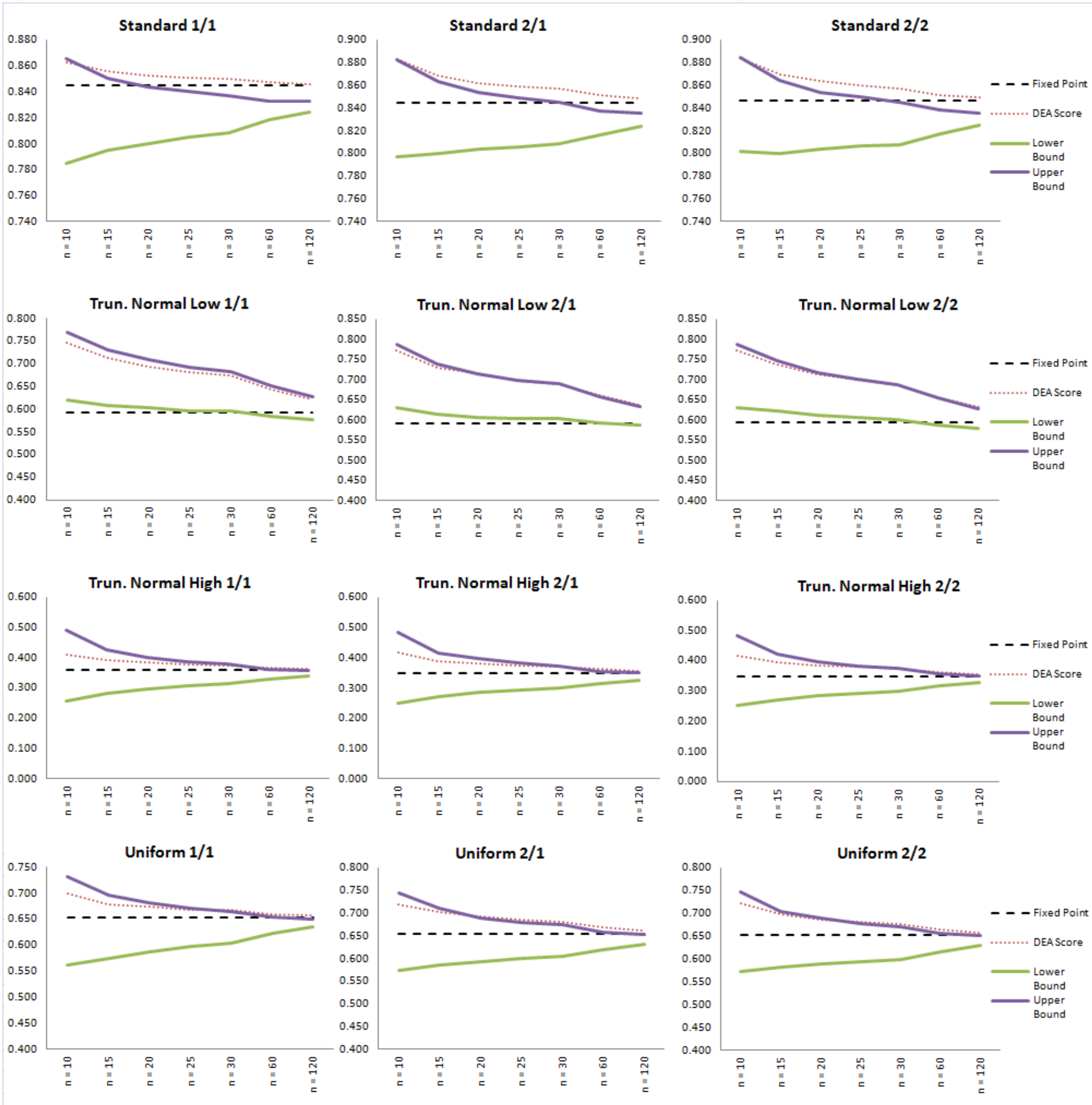


Figure III.4. Simar and Wilson (2000) confidence intervals – SJ smoothing

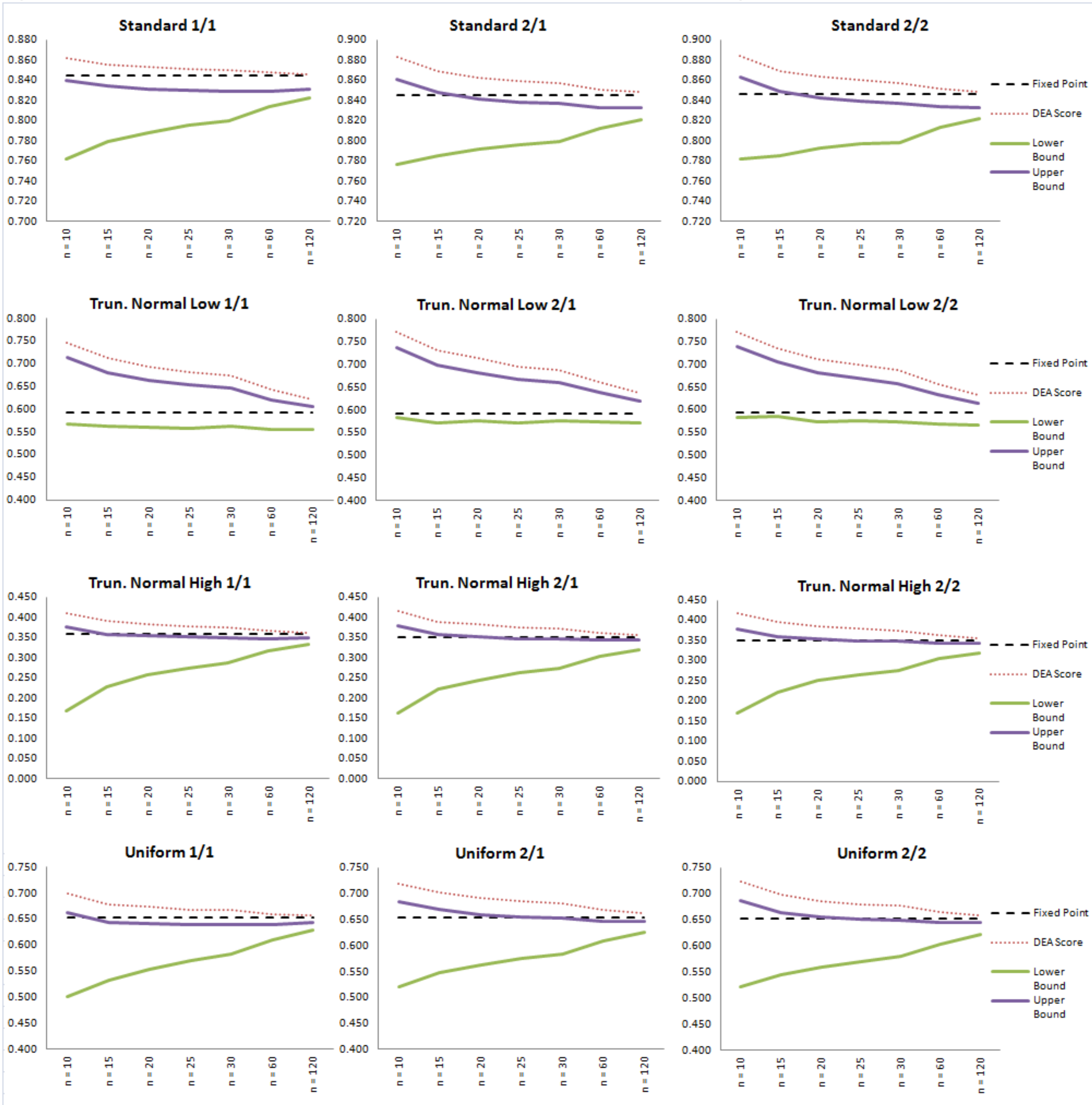


Figure III.5. Simar and Wilson (1998) confidence intervals – Naïve bootstrap

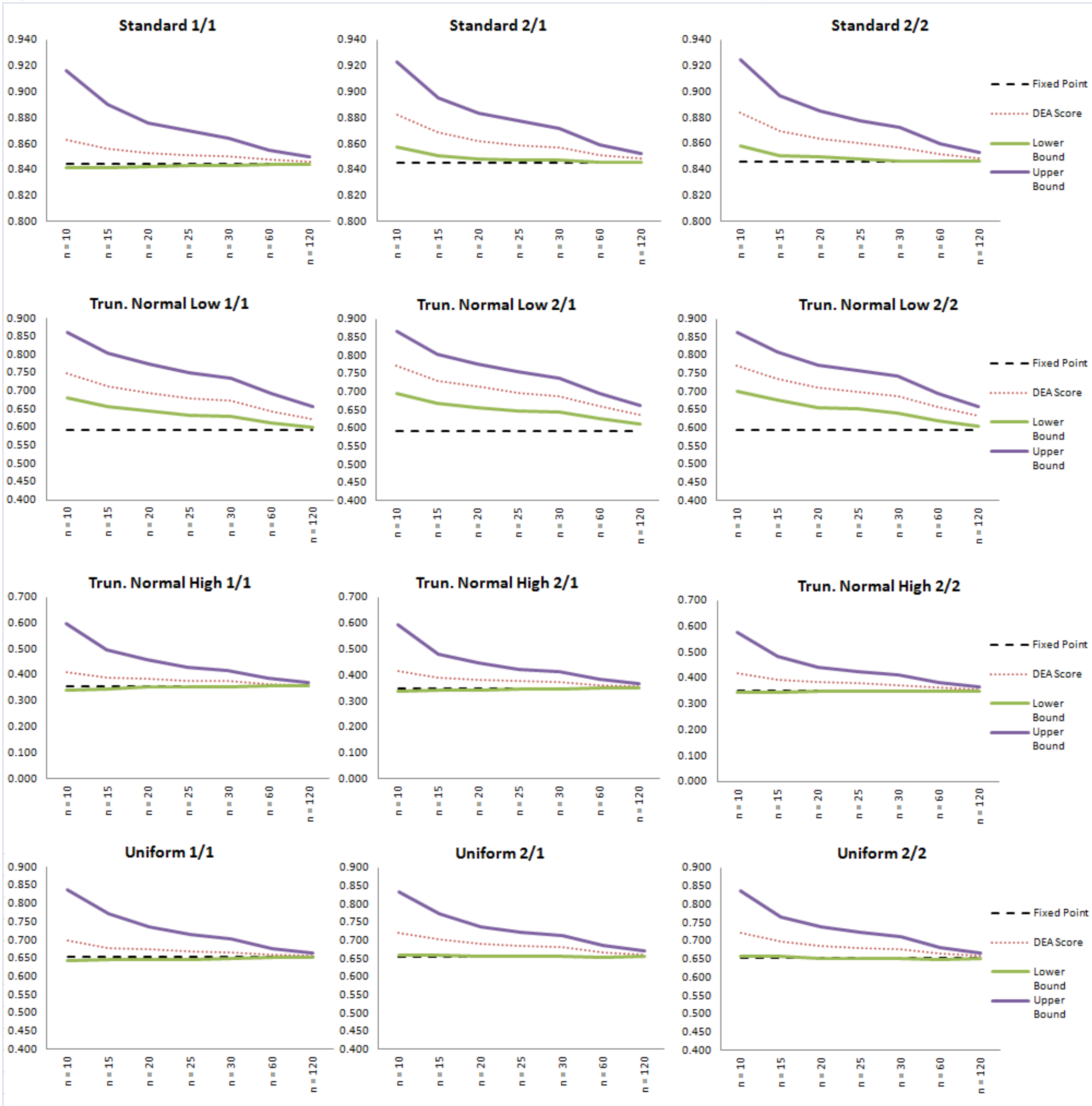
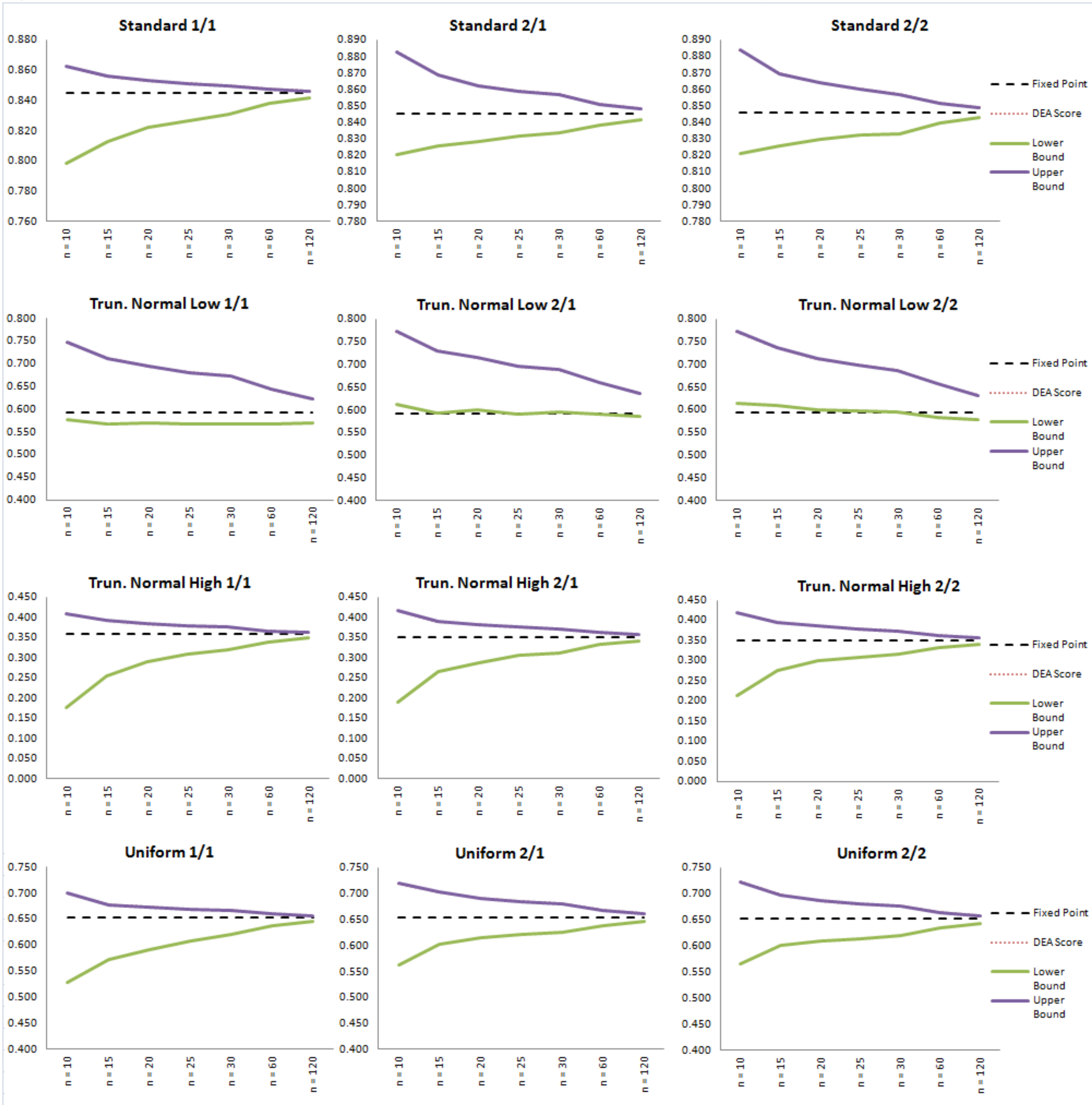


Figure III.6. Simar and Wilson (2000) confidence intervals – Naïve bootstrap



IV. Appendix IV: Skewness and effect on Simar and Wilson's confidence intervals

We show that in the presence of positive skewness (as evidenced in all simulations), the upper and lower boundaries of the Simar and Wilson's (1998) intervals are higher than the Simar and Wilson's (2000a) intervals. The implication of this is that when the bootstrap bias is greater than the DEA bias then SW1998 are expected to perform better while in the opposite case SW2000 will perform better. Hence, we theoretically justify the observed behaviour of coverage probabilities and confidence intervals in chapter 2.

For the SW1998 intervals we know from (2.26) that $\tilde{\theta}_k^* = \hat{\theta}_k^* - 2\widehat{bias}_k$ and that the $(1-a)\%$ SW1998 intervals are $\theta \in \left(\tilde{\theta}_k^{*,(a/2)}\ ,\ \tilde{\theta}_k^{*,(1-a/2)}\right)$. Taking into account (2.24) and (2.26) the $j\%$ SW1998 percentile satisfies:

$$\begin{aligned} \tilde{\theta}_k^{*,(j)} &= \left[\hat{\theta}_k^* - 2\widehat{bias}_k\right]^{(j)} = \left[\hat{\theta}_k^* - 2\left(\overline{\hat{\theta}_k^*} - \hat{\theta}_k\right)\right]^{(j)} = \hat{\theta}_k^{*,(j)} - 2\left(\overline{\hat{\theta}_k^*} - \hat{\theta}_k\right) \\ &= \hat{\theta}_k^{*,(j)} + 2\hat{\theta}_k - 2\overline{\hat{\theta}_k^*} \end{aligned} \tag{IV.1}$$

Note that we can take the term $2\left(\overline{\hat{\theta}_k^*} - \hat{\theta}_k\right)$ out of the bracket since it is a constant which shifts the distribution of $\hat{\theta}_k^*$ without affecting its shape. Regarding the SW2000 intervals we know from (2.30) and (2.31) that $\Pr\left(\Delta\hat{\theta}_k^{*(a)} < \hat{\theta}_k^* - \hat{\theta}_k < \Delta\hat{\theta}_k^{*(1-a/2)}\right) = 1-a$, hence the associated percentiles satisfy:

$$\Delta\hat{\theta}_k^{*(j)} = \left[\hat{\theta}_k^* - \hat{\theta}_k\right]^{(j)} = \hat{\theta}_k^{*,(j)} - \hat{\theta}_k \tag{IV.2}$$

And we already know from (2.31) that the associated confidence intervals are $\theta_k \in \left(\hat{\theta}_k - \Delta\hat{\theta}_k^{*\,(1-a/2)},\ \hat{\theta}_k - \Delta\hat{\theta}_k^{*\,(a/2)}\right)$.

We will show first that under reasonable conditions the upper boundary of the SW1998 intervals lies higher compared to that of the SW2000 intervals. We have:

$$\tilde{\theta}_k^{*,\left(1-\frac{a}{2}\right)} > \hat{\theta}_k - \Delta\hat{\theta}_k^{*\left(\frac{a}{2}\right)} \Rightarrow$$

$$\hat{\theta}_k^{*,\left(1-\frac{a}{2}\right)} + 2\hat{\theta}_k - 2\overline{\hat{\theta}_k^*} > \hat{\theta}_k - \left[\hat{\theta}_k^{*,\left(\frac{a}{2}\right)} - \hat{\theta}_k\right] \Rightarrow \tag{IV.3}$$

$$\hat{\theta}_k^{*,\left(1-\frac{a}{2}\right)} - \overline{\hat{\theta}_k^*} > \overline{\hat{\theta}_k^*} - \hat{\theta}_k^{*,\left(\frac{a}{2}\right)}$$

Note that $\overline{\hat{\theta}_k^*}$ is the centre of the distribution of $\hat{\theta}_k^*$, and therefore $\hat{\theta}_k^{*,(a/2)} < \overline{\hat{\theta}_k^*} < \hat{\theta}_k^{*,(1-a/2)}$. If the distribution is positively skewed, as this seems to be on average the cases from our simulations in subsection 2.9.5, then the last inequality is almost certain to apply. If the distribution is also leptokurtic (which also seems to be true on average from our simulations), then $\overline{\hat{\theta}_k^*}$ should lie closer to $\hat{\theta}_k^{*,(a/2)}$ than $\hat{\theta}_k^{*,(1-a/2)}$ as there would be a high concentration of values towards the lower end of the distribution and very close to $\overline{\hat{\theta}_k^*}$. Therefore, we have shown that under the usually observed conditions $\hat{\theta}_k^{*,(1-a/2)} - \overline{\hat{\theta}_k^*} > \overline{\hat{\theta}_k^*} - \hat{\theta}_k^{*,(a/2)}$. Following the same approach for the lower bounds of the two confidence intervals we reach exactly the same inequality. Hence, in these cases the SW1998 endpoints should lie higher than the SW2000 ones which is confirmed in our simulations for all cases.

We also need to note, that Simar and Wilson (1998) have suggested that in the presence of skewness that the median should be preferred in bias corrections instead of the mean and they suggested using Efron's (1982) bias-corrected intervals. If the median is used to compute $\overline{\hat{\theta}_k^*}$ instead of the mean, then the only condition necessary for the previous inequality to apply would be that the distribution be positively skewed. Given that all simulations exhibit skewness and given that we should be using the median instead in these cases, we deduce that the endpoints of Simar and Wilson's (1998) lie higher compared to those of Simar and Wilson (2000).

The implications of this are quite useful as they explain why the SW1998 intervals perform better when the bootstrap bias is larger than the DEA bias and why the opposite is true when the DEA bias is larger than the bootstrap bias (as in the "Trun.Normal Low" case or under all naïve bootstraps). Consider the case where the bootstrap bias is larger than the DEA bias, suggesting that the true efficiency score is underestimated and that the associated confidence intervals target at a value below θ_k. That is, in both cases the lower bounds of the intervals will be well below θ_k while the extent to which the upper bounds will cover θ_k will depend upon the magnitude of the bias (see also subsection 3.2.1). Since SW1998 upper bound lies further up compared to the SW2000 one, then there is a higher probability for θ_k to be included in SW1998 intervals rather than the SW2000 ones. This is confirmed in all of our simulation results in subsections 2.9.3 and 2.9.4.

Likewise, when the DEA bias is greater than the bootstrap bias then θ_k is overestimated and the upper bounds of the intervals lie well-above θ_k. Moreover, the larger the DEA bias is compared to the bootstrap bias (see also subsection 3.2.3) the higher is the probability for the intervals to overestimate θ_k as well. Since the SW2000 lower bound lies below the SW1998, the probability of including θ_k is greater. Again, the simulations for the "Trun.Normal Low" case and for samples up to $n = 120$, confirm this argument.

Hence, in the presence of positive skewness the SW1998 confidence intervals perform better when the bootstrap bias is greater than the DEA bias while SW2000 perform better when the bootstrap bias is smaller than the DEA bias. However, it is reminded that this case has been associated with technological heterogeneity which might suggest that SW2000 should not be preferred if there is positive skewness.

V. Appendix V: Moments of the fixed DMU's bootstrap distribution

Table V.1. Moments for the fixed DMU: "Standard" case

	Standard 1/1				Standard 2/1				Standard 2/2			
Population	**Eff. Score**				**Eff. Score**				**Eff. Score**			
N = 10,000	0.845				0.845				0.846			
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.869	0.023	1.823	7.241	0.887	0.026	0.989	4.330	0.889	0.027	1.043	4.228
n = 15	0.860	0.014	1.536	5.603	0.872	0.018	1.080	4.799	0.874	0.019	1.151	4.453
n = 20	0.856	0.011	1.724	6.616	0.866	0.014	1.042	3.974	0.867	0.014	1.146	4.535
n = 25	0.854	0.008	1.404	5.088	0.862	0.012	1.233	4.926	0.863	0.011	1.115	4.500
n = 30	0.852	0.007	2.171	12.005	0.860	0.010	1.368	5.556	0.859	0.009	1.111	4.198
n = 60	0.849	0.004	2.050	9.177	0.852	0.005	1.514	6.117	0.853	0.005	1.365	5.097
n = 120	0.847	0.002	1.855	7.513	0.849	0.002	1.069	4.544	0.850	0.002	1.532	7.484
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.931	0.022	1.010	4.999	0.948	0.023	0.870	4.305	0.950	0.023	0.852	4.210
n = 15	0.914	0.015	0.986	5.002	0.927	0.017	0.838	4.211	0.928	0.018	0.819	4.141
n = 20	0.904	0.012	0.972	5.020	0.915	0.014	0.817	4.154	0.915	0.014	0.820	4.153
n = 25	0.898	0.010	0.964	5.036	0.907	0.012	0.826	4.207	0.909	0.012	0.814	4.150
n = 30	0.893	0.008	0.954	5.064	0.902	0.010	0.798	4.117	0.902	0.010	0.802	4.116
n = 60	0.880	0.004	0.875	4.821	0.886	0.006	0.789	4.059	0.886	0.006	0.784	4.093
n = 120	0.871	0.002	0.825	4.700	0.875	0.003	0.763	4.008	0.875	0.003	0.765	4.041
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.917	0.020	1.373	6.051	0.937	0.022	1.071	4.747	0.938	0.022	1.045	4.684
n = 15	0.902	0.014	1.363	6.069	0.916	0.016	0.999	4.545	0.918	0.017	0.977	4.435
n = 20	0.894	0.011	1.353	6.099	0.906	0.013	0.977	4.467	0.906	0.013	0.967	4.409
n = 25	0.888	0.009	1.345	6.068	0.899	0.011	0.971	4.456	0.900	0.011	0.955	4.390
n = 30	0.884	0.007	1.328	6.142	0.894	0.010	0.955	4.449	0.894	0.010	0.932	4.345
n = 60	0.873	0.004	1.207	5.792	0.879	0.005	0.918	4.301	0.880	0.005	0.909	4.285
n = 120	0.866	0.002	1.138	5.635	0.869	0.003	0.897	4.273	0.870	0.003	0.903	4.267
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.876	0.020	2.454	9.919	0.898	0.018	2.317	9.779	0.898	0.018	2.267	9.734
n = 15	0.865	0.013	2.538	10.918	0.879	0.013	2.303	9.743	0.881	0.013	2.133	8.856
n = 20	0.860	0.010	2.491	10.446	0.871	0.010	2.193	9.226	0.872	0.010	2.285	10.025
n = 25	0.857	0.008	2.472	10.496	0.866	0.008	2.258	9.877	0.867	0.008	2.188	9.365
n = 30	0.854	0.007	2.438	10.098	0.863	0.007	2.174	9.058	0.863	0.007	2.248	9.595
n = 60	0.850	0.003	2.494	10.638	0.854	0.004	2.377	10.617	0.855	0.004	2.321	10.098
n = 120	0.847	0.002	2.501	10.909	0.850	0.002	2.471	11.138	0.850	0.002	2.613	11.846

Table V.2. Moments for the fixed DMU: "Trun. Normal Low" case

	Trun. Normal Low 1/1				Trun. Normal Low 2/1				Trun. Normal Low 2/2			
Population	**Eff. Score**				**Eff. Score**				**Eff. Score**			
N = 10,000	0.592				0.591				0.593			
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.749	0.083	0.329	2.806	0.771	0.080	0.219	2.730	0.774	0.083	0.251	2.459
n = 15	0.716	0.066	0.237	2.406	0.734	0.068	0.274	2.574	0.738	0.067	0.212	2.796
n = 20	0.700	0.059	0.370	2.762	0.716	0.060	0.257	2.607	0.715	0.062	0.398	2.780
n = 25	0.685	0.053	0.452	2.774	0.701	0.053	0.357	2.690	0.703	0.054	0.337	2.640
n = 30	0.677	0.050	0.454	2.687	0.691	0.048	0.341	2.728	0.694	0.052	0.391	2.675
n = 60	0.647	0.036	0.579	2.919	0.662	0.036	0.478	2.925	0.660	0.038	0.607	3.033
n = 120	0.626	0.026	0.657	2.706	0.640	0.025	0.510	3.022	0.637	0.026	0.766	3.523
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.833	0.038	1.312	5.432	0.861	0.039	1.097	4.763	0.860	0.039	1.088	4.723
n = 15	0.785	0.031	1.462	5.788	0.808	0.032	1.171	4.828	0.815	0.031	1.143	4.810
n = 20	0.759	0.027	1.528	5.983	0.782	0.027	1.178	4.859	0.781	0.027	1.163	4.824
n = 25	0.739	0.024	1.596	6.247	0.760	0.025	1.218	5.033	0.763	0.025	1.208	4.913
n = 30	0.726	0.023	1.600	6.277	0.746	0.023	1.235	5.059	0.746	0.023	1.195	4.879
n = 60	0.684	0.017	1.685	6.512	0.703	0.017	1.215	4.960	0.701	0.017	1.201	4.830
n = 120	0.653	0.013	1.778	6.963	0.672	0.012	1.185	4.850	0.669	0.013	1.194	4.817
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.827	0.038	1.471	5.965	0.858	0.039	1.229	5.125	0.853	0.038	1.211	5.059
n = 15	0.783	0.031	1.536	6.088	0.806	0.032	1.218	5.020	0.812	0.031	1.205	4.936
n = 20	0.758	0.027	1.566	6.161	0.784	0.027	1.207	4.947	0.780	0.028	1.192	4.910
n = 25	0.738	0.025	1.627	6.397	0.761	0.025	1.238	5.046	0.762	0.024	1.220	4.989
n = 30	0.727	0.023	1.634	6.427	0.748	0.023	1.235	4.991	0.748	0.023	1.216	4.971
n = 60	0.685	0.017	1.695	6.589	0.705	0.017	1.212	4.922	0.703	0.017	1.178	4.826
n = 120	0.654	0.013	1.753	6.851	0.673	0.012	1.181	4.835	0.669	0.013	1.152	4.720
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.780	0.053	1.585	5.088	0.807	0.046	1.803	6.097	0.808	0.045	1.786	5.959
n = 15	0.744	0.044	1.665	5.552	0.764	0.039	1.755	6.006	0.771	0.038	1.857	6.319
n = 20	0.724	0.040	1.591	5.191	0.745	0.034	1.755	6.010	0.741	0.034	1.760	6.117
n = 25	0.706	0.036	1.680	5.499	0.724	0.031	1.800	6.395	0.727	0.030	1.792	6.339
n = 30	0.696	0.034	1.665	5.400	0.715	0.027	1.777	6.245	0.714	0.027	1.823	6.521
n = 60	0.661	0.025	1.756	5.762	0.678	0.020	1.709	6.283	0.676	0.021	1.715	6.113
n = 120	0.635	0.018	1.810	6.055	0.653	0.014	1.720	6.365	0.648	0.015	1.694	6.220

Table V.3. Moments for the fixed DMU: "Trun. Normal High" case

	Trun. Normal High 1/1				Trun. Normal High 2/1				Trun. Normal High 2/2			
Population	**Eff. Score**				**Eff. Score**				**Eff. Score**			
N = 10,000	0.358				0.350				0.349			
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.434	0.076	2.118	9.663	0.440	0.078	1.514	5.358	0.440	0.076	1.898	8.496
n = 15	0.405	0.046	1.916	8.254	0.407	0.052	2.339	11.288	0.409	0.048	1.773	8.456
n = 20	0.394	0.034	1.620	6.318	0.393	0.033	1.545	5.839	0.394	0.034	1.510	6.450
n = 25	0.386	0.027	1.878	7.777	0.384	0.029	2.056	9.864	0.386	0.028	1.571	6.208
n = 30	0.381	0.024	2.233	11.064	0.380	0.024	2.003	8.814	0.381	0.025	1.898	8.371
n = 60	0.369	0.011	1.895	8.763	0.366	0.012	1.846	8.037	0.366	0.013	1.657	7.525
n = 120	0.364	0.006	1.794	6.801	0.358	0.006	1.530	7.053	0.358	0.006	1.106	4.315
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.515	0.059	1.971	8.895	0.526	0.059	1.908	8.812	0.533	0.058	1.836	8.443
n = 15	0.465	0.036	1.983	8.820	0.473	0.038	1.798	8.184	0.479	0.038	1.770	8.024
n = 20	0.446	0.026	2.022	9.219	0.453	0.028	1.712	7.773	0.453	0.027	1.646	7.330
n = 25	0.430	0.020	1.994	9.069	0.435	0.022	1.661	7.501	0.437	0.022	1.581	6.991
n = 30	0.419	0.017	2.030	9.204	0.425	0.019	1.626	7.139	0.428	0.019	1.548	6.891
n = 60	0.395	0.008	1.979	8.927	0.395	0.010	1.426	6.208	0.395	0.010	1.364	5.914
n = 120	0.380	0.005	1.964	8.686	0.378	0.006	1.291	5.536	0.377	0.006	1.253	5.372
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.512	0.058	2.113	9.874	0.528	0.058	1.969	9.170	0.527	0.057	1.938	9.182
n = 15	0.468	0.035	2.052	9.418	0.474	0.037	1.814	8.378	0.479	0.038	1.777	8.112
n = 20	0.446	0.026	2.054	9.579	0.454	0.028	1.725	7.758	0.454	0.027	1.671	7.564
n = 25	0.430	0.020	2.030	9.385	0.435	0.022	1.654	7.438	0.437	0.022	1.598	7.152
n = 30	0.422	0.017	2.028	9.246	0.425	0.019	1.611	7.166	0.427	0.019	1.540	6.772
n = 60	0.395	0.009	1.992	8.987	0.397	0.010	1.401	5.996	0.397	0.010	1.347	5.758
n = 120	0.381	0.005	1.944	8.574	0.380	0.006	1.262	5.361	0.379	0.006	1.238	5.297
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.458	0.073	2.662	11.409	0.467	0.068	2.793	13.291	0.467	0.063	2.862	14.018
n = 15	0.420	0.043	2.504	10.680	0.424	0.038	2.759	13.425	0.427	0.039	2.729	13.664
n = 20	0.404	0.031	2.675	11.926	0.409	0.030	2.599	12.554	0.408	0.026	2.633	12.789
n = 25	0.393	0.023	2.624	12.151	0.396	0.022	2.630	12.458	0.397	0.022	2.544	12.126
n = 30	0.388	0.019	2.593	11.630	0.388	0.019	2.653	12.824	0.391	0.018	2.513	11.605
n = 60	0.373	0.009	2.678	11.893	0.370	0.009	2.595	12.112	0.371	0.009	2.353	10.600
n = 120	0.365	0.005	2.501	10.774	0.361	0.005	2.417	11.111	0.361	0.005	2.253	9.982

Table V.4. Moments for the fixed DMU: "Uniform" case

	Uniform 1/1				Uniform 2/1				Uniform 2/2			
Population	**Eff. Score**				**Eff. Score**				**Eff. Score**			
N = 10,000	0.653				0.655				0.652			
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.716	0.061	1.755	6.513	0.736	0.060	1.578	5.975	0.737	0.063	1.323	4.916
n = 15	0.691	0.038	1.885	7.272	0.713	0.040	1.298	4.953	0.710	0.044	1.846	8.293
n = 20	0.681	0.028	2.133	10.499	0.698	0.031	1.594	6.971	0.696	0.034	1.662	7.130
n = 25	0.676	0.023	1.814	7.115	0.690	0.024	1.288	5.447	0.686	0.025	1.557	6.459
n = 30	0.672	0.017	1.527	5.453	0.686	0.022	1.193	4.608	0.681	0.021	1.247	5.109
n = 60	0.662	0.009	2.035	8.958	0.671	0.012	1.276	5.102	0.667	0.011	1.364	5.482
n = 120	0.658	0.004	2.017	9.309	0.663	0.006	1.831	8.819	0.660	0.006	1.509	6.176
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.790	0.041	1.489	6.223	0.820	0.043	1.280	5.488	0.822	0.042	1.235	5.355
n = 15	0.751	0.028	1.730	7.148	0.782	0.031	1.386	5.942	0.781	0.031	1.302	5.615
n = 20	0.733	0.022	1.816	7.792	0.758	0.024	1.355	5.812	0.754	0.025	1.390	6.063
n = 25	0.718	0.018	1.899	8.251	0.741	0.020	1.385	5.929	0.736	0.021	1.372	5.894
n = 30	0.709	0.014	1.877	8.305	0.729	0.017	1.368	5.880	0.725	0.018	1.371	5.887
n = 60	0.681	0.008	1.968	8.717	0.697	0.009	1.278	5.487	0.692	0.010	1.264	5.463
n = 120	0.667	0.004	1.961	8.657	0.677	0.005	1.218	5.213	0.674	0.006	1.185	5.062
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.787	0.041	1.630	6.841	0.817	0.042	1.381	5.758	0.818	0.042	1.319	5.635
n = 15	0.752	0.029	1.726	7.376	0.782	0.031	1.376	5.875	0.778	0.031	1.323	5.704
n = 20	0.735	0.023	1.781	7.747	0.760	0.024	1.319	5.729	0.755	0.025	1.331	5.767
n = 25	0.722	0.018	1.843	8.072	0.744	0.020	1.322	5.748	0.741	0.021	1.290	5.597
n = 30	0.713	0.015	1.837	8.081	0.735	0.018	1.311	5.690	0.731	0.018	1.276	5.502
n = 60	0.689	0.008	1.871	8.226	0.704	0.010	1.211	5.182	0.700	0.010	1.178	5.058
n = 120	0.675	0.004	1.906	8.409	0.684	0.006	1.156	4.961	0.681	0.006	1.129	4.818
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.731	0.052	2.354	8.278	0.758	0.047	2.412	9.201	0.763	0.045	2.348	9.415
n = 15	0.703	0.034	2.531	10.066	0.729	0.030	2.531	11.888	0.724	0.030	2.507	11.667
n = 20	0.691	0.026	2.462	10.201	0.711	0.023	2.517	11.435	0.708	0.023	2.525	11.731
n = 25	0.683	0.021	2.565	11.379	0.701	0.018	2.368	10.696	0.697	0.020	2.403	10.765
n = 30	0.677	0.016	2.635	11.749	0.695	0.016	2.321	10.060	0.691	0.017	2.311	10.238
n = 60	0.665	0.008	2.608	12.079	0.676	0.008	2.244	9.781	0.672	0.009	2.217	9.620
n = 120	0.659	0.004	2.627	11.833	0.666	0.004	2.356	10.246	0.662	0.005	2.262	9.783

VI. Appendix VI: SW1998 and SW2000 intervals in large samples

We observe that intervals narrow down with sample size and there is an obvious asymptotic trend to converge to the fixed point. The convergence slows down due to the fact that the bootstrap bias is not exactly the same as the DEA bias; this will only occur asymptotically where both will be equal to zero.

Figure VI.1. Simar and Wilson's (1998) confidence intervals: large samples

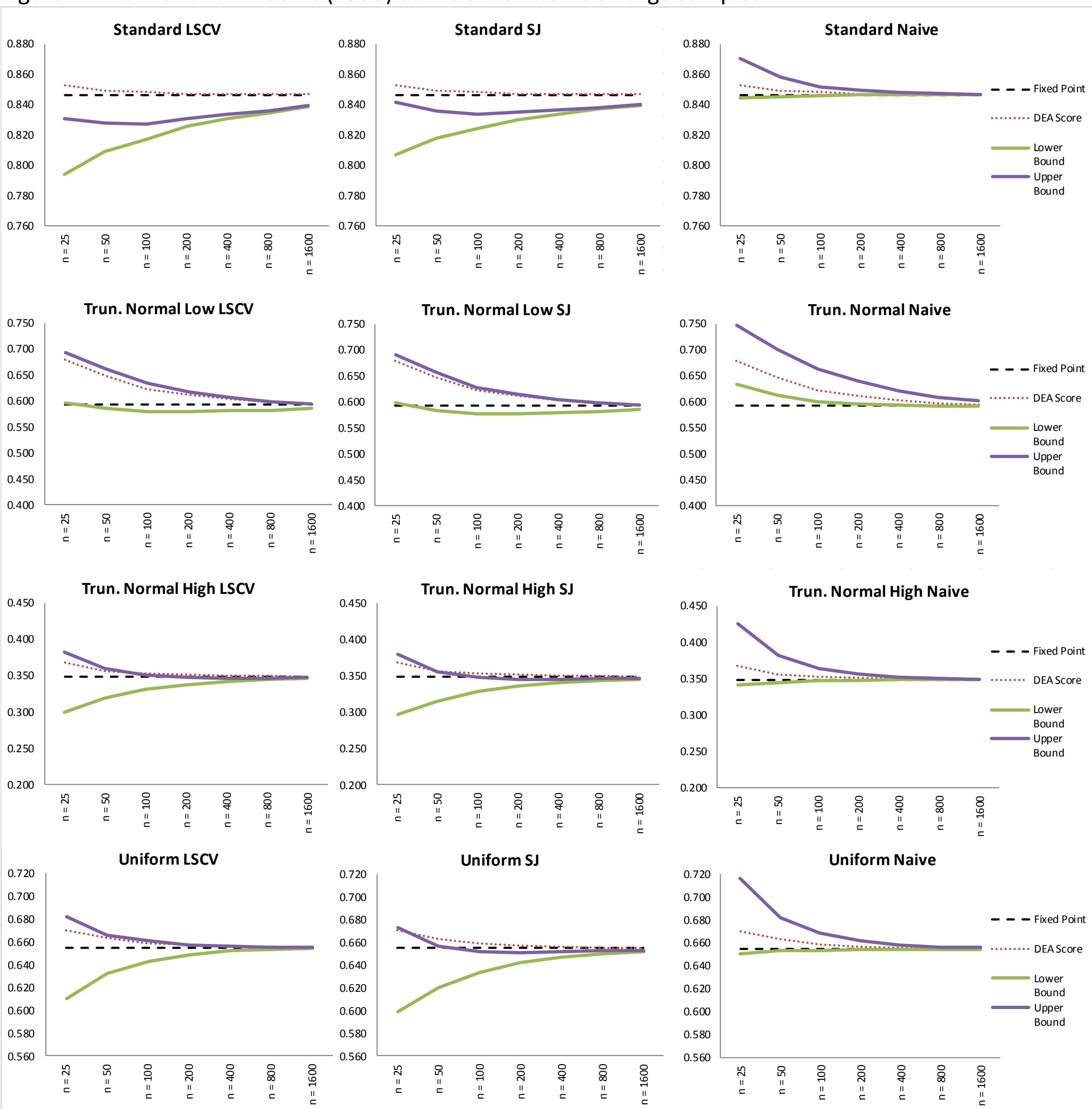
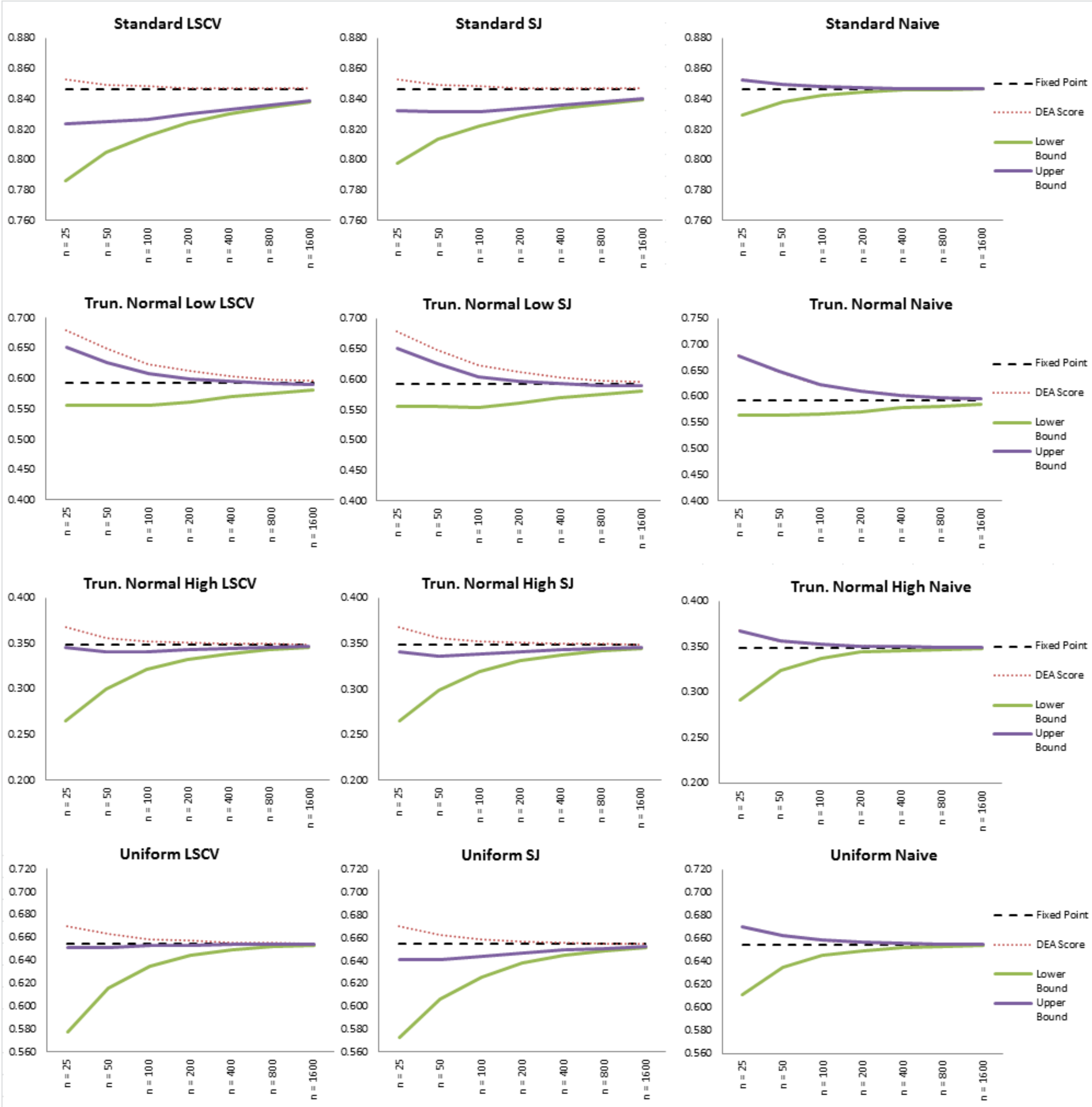


Figure VI.2. Simar and Wilson's (2000a) confidence intervals: large samples



VII. Appendix VII: Bias corrected and accelerated confidence intervals

Skewness may affect the validity of hypothesis testing and the performance of bootstrap DEA in general. As already mentioned in chapter 3, Simar and Wilson (1998) propose using the bias-corrected intervals of Efron (1982) and in an empirical illustration it is shown that the bias-corrected intervals are wider towards the upper bound (due to input orientation and positive skewness). However, Efron (1987) proposed a better technique for accounting for skewness: the bias corrected and accelerated intervals. In fact Efron's (1982) bias-corrected intervals (BC) are a special case of Efron's (1987) bias-corrected and accelerated intervals (BC_a) where the "acceleration parameter" is equal to zero. However, the estimation of the acceleration parameter can be very challenging when the problem in hand is complicated (Shao and Tu, 1995) as in the case of bootstrap DEA. In this appendix we outline some ideas on how the acceleration parameter could be computed, which comprises work in progress by the author.

Let us first explain how the BC_a intervals could be computed in the case of bootstrap DEA by employing a straight application from Efron (1987). The logic is similar with implementing the BC intervals: instead of using the SW1998 intervals $\left(\tilde{\theta}_k^{*,a/2}, \tilde{\theta}_k^{*,1-a/2}\right)$, two endpoints a_1 and a_2 are determined and the following intervals are estimated $\theta \in \left(\tilde{\theta}_k^{*,a_1}, \tilde{\theta}_k^{*,a_2}\right)$, where

$$a_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(a/_2)}}{1 - \hat{\alpha}\left(\hat{z}_0 + z^{(a/_2)}\right)}\right) \tag{VII.1}$$

and:

$$a_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-a/_2)}}{1 - \hat{\alpha}\left(\hat{z}_0 + z^{(1-a/_2)}\right)}\right) \tag{VII.2}$$

As explained in chapter 3, Φ is the standard normal cumulative density function and $z^{(a/_2)}$ is the normalized value that corresponds to the $\left(a/_2\right)^{\text{th}}$ percentile of the standard normal distribution, so that $\Phi\left(z^{(a/_2)}\right) = a/_2$. The parameter $\hat{z}_0$ is called the bias-correction parameter and is computed as $\hat{z}_0 = \Phi^{-1}\left[G\left(\tilde{\theta}_k^*\right)\right]$ where $G\left(\tilde{\theta}_k^*\right) = \Pr\left(\tilde{\theta}_k^* < \overline{\tilde{\theta}_k^*}\right)$.

We would like to note at this point that Efron (1987) suggests for a general estimator $\hat{\theta}$ that $G\left(\hat{\theta}\ \right) = \Pr\left(\hat{\theta}^* < \hat{\theta}\ \right)$; this involves the proportion of bootstrap estimates that are smaller the sample estimate. However in bootstrap DEA we know that by definition $\hat{\theta}^* > \hat{\theta}$ and hence the point $\overline{\tilde{\theta}_k^*}$ is chosen which serves as an estimator for θ_k, as in Simar and Wilson (1998). One may think that we could correct the bootstrap distribution once so that $\hat{\theta}_k^{*,c} = \hat{\theta}_k^* - \widehat{bias_k}$ which would centre the distribution on $\hat{\theta}_k$ [149]and therefore we could compute $G\left(\hat{\theta}_k\ \right) = \Pr\left(\hat{\theta}_k^{*,c} < \hat{\theta}_k\ \right)$ instead. However, it can be easily shown that $G\left(\tilde{\theta}_k^*\right) = \Pr\left(\tilde{\theta}_k^* < \overline{\tilde{\theta}_k^*}\right) = \Pr\left(\hat{\theta}_k^{*,c} < \hat{\theta}_k\ \right) = G\left(\hat{\theta}_k\ \right)$; we just need to observe that $\Pr\left(\tilde{\theta}_k^* < \overline{\tilde{\theta}_k^*}\right) = \Pr\left(\hat{\theta}_k^{*,c} - \widehat{bias_k} < \overline{\hat{\theta}_k^{*,c}} - \widehat{bias_k}\right) = \Pr\left(\hat{\theta}_k^{*,c} < \hat{\theta}_k\right)$. Hence, the choice of $\left(\tilde{\theta}_k^*\right) = \Pr\left(\tilde{\theta}_k^* < \overline{\tilde{\theta}_k^*}\right)$ by Simar and Wilson (1998) is appropriate for the estimation of the bias-correction parameter $\hat{z}_0$.

---

[149] Note that $\overline{\hat{\theta}_k^{*,c}} = \overline{\hat{\theta}_k^*} - \widehat{bias_k} = \overline{\hat{\theta}_k^*} - \left(\overline{\hat{\theta}_k^*} - \hat{\theta}_k\right) = \hat{\theta}_k$.

The acceleration parameter for the non-parametric case can be calculated in various ways[150] one of which involves using the jackknife. We will first explain how it can be computed in a general setup (non-specific to DEA) and we will try then to apply it on DEA. We follow closely the analysis in Efron and Tibshirani (1993; pp.186) and the interested reader may refer there for more information. Suppose that θ is estimated by the model $\hat{\theta} = s(\mathbf{x})$. Denote with $\mathbf{x}_{(i)}$ the original data with the i^{th} observation deleted and let $\hat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{(i)}$, where $\hat{\theta}_{(i)} = s\left(\mathbf{x}_{(i)}\right)$. Then the acceleration parameter can be estimated as:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right]^{3/2}} \tag{VII.3}$$

Note that the acceleration parameter, as opposed to the bias-correction parameter $\hat{z}_0$, is not computed on the basis of the bootstrap distribution but on the basis of the empirical distribution. Attaching an intuitive interpretation to the acceleration parameter is not straightforward. We could state though that it tries to capture the effect of skewness in the distribution of $\hat{\theta}$ on the estimation of bootstrap confidence intervals that have been generated using the empirical distribution $\hat{\theta}$. To some extent it measures how the standard error of $\hat{\theta}$ changes by moving along its distribution.

There is a challenge in applying this estimator on DEA: $\hat{\theta}_{(i)} = s\left(\mathbf{x}_{(i)}\right)$ cannot be estimated since it would require deleting DMU i to compute the efficiency score of DMU i, which is logically inconsistent. We propose two alternative approaches: either

[150] See Efron and Tibshirani (1993) for more information and in particular section 14.3 in pp.184.

applying the jackknife on the means of efficiency scores or using a form of leave-one-out cross validation to estimate the acceleration parameter for DMU k.

The first suggestion of applying jackknife on the means can be easily implemented; instead of using $\hat{\theta}_{(i)}$ one could use $\overline{\hat{\theta}_{(\imath)}}$, and instead of $\hat{\theta}_{(\cdot)}$ one could use $\overline{\hat{\theta}_{(\cdot)}}$. To be more specific, $\overline{\hat{\theta}_{(\imath)}}$ involves deleting the i^{th} DMU from the sample, applying DEA on the $n-1$ DMUs and calculating their mean, whereas $\overline{\hat{\theta}_{(\cdot)}} = \frac{1}{n}\sum_{i=1}^{n} \overline{\hat{\theta}_{(\imath)}}$ involves computing the mean of these means. The acceleration parameter would then be:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n}\left(\overline{\hat{\theta}_{(\cdot)}} - \overline{\hat{\theta}_{(\imath)}}\right)^{3}}{6\left[\sum_{i=1}^{n}\left(\overline{\hat{\theta}_{(\cdot)}} - \overline{\hat{\theta}_{(\imath)}}\right)^{2}\right]^{3/2}} \tag{VII.4}$$

What we find less attractive in this approach is that the estimated acceleration parameter is not specific to some DMU but to the whole dataset. This means that computing the BC_a intervals for each DMUs would involve all using the same acceleration parameter which does not seem ideal in the case of bootstrap DEA.

An alternative approach would be to use a form of leave-one-out cross validation (CV) which would return an acceleration parameter for each DMU. The idea here is that instead of $\hat{\theta}_{(i)}$ we could proceed with our analysis for some DMU k by deleting DMU $i \neq k$ which we denote as $\hat{\theta}_{k,(i)}$. And instead of using $\hat{\theta}_{(\cdot)}$, we propose using $\hat{\theta}_{k,(\cdot)} = \frac{1}{n-1}\sum_{i \neq k=1}^{n} \hat{\theta}_{k,(i)}$. This means that the acceleration parameter is now specific to each DMU, which seems to be more relevant for the case of bootstrap DEA where each DMU has its own bootstrap distribution and on which confidence intervals are estimated. Hence, the acceleration parameter could be estimated as:

$$\hat{\alpha}_k = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{k,(\cdot)} - \hat{\theta}_{k,(i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\theta}_{k,(\cdot)} - \hat{\theta}_{k,(i)}\right)^2\right]^{3/2}} \qquad \text{(VII.5)}$$

To summarise, in the presence of skewness it might be a good idea to consider confidence intervals which account for it. Despite Efron's (1982) intervals, suggested by Simar and Wilson (1998) provide median-corrected intervals, one would need to use Efron's (1987) BC_a intervals which account for skewness. However, for the case of bootstrap DEA they are not straightforward to apply and we therefore suggested two potential ways, although we favour the latter which employs cross validation. Some simulations would be required to estimate the benefit of employing this procedure while a deeper exploration on the suitability of the proposed estimator of the acceleration parameter would be necessary. This is work in progress of the author and it seems an interesting area of research with potential benefits for researchers and practitioners.

VIII. Appendix VIII: Truncating the moments bootstrap at 1

We explained in Chapter 3 that the moments bootstrap uses the sample moments of the empirical distribution of DEA scores to generate pseudo-population values which can be used to perform the bootstrap draws. It is possible that these values violate the requirement that $\theta \in (0,1]$[151], though not to a considerable extent, and we therefore proposed truncating the generated random numbers to satisfy $\theta \in (0,1]$. Approaches such as reflection, used in Simar and Wilson (1998), were avoided since it would impose a symmetric structure and perhaps introduce excess noise as in the case of the smooth bootstrap (Simar and Wilson, 2002). Furthermore it might not be possible to employ this technique under certain types of the Pearson family distributions.

In this section we provide evidence that truncating the pseudo-population does not affect results, especially in larger samples. To perform this task we used the DGPs in the Monte Carlo simulations to generate pseudo-populations with and without truncation. Then we computed the moments of the two pseudo-populations that corresponds to each DEA sample and DGP and we calculated their median absolute differences (MAD) which serves our comparison purposes.

Table VIII.1 reports the results of our comparison exercise. We observe that the absolute differences become very small as sample size increases and especially for $n = 120$ which is associated with converging coverage probabilities to their nominal

[151] However, we only observed a few cases that violated the upper bound.

values. The absolute differences are too small to change the characterisation of the Pearson Type of distribution. Focusing on $n = 120$ and under the 2-input/2-output cases we observe that the displacement of the mean is negligible, there is almost no excess variability introduced, while the distribution preserves its shape as evident from the small differences in skewness and kurtosis. Taking into account these results and the good behaviour of the coverage probabilities we conclude that truncating the pseudo-populations in the moments bootstrap does not affect the validity of the results.

Table VIII.1. Median Absolute Differences (MAD) of the two pseudo-populations

	Standard 1/1				**Standard 2/1**				**Standard 2/2**			
	Mean	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.008	0.004	0.022	0.052	0.016	0.007	0.029	0.078	0.016	0.007	0.029	0.077
n = 15	0.006	0.003	0.014	0.030	0.011	0.005	0.025	0.066	0.012	0.005	0.026	0.069
n = 20	0.004	0.002	0.011	0.025	0.010	0.005	0.022	0.055	0.011	0.005	0.024	0.060
n = 25	0.003	0.001	0.009	0.020	0.008	0.004	0.020	0.047	0.009	0.004	0.019	0.050
n = 30	0.003	0.001	0.008	0.018	0.008	0.003	0.018	0.043	0.008	0.004	0.020	0.048
n = 60	0.002	0.001	0.006	0.013	0.005	0.003	0.015	0.033	0.006	0.003	0.015	0.036
n = 120	0.003	0.001	0.008	0.016	0.004	0.002	0.013	0.030	0.004	0.002	0.013	0.030
	Trun. Normal Low 1/1				**Trun. Normal Low 2/1**				**Trun. Normal Low 2/2**			
	Mean	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.014	0.011	0.132	0.199	0.016	0.011	0.121	0.212	0.017	0.012	0.117	0.219
n = 15	0.009	0.008	0.121	0.152	0.012	0.009	0.116	0.185	0.013	0.010	0.118	0.183
n = 20	0.007	0.006	0.102	0.117	0.009	0.008	0.113	0.157	0.010	0.008	0.110	0.164
n = 25	0.006	0.005	0.097	0.110	0.009	0.007	0.109	0.136	0.009	0.007	0.106	0.144
n = 30	0.005	0.005	0.087	0.084	0.007	0.007	0.105	0.127	0.008	0.007	0.107	0.135
n = 60	0.003	0.003	0.066	0.050	0.005	0.004	0.083	0.085	0.005	0.005	0.088	0.084
n = 120	0.002	0.002	0.052	0.037	0.003	0.003	0.070	0.064	0.003	0.003	0.072	0.063
	Trun. Normal High 1/1				**Trun. Normal High 2/1**				**Trun. Normal High 2/2**			
	Mean	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.031	0.030	0.094	0.114	0.035	0.032	0.100	0.112	0.039	0.033	0.093	0.111
n = 15	0.020	0.020	0.068	0.074	0.027	0.024	0.067	0.075	0.028	0.024	0.073	0.076
n = 20	0.016	0.016	0.057	0.058	0.021	0.019	0.064	0.068	0.023	0.020	0.052	0.060
n = 25	0.014	0.014	0.054	0.049	0.019	0.018	0.055	0.058	0.020	0.018	0.046	0.048
n = 30	0.013	0.012	0.048	0.043	0.018	0.017	0.058	0.055	0.018	0.017	0.051	0.049
n = 60	0.009	0.009	0.043	0.044	0.013	0.013	0.054	0.052	0.013	0.012	0.043	0.035
n = 120	0.006	0.007	0.033	0.035	0.010	0.011	0.050	0.055	0.009	0.009	0.040	0.039
	Uniform 1/1				**Uniform 2/1**				**Uniform 2/2**			
	Mean	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.019	0.015	0.130	0.245	0.025	0.017	0.133	0.298	0.025	0.017	0.129	0.305
n = 15	0.014	0.011	0.108	0.183	0.020	0.014	0.105	0.230	0.020	0.014	0.106	0.229
n = 20	0.010	0.009	0.085	0.128	0.016	0.012	0.089	0.180	0.018	0.013	0.099	0.208
n = 25	0.009	0.007	0.070	0.108	0.014	0.010	0.079	0.149	0.015	0.011	0.095	0.181
n = 30	0.008	0.007	0.059	0.098	0.012	0.009	0.074	0.141	0.014	0.010	0.084	0.170
n = 60	0.005	0.004	0.040	0.056	0.009	0.006	0.053	0.092	0.009	0.007	0.061	0.102
n = 120	0.003	0.003	0.024	0.033	0.006	0.004	0.037	0.062	0.007	0.005	0.042	0.072

IX. Appendix IX : Population, sample and bootstrap moments

Table IX.1. Population, sample and bootstrap moments: Standard

	Standard 1/1				Standard 2/1				Standard 2/2			
Population	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
N = 10,000	0.857	0.098	-0.686	2.929	0.858	0.097	-0.683	2.946	0.859	0.097	-0.675	2.893
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.865	0.102	-0.392	2.181	0.888	0.098	-0.466	2.102	0.891	0.097	-0.494	2.117
n = 15	0.863	0.101	-0.450	2.279	0.881	0.098	-0.468	2.200	0.882	0.098	-0.472	2.185
n = 20	0.861	0.101	-0.467	2.303	0.878	0.098	-0.493	2.243	0.880	0.098	-0.501	2.235
n = 25	0.858	0.100	-0.535	2.419	0.875	0.099	-0.530	2.361	0.876	0.099	-0.536	2.344
n = 30	0.859	0.100	-0.518	2.412	0.873	0.099	-0.547	2.420	0.873	0.099	-0.519	2.370
n = 60	0.859	0.099	-0.611	2.627	0.867	0.098	-0.583	2.565	0.869	0.098	-0.597	2.579
n = 120	0.858	0.099	-0.644	2.744	0.864	0.098	-0.641	2.758	0.865	0.098	-0.650	2.753
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.930	0.109	-0.465	2.516	0.958	0.108	-0.440	2.399	0.959	0.108	-0.476	2.470
n = 15	0.916	0.107	-0.501	2.503	0.941	0.106	-0.455	2.411	0.943	0.107	-0.450	2.383
n = 20	0.908	0.106	-0.506	2.464	0.931	0.105	-0.475	2.399	0.933	0.105	-0.473	2.395
n = 25	0.903	0.105	-0.569	2.569	0.923	0.105	-0.527	2.499	0.927	0.106	-0.521	2.479
n = 30	0.900	0.104	-0.546	2.531	0.919	0.104	-0.537	2.542	0.921	0.105	-0.509	2.488
n = 60	0.891	0.102	-0.627	2.701	0.903	0.103	-0.582	2.623	0.905	0.103	-0.590	2.645
n = 120	0.881	0.101	-0.652	2.785	0.891	0.101	-0.638	2.796	0.893	0.101	-0.645	2.786
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.913	0.108	-0.465	2.516	0.942	0.107	-0.439	2.402	0.944	0.107	-0.468	2.476
n = 15	0.903	0.106	-0.501	2.503	0.928	0.105	-0.457	2.417	0.931	0.106	-0.449	2.381
n = 20	0.897	0.105	-0.506	2.464	0.920	0.104	-0.474	2.401	0.924	0.105	-0.469	2.399
n = 25	0.892	0.104	-0.569	2.569	0.915	0.104	-0.526	2.499	0.917	0.105	-0.519	2.478
n = 30	0.890	0.104	-0.546	2.531	0.911	0.104	-0.535	2.542	0.913	0.105	-0.507	2.488
n = 60	0.883	0.102	-0.627	2.701	0.896	0.102	-0.580	2.624	0.898	0.103	-0.590	2.645
n = 120	0.876	0.101	-0.652	2.785	0.886	0.101	-0.637	2.796	0.887	0.101	-0.645	2.786
Moments	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.881	0.104	-0.465	2.516	0.914	0.105	-0.427	2.410	0.919	0.105	-0.455	2.504
n = 15	0.875	0.103	-0.501	2.503	0.903	0.102	-0.450	2.420	0.906	0.103	-0.443	2.397
n = 20	0.871	0.102	-0.506	2.464	0.897	0.102	-0.471	2.410	0.900	0.103	-0.465	2.409
n = 25	0.868	0.101	-0.569	2.569	0.892	0.102	-0.523	2.497	0.895	0.103	-0.517	2.485
n = 30	0.867	0.101	-0.546	2.531	0.889	0.101	-0.535	2.544	0.891	0.102	-0.505	2.487
n = 60	0.864	0.100	-0.627	2.701	0.878	0.100	-0.578	2.625	0.880	0.101	-0.587	2.642
n = 120	0.861	0.099	-0.652	2.785	0.872	0.099	-0.634	2.798	0.873	0.099	-0.642	2.785
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.865	0.102	-0.465	2.516	0.898	0.103	-0.449	2.412	0.904	0.102	-0.483	2.491
n = 15	0.863	0.101	-0.501	2.503	0.890	0.101	-0.461	2.418	0.893	0.101	-0.461	2.385
n = 20	0.861	0.101	-0.506	2.464	0.885	0.100	-0.478	2.400	0.889	0.101	-0.481	2.411
n = 25	0.858	0.100	-0.569	2.569	0.881	0.101	-0.531	2.503	0.884	0.101	-0.526	2.485
n = 30	0.859	0.100	-0.546	2.531	0.879	0.100	-0.535	2.544	0.880	0.101	-0.511	2.488
n = 60	0.859	0.099	-0.627	2.701	0.871	0.099	-0.584	2.625	0.873	0.099	-0.590	2.650
n = 120	0.858	0.099	-0.652	2.785	0.866	0.099	-0.639	2.797	0.868	0.099	-0.647	2.788

Table IX.2. Population, sample and bootstrap moments: Trun. Normal Low

	Trun. Normal Low 1/1				Trun. Normal Low 2/1				Trun. Normal Low 2/2			
Population	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
N = 10,000	0.615	0.120	0.397	2.957	0.616	0.120	0.427	2.973	0.617	0.121	0.412	3.003
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.757	0.142	0.270	2.308	0.782	0.139	0.189	2.143	0.786	0.139	0.146	2.125
n = 15	0.730	0.140	0.341	2.461	0.752	0.138	0.275	2.322	0.759	0.138	0.240	2.266
n = 20	0.714	0.136	0.359	2.568	0.739	0.137	0.288	2.403	0.741	0.139	0.268	2.330
n = 25	0.702	0.135	0.395	2.610	0.724	0.137	0.306	2.435	0.730	0.138	0.270	2.407
n = 30	0.693	0.134	0.377	2.681	0.720	0.137	0.329	2.473	0.720	0.138	0.317	2.505
n = 60	0.667	0.130	0.401	2.846	0.691	0.133	0.374	2.667	0.688	0.134	0.356	2.676
n = 120	0.646	0.126	0.397	2.915	0.670	0.130	0.391	2.780	0.667	0.131	0.381	2.802
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.828	0.156	0.321	2.740	0.867	0.162	0.344	2.640	0.876	0.161	0.318	2.601
n = 15	0.789	0.152	0.379	2.765	0.824	0.156	0.397	2.723	0.834	0.157	0.367	2.643
n = 20	0.766	0.146	0.389	2.808	0.803	0.153	0.401	2.747	0.808	0.156	0.388	2.667
n = 25	0.751	0.145	0.420	2.804	0.784	0.152	0.408	2.740	0.791	0.154	0.368	2.704
n = 30	0.740	0.143	0.397	2.851	0.774	0.150	0.411	2.734	0.778	0.152	0.405	2.791
n = 60	0.702	0.137	0.411	2.940	0.736	0.142	0.413	2.840	0.735	0.144	0.409	2.877
n = 120	0.674	0.132	0.402	2.963	0.705	0.137	0.417	2.886	0.703	0.139	0.410	2.918
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.826	0.156	0.321	2.740	0.863	0.161	0.343	2.655	0.870	0.161	0.323	2.602
n = 15	0.790	0.151	0.379	2.765	0.825	0.156	0.398	2.734	0.832	0.157	0.369	2.648
n = 20	0.767	0.147	0.389	2.808	0.805	0.153	0.402	2.750	0.808	0.156	0.389	2.672
n = 25	0.751	0.146	0.420	2.804	0.785	0.152	0.407	2.743	0.792	0.154	0.371	2.699
n = 30	0.740	0.144	0.397	2.851	0.777	0.150	0.411	2.732	0.778	0.152	0.406	2.796
n = 60	0.705	0.137	0.411	2.940	0.737	0.143	0.413	2.838	0.735	0.144	0.408	2.877
n = 120	0.676	0.132	0.402	2.963	0.706	0.138	0.417	2.886	0.705	0.139	0.410	2.918
Moments	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.800	0.152	0.321	2.739	0.837	0.157	0.356	2.666	0.843	0.159	0.335	2.633
n = 15	0.772	0.148	0.381	2.758	0.807	0.153	0.411	2.755	0.813	0.154	0.382	2.663
n = 20	0.753	0.144	0.389	2.807	0.790	0.151	0.407	2.766	0.793	0.154	0.397	2.694
n = 25	0.740	0.143	0.424	2.799	0.774	0.150	0.418	2.756	0.780	0.152	0.381	2.714
n = 30	0.730	0.142	0.400	2.844	0.766	0.149	0.416	2.749	0.768	0.150	0.417	2.814
n = 60	0.700	0.136	0.414	2.938	0.732	0.142	0.421	2.848	0.731	0.144	0.413	2.898
n = 120	0.673	0.131	0.402	2.963	0.704	0.137	0.422	2.895	0.703	0.139	0.416	2.931
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.757	0.142	0.321	2.740	0.804	0.151	0.342	2.638	0.809	0.152	0.324	2.587
n = 15	0.730	0.140	0.379	2.765	0.775	0.147	0.393	2.732	0.783	0.149	0.374	2.642
n = 20	0.714	0.136	0.389	2.808	0.759	0.144	0.400	2.751	0.762	0.148	0.390	2.687
n = 25	0.702	0.135	0.420	2.804	0.743	0.144	0.409	2.741	0.751	0.146	0.377	2.707
n = 30	0.693	0.134	0.397	2.851	0.738	0.143	0.410	2.733	0.740	0.144	0.409	2.790
n = 60	0.667	0.130	0.411	2.940	0.706	0.137	0.414	2.836	0.705	0.138	0.406	2.880
n = 120	0.646	0.126	0.402	2.963	0.682	0.133	0.417	2.892	0.680	0.135	0.413	2.918

Table IX.3. Population, sample and bootstrap moments: Trun. Normal High

	Trun. Normal High 1/1				Trun. Normal High 2/1				Trun. Normal High 2/2			
Population	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
N = 10,000	0.495	0.238	0.278	2.095	0.490	0.239	0.326	2.144	0.493	0.241	0.284	2.074
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.535	0.269	0.459	2.084	0.547	0.274	0.394	1.993	0.555	0.278	0.347	1.921
n = 15	0.524	0.259	0.386	2.094	0.536	0.268	0.360	1.999	0.538	0.270	0.355	1.975
n = 20	0.520	0.255	0.372	2.133	0.525	0.262	0.370	2.074	0.534	0.267	0.303	1.947
n = 25	0.514	0.254	0.360	2.107	0.524	0.259	0.363	2.054	0.526	0.263	0.338	2.000
n = 30	0.510	0.250	0.361	2.119	0.517	0.257	0.367	2.089	0.524	0.258	0.321	2.026
n = 60	0.505	0.245	0.330	2.117	0.509	0.251	0.338	2.108	0.515	0.255	0.309	2.033
n = 120	0.497	0.241	0.298	2.108	0.503	0.246	0.338	2.122	0.504	0.248	0.305	2.066
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.630	0.318	0.544	2.344	0.662	0.341	0.535	2.337	0.676	0.348	0.493	2.181
n = 15	0.596	0.298	0.430	2.238	0.622	0.319	0.446	2.237	0.630	0.322	0.443	2.181
n = 20	0.580	0.287	0.403	2.243	0.598	0.303	0.442	2.274	0.613	0.310	0.370	2.085
n = 25	0.566	0.280	0.384	2.184	0.589	0.294	0.417	2.180	0.596	0.301	0.392	2.118
n = 30	0.557	0.275	0.380	2.183	0.578	0.288	0.420	2.197	0.584	0.293	0.368	2.131
n = 60	0.536	0.260	0.338	2.146	0.550	0.272	0.361	2.162	0.555	0.276	0.332	2.082
n = 120	0.519	0.251	0.302	2.121	0.529	0.260	0.347	2.150	0.531	0.263	0.316	2.090
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.635	0.319	0.544	2.344	0.664	0.342	0.537	2.336	0.674	0.348	0.495	2.190
n = 15	0.603	0.300	0.430	2.238	0.632	0.322	0.448	2.238	0.637	0.325	0.444	2.184
n = 20	0.588	0.289	0.403	2.243	0.606	0.306	0.442	2.276	0.619	0.312	0.372	2.088
n = 25	0.573	0.283	0.384	2.184	0.596	0.297	0.418	2.179	0.601	0.303	0.393	2.116
n = 30	0.563	0.277	0.380	2.183	0.582	0.292	0.421	2.199	0.591	0.294	0.368	2.131
n = 60	0.539	0.262	0.338	2.146	0.554	0.274	0.362	2.162	0.560	0.279	0.332	2.082
n = 120	0.521	0.252	0.302	2.121	0.534	0.262	0.347	2.150	0.535	0.264	0.316	2.090
Moments	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.591	0.300	0.563	2.291	0.619	0.321	0.552	2.296	0.629	0.325	0.513	2.174
n = 15	0.564	0.280	0.451	2.198	0.592	0.301	0.461	2.212	0.597	0.305	0.469	2.142
n = 20	0.552	0.271	0.427	2.199	0.574	0.289	0.465	2.225	0.583	0.294	0.390	2.068
n = 25	0.539	0.266	0.399	2.154	0.567	0.282	0.433	2.166	0.570	0.287	0.404	2.108
n = 30	0.535	0.262	0.401	2.152	0.556	0.279	0.437	2.185	0.563	0.281	0.383	2.117
n = 60	0.519	0.252	0.340	2.133	0.536	0.265	0.367	2.162	0.541	0.269	0.337	2.082
n = 120	0.506	0.245	0.303	2.118	0.521	0.256	0.350	2.152	0.523	0.258	0.318	2.093
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.535	0.269	0.544	2.344	0.566	0.292	0.530	2.314	0.577	0.299	0.487	2.176
n = 15	0.524	0.259	0.430	2.238	0.554	0.282	0.438	2.226	0.560	0.284	0.443	2.168
n = 20	0.520	0.255	0.403	2.243	0.541	0.272	0.438	2.261	0.551	0.278	0.365	2.078
n = 25	0.514	0.254	0.384	2.184	0.539	0.268	0.415	2.179	0.542	0.274	0.386	2.107
n = 30	0.510	0.250	0.380	2.183	0.529	0.265	0.418	2.194	0.537	0.267	0.365	2.125
n = 60	0.505	0.245	0.338	2.146	0.517	0.255	0.359	2.158	0.523	0.260	0.330	2.081
n = 120	0.497	0.241	0.302	2.121	0.507	0.249	0.347	2.149	0.510	0.252	0.316	2.089

Table IX.4. Population, sample and bootstrap moments: Uniform

	Uniform 1/1				Uniform 2/1				Uniform 2/2			
Population	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
N = 10,000	0.689	0.158	0.273	1.885	0.691	0.158	0.249	1.864	0.688	0.158	0.286	1.885
DEA	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.729	0.163	0.359	2.001	0.751	0.163	0.261	1.841	0.756	0.163	0.250	1.829
n = 15	0.712	0.162	0.377	2.002	0.737	0.165	0.265	1.855	0.741	0.164	0.256	1.855
n = 20	0.703	0.161	0.375	2.032	0.732	0.163	0.256	1.861	0.729	0.164	0.307	1.862
n = 25	0.701	0.160	0.354	1.985	0.725	0.163	0.264	1.857	0.724	0.163	0.304	1.898
n = 30	0.701	0.162	0.323	1.941	0.721	0.162	0.275	1.890	0.717	0.164	0.314	1.887
n = 60	0.694	0.159	0.303	1.938	0.709	0.161	0.269	1.882	0.707	0.162	0.293	1.888
n = 120	0.692	0.160	0.281	1.893	0.702	0.160	0.261	1.871	0.700	0.161	0.291	1.889
LSCV	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.801	0.180	0.425	2.198	0.839	0.189	0.407	2.077	0.845	0.189	0.410	2.062
n = 15	0.765	0.176	0.420	2.105	0.806	0.184	0.349	1.982	0.814	0.185	0.351	1.990
n = 20	0.749	0.172	0.406	2.110	0.789	0.178	0.321	1.958	0.789	0.180	0.376	1.966
n = 25	0.736	0.170	0.377	2.032	0.775	0.177	0.318	1.933	0.774	0.177	0.359	1.996
n = 30	0.733	0.170	0.340	1.971	0.763	0.174	0.322	1.961	0.761	0.176	0.365	1.966
n = 60	0.712	0.164	0.311	1.951	0.738	0.169	0.291	1.919	0.736	0.169	0.320	1.928
n = 120	0.701	0.162	0.285	1.898	0.718	0.165	0.272	1.887	0.717	0.165	0.303	1.909
SJ	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.805	0.181	0.425	2.198	0.842	0.190	0.409	2.077	0.847	0.190	0.417	2.068
n = 15	0.776	0.176	0.420	2.105	0.813	0.185	0.350	1.987	0.819	0.185	0.352	1.991
n = 20	0.756	0.173	0.406	2.110	0.799	0.179	0.318	1.959	0.797	0.182	0.377	1.970
n = 25	0.746	0.171	0.377	2.032	0.783	0.179	0.320	1.933	0.782	0.179	0.362	1.996
n = 30	0.742	0.172	0.340	1.971	0.773	0.175	0.325	1.962	0.772	0.178	0.365	1.969
n = 60	0.721	0.166	0.311	1.951	0.746	0.171	0.292	1.920	0.745	0.171	0.321	1.929
n = 120	0.710	0.164	0.285	1.898	0.727	0.166	0.273	1.887	0.725	0.167	0.303	1.909
Moments	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.769	0.174	0.425	2.198	0.807	0.183	0.416	2.104	0.815	0.183	0.430	2.107
n = 15	0.741	0.169	0.420	2.105	0.782	0.178	0.357	1.995	0.787	0.179	0.362	2.007
n = 20	0.727	0.166	0.406	2.110	0.768	0.174	0.325	1.969	0.769	0.176	0.382	1.988
n = 25	0.719	0.165	0.377	2.032	0.757	0.173	0.322	1.939	0.758	0.174	0.366	2.006
n = 30	0.717	0.165	0.340	1.971	0.750	0.170	0.327	1.967	0.750	0.173	0.367	1.981
n = 60	0.703	0.161	0.311	1.951	0.730	0.167	0.293	1.924	0.730	0.168	0.322	1.932
n = 120	0.697	0.161	0.285	1.898	0.716	0.164	0.273	1.888	0.715	0.164	0.304	1.912
Naïve	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**	**Mean**	**Std**	**Skew**	**Kurt**
n = 10	0.729	0.163	0.425	2.198	0.772	0.175	0.401	2.074	0.778	0.175	0.411	2.062
n = 15	0.712	0.162	0.420	2.105	0.753	0.171	0.344	1.979	0.760	0.171	0.345	1.990
n = 20	0.703	0.161	0.406	2.110	0.746	0.167	0.318	1.952	0.743	0.170	0.375	1.961
n = 25	0.701	0.160	0.377	2.032	0.737	0.168	0.315	1.933	0.736	0.168	0.360	1.992
n = 30	0.701	0.162	0.340	1.971	0.731	0.165	0.321	1.960	0.730	0.168	0.365	1.965
n = 60	0.694	0.159	0.311	1.951	0.717	0.163	0.290	1.919	0.715	0.164	0.317	1.927
n = 120	0.692	0.160	0.285	1.898	0.707	0.162	0.272	1.886	0.705	0.162	0.302	1.907

X. Appendix X: Coverage probabilities – Moments bootstrap

Table X.1. Coverage probabilities of moments-bootstrap – "Standard" case

	Standard 1/1				Standard 2/1				Standard 2/2			
SW1998	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.660	0.688	0.694	0.703	0.591	0.615	0.626	0.635	0.596	0.628	0.637	0.644
n = 15	0.652	0.692	0.713	0.743	0.663	0.710	0.715	0.722	0.646	0.713	0.727	0.737
n = 20	0.649	0.701	0.729	0.762	0.669	0.744	0.764	0.773	0.663	0.740	0.747	0.760
n = 25	0.659	0.708	0.731	0.775	0.686	0.765	0.786	0.797	0.685	0.761	0.779	0.794
n = 30	0.663	0.709	0.745	0.780	0.679	0.786	0.813	0.824	0.694	0.808	0.823	0.842
n = 60	0.681	0.719	0.750	0.785	0.719	0.830	0.853	0.880	0.695	0.831	0.866	0.890
n = 120	0.733	0.770	0.787	0.820	0.744	0.888	0.917	0.936	0.742	0.897	0.929	0.946
SW2000	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.629	0.735	0.792	0.872	0.594	0.715	0.800	0.893	0.601	0.722	0.806	0.892
n = 15	0.609	0.715	0.774	0.839	0.638	0.742	0.825	0.911	0.636	0.753	0.823	0.912
n = 20	0.579	0.671	0.733	0.791	0.657	0.762	0.835	0.920	0.638	0.751	0.825	0.915
n = 25	0.595	0.679	0.727	0.774	0.634	0.756	0.825	0.902	0.635	0.762	0.824	0.907
n = 30	0.591	0.687	0.735	0.776	0.658	0.754	0.832	0.914	0.630	0.764	0.842	0.927
n = 60	0.613	0.688	0.730	0.758	0.645	0.767	0.829	0.903	0.627	0.745	0.814	0.900
n = 120	0.649	0.719	0.754	0.792	0.669	0.768	0.837	0.908	0.615	0.749	0.817	0.896

Table X.2. Coverage probabilities of moments-bootstrap – "Truncated Normal Low" case

	Trun. Normal Low 1/1				Trun. Normal Low 2/1				Trun. Normal Low 2/2			
SW1998	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.348	0.359	0.363	0.366	0.281	0.305	0.312	0.326	0.295	0.321	0.337	0.347
n = 15	0.381	0.400	0.408	0.415	0.336	0.370	0.384	0.401	0.314	0.344	0.358	0.377
n = 20	0.399	0.427	0.434	0.441	0.360	0.399	0.423	0.435	0.370	0.408	0.417	0.426
n = 25	0.440	0.471	0.482	0.487	0.379	0.435	0.451	0.469	0.367	0.417	0.438	0.452
n = 30	0.442	0.482	0.489	0.499	0.365	0.430	0.450	0.468	0.377	0.438	0.466	0.486
n = 60	0.544	0.603	0.610	0.613	0.444	0.502	0.539	0.556	0.464	0.540	0.574	0.607
n = 120	0.600	0.676	0.690	0.699	0.526	0.614	0.645	0.670	0.568	0.632	0.674	0.703
SW2000	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.363	0.452	0.528	0.667	0.310	0.411	0.483	0.621	0.327	0.423	0.487	0.608
n = 15	0.405	0.491	0.557	0.678	0.359	0.460	0.534	0.645	0.323	0.420	0.492	0.623
n = 20	0.403	0.506	0.583	0.687	0.366	0.455	0.511	0.619	0.380	0.470	0.533	0.636
n = 25	0.442	0.531	0.605	0.730	0.404	0.498	0.549	0.676	0.382	0.467	0.534	0.657
n = 30	0.450	0.541	0.611	0.722	0.385	0.475	0.565	0.691	0.403	0.495	0.562	0.674
n = 60	0.531	0.634	0.705	0.798	0.455	0.547	0.619	0.745	0.459	0.579	0.640	0.742
n = 120	0.559	0.674	0.746	0.815	0.543	0.656	0.721	0.825	0.565	0.651	0.702	0.818

Table X.3. Coverage probabilities of moments-bootstrap – "Truncated Normal High" case

	Trun. Normal High 1/1				Trun. Normal High 2/1				Trun. Normal High 2/2			
SW1998	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.761	0.774	0.780	0.785	0.761	0.781	0.787	0.796	0.758	0.771	0.782	0.789
n = 15	0.764	0.783	0.794	0.805	0.769	0.791	0.796	0.809	0.774	0.802	0.813	0.822
n = 20	0.768	0.799	0.803	0.810	0.795	0.838	0.844	0.853	0.748	0.791	0.800	0.808
n = 25	0.780	0.808	0.814	0.819	0.808	0.856	0.866	0.877	0.763	0.809	0.818	0.825
n = 30	0.795	0.816	0.827	0.835	0.805	0.853	0.862	0.869	0.754	0.825	0.836	0.841
n = 60	0.818	0.849	0.859	0.866	0.810	0.903	0.918	0.925	0.785	0.879	0.885	0.893
n = 120	0.856	0.881	0.897	0.907	0.797	0.944	0.956	0.963	0.799	0.949	0.960	0.967
SW2000	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.736	0.850	0.920	0.970	0.730	0.838	0.915	0.971	0.751	0.843	0.909	0.962
n = 15	0.709	0.824	0.872	0.952	0.745	0.840	0.901	0.965	0.752	0.862	0.916	0.970
n = 20	0.735	0.833	0.894	0.959	0.753	0.861	0.915	0.974	0.722	0.838	0.913	0.965
n = 25	0.710	0.837	0.897	0.961	0.745	0.851	0.912	0.969	0.726	0.842	0.895	0.966
n = 30	0.711	0.836	0.884	0.944	0.741	0.852	0.911	0.969	0.718	0.834	0.901	0.966
n = 60	0.742	0.853	0.907	0.945	0.704	0.815	0.903	0.974	0.703	0.818	0.886	0.958
n = 120	0.781	0.882	0.922	0.949	0.664	0.788	0.881	0.959	0.700	0.814	0.880	0.959

Table X.4. Coverage probabilities of moments-bootstrap – "Uniform" case

	Uniform 1/1				Uniform 2/1				Uniform 2/2			
SW1998	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.743	0.758	0.760	0.765	0.727	0.749	0.759	0.769	0.672	0.696	0.702	0.712
n = 15	0.788	0.806	0.813	0.816	0.718	0.747	0.756	0.767	0.706	0.747	0.753	0.764
n = 20	0.787	0.808	0.819	0.827	0.721	0.765	0.774	0.787	0.738	0.795	0.809	0.815
n = 25	0.775	0.798	0.811	0.818	0.746	0.797	0.804	0.816	0.755	0.830	0.840	0.847
n = 30	0.771	0.800	0.815	0.829	0.720	0.789	0.800	0.812	0.752	0.838	0.847	0.858
n = 60	0.798	0.840	0.859	0.885	0.734	0.824	0.838	0.849	0.752	0.901	0.906	0.916
n = 120	0.764	0.793	0.828	0.859	0.790	0.884	0.902	0.915	0.728	0.915	0.930	0.946
SW2000	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**	**p = 0.20**	**p = 0.10**	**p = 0.05**	**p = 0.01**
n = 10	0.711	0.806	0.866	0.935	0.715	0.821	0.883	0.937	0.645	0.778	0.855	0.924
n = 15	0.699	0.805	0.866	0.942	0.720	0.799	0.879	0.938	0.689	0.792	0.864	0.944
n = 20	0.688	0.799	0.869	0.931	0.694	0.823	0.899	0.961	0.688	0.816	0.878	0.951
n = 25	0.686	0.779	0.845	0.917	0.710	0.835	0.890	0.963	0.704	0.813	0.884	0.964
n = 30	0.684	0.793	0.854	0.913	0.668	0.802	0.875	0.953	0.694	0.818	0.887	0.971
n = 60	0.670	0.769	0.824	0.867	0.664	0.803	0.878	0.948	0.650	0.774	0.860	0.958
n = 120	0.649	0.738	0.783	0.820	0.724	0.832	0.904	0.948	0.627	0.749	0.838	0.940

XI. Appendix XI: A note on the compatibility of Simar and Wilson's (1999) bootstrap Malmquist with unbalanced panels

The Malmquist index, as explained by Färe et al., (1994) in their seminal paper, can be applied on unbalanced panels but with the index being undefined for the missing observations (see footnote 14, pp. 73 of their paper). A reasonable implication is that the Bootstrap Malmquist Index of Simar and Wilson (1999) can cope with unbalanced panels as well. Simar and Wilson (1999) do comment on the applicability of their approach on unbalanced panels. However, according to the manual of the *FEAR* software package of Prof Paul Wilson, the bootstrap Malmquist index is presented as compatible with unbalanced panels, but "*with some small modifications*"[152]. In this note I will explain a potential problem with Simar and Wilson's (1999) bootstrap approach on the Malmquist index when dealing with unbalanced panels.

The approach of Simar and Wilson (1999) is an extension of the univariate case in Simar and Wilson (1998). In particular, smoothing is applied by fitting a bivariate (instead of univariate) kernel density to the efficiency score distributions of the two examined periods, which maintains the correlation structure between the DMUs in the two periods under examination when bootstrapping. Our understanding of Eq.18 through Eq.24 in Simar and Wilson (1999) is that to estimate a bivariate kernel density, to preserve the correlation structure as well as to reflect bootstrap values, all require

[152] Look at the last sentence of the "Details" section on pp. 39 in the *FEAR* manual: http://www.clemson.edu/economics/faculty/wilson/Software/FEAR/Compiled/2.0.1/FEAR-manual.pdf

$2N$-dimensional vectors; this is the first indication that their approach may not be able to deal with unbalanced panels appropriately.

It is not clear which are the “modifications” mentioned in the *FEAR* manual and whether these require performing smoothing and reflecting only on the common observations between two reference sets, while including the non-common observations in the computation of Malmquist indices. If the latter is true then we would expect some degree of bias due to possible errors in the computation of the smoothing parameter and of the covariance matrix.

Studying carefully Simar and Wilson’s (1999) work we find that that the code for bootstrapping the Malmquist index with unbalanced panels probably works as we have just suggested. To support our argument we will discuss four relevant parts from their paper. In section 3 and Eq.10 (pp. 462), the (naïve) bootstrap Malmquist indices can be easily adjusted to account for unbalanced panels: hence, this ensures the feasibility of the task. In section 3 and Eq. 19-21 (pp. 465) they perform reflection as follows (using the therein notation):

$$\boldsymbol{\Delta}_{(4N\times 2)} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ 2-\boldsymbol{A} & \boldsymbol{B} \\ 2-\boldsymbol{A} & 2-\boldsymbol{B} \\ \boldsymbol{A} & 2-\boldsymbol{B} \end{bmatrix} \text{ where} \tag{XI.1}$$

$$\boldsymbol{A} = \left(\widehat{D}_1^{t_1,t_1} \dots \widehat{D}_N^{t_1,t_1}\right)' \text{ and } \boldsymbol{B} = \left(\widehat{D}_1^{t_2,t_2} \dots \widehat{D}_N^{t_2,t_2}\right)'$$

Note that $\widehat{D}_i^{t_j,t_j}$ corresponds to the distance function estimated for DMU i for the reference set of period j. The fact that $\boldsymbol{\Delta}$ is a $(4N \times 2)$ matrix indicates that reflection is performed on the common elements of two reference sets.

Even if this can be modified, there is a third point in Simar and Wilson (1999) that suggests otherwise. In particular, to preserve the intertemporal correlation between two DMUs a covariance matrix is computed as $\widehat{\boldsymbol{\Sigma}} = \mathrm{cov}(\boldsymbol{A}, \boldsymbol{B})$, which is by definition a square matrix and requires both $\boldsymbol{A}$ and $\boldsymbol{B}$ to have the same dimensions. The intertemporal correlation is then accounted for in the following bivariate kernel density estimator (Eq.24):

$$\hat{g}(z) = \frac{1}{4Nh^2} \sum_{j=1}^{4N} K_j \left(\frac{\boldsymbol{z} - \boldsymbol{\Delta}_j}{h} \right) \tag{XI.2}$$

This suggests that the bootstrap procedure produces the smoothed bootstrap distribution on the basis of common observations, disregarding the non-common elements.

Finally, Simar and Wilson (1999) state in pp.466 that the smoothing parameter (h) is chosen by the approximation rule $h = (4/5\,N)^{1/6}$ which corresponds to the number of observations in each sample. Hence, if the panels are unbalanced, then the larger the size difference, the higher the degree of discrepancy in computing h would be. If, on the other hand, someone wanted to use a smoothing process (such as LSCV and SJ) or an alternative distribution enrichment approach (such as the moments bootstrap) it is not clear how this task could be performed.

The discussion here shows Simar and Wilson' (1999) method can accommodate unbalanced panels in the first step of computing the required distance functions for the computation of the Malmquist index. However, in implementing the bootstrap and generating bootstrap values, only the common observations are taken into account. Our understanding is that the processes of reflection, of random number generation from a bivariate kernel density (that accounts for intertemporal correlation) and of smoothing are all based on the “balanced” part of the dataset. This might cause inaccuracies in computing bootstrap Malmquist indices which will be more important as the number of non-common observations increases. It is within the future plans of the author to extend this note by including numerical examples which will illustrate the extent to which results can be affected by such discrepancies.

XII. Appendix XII: Moments and confidence intervals for the empirical illustration

This appendix provides analytical results for the distributional aspects and confidence intervals of the bootstrap DEA distributions for each bank in each year and under for each smoothing method (LSCV, SJ and moments bootstrap). Each table has three sections. The first lists the DMUs (banks per year of operation) and their DEA score. The second section reports the mean, median, standard deviation, skewness and kurtosis of the bias-corrected bootstrap distributions (the last three are the same as with the non-bias-corrected distributions as we shift the distribution twice to left for bias). The third section reports the 95% confidence intervals under the percentile method (adopted by Simar and Wilson (1998)) and under the bias-corrected and accelerated intervals method of Efron (1987) which was proposed and adapted in Appendix VII for bootstrap DEA. Although the analysis is not based on the BCa intervals, we present them here as we believe that there is a good potential for the enhancement of the performance of bootstrap DEA which needs to be confirmed by future research.

Table XII.1. Confidence intervals under the LSCV bootstrap

Bank	DEA Score	Eff. bc mean	Eff. bc median	Stand. Dev.	Skew.	Kurt.	SW98 Low	SW98 High	BCa Low	BCa High
Agricultural_1992	1.000	0.837	0.841	0.065	-0.313	2.891	0.696	0.954	0.689	0.939
Agricultural_1993	0.976	0.929	0.926	0.017	0.701	3.882	0.899	0.967	0.895	0.962
Agricultural_1994	0.927	0.871	0.869	0.020	0.699	3.935	0.836	0.918	0.830	0.912
Agricultural_1995	0.919	0.845	0.839	0.036	0.704	3.163	0.791	0.927	0.788	0.923
Agricultural_1996	0.906	0.838	0.832	0.035	0.703	3.189	0.785	0.917	0.783	0.912
Agricultural_1997	0.936	0.851	0.846	0.038	0.566	3.060	0.789	0.937	0.783	0.927
Agricultural_1998	0.937	0.809	0.814	0.059	-0.006	2.561	0.701	0.919	0.691	0.890
Agricultural_1999	1.000	0.833	0.843	0.073	-0.171	2.525	0.689	0.963	0.677	0.925
Alpha_1987	0.784	0.751	0.749	0.014	1.048	5.038	0.730	0.784	0.729	0.781
Alpha_1988	0.796	0.766	0.765	0.013	1.063	5.033	0.748	0.796	0.747	0.791
Alpha_1989	0.803	0.774	0.772	0.013	1.180	5.532	0.756	0.805	0.755	0.803
Alpha_1990	0.893	0.845	0.842	0.021	0.704	3.373	0.812	0.893	0.808	0.886
Alpha_1991	0.801	0.757	0.755	0.019	0.692	3.339	0.729	0.798	0.724	0.791
Alpha_1992	0.709	0.674	0.672	0.016	0.929	4.826	0.649	0.710	0.648	0.705
Alpha_1993	0.815	0.766	0.763	0.023	0.882	3.947	0.732	0.821	0.729	0.814
Alpha_1994	0.750	0.724	0.722	0.012	1.126	5.320	0.707	0.751	0.706	0.748
Alpha_1995	0.811	0.782	0.780	0.013	1.140	5.477	0.764	0.812	0.763	0.809
Alpha_1996	0.953	0.893	0.890	0.028	0.744	3.623	0.848	0.959	0.843	0.948
Alpha_1997	1.000	0.922	0.920	0.032	0.463	3.271	0.864	0.991	0.855	0.976
Alpha_1998	0.892	0.778	0.780	0.052	0.121	2.655	0.685	0.883	0.674	0.854
Alpha_1999	1.000	0.763	0.774	0.101	-0.380	2.734	0.549	0.939	0.542	0.920
Bank of Athens_1988	0.783	0.763	0.762	0.006	1.084	5.266	0.753	0.778	0.752	0.777
Bank of Athens_1989	0.805	0.784	0.783	0.007	1.207	5.629	0.775	0.801	0.774	0.799
Bank of Athens_1990	0.844	0.817	0.816	0.009	0.711	3.668	0.803	0.837	0.801	0.834
Bank of Athens_1991	0.855	0.776	0.777	0.030	0.132	2.998	0.718	0.835	0.707	0.820
Bank of Athens_1992	0.746	0.725	0.724	0.007	0.879	4.251	0.714	0.742	0.713	0.741
Bank of Athens_1993	0.733	0.712	0.711	0.007	1.012	4.533	0.701	0.729	0.701	0.729
Bank of Athens_1994	0.543	0.510	0.509	0.014	0.462	2.953	0.487	0.541	0.483	0.533
Bank of Athens_1995	0.635	0.603	0.601	0.013	0.921	4.164	0.583	0.636	0.580	0.630
Bank of Athens_1996	0.653	0.621	0.619	0.013	1.058	4.630	0.602	0.656	0.600	0.651
Bank of Athens_1997	0.753	0.685	0.680	0.036	0.715	3.148	0.632	0.770	0.630	0.763
Bank of Attica_1987	0.800	0.779	0.779	0.007	1.184	5.586	0.770	0.796	0.769	0.794
Bank of Attica_1988	0.742	0.724	0.723	0.006	1.176	5.645	0.715	0.738	0.715	0.737
Bank of Attica_1989	0.660	0.636	0.635	0.008	0.437	2.998	0.622	0.654	0.620	0.650
Bank of Attica_1990	0.744	0.726	0.725	0.007	1.090	5.126	0.716	0.741	0.715	0.739
Bank of Attica_1991	0.910	0.878	0.877	0.012	0.562	3.182	0.859	0.903	0.856	0.899
Bank of Attica_1992	1.000	0.847	0.858	0.051	-1.055	3.816	0.717	0.920	0.704	0.898
Bank of Attica_1993	0.945	0.891	0.889	0.022	0.551	3.567	0.851	0.940	0.849	0.934
Bank of Attica_1994	0.820	0.800	0.799	0.007	1.259	6.084	0.790	0.816	0.790	0.815
Bank of Attica_1995	0.793	0.734	0.737	0.019	-0.420	2.813	0.691	0.768	0.682	0.756
Bank of Attica_1996	0.747	0.692	0.693	0.019	-0.168	2.698	0.653	0.727	0.644	0.717
Bank of Attica_1997	0.810	0.761	0.759	0.017	0.364	3.072	0.730	0.798	0.724	0.795
Bank of Attica_1998	0.819	0.749	0.743	0.038	0.698	2.947	0.697	0.836	0.694	0.828
Bank of Attica_1999	0.764	0.702	0.701	0.028	0.369	2.770	0.656	0.762	0.649	0.747

Central Greece_1987	1.000	0.627	0.590	0.252	1.160	4.969	0.278	1.298	0.297	1.464
Central Greece_1989	0.706	0.685	0.684	0.007	0.744	3.544	0.674	0.702	0.674	0.700
Central Greece_1990	0.668	0.648	0.647	0.007	0.803	3.653	0.637	0.665	0.637	0.664
Central Greece_1991	0.667	0.635	0.635	0.009	0.076	3.105	0.617	0.653	0.613	0.650
Central Greece_1992	0.636	0.610	0.610	0.008	0.376	3.193	0.596	0.627	0.593	0.624
Central Greece_1993	0.705	0.666	0.663	0.017	0.811	3.791	0.640	0.706	0.636	0.700
Central Greece_1994	0.700	0.663	0.660	0.016	0.970	4.184	0.639	0.703	0.637	0.699
Central Greece_1995	0.660	0.629	0.628	0.012	0.661	3.951	0.608	0.656	0.606	0.653
Central Greece_1996	0.675	0.641	0.639	0.013	0.653	4.036	0.618	0.669	0.616	0.665
Central Greece_1997	0.626	0.575	0.571	0.025	0.655	2.987	0.537	0.631	0.535	0.628
Central Greece_1998	0.630	0.581	0.577	0.025	0.862	3.595	0.544	0.643	0.542	0.639
Cretabank_1987	0.655	0.640	0.639	0.005	1.241	6.059	0.632	0.652	0.632	0.652
Cretabank_1989	0.449	0.412	0.409	0.019	0.695	3.334	0.383	0.455	0.381	0.450
Cretabank_1990	0.526	0.496	0.494	0.014	1.077	4.887	0.475	0.529	0.473	0.523
Cretabank_1991	0.578	0.557	0.555	0.009	1.233	5.164	0.545	0.580	0.544	0.577
Cretabank_1992	0.643	0.620	0.618	0.009	1.154	4.844	0.607	0.644	0.606	0.643
Cretabank_1993	0.701	0.678	0.677	0.008	0.714	3.652	0.666	0.696	0.665	0.694
Cretabank_1994	0.589	0.558	0.556	0.014	0.794	3.520	0.537	0.591	0.535	0.588
Cretabank_1995	0.605	0.544	0.542	0.026	0.386	2.924	0.498	0.599	0.494	0.593
Cretabank_1996	0.726	0.701	0.700	0.009	0.717	3.314	0.687	0.722	0.686	0.721
Cretabank_1997	0.740	0.712	0.711	0.009	0.415	3.057	0.695	0.732	0.693	0.730
Cretabank_1998	0.814	0.772	0.772	0.014	0.311	2.955	0.747	0.802	0.742	0.798
Egnatia_1993	0.628	0.553	0.557	0.030	-0.246	2.648	0.492	0.608	0.483	0.587
Egnatia_1994	0.484	0.450	0.448	0.016	0.835	4.159	0.425	0.486	0.423	0.481
Egnatia_1995	0.470	0.433	0.431	0.020	0.701	3.336	0.404	0.478	0.403	0.471
Egnatia_1996	0.685	0.619	0.617	0.033	0.434	3.069	0.563	0.690	0.560	0.675
Egnatia_1997	0.779	0.711	0.709	0.033	0.622	3.329	0.660	0.784	0.655	0.774
Egnatia_1998	0.719	0.678	0.674	0.021	0.985	4.064	0.649	0.729	0.646	0.721
Egnatia_1999	0.715	0.669	0.665	0.022	0.863	3.685	0.636	0.720	0.632	0.713
Emporiki_1987	0.750	0.716	0.714	0.013	0.799	3.698	0.696	0.748	0.694	0.745
Emporiki_1988	0.738	0.717	0.715	0.008	1.090	4.989	0.706	0.735	0.705	0.733
Emporiki_1989	0.718	0.697	0.695	0.008	1.109	5.225	0.685	0.716	0.685	0.715
Emporiki_1990	0.753	0.727	0.725	0.010	1.008	4.634	0.712	0.750	0.712	0.748
Emporiki_1991	0.722	0.696	0.695	0.010	0.934	4.296	0.682	0.719	0.681	0.717
Emporiki_1992	0.860	0.813	0.811	0.019	0.731	3.695	0.784	0.858	0.778	0.848
Emporiki_1993	0.956	0.899	0.896	0.023	0.639	3.390	0.861	0.952	0.855	0.943
Emporiki_1994	0.928	0.867	0.865	0.025	0.488	3.141	0.825	0.921	0.820	0.914
Emporiki_1995	0.851	0.808	0.807	0.016	0.600	3.839	0.779	0.845	0.775	0.835
Emporiki_1996	0.790	0.767	0.766	0.008	0.855	3.901	0.754	0.786	0.754	0.785
Emporiki_1997	0.811	0.786	0.784	0.009	0.787	3.587	0.773	0.806	0.772	0.805
Emporiki_1998	0.769	0.732	0.731	0.013	0.415	3.257	0.710	0.760	0.706	0.754
Emporiki_1999	0.916	0.864	0.861	0.024	0.698	3.220	0.829	0.919	0.825	0.909

Ergobank_1987	0.683	0.649	0.647	0.014	0.728	4.084	0.626	0.680	0.624	0.674
Ergobank_1988	0.727	0.695	0.693	0.014	1.016	5.071	0.675	0.727	0.674	0.723
Ergobank_1989	0.688	0.645	0.644	0.017	0.451	3.457	0.614	0.681	0.610	0.673
Ergobank_1990	0.749	0.697	0.695	0.022	0.772	4.251	0.661	0.746	0.658	0.739
Ergobank_1991	0.875	0.826	0.822	0.025	0.862	3.855	0.789	0.885	0.787	0.876
Ergobank_1992	0.708	0.659	0.656	0.021	0.912	4.516	0.627	0.707	0.623	0.702
Ergobank_1993	0.658	0.619	0.617	0.017	0.850	4.527	0.592	0.658	0.590	0.654
Ergobank_1994	0.567	0.543	0.542	0.009	0.815	4.229	0.529	0.563	0.527	0.561
Ergobank_1995	0.600	0.576	0.576	0.008	0.586	3.732	0.563	0.594	0.561	0.591
Ergobank_1996	0.657	0.630	0.629	0.010	0.700	3.733	0.614	0.654	0.611	0.649
Ergobank_1997	0.723	0.682	0.681	0.016	0.356	3.122	0.654	0.715	0.649	0.707
Ergobank_1998	0.639	0.559	0.557	0.036	0.259	2.657	0.493	0.634	0.487	0.619
Ergobank_1999	0.682	0.608	0.607	0.034	0.262	2.694	0.546	0.676	0.542	0.663
Eurobank_1997	0.512	0.462	0.457	0.026	0.626	2.808	0.422	0.521	0.419	0.515
Eurobank_1998	0.990	0.867	0.857	0.079	0.467	2.268	0.759	1.026	0.755	1.003
Eurobank_1999	0.747	0.706	0.704	0.018	0.831	4.258	0.678	0.746	0.674	0.739
General_1987	0.731	0.707	0.706	0.008	0.591	3.427	0.694	0.724	0.693	0.722
General_1988	0.754	0.733	0.732	0.007	0.864	4.210	0.722	0.749	0.721	0.748
General_1989	0.782	0.756	0.755	0.009	0.440	3.412	0.740	0.774	0.739	0.771
General_1990	0.791	0.770	0.769	0.007	1.126	5.314	0.760	0.785	0.760	0.784
General_1991	0.690	0.663	0.663	0.008	0.267	3.160	0.647	0.680	0.645	0.677
General_1992	0.677	0.660	0.659	0.006	1.161	5.503	0.652	0.674	0.651	0.673
General_1993	0.577	0.561	0.560	0.005	1.032	4.654	0.553	0.573	0.553	0.573
General_1994	0.680	0.659	0.658	0.007	0.773	3.839	0.648	0.675	0.647	0.673
General_1995	0.779	0.753	0.752	0.009	0.616	3.409	0.738	0.773	0.736	0.769
General_1996	0.714	0.685	0.684	0.010	0.350	2.890	0.668	0.705	0.665	0.701
General_1997	0.716	0.691	0.691	0.008	0.592	3.319	0.678	0.710	0.676	0.707
General_1998	0.714	0.682	0.682	0.010	0.259	3.147	0.663	0.702	0.659	0.699
General_1999	0.793	0.757	0.757	0.010	0.134	3.136	0.738	0.778	0.733	0.774
Interbank_1995	0.558	0.538	0.537	0.008	0.950	4.058	0.527	0.557	0.526	0.555
Interbank_1996	0.557	0.527	0.526	0.013	0.790	3.653	0.507	0.557	0.506	0.554
Ionian and Popular_1987	0.780	0.721	0.718	0.027	0.549	3.004	0.678	0.779	0.671	0.768
Ionian and Popular_1988	0.790	0.751	0.750	0.014	0.594	3.641	0.727	0.785	0.722	0.775
Ionian and Popular_1989	0.725	0.691	0.690	0.014	0.517	3.172	0.669	0.722	0.665	0.714
Ionian and Popular_1990	0.761	0.735	0.734	0.009	0.755	3.974	0.720	0.756	0.718	0.752
Ionian and Popular_1991	0.807	0.757	0.753	0.023	0.918	3.801	0.723	0.815	0.719	0.805
Ionian and Popular_1992	0.846	0.803	0.801	0.016	0.911	4.403	0.779	0.841	0.773	0.835
Ionian and Popular_1993	0.748	0.696	0.693	0.023	0.549	3.099	0.659	0.747	0.653	0.737
Ionian and Popular_1994	1.000	0.897	0.899	0.040	-0.072	2.829	0.815	0.973	0.808	0.963
Ionian and Popular_1995	1.000	0.880	0.876	0.047	0.217	3.178	0.782	0.978	0.779	0.973
Ionian and Popular_1996	0.930	0.837	0.833	0.041	0.362	3.032	0.763	0.923	0.756	0.909
Ionian and Popular_1997	1.000	0.878	0.871	0.056	0.635	3.542	0.777	1.004	0.770	0.992
Ionian and Popular_1998	0.930	0.848	0.842	0.039	0.928	4.103	0.789	0.940	0.785	0.936

Laiki (Hellas)_1993	0.495	0.470	0.469	0.012	0.783	3.515	0.453	0.497	0.451	0.492
Laiki (Hellas)_1994	0.294	0.250	0.250	0.023	0.179	2.317	0.211	0.295	0.210	0.285
Laiki (Hellas)_1995	0.481	0.409	0.407	0.039	0.227	2.355	0.345	0.485	0.341	0.470
Laiki (Hellas)_1996	0.704	0.562	0.572	0.072	-0.083	2.131	0.437	0.690	0.428	0.657
Laiki (Hellas)_1997	0.864	0.759	0.757	0.046	0.381	3.022	0.678	0.860	0.671	0.844
Laiki (Hellas)_1998	0.798	0.684	0.680	0.059	0.382	2.591	0.589	0.808	0.581	0.786
Laiki (Hellas)_1999	0.857	0.738	0.733	0.059	0.406	2.656	0.640	0.861	0.630	0.840
Macedonia Thrace_1987	0.786	0.753	0.752	0.010	0.456	3.540	0.735	0.775	0.732	0.772
Macedonia Thrace_1988	0.741	0.712	0.712	0.009	0.275	3.146	0.696	0.730	0.692	0.727
Macedonia Thrace_1989	0.681	0.655	0.655	0.008	0.421	3.500	0.641	0.671	0.639	0.669
Macedonia Thrace_1990	0.744	0.723	0.722	0.006	1.042	4.683	0.714	0.739	0.713	0.738
Macedonia Thrace_1991	0.603	0.580	0.579	0.008	0.381	3.021	0.566	0.597	0.564	0.594
Macedonia Thrace_1992	0.694	0.675	0.674	0.007	1.026	4.904	0.664	0.692	0.663	0.688
Macedonia Thrace_1993	0.683	0.662	0.661	0.007	0.868	3.894	0.652	0.679	0.651	0.679
Macedonia Thrace_1994	0.591	0.566	0.566	0.007	0.271	3.142	0.554	0.581	0.550	0.579
Macedonia Thrace_1995	0.619	0.580	0.579	0.017	0.493	3.105	0.552	0.616	0.549	0.608
Macedonia Thrace_1996	0.662	0.638	0.636	0.010	0.733	3.397	0.623	0.659	0.622	0.658
Macedonia Thrace_1997	0.635	0.612	0.611	0.008	0.642	3.206	0.599	0.631	0.599	0.629
Macedonia Thrace_1998	0.635	0.610	0.609	0.008	0.513	3.506	0.596	0.627	0.595	0.625
Macedonia Thrace_1999	0.733	0.702	0.701	0.013	0.753	3.435	0.683	0.734	0.681	0.728
National_1987	0.723	0.644	0.641	0.036	0.359	2.599	0.582	0.721	0.576	0.712
National_1988	0.664	0.625	0.621	0.020	1.040	4.344	0.597	0.672	0.596	0.667
National_1989	0.679	0.631	0.627	0.023	0.891	4.020	0.596	0.683	0.595	0.681
National_1990	0.674	0.620	0.617	0.023	0.673	3.501	0.581	0.672	0.577	0.668
National_1991	0.628	0.574	0.568	0.029	0.655	2.785	0.531	0.639	0.529	0.633
National_1992	0.850	0.776	0.773	0.036	0.624	3.468	0.719	0.855	0.712	0.838
National_1993	1.000	0.805	0.813	0.081	-0.296	2.690	0.636	0.946	0.628	0.931
National_1994	0.913	0.796	0.792	0.059	0.333	2.445	0.702	0.914	0.690	0.893
National_1995	0.909	0.808	0.802	0.055	0.518	2.587	0.727	0.922	0.721	0.914
National_1996	0.817	0.738	0.734	0.041	0.593	2.887	0.676	0.827	0.672	0.823
National_1997	1.000	0.864	0.858	0.068	0.435	2.653	0.750	1.005	0.746	1.001
National_1998	0.962	0.866	0.858	0.047	0.686	3.262	0.789	0.970	0.788	0.969
National_1999	1.000	0.900	0.894	0.045	0.638	3.478	0.824	0.999	0.833	1.018
Piraeus_1987	0.748	0.715	0.712	0.014	0.897	4.269	0.693	0.748	0.690	0.742
Piraeus_1988	0.788	0.764	0.763	0.009	0.949	4.593	0.751	0.786	0.749	0.782
Piraeus_1989	0.747	0.722	0.721	0.010	0.862	4.245	0.706	0.745	0.705	0.740
Piraeus_1990	0.828	0.798	0.797	0.012	0.844	4.194	0.779	0.826	0.777	0.821
Piraeus_1991	0.706	0.687	0.686	0.007	0.892	4.117	0.676	0.702	0.676	0.702
Piraeus_1992	0.758	0.726	0.724	0.012	0.830	4.353	0.705	0.754	0.702	0.749
Piraeus_1993	0.870	0.820	0.818	0.020	0.556	3.532	0.785	0.865	0.783	0.861
Piraeus_1994	0.899	0.875	0.874	0.008	1.062	5.150	0.864	0.893	0.863	0.891
Piraeus_1995	0.946	0.908	0.907	0.012	0.433	3.211	0.887	0.933	0.882	0.929
Piraeus_1996	0.768	0.740	0.738	0.011	0.778	3.503	0.724	0.765	0.723	0.763
Piraeus_1997	0.924	0.870	0.869	0.020	0.376	3.061	0.835	0.913	0.828	0.904
Piraeus_1998	0.966	0.823	0.832	0.068	-0.135	2.255	0.701	0.944	0.690	0.905
Piraeus_1999	0.780	0.702	0.702	0.035	0.126	2.578	0.639	0.767	0.632	0.754

T Bank_1993	0.225	0.195	0.195	0.015	0.123	2.318	0.169	0.224	0.167	0.216
T Bank_1994	0.590	0.569	0.568	0.008	1.024	4.289	0.557	0.589	0.557	0.589
T Bank_1995	0.867	0.836	0.835	0.011	0.444	3.056	0.817	0.858	0.815	0.855
T Bank_1996	0.760	0.693	0.690	0.030	0.591	3.489	0.642	0.759	0.637	0.747
T Bank_1997	0.749	0.660	0.654	0.048	0.403	2.420	0.586	0.756	0.580	0.740
T Bank_1998	0.758	0.653	0.647	0.059	0.395	2.348	0.564	0.775	0.558	0.750
T Bank_1999	0.708	0.567	0.574	0.079	0.071	2.134	0.441	0.710	0.433	0.671
Xiosbank_1991	0.561	0.545	0.544	0.006	0.758	3.919	0.536	0.557	0.535	0.555
Xiosbank_1992	0.846	0.773	0.768	0.032	0.756	3.958	0.720	0.846	0.716	0.840
Xiosbank_1993	0.639	0.618	0.617	0.009	1.126	5.116	0.606	0.639	0.606	0.638
Xiosbank_1994	0.466	0.440	0.436	0.014	1.091	4.088	0.422	0.473	0.421	0.472
Xiosbank_1995	0.499	0.455	0.451	0.024	0.615	2.816	0.422	0.506	0.418	0.498
Xiosbank_1996	0.597	0.556	0.553	0.019	0.864	3.677	0.528	0.602	0.524	0.597
Xiosbank_1997	0.700	0.653	0.649	0.020	0.796	3.590	0.621	0.702	0.617	0.697
Xiosbank_1998	0.667	0.577	0.572	0.045	0.365	2.654	0.501	0.671	0.496	0.662
Average_1987	0.710	0.660	0.658	0.022	0.405	2.936	0.621	0.708	0.618	0.696
Average_1988	0.681	0.648	0.647	0.013	0.738	3.978	0.626	0.677	0.623	0.673
Average_1989	0.681	0.653	0.652	0.011	1.069	5.076	0.637	0.680	0.636	0.678
Average_1990	0.704	0.672	0.670	0.012	0.772	4.105	0.651	0.702	0.648	0.695
Average_1991	0.668	0.638	0.637	0.012	0.817	3.796	0.619	0.667	0.618	0.664
Average_1992	0.741	0.717	0.716	0.009	1.087	5.042	0.704	0.739	0.704	0.737
Average_1993	0.830	0.777	0.775	0.020	0.563	3.449	0.743	0.821	0.738	0.813
Average_1994	0.782	0.743	0.741	0.015	0.724	3.721	0.720	0.776	0.717	0.773
Average_1995	0.769	0.731	0.729	0.015	0.740	3.559	0.708	0.766	0.704	0.764
Average_1996	0.748	0.712	0.710	0.016	0.716	3.347	0.688	0.749	0.686	0.744
Average_1997	0.806	0.760	0.757	0.020	0.669	3.218	0.727	0.806	0.725	0.799
Average_1998	0.748	0.672	0.674	0.039	0.072	2.180	0.606	0.746	0.603	0.724
Average_1999	0.812	0.723	0.726	0.044	0.070	2.204	0.648	0.806	0.642	0.783
Average W_1987	0.718	0.648	0.646	0.033	0.390	2.628	0.594	0.715	0.589	0.707
Average W_1988	0.657	0.616	0.614	0.019	0.866	3.875	0.588	0.662	0.586	0.656
Average W_1989	0.664	0.615	0.612	0.021	0.750	3.743	0.580	0.663	0.577	0.657
Average W_1990	0.677	0.627	0.625	0.022	0.511	3.225	0.590	0.673	0.585	0.666
Average W_1991	0.639	0.594	0.589	0.024	0.734	2.919	0.561	0.648	0.559	0.643
Average W_1992	0.733	0.681	0.681	0.020	0.395	3.178	0.646	0.725	0.638	0.713
Average W_1993	0.937	0.848	0.843	0.040	0.504	2.942	0.780	0.934	0.772	0.922
Average W_1994	0.843	0.779	0.775	0.028	0.665	3.274	0.734	0.843	0.733	0.842
Average W_1995	0.816	0.747	0.744	0.028	0.493	3.210	0.698	0.809	0.692	0.803
Average W_1996	0.760	0.700	0.698	0.025	0.462	3.258	0.655	0.753	0.650	0.747
Average W_1997	0.885	0.806	0.802	0.034	0.519	3.204	0.747	0.881	0.742	0.874
Average W_1998	0.857	0.775	0.771	0.037	0.575	3.376	0.714	0.856	0.705	0.840
Average W_1999	0.922	0.837	0.834	0.041	0.525	3.079	0.771	0.927	0.763	0.907

Table XII.2. Confidence intervals under the SJ bootstrap

Bank	DEA Score	Eff. bc mean	Eff. bc median	Stand. Dev.	Skew.	Kurt.	SW98 Low	SW98 High	BCa Low	BCa High
Agricultural_1992	1.000	0.831	0.836	0.065	-0.241	3.051	0.686	0.953	0.676	0.927
Agricultural_1993	0.976	0.928	0.926	0.017	0.681	3.941	0.898	0.968	0.894	0.962
Agricultural_1994	0.927	0.870	0.868	0.020	0.714	4.098	0.835	0.916	0.830	0.911
Agricultural_1995	0.919	0.845	0.839	0.037	0.689	3.293	0.789	0.926	0.785	0.920
Agricultural_1996	0.906	0.837	0.832	0.035	0.708	3.347	0.784	0.916	0.783	0.909
Agricultural_1997	0.936	0.851	0.846	0.039	0.516	3.046	0.786	0.935	0.781	0.931
Agricultural_1998	0.937	0.809	0.813	0.058	-0.059	2.590	0.702	0.918	0.690	0.883
Agricultural_1999	1.000	0.832	0.841	0.070	-0.230	2.633	0.687	0.958	0.674	0.921
Alpha_1987	0.784	0.751	0.749	0.014	0.855	4.301	0.729	0.782	0.728	0.778
Alpha_1988	0.796	0.767	0.765	0.013	0.942	4.251	0.749	0.797	0.748	0.793
Alpha_1989	0.803	0.774	0.772	0.013	1.009	4.733	0.756	0.803	0.755	0.800
Alpha_1990	0.893	0.845	0.842	0.021	0.644	3.325	0.812	0.893	0.808	0.884
Alpha_1991	0.801	0.757	0.755	0.018	0.622	3.294	0.728	0.798	0.723	0.792
Alpha_1992	0.709	0.674	0.672	0.016	0.763	3.722	0.649	0.711	0.648	0.708
Alpha_1993	0.815	0.767	0.763	0.024	0.830	3.704	0.731	0.823	0.728	0.812
Alpha_1994	0.750	0.724	0.722	0.012	0.973	4.420	0.708	0.750	0.707	0.748
Alpha_1995	0.811	0.782	0.781	0.013	0.953	4.571	0.764	0.811	0.763	0.808
Alpha_1996	0.953	0.893	0.890	0.028	0.707	3.520	0.847	0.959	0.843	0.944
Alpha_1997	1.000	0.922	0.920	0.032	0.405	3.129	0.866	0.992	0.854	0.974
Alpha_1998	0.892	0.778	0.779	0.053	0.062	2.388	0.684	0.881	0.673	0.857
Alpha_1999	1.000	0.763	0.777	0.102	-0.416	2.605	0.550	0.936	0.544	0.914
Bank of Athens_1988	0.783	0.763	0.762	0.006	0.951	4.238	0.754	0.778	0.753	0.777
Bank of Athens_1989	0.805	0.785	0.784	0.006	1.018	4.375	0.775	0.800	0.775	0.799
Bank of Athens_1990	0.844	0.818	0.817	0.009	0.549	3.180	0.803	0.836	0.801	0.833
Bank of Athens_1991	0.855	0.775	0.775	0.031	0.121	3.040	0.714	0.837	0.703	0.820
Bank of Athens_1992	0.746	0.726	0.725	0.007	0.691	3.552	0.714	0.741	0.713	0.739
Bank of Athens_1993	0.733	0.712	0.711	0.007	0.933	4.225	0.701	0.729	0.701	0.728
Bank of Athens_1994	0.543	0.510	0.510	0.014	0.326	2.829	0.486	0.540	0.483	0.532
Bank of Athens_1995	0.635	0.604	0.602	0.013	0.849	4.191	0.583	0.633	0.580	0.630
Bank of Athens_1996	0.653	0.622	0.620	0.014	1.112	5.150	0.601	0.654	0.599	0.650
Bank of Athens_1997	0.753	0.686	0.680	0.036	0.751	3.235	0.633	0.768	0.630	0.761
Bank of Attica_1987	0.800	0.780	0.779	0.006	1.008	4.391	0.770	0.795	0.770	0.795
Bank of Attica_1988	0.742	0.725	0.724	0.006	0.981	4.266	0.716	0.738	0.716	0.737
Bank of Attica_1989	0.660	0.636	0.636	0.008	0.363	2.847	0.622	0.654	0.620	0.651
Bank of Attica_1990	0.744	0.726	0.725	0.006	0.917	4.065	0.717	0.741	0.716	0.739
Bank of Attica_1991	0.910	0.878	0.878	0.012	0.462	2.899	0.860	0.903	0.857	0.898
Bank of Attica_1992	1.000	0.846	0.858	0.051	-1.111	3.951	0.713	0.921	0.704	0.898
Bank of Attica_1993	0.945	0.892	0.891	0.022	0.491	3.409	0.855	0.938	0.851	0.931
Bank of Attica_1994	0.820	0.800	0.799	0.007	1.112	4.689	0.791	0.816	0.790	0.815
Bank of Attica_1995	0.793	0.735	0.737	0.020	-0.365	2.734	0.692	0.768	0.685	0.758
Bank of Attica_1996	0.747	0.692	0.694	0.020	-0.151	2.607	0.653	0.729	0.646	0.719
Bank of Attica_1997	0.810	0.761	0.759	0.018	0.520	3.525	0.727	0.802	0.723	0.796
Bank of Attica_1998	0.819	0.750	0.742	0.038	0.738	3.088	0.697	0.836	0.693	0.828
Bank of Attica_1999	0.764	0.702	0.701	0.027	0.301	2.639	0.656	0.759	0.649	0.747

Central Greece_1987	1.000	0.639	0.608	0.246	1.213	5.358	0.302	1.265	0.318	1.440
Central Greece_1989	0.706	0.685	0.684	0.007	0.782	3.831	0.674	0.702	0.674	0.701
Central Greece_1990	0.668	0.648	0.647	0.007	0.820	3.829	0.637	0.665	0.637	0.664
Central Greece_1991	0.667	0.636	0.636	0.009	0.155	3.378	0.617	0.656	0.612	0.651
Central Greece_1992	0.636	0.610	0.610	0.008	0.454	3.428	0.596	0.627	0.593	0.625
Central Greece_1993	0.705	0.666	0.665	0.016	0.665	3.551	0.640	0.703	0.636	0.696
Central Greece_1994	0.700	0.664	0.661	0.016	0.919	4.270	0.640	0.701	0.637	0.696
Central Greece_1995	0.660	0.629	0.628	0.012	0.787	4.199	0.610	0.657	0.607	0.653
Central Greece_1996	0.675	0.640	0.639	0.013	0.829	4.579	0.619	0.672	0.615	0.667
Central Greece_1997	0.626	0.574	0.570	0.026	0.713	3.209	0.535	0.633	0.534	0.631
Central Greece_1998	0.630	0.582	0.577	0.026	0.895	3.745	0.544	0.641	0.543	0.637
Cretabank_1987	0.655	0.640	0.639	0.005	1.071	4.634	0.633	0.652	0.633	0.652
Cretabank_1989	0.449	0.411	0.408	0.020	0.784	3.492	0.382	0.456	0.381	0.453
Cretabank_1990	0.526	0.495	0.493	0.015	1.009	4.307	0.474	0.530	0.472	0.526
Cretabank_1991	0.578	0.557	0.555	0.010	1.246	5.139	0.544	0.581	0.543	0.578
Cretabank_1992	0.643	0.620	0.618	0.010	1.152	5.027	0.605	0.645	0.604	0.642
Cretabank_1993	0.701	0.678	0.677	0.008	0.807	3.884	0.665	0.697	0.664	0.695
Cretabank_1994	0.589	0.557	0.554	0.015	0.989	4.042	0.536	0.595	0.535	0.591
Cretabank_1995	0.605	0.543	0.540	0.027	0.483	3.194	0.494	0.603	0.490	0.595
Cretabank_1996	0.726	0.701	0.700	0.009	0.744	3.636	0.687	0.723	0.685	0.720
Cretabank_1997	0.740	0.712	0.711	0.010	0.450	3.155	0.696	0.732	0.693	0.730
Cretabank_1998	0.814	0.773	0.772	0.015	0.255	3.005	0.745	0.804	0.739	0.797
Egnatia_1993	0.628	0.553	0.557	0.030	-0.280	2.650	0.491	0.606	0.483	0.587
Egnatia_1994	0.484	0.450	0.447	0.017	0.829	3.871	0.425	0.489	0.423	0.485
Egnatia_1995	0.470	0.433	0.430	0.021	0.715	3.305	0.402	0.481	0.401	0.473
Egnatia_1996	0.685	0.618	0.615	0.033	0.505	3.151	0.561	0.689	0.557	0.676
Egnatia_1997	0.779	0.710	0.706	0.034	0.675	3.445	0.657	0.787	0.653	0.779
Egnatia_1998	0.719	0.676	0.672	0.023	0.933	3.630	0.645	0.729	0.644	0.725
Egnatia_1999	0.715	0.667	0.663	0.024	0.951	3.808	0.634	0.727	0.629	0.720
Emporiki_1987	0.750	0.717	0.715	0.014	0.815	3.956	0.695	0.747	0.694	0.745
Emporiki_1988	0.738	0.717	0.716	0.008	0.931	3.978	0.707	0.736	0.705	0.733
Emporiki_1989	0.718	0.697	0.696	0.008	0.972	4.397	0.685	0.716	0.685	0.715
Emporiki_1990	0.753	0.727	0.726	0.010	0.929	4.302	0.713	0.750	0.712	0.749
Emporiki_1991	0.722	0.696	0.695	0.010	0.903	4.218	0.682	0.720	0.681	0.717
Emporiki_1992	0.860	0.814	0.812	0.018	0.725	3.754	0.786	0.855	0.779	0.848
Emporiki_1993	0.956	0.900	0.898	0.023	0.600	3.391	0.862	0.950	0.856	0.942
Emporiki_1994	0.928	0.868	0.866	0.025	0.455	3.143	0.826	0.921	0.821	0.913
Emporiki_1995	0.851	0.809	0.808	0.015	0.574	3.838	0.781	0.843	0.776	0.834
Emporiki_1996	0.790	0.767	0.766	0.008	0.857	3.967	0.755	0.785	0.754	0.785
Emporiki_1997	0.811	0.786	0.785	0.009	0.814	3.771	0.773	0.806	0.772	0.805
Emporiki_1998	0.769	0.733	0.732	0.013	0.353	3.085	0.709	0.761	0.706	0.756
Emporiki_1999	0.916	0.865	0.863	0.023	0.705	3.540	0.830	0.918	0.825	0.906

Ergobank_1987	0.683	0.649	0.648	0.014	0.670	3.707	0.626	0.681	0.622	0.673
Ergobank_1988	0.727	0.696	0.694	0.014	0.888	4.275	0.675	0.726	0.674	0.722
Ergobank_1989	0.688	0.645	0.644	0.017	0.538	3.380	0.615	0.683	0.610	0.676
Ergobank_1990	0.749	0.697	0.695	0.022	0.632	3.421	0.661	0.749	0.656	0.740
Ergobank_1991	0.875	0.826	0.823	0.025	0.837	3.769	0.789	0.885	0.786	0.874
Ergobank_1992	0.708	0.660	0.657	0.021	0.755	3.692	0.625	0.709	0.621	0.703
Ergobank_1993	0.658	0.620	0.618	0.018	0.687	3.497	0.592	0.660	0.590	0.655
Ergobank_1994	0.567	0.543	0.542	0.009	0.785	4.073	0.528	0.563	0.527	0.560
Ergobank_1995	0.600	0.576	0.575	0.008	0.600	3.753	0.562	0.593	0.560	0.590
Ergobank_1996	0.657	0.629	0.628	0.010	0.705	3.934	0.612	0.653	0.609	0.648
Ergobank_1997	0.723	0.682	0.681	0.016	0.315	2.960	0.653	0.715	0.649	0.708
Ergobank_1998	0.639	0.559	0.557	0.037	0.247	2.722	0.495	0.633	0.486	0.618
Ergobank_1999	0.682	0.607	0.605	0.033	0.375	2.983	0.547	0.676	0.541	0.665
Eurobank_1997	0.512	0.461	0.457	0.027	0.667	3.125	0.421	0.521	0.417	0.513
Eurobank_1998	0.990	0.868	0.853	0.080	0.488	2.214	0.760	1.029	0.755	1.017
Eurobank_1999	0.747	0.706	0.705	0.017	0.713	3.681	0.678	0.745	0.675	0.739
General_1987	0.731	0.707	0.706	0.008	0.647	3.755	0.694	0.725	0.693	0.723
General_1988	0.754	0.734	0.733	0.007	0.785	3.853	0.723	0.750	0.722	0.748
General_1989	0.782	0.756	0.756	0.008	0.453	3.302	0.742	0.774	0.740	0.771
General_1990	0.791	0.770	0.769	0.006	0.955	4.220	0.761	0.785	0.761	0.784
General_1991	0.690	0.664	0.663	0.008	0.272	3.093	0.648	0.681	0.645	0.678
General_1992	0.677	0.660	0.659	0.005	1.020	4.417	0.652	0.673	0.652	0.673
General_1993	0.577	0.561	0.560	0.005	0.958	4.286	0.553	0.573	0.553	0.573
General_1994	0.680	0.660	0.659	0.007	0.587	3.255	0.649	0.675	0.647	0.672
General_1995	0.779	0.753	0.753	0.009	0.491	3.014	0.739	0.772	0.736	0.769
General_1996	0.714	0.685	0.685	0.010	0.291	2.781	0.668	0.705	0.665	0.701
General_1997	0.716	0.692	0.691	0.008	0.476	2.972	0.679	0.709	0.676	0.706
General_1998	0.714	0.682	0.682	0.010	0.357	3.367	0.664	0.702	0.659	0.699
General_1999	0.793	0.758	0.757	0.010	0.294	3.449	0.739	0.779	0.733	0.775
Interbank_1995	0.558	0.538	0.536	0.008	1.073	4.526	0.526	0.558	0.526	0.557
Interbank_1996	0.557	0.527	0.524	0.014	1.008	4.248	0.507	0.561	0.506	0.558
Ionian and Popular_1987	0.780	0.721	0.719	0.027	0.564	3.129	0.678	0.783	0.671	0.770
Ionian and Popular_1988	0.790	0.752	0.751	0.014	0.515	3.666	0.726	0.781	0.722	0.774
Ionian and Popular_1989	0.725	0.691	0.690	0.014	0.501	3.105	0.669	0.720	0.665	0.714
Ionian and Popular_1990	0.761	0.735	0.734	0.009	0.774	3.938	0.721	0.756	0.719	0.754
Ionian and Popular_1991	0.807	0.756	0.751	0.025	1.028	4.045	0.720	0.821	0.718	0.814
Ionian and Popular_1992	0.846	0.804	0.802	0.016	0.796	4.085	0.780	0.840	0.775	0.833
Ionian and Popular_1993	0.748	0.697	0.694	0.022	0.703	3.527	0.661	0.747	0.656	0.739
Ionian and Popular_1994	1.000	0.898	0.899	0.039	-0.032	2.953	0.819	0.975	0.811	0.964
Ionian and Popular_1995	1.000	0.881	0.877	0.048	0.395	3.513	0.786	0.987	0.781	0.980
Ionian and Popular_1996	0.930	0.838	0.835	0.042	0.570	3.525	0.764	0.934	0.757	0.916
Ionian and Popular_1997	1.000	0.877	0.871	0.058	0.779	3.917	0.778	1.013	0.768	0.993
Ionian and Popular_1998	0.930	0.848	0.841	0.041	1.053	4.479	0.787	0.947	0.786	0.944

Laiki (Hellas)_1993	0.495	0.470	0.469	0.012	0.688	3.389	0.453	0.496	0.451	0.493
Laiki (Hellas)_1994	0.294	0.250	0.250	0.024	0.199	2.143	0.212	0.295	0.210	0.288
Laiki (Hellas)_1995	0.481	0.409	0.405	0.040	0.265	2.267	0.344	0.486	0.341	0.476
Laiki (Hellas)_1996	0.704	0.564	0.573	0.073	-0.061	2.067	0.438	0.693	0.432	0.660
Laiki (Hellas)_1997	0.864	0.761	0.760	0.048	0.249	2.757	0.674	0.857	0.668	0.841
Laiki (Hellas)_1998	0.798	0.684	0.681	0.061	0.372	2.622	0.587	0.809	0.579	0.788
Laiki (Hellas)_1999	0.857	0.737	0.731	0.060	0.434	2.727	0.637	0.864	0.629	0.845
Macedonia Thrace_1987	0.786	0.753	0.752	0.010	0.498	3.377	0.735	0.775	0.732	0.773
Macedonia Thrace_1988	0.741	0.712	0.712	0.009	0.407	3.385	0.697	0.732	0.694	0.728
Macedonia Thrace_1989	0.681	0.655	0.655	0.008	0.463	3.572	0.642	0.672	0.639	0.670
Macedonia Thrace_1990	0.744	0.724	0.723	0.006	0.879	4.076	0.714	0.739	0.714	0.738
Macedonia Thrace_1991	0.603	0.580	0.579	0.008	0.430	3.056	0.567	0.597	0.565	0.595
Macedonia Thrace_1992	0.694	0.675	0.674	0.007	0.922	4.082	0.665	0.691	0.664	0.688
Macedonia Thrace_1993	0.683	0.663	0.661	0.007	0.865	3.988	0.652	0.679	0.651	0.679
Macedonia Thrace_1994	0.591	0.567	0.567	0.007	0.285	3.188	0.554	0.582	0.551	0.580
Macedonia Thrace_1995	0.619	0.579	0.578	0.017	0.543	3.153	0.551	0.618	0.548	0.609
Macedonia Thrace_1996	0.662	0.637	0.636	0.010	0.775	3.651	0.623	0.660	0.622	0.658
Macedonia Thrace_1997	0.635	0.612	0.611	0.008	0.654	3.463	0.599	0.630	0.598	0.628
Macedonia Thrace_1998	0.635	0.610	0.609	0.008	0.562	3.381	0.596	0.628	0.594	0.626
Macedonia Thrace_1999	0.733	0.703	0.701	0.013	0.761	3.369	0.683	0.734	0.681	0.730
National_1987	0.723	0.644	0.641	0.037	0.262	2.405	0.580	0.715	0.574	0.708
National_1988	0.664	0.625	0.621	0.020	0.992	4.058	0.597	0.674	0.595	0.667
National_1989	0.679	0.631	0.627	0.023	0.724	3.448	0.594	0.683	0.592	0.677
National_1990	0.674	0.619	0.617	0.023	0.553	3.191	0.581	0.670	0.575	0.663
National_1991	0.628	0.573	0.568	0.030	0.707	3.124	0.529	0.639	0.526	0.632
National_1992	0.850	0.775	0.772	0.036	0.523	2.945	0.718	0.854	0.711	0.838
National_1993	1.000	0.805	0.811	0.080	-0.274	2.830	0.632	0.954	0.627	0.936
National_1994	0.913	0.797	0.796	0.060	0.372	2.655	0.701	0.919	0.690	0.891
National_1995	0.909	0.808	0.800	0.056	0.640	3.011	0.726	0.930	0.721	0.920
National_1996	0.817	0.739	0.733	0.043	0.777	3.518	0.677	0.837	0.675	0.829
National_1997	1.000	0.865	0.854	0.069	0.557	2.982	0.751	1.013	0.750	1.011
National_1998	0.962	0.866	0.859	0.048	0.787	3.720	0.790	0.977	0.789	0.972
National_1999	1.000	0.900	0.894	0.047	0.684	3.699	0.820	1.003	0.832	1.028
Piraeus_1987	0.748	0.715	0.714	0.014	0.848	4.155	0.694	0.748	0.691	0.741
Piraeus_1988	0.788	0.765	0.764	0.009	0.963	4.348	0.752	0.786	0.751	0.783
Piraeus_1989	0.747	0.722	0.721	0.010	0.822	4.100	0.707	0.746	0.706	0.741
Piraeus_1990	0.828	0.799	0.797	0.012	0.838	4.175	0.780	0.826	0.778	0.821
Piraeus_1991	0.706	0.687	0.686	0.007	0.886	4.086	0.677	0.702	0.677	0.702
Piraeus_1992	0.758	0.726	0.724	0.012	0.773	4.231	0.705	0.753	0.703	0.748
Piraeus_1993	0.870	0.821	0.820	0.020	0.513	3.493	0.787	0.863	0.785	0.857
Piraeus_1994	0.899	0.876	0.874	0.007	0.898	4.084	0.865	0.893	0.864	0.892
Piraeus_1995	0.946	0.909	0.908	0.012	0.338	2.980	0.887	0.934	0.882	0.930
Piraeus_1996	0.768	0.740	0.738	0.011	0.888	4.044	0.724	0.766	0.723	0.764
Piraeus_1997	0.924	0.870	0.869	0.020	0.313	2.914	0.836	0.912	0.827	0.904
Piraeus_1998	0.966	0.822	0.832	0.068	-0.179	2.205	0.698	0.936	0.688	0.906
Piraeus_1999	0.780	0.702	0.702	0.035	0.115	2.654	0.638	0.771	0.631	0.752

T Bank_1993	0.225	0.195	0.195	0.015	0.133	2.327	0.170	0.222	0.167	0.216
T Bank_1994	0.590	0.569	0.567	0.009	1.005	4.284	0.557	0.590	0.557	0.590
T Bank_1995	0.867	0.836	0.835	0.010	0.493	3.092	0.819	0.859	0.816	0.855
T Bank_1996	0.760	0.691	0.689	0.032	0.572	3.380	0.637	0.763	0.630	0.749
T Bank_1997	0.749	0.657	0.652	0.049	0.332	2.273	0.580	0.755	0.573	0.739
T Bank_1998	0.758	0.651	0.647	0.060	0.350	2.313	0.560	0.767	0.552	0.747
T Bank_1999	0.708	0.566	0.574	0.079	0.105	2.278	0.441	0.715	0.431	0.665
Xiosbank_1991	0.561	0.545	0.544	0.005	0.725	3.661	0.536	0.557	0.536	0.556
Xiosbank_1992	0.846	0.773	0.768	0.033	0.767	3.851	0.717	0.851	0.715	0.844
Xiosbank_1993	0.639	0.619	0.617	0.009	1.071	4.867	0.607	0.639	0.606	0.637
Xiosbank_1994	0.466	0.439	0.435	0.014	1.064	4.042	0.419	0.473	0.418	0.471
Xiosbank_1995	0.499	0.453	0.450	0.025	0.563	2.603	0.419	0.506	0.415	0.500
Xiosbank_1996	0.597	0.555	0.551	0.021	0.900	3.717	0.524	0.604	0.521	0.600
Xiosbank_1997	0.700	0.652	0.648	0.021	0.852	4.018	0.617	0.701	0.615	0.698
Xiosbank_1998	0.667	0.577	0.574	0.043	0.331	2.776	0.500	0.667	0.495	0.656
Average_1987	0.710	0.660	0.658	0.022	0.367	2.779	0.621	0.707	0.618	0.698
Average_1988	0.681	0.649	0.647	0.013	0.653	3.656	0.627	0.677	0.624	0.672
Average_1989	0.681	0.653	0.652	0.011	0.878	4.332	0.636	0.679	0.635	0.676
Average_1990	0.704	0.672	0.671	0.012	0.688	3.901	0.652	0.699	0.648	0.693
Average_1991	0.668	0.638	0.636	0.012	0.810	4.012	0.618	0.666	0.617	0.663
Average_1992	0.741	0.717	0.716	0.009	0.962	4.388	0.704	0.739	0.703	0.737
Average_1993	0.830	0.778	0.776	0.019	0.602	3.808	0.743	0.820	0.736	0.812
Average_1994	0.782	0.744	0.742	0.014	0.795	4.209	0.720	0.776	0.716	0.773
Average_1995	0.769	0.731	0.729	0.015	0.828	3.880	0.707	0.767	0.705	0.764
Average_1996	0.748	0.713	0.711	0.016	0.767	3.502	0.688	0.750	0.686	0.745
Average_1997	0.806	0.760	0.758	0.020	0.758	3.675	0.729	0.808	0.726	0.799
Average_1998	0.748	0.671	0.672	0.038	0.051	2.295	0.604	0.742	0.601	0.724
Average_1999	0.812	0.722	0.725	0.043	0.029	2.300	0.646	0.804	0.640	0.781
Average W_1987	0.718	0.647	0.645	0.033	0.284	2.411	0.592	0.710	0.586	0.704
Average W_1988	0.657	0.616	0.613	0.019	0.823	3.677	0.588	0.663	0.584	0.654
Average W_1989	0.664	0.615	0.613	0.021	0.643	3.391	0.580	0.664	0.575	0.656
Average W_1990	0.677	0.626	0.625	0.021	0.452	3.109	0.591	0.674	0.584	0.662
Average W_1991	0.639	0.593	0.589	0.024	0.776	3.290	0.558	0.648	0.556	0.644
Average W_1992	0.733	0.682	0.680	0.020	0.342	3.110	0.645	0.724	0.638	0.715
Average W_1993	0.937	0.849	0.844	0.041	0.534	3.039	0.781	0.938	0.773	0.927
Average W_1994	0.843	0.780	0.777	0.028	0.776	3.677	0.737	0.845	0.737	0.844
Average W_1995	0.816	0.748	0.746	0.029	0.510	3.586	0.695	0.811	0.690	0.801
Average W_1996	0.760	0.701	0.700	0.025	0.481	3.457	0.657	0.755	0.652	0.744
Average W_1997	0.885	0.807	0.803	0.035	0.674	3.813	0.747	0.886	0.744	0.878
Average W_1998	0.857	0.774	0.772	0.038	0.520	3.282	0.709	0.860	0.700	0.839
Average W_1999	0.922	0.836	0.835	0.042	0.453	2.948	0.768	0.929	0.760	0.901

Table XII.3. Confidence intervals under the Moments bootstrap

Bank	DEA Score	Eff. bc mean	Eff. bc median	Stand. Dev.	Skew.	Kurt.	SW98 Low	SW98 High	BCa Low	BCa High
Agricultural_1992	1.000	0.812	0.816	0.066	-0.247	2.947	0.670	0.931	0.637	0.909
Agricultural_1993	0.976	0.912	0.911	0.024	0.107	2.796	0.865	0.958	0.853	0.950
Agricultural_1994	0.927	0.853	0.853	0.027	0.104	2.709	0.805	0.907	0.785	0.896
Agricultural_1995	0.919	0.823	0.821	0.039	0.087	2.661	0.749	0.901	0.735	0.889
Agricultural_1996	0.906	0.816	0.815	0.038	0.052	2.645	0.739	0.890	0.731	0.878
Agricultural_1997	0.936	0.828	0.827	0.041	0.181	2.767	0.753	0.911	0.739	0.897
Agricultural_1998	0.937	0.787	0.791	0.058	-0.155	2.830	0.664	0.894	0.642	0.866
Agricultural_1999	1.000	0.810	0.817	0.071	-0.241	2.807	0.664	0.937	0.624	0.901
Alpha_1987	0.784	0.738	0.736	0.021	0.445	3.046	0.702	0.782	0.699	0.777
Alpha_1988	0.796	0.756	0.754	0.021	0.533	3.001	0.723	0.803	0.719	0.794
Alpha_1989	0.803	0.763	0.761	0.021	0.606	3.140	0.730	0.809	0.727	0.803
Alpha_1990	0.893	0.828	0.827	0.029	0.281	2.740	0.776	0.887	0.767	0.874
Alpha_1991	0.801	0.742	0.742	0.026	0.277	2.753	0.697	0.795	0.688	0.783
Alpha_1992	0.709	0.662	0.660	0.023	0.412	3.063	0.622	0.712	0.618	0.704
Alpha_1993	0.815	0.749	0.747	0.031	0.350	2.745	0.695	0.814	0.687	0.801
Alpha_1994	0.750	0.714	0.712	0.020	0.573	3.038	0.683	0.758	0.681	0.750
Alpha_1995	0.811	0.771	0.769	0.021	0.580	3.101	0.738	0.817	0.735	0.811
Alpha_1996	0.953	0.872	0.870	0.036	0.250	2.667	0.808	0.946	0.794	0.929
Alpha_1997	1.000	0.899	0.898	0.038	0.065	2.825	0.824	0.977	0.804	0.957
Alpha_1998	0.892	0.758	0.761	0.053	-0.091	2.835	0.650	0.858	0.626	0.831
Alpha_1999	1.000	0.744	0.754	0.101	-0.335	2.730	0.525	0.923	0.515	0.909
Bank of Athens_1988	0.783	0.758	0.756	0.013	0.870	3.767	0.740	0.788	0.739	0.785
Bank of Athens_1989	0.805	0.779	0.777	0.013	0.873	3.751	0.761	0.811	0.760	0.809
Bank of Athens_1990	0.844	0.811	0.810	0.015	0.590	3.227	0.787	0.844	0.784	0.839
Bank of Athens_1991	0.855	0.758	0.758	0.035	0.061	3.074	0.689	0.826	0.666	0.809
Bank of Athens_1992	0.746	0.721	0.719	0.013	0.759	3.543	0.702	0.750	0.700	0.746
Bank of Athens_1993	0.733	0.705	0.704	0.014	0.591	3.111	0.684	0.736	0.682	0.732
Bank of Athens_1994	0.543	0.498	0.498	0.018	0.041	2.671	0.464	0.533	0.456	0.525
Bank of Athens_1995	0.635	0.591	0.590	0.019	0.310	2.793	0.557	0.630	0.550	0.622
Bank of Athens_1996	0.653	0.608	0.607	0.019	0.381	2.754	0.576	0.649	0.568	0.642
Bank of Athens_1997	0.753	0.667	0.666	0.037	0.204	2.704	0.602	0.742	0.585	0.728
Bank of Attica_1987	0.800	0.775	0.773	0.013	0.868	3.752	0.756	0.806	0.755	0.804
Bank of Attica_1988	0.742	0.720	0.719	0.012	0.949	3.953	0.704	0.749	0.703	0.746
Bank of Attica_1989	0.660	0.630	0.629	0.013	0.349	2.947	0.607	0.657	0.604	0.652
Bank of Attica_1990	0.744	0.722	0.720	0.012	0.917	3.841	0.704	0.751	0.704	0.748
Bank of Attica_1991	0.910	0.870	0.869	0.018	0.404	2.970	0.839	0.909	0.835	0.900
Bank of Attica_1992	1.000	0.826	0.833	0.051	-0.958	4.363	0.683	0.905	0.657	0.892
Bank of Attica_1993	0.945	0.875	0.874	0.030	0.247	2.914	0.820	0.936	0.811	0.926
Bank of Attica_1994	0.820	0.795	0.792	0.014	0.921	3.785	0.776	0.828	0.775	0.825
Bank of Attica_1995	0.793	0.723	0.724	0.023	-0.294	3.227	0.675	0.766	0.657	0.753
Bank of Attica_1996	0.747	0.680	0.681	0.023	-0.183	3.118	0.630	0.724	0.615	0.712
Bank of Attica_1997	0.810	0.746	0.745	0.023	0.300	3.268	0.706	0.794	0.697	0.790
Bank of Attica_1998	0.819	0.730	0.729	0.039	0.201	2.657	0.661	0.809	0.647	0.795
Bank of Attica_1999	0.764	0.683	0.683	0.031	0.043	2.797	0.622	0.745	0.605	0.730

Central Greece_1987	1.000	0.620	0.577	0.246	1.097	4.482	0.281	1.243	0.313	1.415
Central Greece_1989	0.706	0.679	0.677	0.013	0.478	2.971	0.658	0.706	0.657	0.704
Central Greece_1990	0.668	0.642	0.641	0.013	0.445	2.855	0.621	0.668	0.619	0.665
Central Greece_1991	0.667	0.627	0.627	0.014	0.131	2.988	0.602	0.654	0.593	0.649
Central Greece_1992	0.636	0.602	0.602	0.012	0.274	3.143	0.580	0.628	0.575	0.624
Central Greece_1993	0.705	0.651	0.649	0.022	0.269	2.783	0.611	0.696	0.603	0.689
Central Greece_1994	0.700	0.649	0.648	0.022	0.315	2.729	0.611	0.695	0.603	0.684
Central Greece_1995	0.660	0.617	0.616	0.018	0.326	2.811	0.587	0.655	0.580	0.648
Central Greece_1996	0.675	0.628	0.626	0.019	0.366	2.953	0.595	0.666	0.590	0.663
Central Greece_1997	0.626	0.560	0.559	0.027	0.216	2.777	0.511	0.615	0.504	0.605
Central Greece_1998	0.630	0.566	0.565	0.027	0.254	2.755	0.519	0.623	0.510	0.614
Cretabank_1987	0.655	0.637	0.635	0.011	0.990	3.993	0.622	0.662	0.622	0.661
Cretabank_1989	0.449	0.401	0.399	0.022	0.323	2.880	0.362	0.445	0.359	0.441
Cretabank_1990	0.526	0.485	0.484	0.018	0.372	3.088	0.453	0.523	0.446	0.514
Cretabank_1991	0.578	0.548	0.547	0.014	0.475	2.959	0.525	0.578	0.522	0.574
Cretabank_1992	0.643	0.611	0.609	0.015	0.455	2.921	0.586	0.643	0.583	0.638
Cretabank_1993	0.701	0.670	0.669	0.013	0.381	2.963	0.646	0.698	0.644	0.694
Cretabank_1994	0.589	0.546	0.545	0.018	0.358	3.042	0.514	0.586	0.509	0.577
Cretabank_1995	0.605	0.530	0.529	0.028	0.138	2.980	0.478	0.587	0.460	0.574
Cretabank_1996	0.726	0.692	0.691	0.015	0.402	2.881	0.667	0.723	0.664	0.719
Cretabank_1997	0.740	0.703	0.703	0.014	0.318	2.969	0.677	0.734	0.674	0.729
Cretabank_1998	0.814	0.761	0.761	0.019	0.232	2.968	0.726	0.802	0.713	0.794
Egnatia_1993	0.628	0.539	0.543	0.031	-0.367	2.935	0.470	0.594	0.451	0.578
Egnatia_1994	0.484	0.440	0.438	0.019	0.379	3.097	0.405	0.479	0.400	0.473
Egnatia_1995	0.470	0.422	0.421	0.022	0.291	2.810	0.382	0.469	0.378	0.461
Egnatia_1996	0.685	0.602	0.600	0.036	0.274	2.941	0.536	0.675	0.521	0.661
Egnatia_1997	0.779	0.692	0.689	0.037	0.302	2.927	0.626	0.768	0.619	0.759
Egnatia_1998	0.719	0.663	0.662	0.025	0.319	2.687	0.620	0.716	0.614	0.704
Egnatia_1999	0.715	0.654	0.652	0.026	0.370	2.960	0.608	0.708	0.597	0.696
Emporiki_1987	0.750	0.703	0.702	0.020	0.192	2.751	0.666	0.743	0.660	0.736
Emporiki_1988	0.738	0.710	0.708	0.015	0.695	3.180	0.689	0.744	0.687	0.740
Emporiki_1989	0.718	0.689	0.688	0.014	0.579	3.040	0.666	0.721	0.665	0.719
Emporiki_1990	0.753	0.717	0.716	0.016	0.439	2.814	0.690	0.752	0.687	0.747
Emporiki_1991	0.722	0.686	0.685	0.016	0.341	2.773	0.659	0.719	0.656	0.716
Emporiki_1992	0.860	0.798	0.797	0.025	0.237	2.780	0.752	0.848	0.741	0.836
Emporiki_1993	0.956	0.881	0.880	0.030	0.180	2.754	0.827	0.940	0.813	0.929
Emporiki_1994	0.928	0.850	0.849	0.031	0.104	2.752	0.790	0.911	0.780	0.900
Emporiki_1995	0.851	0.794	0.794	0.023	0.215	2.857	0.754	0.841	0.740	0.830
Emporiki_1996	0.790	0.759	0.758	0.015	0.488	2.932	0.735	0.791	0.734	0.788
Emporiki_1997	0.811	0.778	0.776	0.015	0.426	2.813	0.752	0.810	0.750	0.806
Emporiki_1998	0.769	0.721	0.721	0.018	0.063	2.805	0.685	0.757	0.677	0.749
Emporiki_1999	0.916	0.844	0.844	0.030	0.157	2.764	0.787	0.904	0.778	0.892

Ergobank_1987	0.683	0.638	0.637	0.019	0.346	2.995	0.602	0.678	0.596	0.671
Ergobank_1988	0.727	0.684	0.682	0.020	0.515	3.133	0.650	0.729	0.646	0.722
Ergobank_1989	0.688	0.633	0.632	0.022	0.258	2.943	0.594	0.677	0.582	0.669
Ergobank_1990	0.749	0.682	0.681	0.028	0.311	2.999	0.634	0.741	0.621	0.731
Ergobank_1991	0.875	0.807	0.805	0.033	0.323	2.713	0.750	0.877	0.743	0.862
Ergobank_1992	0.708	0.645	0.643	0.028	0.363	2.968	0.596	0.702	0.587	0.693
Ergobank_1993	0.658	0.607	0.607	0.023	0.329	3.001	0.565	0.655	0.559	0.647
Ergobank_1994	0.567	0.535	0.534	0.014	0.437	2.944	0.512	0.566	0.508	0.560
Ergobank_1995	0.600	0.568	0.568	0.012	0.254	2.895	0.547	0.594	0.541	0.589
Ergobank_1996	0.657	0.620	0.619	0.015	0.234	2.865	0.593	0.651	0.585	0.644
Ergobank_1997	0.723	0.669	0.669	0.020	-0.076	2.671	0.629	0.707	0.618	0.699
Ergobank_1998	0.639	0.544	0.544	0.035	0.008	3.004	0.475	0.608	0.453	0.596
Ergobank_1999	0.682	0.591	0.591	0.034	-0.029	2.958	0.520	0.656	0.501	0.640
Eurobank_1997	0.512	0.450	0.450	0.027	0.194	2.928	0.399	0.505	0.390	0.491
Eurobank_1998	0.990	0.845	0.841	0.075	0.220	2.447	0.713	0.992	0.703	0.963
Eurobank_1999	0.747	0.692	0.691	0.024	0.321	2.894	0.648	0.743	0.642	0.732
General_1987	0.731	0.700	0.699	0.013	0.379	2.977	0.677	0.727	0.674	0.724
General_1988	0.754	0.728	0.726	0.012	0.641	3.376	0.707	0.756	0.706	0.753
General_1989	0.782	0.749	0.748	0.013	0.364	3.168	0.725	0.778	0.721	0.774
General_1990	0.791	0.765	0.763	0.013	0.851	3.695	0.747	0.795	0.746	0.792
General_1991	0.690	0.656	0.656	0.013	0.272	3.069	0.633	0.683	0.628	0.679
General_1992	0.677	0.656	0.654	0.011	0.874	3.652	0.640	0.682	0.640	0.680
General_1993	0.577	0.556	0.555	0.010	0.649	3.258	0.541	0.579	0.540	0.577
General_1994	0.680	0.654	0.653	0.012	0.633	3.282	0.636	0.681	0.634	0.677
General_1995	0.779	0.747	0.746	0.014	0.485	3.122	0.723	0.779	0.719	0.772
General_1996	0.714	0.678	0.677	0.014	0.273	2.931	0.652	0.708	0.646	0.702
General_1997	0.716	0.686	0.685	0.013	0.465	3.079	0.664	0.715	0.660	0.708
General_1998	0.714	0.673	0.672	0.014	0.323	3.200	0.647	0.704	0.642	0.699
General_1999	0.793	0.748	0.748	0.015	0.176	3.011	0.720	0.779	0.711	0.773
Interbank_1995	0.558	0.530	0.529	0.012	0.506	3.040	0.510	0.557	0.507	0.554
Interbank_1996	0.557	0.516	0.515	0.017	0.347	3.093	0.487	0.552	0.482	0.545
Ionian and Popular_1987	0.780	0.701	0.701	0.030	0.142	2.789	0.645	0.760	0.625	0.749
Ionian and Popular_1988	0.790	0.738	0.737	0.021	0.195	2.914	0.697	0.779	0.689	0.771
Ionian and Popular_1989	0.725	0.679	0.679	0.019	0.035	2.675	0.643	0.714	0.634	0.707
Ionian and Popular_1990	0.761	0.726	0.725	0.015	0.309	2.841	0.699	0.756	0.696	0.752
Ionian and Popular_1991	0.807	0.740	0.739	0.028	0.343	3.035	0.691	0.801	0.678	0.786
Ionian and Popular_1992	0.846	0.789	0.788	0.023	0.374	2.929	0.749	0.838	0.738	0.828
Ionian and Popular_1993	0.748	0.678	0.678	0.027	0.144	2.819	0.629	0.731	0.612	0.720
Ionian and Popular_1994	1.000	0.873	0.875	0.044	-0.204	2.952	0.776	0.955	0.767	0.946
Ionian and Popular_1995	1.000	0.853	0.854	0.051	-0.034	3.121	0.746	0.953	0.722	0.939
Ionian and Popular_1996	0.930	0.813	0.813	0.045	0.024	3.069	0.718	0.900	0.698	0.879
Ionian and Popular_1997	1.000	0.852	0.849	0.057	0.311	3.360	0.749	0.972	0.708	0.943
Ionian and Popular_1998	0.930	0.824	0.823	0.042	0.343	3.244	0.749	0.910	0.727	0.894

Laiki (Hellas)_1993	0.495	0.461	0.461	0.016	0.322	2.727	0.434	0.494	0.430	0.488
Laiki (Hellas)_1994	0.294	0.244	0.245	0.022	-0.013	2.583	0.199	0.287	0.194	0.275
Laiki (Hellas)_1995	0.481	0.398	0.397	0.037	0.117	2.667	0.328	0.473	0.317	0.454
Laiki (Hellas)_1996	0.704	0.550	0.551	0.070	0.020	2.357	0.418	0.680	0.400	0.656
Laiki (Hellas)_1997	0.864	0.739	0.738	0.046	0.040	2.997	0.644	0.828	0.628	0.815
Laiki (Hellas)_1998	0.798	0.666	0.665	0.057	0.182	2.927	0.559	0.779	0.538	0.753
Laiki (Hellas)_1999	0.857	0.718	0.715	0.057	0.234	3.058	0.611	0.832	0.587	0.808
Macedonia Thrace_1987	0.786	0.742	0.742	0.015	0.193	2.992	0.713	0.774	0.707	0.769
Macedonia Thrace_1988	0.741	0.704	0.704	0.013	0.236	2.982	0.680	0.732	0.675	0.727
Macedonia Thrace_1989	0.681	0.649	0.647	0.012	0.332	3.026	0.627	0.673	0.623	0.671
Macedonia Thrace_1990	0.744	0.718	0.717	0.012	0.739	3.485	0.700	0.746	0.699	0.743
Macedonia Thrace_1991	0.603	0.573	0.573	0.012	0.277	2.924	0.552	0.598	0.548	0.593
Macedonia Thrace_1992	0.694	0.669	0.667	0.013	0.704	3.256	0.650	0.700	0.648	0.694
Macedonia Thrace_1993	0.683	0.656	0.654	0.013	0.473	2.936	0.634	0.683	0.633	0.680
Macedonia Thrace_1994	0.591	0.561	0.560	0.011	0.272	3.002	0.541	0.583	0.538	0.580
Macedonia Thrace_1995	0.619	0.568	0.567	0.021	0.236	3.015	0.529	0.610	0.519	0.602
Macedonia Thrace_1996	0.662	0.629	0.628	0.015	0.408	2.972	0.604	0.659	0.600	0.656
Macedonia Thrace_1997	0.635	0.605	0.604	0.013	0.370	2.916	0.583	0.632	0.579	0.628
Macedonia Thrace_1998	0.635	0.602	0.601	0.012	0.175	2.880	0.578	0.627	0.573	0.623
Macedonia Thrace_1999	0.733	0.690	0.690	0.019	0.140	2.577	0.655	0.728	0.650	0.719
National_1987	0.723	0.626	0.628	0.037	-0.004	2.739	0.553	0.697	0.535	0.684
National_1988	0.664	0.610	0.608	0.026	0.435	2.847	0.567	0.667	0.563	0.657
National_1989	0.679	0.615	0.613	0.027	0.332	2.918	0.566	0.672	0.559	0.664
National_1990	0.674	0.604	0.602	0.026	0.204	2.911	0.557	0.658	0.538	0.646
National_1991	0.628	0.559	0.559	0.030	0.166	2.841	0.502	0.619	0.492	0.604
National_1992	0.850	0.756	0.754	0.040	0.210	2.921	0.680	0.837	0.665	0.816
National_1993	1.000	0.780	0.788	0.078	-0.374	2.994	0.604	0.920	0.588	0.903
National_1994	0.913	0.774	0.775	0.058	0.075	2.705	0.662	0.891	0.639	0.860
National_1995	0.909	0.786	0.785	0.055	0.243	2.747	0.690	0.899	0.673	0.880
National_1996	0.817	0.718	0.717	0.042	0.277	2.934	0.642	0.806	0.632	0.791
National_1997	1.000	0.840	0.839	0.068	0.188	2.762	0.717	0.978	0.708	0.966
National_1998	0.962	0.841	0.840	0.048	0.245	3.102	0.749	0.941	0.732	0.926
National_1999	1.000	0.874	0.873	0.047	0.161	3.053	0.782	0.966	0.784	0.967
Piraeus_1987	0.748	0.703	0.702	0.020	0.369	2.914	0.668	0.745	0.661	0.737
Piraeus_1988	0.788	0.758	0.756	0.016	0.628	3.254	0.733	0.793	0.731	0.786
Piraeus_1989	0.747	0.714	0.713	0.017	0.519	3.117	0.687	0.751	0.684	0.742
Piraeus_1990	0.828	0.789	0.788	0.019	0.483	3.068	0.757	0.831	0.751	0.821
Piraeus_1991	0.706	0.681	0.679	0.013	0.567	3.085	0.661	0.708	0.659	0.706
Piraeus_1992	0.758	0.716	0.714	0.019	0.445	3.004	0.685	0.756	0.675	0.747
Piraeus_1993	0.870	0.806	0.805	0.027	0.315	2.959	0.757	0.862	0.750	0.854
Piraeus_1994	0.899	0.870	0.868	0.014	0.862	3.789	0.849	0.904	0.848	0.901
Piraeus_1995	0.946	0.899	0.898	0.018	0.361	3.040	0.867	0.938	0.861	0.930
Piraeus_1996	0.768	0.729	0.728	0.016	0.466	3.014	0.702	0.765	0.699	0.761
Piraeus_1997	0.924	0.852	0.853	0.025	0.021	2.631	0.804	0.901	0.792	0.891
Piraeus_1998	0.966	0.800	0.807	0.067	-0.235	2.584	0.662	0.919	0.637	0.884
Piraeus_1999	0.780	0.682	0.684	0.037	-0.156	2.769	0.605	0.749	0.588	0.732

T Bank_1993	0.225	0.190	0.190	0.014	-0.038	2.611	0.162	0.216	0.156	0.210
T Bank_1994	0.590	0.561	0.559	0.013	0.544	3.076	0.540	0.590	0.539	0.587
T Bank_1995	0.867	0.827	0.826	0.016	0.289	2.888	0.798	0.861	0.793	0.854
T Bank_1996	0.760	0.675	0.674	0.034	0.219	2.949	0.611	0.744	0.596	0.730
T Bank_1997	0.749	0.643	0.645	0.047	0.065	2.543	0.555	0.732	0.539	0.710
T Bank_1998	0.758	0.638	0.637	0.056	0.227	2.632	0.540	0.751	0.521	0.727
T Bank_1999	0.708	0.557	0.555	0.075	0.202	2.531	0.422	0.704	0.410	0.676
Xiosbank_1991	0.561	0.541	0.540	0.009	0.563	3.258	0.525	0.561	0.524	0.559
Xiosbank_1992	0.846	0.754	0.750	0.037	0.364	3.096	0.689	0.832	0.673	0.821
Xiosbank_1993	0.639	0.611	0.609	0.015	0.613	3.102	0.587	0.645	0.586	0.640
Xiosbank_1994	0.466	0.430	0.429	0.016	0.418	2.820	0.403	0.463	0.398	0.458
Xiosbank_1995	0.499	0.444	0.444	0.025	0.185	2.626	0.400	0.493	0.390	0.482
Xiosbank_1996	0.597	0.543	0.542	0.022	0.298	3.012	0.505	0.587	0.491	0.578
Xiosbank_1997	0.700	0.637	0.637	0.024	0.202	2.798	0.595	0.686	0.582	0.677
Xiosbank_1998	0.667	0.561	0.561	0.042	0.145	2.880	0.483	0.644	0.460	0.628
Average_1987	0.710	0.644	0.645	0.026	-0.029	2.825	0.590	0.694	0.582	0.683
Average_1988	0.681	0.637	0.636	0.020	0.329	2.950	0.601	0.678	0.597	0.668
Average_1989	0.681	0.643	0.641	0.018	0.480	3.060	0.613	0.680	0.609	0.676
Average_1990	0.704	0.660	0.659	0.019	0.320	2.956	0.626	0.699	0.620	0.691
Average_1991	0.668	0.627	0.626	0.017	0.290	2.904	0.597	0.662	0.593	0.658
Average_1992	0.741	0.708	0.707	0.016	0.516	2.934	0.682	0.743	0.680	0.739
Average_1993	0.830	0.759	0.758	0.025	0.129	2.922	0.711	0.809	0.693	0.799
Average_1994	0.782	0.728	0.727	0.020	0.261	2.835	0.691	0.770	0.684	0.765
Average_1995	0.769	0.716	0.715	0.021	0.192	2.750	0.676	0.760	0.670	0.751
Average_1996	0.748	0.698	0.698	0.022	0.120	2.634	0.656	0.743	0.650	0.731
Average_1997	0.806	0.742	0.740	0.026	0.173	2.832	0.693	0.794	0.682	0.783
Average_1998	0.748	0.652	0.654	0.041	-0.179	2.454	0.567	0.724	0.558	0.707
Average_1999	0.812	0.701	0.705	0.046	-0.182	2.466	0.608	0.783	0.593	0.763
Average W_1987	0.718	0.630	0.631	0.034	-0.012	2.706	0.564	0.694	0.551	0.682
Average W_1988	0.657	0.602	0.600	0.025	0.348	2.739	0.559	0.655	0.551	0.643
Average W_1989	0.664	0.600	0.599	0.026	0.256	2.885	0.554	0.654	0.541	0.642
Average W_1990	0.677	0.611	0.610	0.025	0.087	2.859	0.563	0.662	0.548	0.648
Average W_1991	0.639	0.580	0.580	0.026	0.144	2.728	0.531	0.632	0.524	0.618
Average W_1992	0.733	0.665	0.665	0.025	-0.002	2.896	0.615	0.712	0.599	0.702
Average W_1993	0.937	0.826	0.825	0.043	0.121	2.910	0.742	0.914	0.719	0.897
Average W_1994	0.843	0.758	0.757	0.032	0.213	2.935	0.697	0.825	0.690	0.816
Average W_1995	0.816	0.726	0.726	0.032	0.019	3.138	0.660	0.791	0.640	0.777
Average W_1996	0.760	0.680	0.681	0.028	0.061	2.963	0.625	0.739	0.606	0.723
Average W_1997	0.885	0.783	0.782	0.036	0.183	3.160	0.715	0.857	0.695	0.846
Average W_1998	0.857	0.752	0.752	0.040	-0.010	2.807	0.672	0.828	0.650	0.813
Average W_1999	0.922	0.812	0.813	0.045	0.004	2.598	0.727	0.896	0.705	0.877

XIII. Appendix XIII: Hypothesis testing results

In the tables that follow, "Global Malm" is the Global Mamlquist Index, "Glob.Mal. BC" is the bias-corrected mean of the bootstrapped Global Malmquist index where one or two stars denote significance at the 5% or 1% level, based on the probabilities in the following two columns. The last two columns are the denoted percentiles for the distribution of the bootstrapped and bias corrected values of the index.

Table XIII.1. Results based on the LSCV bootstrap DEA

Bank	Global Malm	Glob.Mal. BC	Prob<1	Prob>1	CI 2.5%	CI 97.5%
Agricultural_92-93	1.024	0.912	0.932	0.069	0.780	1.028
Agricultural_93-94	1.053	1.064**	0.000	1.000	1.041	1.079
Agricultural_94-95	1.008	1.026	0.199	0.802	0.956	1.074
Agricultural_95-96	1.015	1.009**	0.000	1.000	1.003	1.023
Agricultural_96-97	0.968	0.981	0.802	0.199	0.950	1.020
Agricultural_97-98	0.999	1.037	0.112	0.889	0.982	1.103
Agricultural_98-99	0.937	0.960**	0.994	0.007	0.948	0.995
Alpha_87-88	0.985	0.979*	0.953	0.047	0.959	1.006
Alpha_88-89	0.990	0.990	0.806	0.194	0.971	1.013
Alpha_89-90	0.899	0.914**	1.000	0.000	0.872	0.954
Alpha_90-91	1.114	1.115**	0.000	1.000	1.103	1.134
Alpha_91-92	1.130	1.123**	0.000	1.000	1.073	1.169
Alpha_92-93	0.870	0.877**	1.000	0.000	0.844	0.911
Alpha_93-94	1.087	1.061**	0.000	1.000	1.021	1.123
Alpha_94-95	0.925	0.925**	1.000	0.000	0.915	0.931
Alpha_95-96	0.851	0.872**	1.000	0.000	0.826	0.907
Alpha_96-97	0.953	0.966**	1.000	0.000	0.936	0.981
Alpha_97-98	1.121	1.167**	0.000	1.000	1.069	1.270
Alpha_98-99	0.892	0.967	0.753	0.247	0.885	1.056
Bank of Athens_88-89	0.973	0.972**	1.000	0.000	0.965	0.980
Bank of Athens_89-90	0.954	0.959**	1.000	0.000	0.944	0.974
Bank of Athens_90-91	0.987	1.040	0.086	0.915	0.983	1.100
Bank of Athens_91-92	1.145	1.074*	0.026	0.974	0.999	1.154
Bank of Athens_92-93	1.018	1.018	0.053	0.947	0.995	1.038
Bank of Athens_93-94	1.350	1.390**	0.000	1.000	1.325	1.443
Bank of Athens_94-95	0.855	0.845**	1.000	0.000	0.812	0.881
Bank of Athens_95-96	0.972	0.970**	1.000	0.000	0.957	0.982
Bank of Athens_96-97	0.868	0.899**	1.000	0.000	0.829	0.952

Bank of Attica_87-88	1.078	1.076**	0.000	1.000	1.069	1.088
Bank of Attica_88-89	1.124	1.137**	0.000	1.000	1.115	1.158
Bank of Attica_89-90	0.887	0.877**	1.000	0.000	0.863	0.894
Bank of Attica_90-91	0.818	0.826**	1.000	0.000	0.812	0.838
Bank of Attica_91-92	0.910	1.001	0.600	0.401	0.951	1.091
Bank of Attica_92-93	1.058	0.961	0.747	0.254	0.825	1.053
Bank of Attica_93-94	1.153	1.115**	0.000	1.000	1.070	1.174
Bank of Attica_94-95	1.034	1.081**	0.000	1.000	1.043	1.132
Bank of Attica_95-96	1.061	1.061**	0.000	1.000	1.036	1.082
Bank of Attica_96-97	0.922	0.910**	1.000	0.000	0.873	0.937
Bank of Attica_97-98	0.990	1.009	0.384	0.617	0.921	1.076
Bank of Attica_98-99	1.071	1.066	0.100	0.900	0.975	1.174
Central Greece_87-89	1.416	0.944	0.634	0.366	0.465	1.850
Central Greece_89-90	1.057	1.056**	0.000	1.000	1.050	1.061
Central Greece_90-91	1.002	1.018*	0.029	0.972	0.999	1.040
Central Greece_91-92	1.048	1.041**	0.001	0.999	1.019	1.059
Central Greece_92-93	0.903	0.914**	1.000	0.000	0.869	0.950
Central Greece_93-94	1.007	1.004	0.208	0.793	0.989	1.017
Central Greece_94-95	1.060	1.054**	0.000	1.000	1.029	1.098
Central Greece_95-96	0.979	0.981**	1.000	0.000	0.969	0.992
Central Greece_96-97	1.078	1.107**	0.000	1.000	1.034	1.163
Central Greece_97-98	0.993	0.989	0.689	0.312	0.936	1.041
Cretabank_87-89	1.460	1.538**	0.000	1.000	1.422	1.627
Cretabank_89-90	0.853	0.832**	1.000	0.000	0.784	0.883
Cretabank_90-91	0.910	0.891**	1.000	0.000	0.867	0.930
Cretabank_91-92	0.899	0.898**	1.000	0.000	0.886	0.913
Cretabank_92-93	0.918	0.914**	1.000	0.000	0.898	0.939
Cretabank_93-94	1.191	1.212**	0.000	1.000	1.158	1.251
Cretabank_94-95	0.973	1.014	0.283	0.717	0.961	1.059
Cretabank_95-96	0.834	0.779**	1.000	0.000	0.719	0.852
Cretabank_96-97	0.980	0.984*	0.987	0.014	0.972	0.998
Cretabank_97-98	0.909	0.920**	1.000	0.000	0.900	0.938
Egnatia_93-94	1.298	1.237**	0.000	1.000	1.128	1.299
Egnatia_94-95	1.029	1.036	0.087	0.913	0.979	1.093
Egnatia_95-96	0.686	0.697**	1.000	0.000	0.667	0.755
Egnatia_96-97	0.880	0.871**	1.000	0.000	0.823	0.900
Egnatia_97-98	1.083	1.052	0.112	0.888	0.972	1.135
Egnatia_98-99	1.005	1.012	0.287	0.714	0.972	1.060

Emporiki_87-88	1.017	1.000	0.579	0.422	0.980	1.038
Emporiki_88-89	1.028	1.028**	0.000	1.000	1.016	1.039
Emporiki_89-90	0.953	0.958**	1.000	0.000	0.944	0.967
Emporiki_90-91	1.044	1.043**	0.000	1.000	1.034	1.055
Emporiki_91-92	0.839	0.853**	1.000	0.000	0.816	0.885
Emporiki_92-93	0.900	0.904**	1.000	0.000	0.895	0.917
Emporiki_93-94	1.030	1.035**	0.000	1.000	1.031	1.048
Emporiki_94-95	1.090	1.074**	0.000	1.000	1.035	1.108
Emporiki_95-96	1.078	1.055**	0.000	1.000	1.019	1.100
Emporiki_96-97	0.974	0.975**	1.000	0.000	0.971	0.981
Emporiki_97-98	1.054	1.071**	0.000	1.000	1.047	1.090
Emporiki_98-99	0.840	0.845**	1.000	0.000	0.807	0.880
Ergobank_87-88	0.940	0.933**	1.000	0.000	0.922	0.953
Ergobank_88-89	1.056	1.075**	0.000	1.000	1.051	1.103
Ergobank_89-90	0.919	0.924**	1.000	0.000	0.878	0.958
Ergobank_90-91	0.856	0.844**	1.000	0.000	0.808	0.889
Ergobank_91-92	1.236	1.250**	0.000	1.000	1.194	1.293
Ergobank_92-93	1.076	1.065**	0.000	1.000	1.044	1.094
Ergobank_93-94	1.161	1.141**	0.000	1.000	1.103	1.202
Ergobank_94-95	0.945	0.942**	1.000	0.000	0.925	0.965
Ergobank_95-96	0.912	0.914**	1.000	0.000	0.900	0.924
Ergobank_96-97	0.909	0.921**	1.000	0.000	0.896	0.939
Ergobank_97-98	1.131	1.197**	0.000	1.000	1.093	1.303
Ergobank_98-99	0.936	0.922**	1.000	0.000	0.870	0.986
Eurobank_97-98	0.517	0.526**	1.000	0.000	0.452	0.607
Eurobank_98-99	1.326	1.238**	0.000	1.000	1.089	1.437
General_87-88	0.969	0.964**	1.000	0.000	0.959	0.981
General_88-89	0.965	0.969**	1.000	0.000	0.961	0.978
General_89-90	0.989	0.982**	0.990	0.010	0.964	0.998
General_90-91	1.145	1.159**	0.000	1.000	1.139	1.184
General_91-92	1.019	1.005	0.306	0.694	0.982	1.028
General_92-93	1.174	1.176**	0.000	1.000	1.161	1.183
General_93-94	0.848	0.850**	1.000	0.000	0.838	0.864
General_94-95	0.874	0.875**	1.000	0.000	0.870	0.882
General_95-96	1.091	1.098**	0.000	1.000	1.086	1.109
General_96-97	0.997	0.991*	0.963	0.037	0.982	1.001
General_97-98	1.002	1.012	0.124	0.877	0.990	1.036
General_98-99	0.900	0.900**	1.000	0.000	0.883	0.918
Interbank_95-96	1.003	1.018	0.107	0.893	0.985	1.040

Ionian and Pop_87-88	0.988	0.961	0.865	0.135	0.907	1.034
Ionian and Pop_88-89	1.089	1.087**	0.000	1.000	1.046	1.132
Ionian and Pop_89-90	0.953	0.941**	1.000	0.000	0.927	0.966
Ionian and Pop_90-91	0.943	0.966	0.916	0.084	0.908	1.012
Ionian and Pop_91-92	0.954	0.943*	0.951	0.049	0.891	1.012
Ionian and Pop_92-93	1.130	1.150**	0.000	1.000	1.087	1.195
Ionian and Pop_93-94	0.748	0.770**	1.000	0.000	0.755	0.809
Ionian and Pop_94-95	1.000	1.014	0.241	0.759	0.944	1.049
Ionian and Pop_95-96	1.076	1.055**	0.000	1.000	1.024	1.087
Ionian and Pop_96-97	0.930	0.946	0.883	0.118	0.866	1.047
Ionian and Pop_97-98	1.075	1.040	0.090	0.911	0.980	1.098
Laiki (Hellas)_93-94	1.687	1.823**	0.000	1.000	1.627	2.032
Laiki (Hellas)_94-95	0.610	0.610**	1.000	0.000	0.569	0.640
Laiki (Hellas)_95-96	0.684	0.710**	1.000	0.000	0.616	0.797
Laiki (Hellas)_96-97	0.814	0.756**	1.000	0.000	0.653	0.880
Laiki (Hellas)_97-98	1.083	1.100	0.069	0.931	0.974	1.218
Laiki (Hellas)_98-99	0.931	0.928**	1.000	0.000	0.889	0.944
Mac-Thrace_87-88	1.060	1.057**	0.000	1.000	1.040	1.078
Mac-Thrace_88-89	1.088	1.087**	0.000	1.000	1.077	1.101
Mac-Thrace_89-90	0.915	0.906**	1.000	0.000	0.892	0.921
Mac-Thrace_90-91	1.234	1.246**	0.000	1.000	1.219	1.275
Mac-Thrace_91-92	0.868	0.859**	1.000	0.000	0.835	0.886
Mac-Thrace_92-93	1.017	1.018*	0.035	0.965	0.999	1.037
Mac-Thrace_93-94	1.156	1.168**	0.000	1.000	1.145	1.196
Mac-Thrace_94-95	0.955	0.973	0.871	0.129	0.923	1.014
Mac-Thrace_95-96	0.934	0.911**	0.999	0.001	0.884	0.950
Mac-Thrace_96-97	1.044	1.041**	0.000	1.000	1.031	1.055
Mac-Thrace_97-98	1.000	1.003	0.357	0.643	0.981	1.025
Mac-Thrace_98-99	0.865	0.868**	1.000	0.000	0.837	0.892
National_87-88	1.089	1.037	0.251	0.750	0.961	1.143
National_88-89	0.978	0.988	0.843	0.157	0.958	1.007
National_89-90	1.007	1.015	0.149	0.851	0.986	1.050
National_90-91	1.073	1.077	0.084	0.916	0.972	1.183
National_91-92	0.739	0.738**	1.000	0.000	0.662	0.818
National_92-93	0.850	0.923	0.888	0.113	0.831	1.036
National_93-94	1.096	1.030	0.238	0.762	0.951	1.088
National_94-95	1.004	0.989	0.652	0.349	0.948	1.035
National_95-96	1.113	1.097**	0.000	1.000	1.065	1.125
National_96-97	0.817	0.844**	1.000	0.000	0.818	0.892
National_97-98	1.039	1.005	0.448	0.552	0.933	1.077
National_98-99	0.962	0.961*	0.974	0.026	0.934	1.000

Piraeus_87-88	0.949	0.935**	0.999	0.002	0.913	0.968
Piraeus_88-89	1.055	1.058**	0.000	1.000	1.049	1.064
Piraeus_89-90	0.902	0.904**	1.000	0.000	0.898	0.908
Piraeus_90-91	1.173	1.162**	0.000	1.000	1.130	1.204
Piraeus_91-92	0.931	0.944**	1.000	0.001	0.911	0.971
Piraeus_92-93	0.871	0.883**	1.000	0.000	0.863	0.914
Piraeus_93-94	0.968	0.938**	0.992	0.009	0.899	0.988
Piraeus_94-95	0.950	0.963**	1.000	0.000	0.944	0.982
Piraeus_95-96	1.231	1.226**	0.000	1.000	1.190	1.256
Piraeus_96-97	0.832	0.848**	1.000	0.000	0.812	0.883
Piraeus_97-98	0.957	1.028	0.347	0.653	0.943	1.127
Piraeus_98-99	1.238	1.184**	0.000	1.000	1.077	1.305
T Bank_93-94	0.381	0.344**	1.000	0.000	0.302	0.393
T Bank_94-95	0.680	0.681**	1.000	0.000	0.667	0.702
T Bank_95-96	1.140	1.193**	0.000	1.000	1.112	1.268
T Bank_96-97	1.016	1.040	0.196	0.804	0.946	1.105
T Bank_97-98	0.988	1.003	0.373	0.628	0.939	1.033
T Bank_98-99	1.071	1.121	0.066	0.934	0.966	1.198
Xiosbank_91-92	0.663	0.696**	1.000	0.000	0.649	0.736
Xiosbank_92-93	1.324	1.255**	0.000	1.000	1.166	1.367
Xiosbank_93-94	1.372	1.402**	0.000	1.000	1.310	1.465
Xiosbank_94-95	0.933	0.959**	0.992	0.009	0.906	0.996
Xiosbank_95-96	0.836	0.820**	1.000	0.000	0.779	0.872
Xiosbank_96-97	0.853	0.851**	1.000	0.000	0.809	0.894
Xiosbank_97-98	1.049	1.108*	0.025	0.975	1.000	1.222
Average_87-88	1.042	1.019	0.279	0.721	0.964	1.086
Average_88-89	1.000	0.992	0.752	0.249	0.971	1.018
Average_89-90	0.967	0.972**	1.000	0.001	0.956	0.987
Average_90-91	1.053	1.051*	0.014	0.987	1.006	1.095
Average_91-92	0.902	0.891**	1.000	0.000	0.862	0.927
Average_92-93	0.893	0.918**	1.000	0.000	0.878	0.954
Average_93-94	1.062	1.046**	0.000	1.000	1.021	1.076
Average_94-95	1.016	1.016	0.105	0.895	0.990	1.042
Average_95-96	1.028	1.026**	0.000	1.000	1.012	1.045
Average_96-97	0.928	0.936**	1.000	0.000	0.909	0.957
Average_97-98	1.078	1.119**	0.000	1.000	1.051	1.186
Average_98-99	0.921	0.927**	1.000	0.000	0.912	0.934

Average W_87-88	1.093	1.055	0.127	0.874	0.976	1.153
Average W_88-89	0.988	0.999	0.435	0.566	0.967	1.021
Average W_89-90	0.982	0.981	0.929	0.071	0.960	1.009
Average W_90-91	1.059	1.054	0.119	0.881	0.960	1.142
Average W_91-92	0.872	0.870**	1.000	0.001	0.806	0.953
Average W_92-93	0.782	0.798**	1.000	0.000	0.736	0.851
Average W_93-94	1.112	1.092**	0.000	1.000	1.043	1.160
Average W_94-95	1.033	1.041*	0.020	0.981	1.002	1.065
Average W_95-96	1.073	1.067**	0.000	1.000	1.040	1.099
Average W_96-97	0.859	0.866**	1.000	0.000	0.824	0.903
Average W_97-98	1.033	1.038	0.166	0.835	0.971	1.119
Average W_98-99	0.930	0.925**	1.000	0.000	0.905	0.953

Table XIII.2. Results based on the SJ bootstrap DEA

Agricultural_92-93	1.024	0.907	0.941	0.059	0.772	1.026
Agricultural_93-94	1.053	1.064**	0.000	1.000	1.039	1.079
Agricultural_94-95	1.008	1.025	0.206	0.794	0.959	1.074
Agricultural_95-96	1.015	1.009**	0.000	1.000	1.003	1.023
Agricultural_96-97	0.968	0.980	0.806	0.194	0.948	1.018
Agricultural_97-98	0.999	1.038	0.098	0.903	0.983	1.105
Agricultural_98-99	0.937	0.961**	0.991	0.009	0.950	0.995
Alpha_87-88	0.985	0.978	0.949	0.052	0.958	1.007
Alpha_88-89	0.990	0.990	0.788	0.213	0.972	1.013
Alpha_89-90	0.899	0.913**	1.000	0.000	0.871	0.955
Alpha_90-91	1.114	1.115**	0.000	1.000	1.103	1.137
Alpha_91-92	1.130	1.123**	0.000	1.000	1.073	1.171
Alpha_92-93	0.870	0.878**	1.000	0.000	0.844	0.914
Alpha_93-94	1.087	1.060**	0.001	1.000	1.017	1.122
Alpha_94-95	0.925	0.925**	1.000	0.000	0.913	0.931
Alpha_95-96	0.851	0.872**	1.000	0.000	0.829	0.908
Alpha_96-97	0.953	0.966**	1.000	0.000	0.936	0.981
Alpha_97-98	1.121	1.167**	0.000	1.000	1.072	1.268
Alpha_98-99	0.892	0.967	0.750	0.251	0.884	1.054
Bank of Athens_88-89	0.973	0.972**	1.000	0.000	0.964	0.980
Bank of Athens_89-90	0.954	0.959**	1.000	0.000	0.944	0.974
Bank of Athens_90-91	0.987	1.042	0.080	0.920	0.982	1.105
Bank of Athens_91-92	1.145	1.071*	0.037	0.964	0.994	1.154
Bank of Athens_92-93	1.018	1.018	0.051	0.950	0.996	1.039
Bank of Athens_93-94	1.350	1.390**	0.000	1.000	1.327	1.444
Bank of Athens_94-95	0.855	0.845**	1.000	0.000	0.809	0.882
Bank of Athens_95-96	0.972	0.971**	1.000	0.000	0.956	0.983
Bank of Athens_96-97	0.868	0.898**	1.000	0.000	0.830	0.949
Bank of Attica_87-88	1.078	1.076**	0.000	1.000	1.069	1.088
Bank of Attica_88-89	1.124	1.137**	0.000	1.000	1.114	1.159
Bank of Attica_89-90	0.887	0.877**	1.000	0.000	0.862	0.895
Bank of Attica_90-91	0.818	0.825**	1.000	0.000	0.812	0.838
Bank of Attica_91-92	0.910	1.002	0.582	0.419	0.952	1.094
Bank of Attica_92-93	1.058	0.960	0.757	0.244	0.820	1.051
Bank of Attica_93-94	1.153	1.116**	0.000	1.000	1.069	1.173
Bank of Attica_94-95	1.034	1.081**	0.000	1.000	1.042	1.133
Bank of Attica_95-96	1.061	1.060**	0.000	1.000	1.036	1.084
Bank of Attica_96-97	0.922	0.911**	1.000	0.000	0.872	0.939
Bank of Attica_97-98	0.990	1.009	0.382	0.618	0.922	1.077
Bank of Attica_98-99	1.071	1.067	0.091	0.909	0.975	1.176

Central Greece_87-89	1.416	0.960	0.622	0.379	0.495	1.822
Central Greece_89-90	1.057	1.057**	0.000	1.000	1.050	1.061
Central Greece_90-91	1.002	1.018*	0.027	0.974	1.000	1.039
Central Greece_91-92	1.048	1.042**	0.001	0.999	1.019	1.061
Central Greece_92-93	0.903	0.914**	1.000	0.000	0.870	0.950
Central Greece_93-94	1.007	1.004	0.206	0.795	0.990	1.018
Central Greece_94-95	1.060	1.055**	0.000	1.000	1.029	1.100
Central Greece_95-96	0.979	0.982**	1.000	0.001	0.969	0.992
Central Greece_96-97	1.078	1.107**	0.000	1.000	1.037	1.164
Central Greece_97-98	0.993	0.987	0.698	0.303	0.936	1.042
Cretabank_87-89	1.460	1.540**	0.000	1.000	1.418	1.629
Cretabank_89-90	0.853	0.832**	1.000	0.000	0.785	0.881
Cretabank_90-91	0.910	0.891**	1.000	0.000	0.867	0.931
Cretabank_91-92	0.899	0.898**	1.000	0.000	0.886	0.914
Cretabank_92-93	0.918	0.914**	1.000	0.000	0.897	0.941
Cretabank_93-94	1.191	1.213**	0.000	1.000	1.152	1.253
Cretabank_94-95	0.973	1.015	0.283	0.717	0.962	1.061
Cretabank_95-96	0.834	0.777**	1.000	0.000	0.717	0.856
Cretabank_96-97	0.980	0.984*	0.986	0.014	0.972	0.998
Cretabank_97-98	0.909	0.920**	1.000	0.000	0.900	0.939
Egnatia_93-94	1.298	1.237**	0.000	1.000	1.126	1.299
Egnatia_94-95	1.029	1.037	0.087	0.914	0.980	1.095
Egnatia_95-96	0.686	0.697**	1.000	0.000	0.667	0.760
Egnatia_96-97	0.880	0.871**	1.000	0.000	0.817	0.901
Egnatia_97-98	1.083	1.052	0.108	0.893	0.968	1.135
Egnatia_98-99	1.005	1.012	0.290	0.710	0.968	1.062
Emporiki_87-88	1.017	0.999	0.579	0.422	0.979	1.039
Emporiki_88-89	1.028	1.029**	0.001	1.000	1.015	1.039
Emporiki_89-90	0.953	0.957**	1.000	0.000	0.945	0.967
Emporiki_90-91	1.044	1.044**	0.000	1.000	1.034	1.055
Emporiki_91-92	0.839	0.853**	1.000	0.000	0.816	0.884
Emporiki_92-93	0.900	0.904**	1.000	0.000	0.895	0.917
Emporiki_93-94	1.030	1.035**	0.000	1.000	1.031	1.048
Emporiki_94-95	1.090	1.075**	0.000	1.000	1.035	1.108
Emporiki_95-96	1.078	1.055**	0.000	1.000	1.020	1.099
Emporiki_96-97	0.974	0.975**	1.000	0.000	0.971	0.981
Emporiki_97-98	1.054	1.070**	0.000	1.000	1.048	1.091
Emporiki_98-99	0.840	0.845**	1.000	0.000	0.806	0.879

Ergobank_87-88	0.940	0.933**	1.000	0.000	0.922	0.955
Ergobank_88-89	1.056	1.075**	0.000	1.000	1.050	1.102
Ergobank_89-90	0.919	0.923**	1.000	0.000	0.880	0.958
Ergobank_90-91	0.856	0.845**	1.000	0.000	0.807	0.889
Ergobank_91-92	1.236	1.249**	0.000	1.000	1.195	1.294
Ergobank_92-93	1.076	1.065**	0.000	1.000	1.041	1.095
Ergobank_93-94	1.161	1.142**	0.000	1.000	1.103	1.200
Ergobank_94-95	0.945	0.943**	1.000	0.000	0.925	0.964
Ergobank_95-96	0.912	0.914**	1.000	0.000	0.901	0.924
Ergobank_96-97	0.909	0.921**	1.000	0.000	0.895	0.940
Ergobank_97-98	1.131	1.197**	0.000	1.000	1.094	1.304
Ergobank_98-99	0.936	0.923**	1.000	0.000	0.870	0.986
Eurobank_97-98	0.517	0.526**	1.000	0.000	0.449	0.608
Eurobank_98-99	1.326	1.238**	0.000	1.000	1.091	1.441
General_87-88	0.969	0.964**	0.998	0.002	0.958	0.982
General_88-89	0.965	0.969**	1.000	0.000	0.960	0.978
General_89-90	0.989	0.982*	0.980	0.020	0.966	0.999
General_90-91	1.145	1.159**	0.000	1.000	1.138	1.184
General_91-92	1.019	1.005	0.308	0.693	0.981	1.028
General_92-93	1.174	1.177**	0.000	1.000	1.159	1.183
General_93-94	0.848	0.850**	1.000	0.000	0.837	0.864
General_94-95	0.874	0.875**	1.000	0.000	0.870	0.882
General_95-96	1.091	1.098**	0.000	1.000	1.086	1.109
General_96-97	0.997	0.991*	0.953	0.048	0.982	1.001
General_97-98	1.002	1.013	0.118	0.882	0.991	1.036
General_98-99	0.900	0.900**	1.000	0.000	0.883	0.918
Interbank_95-96	1.003	1.019	0.105	0.896	0.980	1.041
Ionian and Pop_87-88	0.988	0.961	0.863	0.138	0.907	1.037
Ionian and Pop_88-89	1.089	1.088**	0.000	1.000	1.047	1.133
Ionian and Pop_89-90	0.953	0.940**	1.000	0.000	0.926	0.965
Ionian and Pop_90-91	0.943	0.968	0.898	0.103	0.906	1.013
Ionian and Pop_91-92	0.954	0.941	0.944	0.056	0.887	1.016
Ionian and Pop_92-93	1.130	1.150**	0.000	1.000	1.084	1.195
Ionian and Pop_93-94	0.748	0.770**	1.000	0.000	0.756	0.807
Ionian and Pop_94-95	1.000	1.014	0.247	0.753	0.939	1.048
Ionian and Pop_95-96	1.076	1.055**	0.000	1.000	1.023	1.088
Ionian and Pop_96-97	0.930	0.948	0.874	0.126	0.869	1.049
Ionian and Pop_97-98	1.075	1.040	0.088	0.913	0.978	1.098

Laiki (Hellas)_93-94	1.687	1.822**	0.000	1.000	1.629	2.025
Laiki (Hellas)_94-95	0.610	0.611**	1.000	0.000	0.569	0.641
Laiki (Hellas)_95-96	0.684	0.708**	1.000	0.000	0.611	0.797
Laiki (Hellas)_96-97	0.814	0.756**	1.000	0.000	0.656	0.892
Laiki (Hellas)_97-98	1.083	1.102	0.071	0.929	0.973	1.227
Laiki (Hellas)_98-99	0.931	0.928**	1.000	0.000	0.887	0.945
Mac-Thrace_87-88	1.060	1.056**	0.000	1.000	1.039	1.078
Mac-Thrace_88-89	1.088	1.086**	0.000	1.000	1.077	1.101
Mac-Thrace_89-90	0.915	0.906**	1.000	0.000	0.891	0.920
Mac-Thrace_90-91	1.234	1.246**	0.000	1.000	1.220	1.276
Mac-Thrace_91-92	0.868	0.859**	1.000	0.000	0.834	0.885
Mac-Thrace_92-93	1.017	1.018*	0.034	0.967	0.999	1.037
Mac-Thrace_93-94	1.156	1.167**	0.000	1.000	1.144	1.195
Mac-Thrace_94-95	0.955	0.975	0.832	0.168	0.923	1.017
Mac-Thrace_95-96	0.934	0.910**	1.000	0.000	0.883	0.954
Mac-Thrace_96-97	1.044	1.041**	0.000	1.000	1.031	1.056
Mac-Thrace_97-98	1.000	1.003	0.342	0.659	0.981	1.025
Mac-Thrace_98-99	0.865	0.867**	1.000	0.000	0.836	0.891
National_87-88	1.089	1.036	0.270	0.730	0.960	1.143
National_88-89	0.978	0.989	0.824	0.176	0.958	1.008
National_89-90	1.007	1.016	0.139	0.862	0.986	1.050
National_90-91	1.073	1.077	0.089	0.911	0.965	1.185
National_91-92	0.739	0.738**	1.000	0.000	0.658	0.826
National_92-93	0.850	0.923	0.892	0.109	0.828	1.033
National_93-94	1.096	1.029	0.242	0.759	0.949	1.085
National_94-95	1.004	0.989	0.641	0.360	0.948	1.036
National_95-96	1.113	1.097**	0.000	1.000	1.065	1.125
National_96-97	0.817	0.844**	1.000	0.000	0.818	0.890
National_97-98	1.039	1.006	0.445	0.555	0.931	1.080
National_98-99	0.962	0.962*	0.970	0.031	0.936	1.001
Piraeus_87-88	0.949	0.935**	1.000	0.000	0.913	0.967
Piraeus_88-89	1.055	1.058**	0.000	1.000	1.049	1.064
Piraeus_89-90	0.902	0.904**	1.000	0.000	0.899	0.908
Piraeus_90-91	1.173	1.163**	0.000	1.000	1.131	1.202
Piraeus_91-92	0.931	0.944**	1.000	0.000	0.913	0.972
Piraeus_92-93	0.871	0.882**	1.000	0.000	0.862	0.914
Piraeus_93-94	0.968	0.939**	0.993	0.007	0.899	0.985
Piraeus_94-95	0.950	0.962**	1.000	0.000	0.944	0.981
Piraeus_95-96	1.231	1.228**	0.000	1.000	1.187	1.258
Piraeus_96-97	0.832	0.847**	1.000	0.000	0.812	0.882
Piraeus_97-98	0.957	1.029	0.343	0.657	0.948	1.132
Piraeus_98-99	1.238	1.183**	0.000	1.000	1.073	1.306

T Bank_93-94	0.381	0.344**	1.000	0.000	0.303	0.390
T Bank_94-95	0.680	0.680**	1.000	0.000	0.667	0.700
T Bank_95-96	1.140	1.196**	0.000	1.000	1.106	1.275
T Bank_96-97	1.016	1.041	0.186	0.815	0.946	1.108
T Bank_97-98	0.988	1.003	0.378	0.623	0.944	1.032
T Bank_98-99	1.071	1.119*	0.049	0.951	0.976	1.193
Xiosbank_91-92	0.663	0.697**	1.000	0.000	0.647	0.738
Xiosbank_92-93	1.324	1.253**	0.000	1.000	1.161	1.376
Xiosbank_93-94	1.372	1.405**	0.000	1.000	1.310	1.469
Xiosbank_94-95	0.933	0.960*	0.979	0.022	0.906	0.999
Xiosbank_95-96	0.836	0.819**	1.000	0.000	0.778	0.871
Xiosbank_96-97	0.853	0.850**	1.000	0.000	0.807	0.894
Xiosbank_97-98	1.049	1.108*	0.020	0.981	1.006	1.222
Average_87-88	1.042	1.019	0.304	0.696	0.964	1.088
Average_88-89	1.000	0.993	0.727	0.273	0.971	1.019
Average_89-90	0.967	0.971**	1.000	0.000	0.956	0.987
Average_90-91	1.053	1.053*	0.015	0.985	1.006	1.098
Average_91-92	0.902	0.889**	1.000	0.000	0.860	0.928
Average_92-93	0.893	0.918**	1.000	0.000	0.879	0.954
Average_93-94	1.062	1.047**	0.000	1.000	1.021	1.076
Average_94-95	1.016	1.016	0.107	0.893	0.989	1.043
Average_95-96	1.028	1.026**	0.000	1.000	1.012	1.045
Average_96-97	0.928	0.936**	1.000	0.000	0.909	0.957
Average_97-98	1.078	1.121**	0.000	1.000	1.053	1.188
Average_98-99	0.921	0.927**	1.000	0.000	0.913	0.934
Average W_87-88	1.093	1.054	0.128	0.872	0.976	1.154
Average W_88-89	0.988	1.000	0.413	0.588	0.967	1.021
Average W_89-90	0.982	0.981	0.924	0.076	0.961	1.011
Average W_90-91	1.059	1.055	0.117	0.884	0.962	1.145
Average W_91-92	0.872	0.869**	0.999	0.002	0.805	0.952
Average W_92-93	0.782	0.798**	1.000	0.000	0.734	0.850
Average W_93-94	1.112	1.091**	0.000	1.000	1.042	1.159
Average W_94-95	1.033	1.041*	0.017	0.983	1.003	1.067
Average W_95-96	1.073	1.067**	0.000	1.000	1.040	1.099
Average W_96-97	0.859	0.866**	1.000	0.000	0.824	0.903
Average W_97-98	1.033	1.040	0.160	0.841	0.975	1.125
Average W_98-99	0.930	0.925**	1.000	0.000	0.906	0.954

Table XIII.3. Results based on the Moments bootstrap DEA

Agricultural_92-93	1.024	0.906	0.933	0.067	0.769	1.029
Agricultural_93-94	1.053	1.066**	0.000	1.000	1.040	1.083
Agricultural_94-95	1.008	1.030	0.167	0.833	0.963	1.084
Agricultural_95-96	1.015	1.009**	0.000	1.000	1.002	1.024
Agricultural_96-97	0.968	0.980	0.808	0.192	0.948	1.021
Agricultural_97-98	0.999	1.036	0.130	0.871	0.982	1.101
Agricultural_98-99	0.937	0.960**	0.999	0.002	0.948	0.994
Alpha_87-88	0.985	0.977	0.921	0.079	0.948	1.013
Alpha_88-89	0.990	0.990	0.757	0.243	0.966	1.024
Alpha_89-90	0.899	0.918**	1.000	0.001	0.866	0.969
Alpha_90-91	1.114	1.115**	0.000	1.000	1.101	1.137
Alpha_91-92	1.130	1.122**	0.000	1.000	1.067	1.177
Alpha_92-93	0.870	0.880**	1.000	0.000	0.847	0.921
Alpha_93-94	1.087	1.052	0.031	0.970	0.998	1.124
Alpha_94-95	0.925	0.925**	1.000	0.000	0.909	0.934
Alpha_95-96	0.851	0.878**	1.000	0.000	0.829	0.923
Alpha_96-97	0.953	0.966**	1.000	0.000	0.935	0.982
Alpha_97-98	1.121	1.167**	0.000	1.000	1.069	1.275
Alpha_98-99	0.892	0.963	0.774	0.226	0.880	1.048
Bank of Athens_88-89	0.973	0.972**	1.000	0.001	0.960	0.984
Bank of Athens_89-90	0.954	0.960**	0.999	0.002	0.940	0.986
Bank of Athens_90-91	0.987	1.052	0.066	0.934	0.982	1.128
Bank of Athens_91-92	1.145	1.057	0.107	0.893	0.966	1.157
Bank of Athens_92-93	1.018	1.021	0.102	0.898	0.986	1.051
Bank of Athens_93-94	1.350	1.404**	0.000	1.000	1.330	1.474
Bank of Athens_94-95	0.855	0.845**	1.000	0.000	0.799	0.888
Bank of Athens_95-96	0.972	0.970**	1.000	0.000	0.954	0.984
Bank of Athens_96-97	0.868	0.901**	1.000	0.000	0.834	0.961
Bank of Attica_87-88	1.078	1.075**	0.000	1.000	1.065	1.093
Bank of Attica_88-89	1.124	1.141**	0.000	1.000	1.111	1.175
Bank of Attica_89-90	0.887	0.873**	1.000	0.000	0.852	0.897
Bank of Attica_90-91	0.818	0.828**	1.000	0.000	0.811	0.847
Bank of Attica_91-92	0.910	1.009	0.431	0.569	0.954	1.103
Bank of Attica_92-93	1.058	0.958	0.754	0.247	0.809	1.064
Bank of Attica_93-94	1.153	1.103**	0.001	1.000	1.034	1.173
Bank of Attica_94-95	1.034	1.088**	0.000	1.000	1.039	1.153
Bank of Attica_95-96	1.061	1.062**	0.000	1.000	1.036	1.087
Bank of Attica_96-97	0.922	0.912**	1.000	0.000	0.870	0.941
Bank of Attica_97-98	0.990	1.013	0.380	0.620	0.930	1.092
Bank of Attica_98-99	1.071	1.067	0.099	0.902	0.969	1.174

Central Greece_87-89	1.416	0.949	0.635	0.365	0.484	1.807
Central Greece_89-90	1.057	1.057**	0.000	1.000	1.048	1.064
Central Greece_90-91	1.002	1.020	0.052	0.949	0.996	1.048
Central Greece_91-92	1.048	1.041**	0.005	0.995	1.015	1.065
Central Greece_92-93	0.903	0.921**	1.000	0.000	0.869	0.970
Central Greece_93-94	1.007	1.004	0.249	0.751	0.988	1.019
Central Greece_94-95	1.060	1.051**	0.000	1.000	1.019	1.099
Central Greece_95-96	0.979	0.982**	0.998	0.003	0.967	0.994
Central Greece_96-97	1.078	1.111**	0.000	1.000	1.042	1.177
Central Greece_97-98	0.993	0.989	0.665	0.335	0.934	1.044
Cretabank_87-89	1.460	1.561**	0.000	1.000	1.440	1.681
Cretabank_89-90	0.853	0.830**	1.000	0.000	0.773	0.880
Cretabank_90-91	0.910	0.887**	1.000	0.000	0.856	0.932
Cretabank_91-92	0.899	0.897**	1.000	0.000	0.880	0.917
Cretabank_92-93	0.918	0.912**	1.000	0.000	0.888	0.943
Cretabank_93-94	1.191	1.220**	0.000	1.000	1.157	1.274
Cretabank_94-95	0.973	1.017	0.257	0.743	0.966	1.063
Cretabank_95-96	0.834	0.771**	1.000	0.000	0.705	0.843
Cretabank_96-97	0.980	0.983	0.970	0.030	0.966	1.001
Cretabank_97-98	0.909	0.921**	1.000	0.000	0.899	0.944
Egnatia_93-94	1.298	1.238**	0.000	1.000	1.115	1.299
Egnatia_94-95	1.029	1.038	0.066	0.935	0.989	1.099
Egnatia_95-96	0.686	0.697**	1.000	0.000	0.666	0.755
Egnatia_96-97	0.880	0.871**	1.000	0.000	0.820	0.900
Egnatia_97-98	1.083	1.048	0.155	0.845	0.957	1.135
Egnatia_98-99	1.005	1.013	0.263	0.738	0.970	1.066
Emporiki_87-88	1.017	0.990	0.696	0.304	0.960	1.040
Emporiki_88-89	1.028	1.030**	0.003	0.997	1.011	1.046
Emporiki_89-90	0.953	0.960**	1.000	0.000	0.943	0.973
Emporiki_90-91	1.044	1.044**	0.000	1.000	1.031	1.060
Emporiki_91-92	0.839	0.857**	1.000	0.000	0.813	0.900
Emporiki_92-93	0.900	0.904**	1.000	0.000	0.896	0.920
Emporiki_93-94	1.030	1.035**	0.000	1.000	1.031	1.051
Emporiki_94-95	1.090	1.071**	0.000	1.000	1.028	1.106
Emporiki_95-96	1.078	1.048	0.036	0.964	0.997	1.105
Emporiki_96-97	0.974	0.976**	1.000	0.000	0.970	0.985
Emporiki_97-98	1.054	1.075**	0.000	1.000	1.045	1.102
Emporiki_98-99	0.840	0.851**	1.000	0.000	0.805	0.894

Ergobank_87-88	0.940	0.932**	1.000	0.000	0.918	0.956
Ergobank_88-89	1.056	1.076**	0.000	1.000	1.050	1.109
Ergobank_89-90	0.919	0.925**	1.000	0.000	0.873	0.967
Ergobank_90-91	0.856	0.846**	1.000	0.000	0.806	0.895
Ergobank_91-92	1.236	1.248**	0.000	1.000	1.190	1.294
Ergobank_92-93	1.076	1.063**	0.000	1.000	1.035	1.094
Ergobank_93-94	1.161	1.137**	0.000	1.000	1.088	1.211
Ergobank_94-95	0.945	0.942**	0.998	0.002	0.918	0.973
Ergobank_95-96	0.912	0.916**	1.000	0.000	0.902	0.929
Ergobank_96-97	0.909	0.924**	1.000	0.000	0.894	0.949
Ergobank_97-98	1.131	1.202**	0.000	1.000	1.107	1.315
Ergobank_98-99	0.936	0.924**	1.000	0.000	0.873	0.987
Eurobank_97-98	0.517	0.526**	1.000	0.000	0.454	0.604
Eurobank_98-99	1.326	1.234**	0.001	0.999	1.069	1.421
General_87-88	0.969	0.961**	0.994	0.006	0.953	0.989
General_88-89	0.965	0.970**	1.000	0.000	0.959	0.983
General_89-90	0.989	0.980	0.961	0.039	0.952	1.002
General_90-91	1.145	1.163**	0.000	1.000	1.134	1.200
General_91-92	1.019	1.001	0.447	0.553	0.964	1.033
General_92-93	1.174	1.179**	0.000	1.000	1.152	1.189
General_93-94	0.848	0.849**	1.000	0.000	0.832	0.871
General_94-95	0.874	0.875**	1.000	0.000	0.869	0.886
General_95-96	1.091	1.100**	0.000	1.000	1.085	1.115
General_96-97	0.997	0.989	0.956	0.045	0.977	1.001
General_97-98	1.002	1.016	0.133	0.867	0.988	1.049
General_98-99	0.900	0.899**	1.000	0.000	0.876	0.922
Interbank_95-96	1.003	1.023	0.100	0.900	0.984	1.050
Ionian and Pop_87-88	0.988	0.954	0.874	0.127	0.889	1.036
Ionian and Pop_88-89	1.089	1.086**	0.000	1.000	1.034	1.145
Ionian and Pop_89-90	0.953	0.936**	1.000	0.000	0.918	0.963
Ionian and Pop_90-91	0.943	0.973	0.781	0.220	0.910	1.035
Ionian and Pop_91-92	0.954	0.939	0.926	0.074	0.863	1.021
Ionian and Pop_92-93	1.130	1.156**	0.000	1.000	1.085	1.209
Ionian and Pop_93-94	0.748	0.770**	1.000	0.000	0.755	0.808
Ionian and Pop_94-95	1.000	1.016	0.226	0.775	0.946	1.053
Ionian and Pop_95-96	1.076	1.055**	0.000	1.000	1.022	1.091
Ionian and Pop_96-97	0.930	0.946	0.880	0.120	0.866	1.046
Ionian and Pop_97-98	1.075	1.042	0.064	0.937	0.980	1.099

Laiki (Hellas)_93-94	1.687	1.827**	0.000	1.000	1.637	2.054
Laiki (Hellas)_94-95	0.610	0.611**	1.000	0.000	0.572	0.640
Laiki (Hellas)_95-96	0.684	0.707**	1.000	0.000	0.615	0.791
Laiki (Hellas)_96-97	0.814	0.761**	1.000	0.001	0.666	0.899
Laiki (Hellas)_97-98	1.083	1.099	0.064	0.936	0.975	1.214
Laiki (Hellas)_98-99	0.931	0.928**	1.000	0.000	0.888	0.945
Mac-Thrace_87-88	1.060	1.054**	0.000	1.000	1.030	1.080
Mac-Thrace_88-89	1.088	1.086**	0.000	1.000	1.072	1.105
Mac-Thrace_89-90	0.915	0.903**	1.000	0.000	0.883	0.924
Mac-Thrace_90-91	1.234	1.251**	0.000	1.000	1.212	1.296
Mac-Thrace_91-92	0.868	0.857**	1.000	0.000	0.815	0.897
Mac-Thrace_92-93	1.017	1.020	0.086	0.915	0.989	1.051
Mac-Thrace_93-94	1.156	1.167**	0.000	1.000	1.134	1.209
Mac-Thrace_94-95	0.955	0.981	0.728	0.272	0.924	1.033
Mac-Thrace_95-96	0.934	0.906**	1.000	0.000	0.873	0.952
Mac-Thrace_96-97	1.044	1.040**	0.000	1.000	1.026	1.058
Mac-Thrace_97-98	1.000	1.004	0.371	0.630	0.971	1.032
Mac-Thrace_98-99	0.865	0.870**	1.000	0.000	0.831	0.907
National_87-88	1.089	1.035	0.290	0.711	0.956	1.141
National_88-89	0.978	0.989	0.799	0.202	0.958	1.011
National_89-90	1.007	1.016	0.138	0.863	0.985	1.051
National_90-91	1.073	1.076	0.093	0.907	0.970	1.197
National_91-92	0.739	0.738**	1.000	0.000	0.653	0.822
National_92-93	0.850	0.923	0.907	0.093	0.833	1.033
National_93-94	1.096	1.030	0.240	0.761	0.951	1.089
National_94-95	1.004	0.989	0.647	0.353	0.947	1.035
National_95-96	1.113	1.098**	0.000	1.000	1.067	1.128
National_96-97	0.817	0.843**	1.000	0.000	0.815	0.891
National_97-98	1.039	1.007	0.425	0.576	0.932	1.083
National_98-99	0.962	0.962	0.959	0.042	0.935	1.004
Piraeus_87-88	0.949	0.929**	1.000	0.001	0.897	0.965
Piraeus_88-89	1.055	1.060**	0.000	1.000	1.049	1.069
Piraeus_89-90	0.902	0.905**	1.000	0.000	0.900	0.911
Piraeus_90-91	1.173	1.159**	0.000	1.000	1.109	1.218
Piraeus_91-92	0.931	0.948**	0.992	0.008	0.901	0.990
Piraeus_92-93	0.871	0.885**	1.000	0.000	0.867	0.923
Piraeus_93-94	0.968	0.928*	0.986	0.015	0.870	0.992
Piraeus_94-95	0.950	0.966*	0.989	0.012	0.941	0.994
Piraeus_95-96	1.231	1.232**	0.000	1.000	1.184	1.273
Piraeus_96-97	0.832	0.851**	1.000	0.000	0.807	0.901
Piraeus_97-98	0.957	1.030	0.308	0.693	0.943	1.133
Piraeus_98-99	1.238	1.187**	0.000	1.000	1.081	1.303

T Bank_93-94	0.381	0.342**	1.000	0.000	0.296	0.388
T Bank_94-95	0.680	0.678**	1.000	0.000	0.658	0.704
T Bank_95-96	1.140	1.206**	0.000	1.000	1.121	1.300
T Bank_96-97	1.016	1.039	0.191	0.810	0.947	1.102
T Bank_97-98	0.988	1.001	0.376	0.624	0.943	1.029
T Bank_98-99	1.071	1.116	0.053	0.948	0.971	1.187
Xiosbank_91-92	0.663	0.705**	1.000	0.000	0.657	0.754
Xiosbank_92-93	1.324	1.241**	0.000	1.000	1.119	1.371
Xiosbank_93-94	1.372	1.411**	0.000	1.000	1.318	1.507
Xiosbank_94-95	0.933	0.961	0.967	0.034	0.908	1.002
Xiosbank_95-96	0.836	0.820**	1.000	0.000	0.774	0.872
Xiosbank_96-97	0.853	0.852**	1.000	0.000	0.806	0.898
Xiosbank_97-98	1.049	1.109*	0.020	0.981	1.005	1.225
Average_87-88	1.042	1.015	0.361	0.640	0.945	1.090
Average_88-89	1.000	0.991	0.721	0.279	0.965	1.023
Average_89-90	0.967	0.973**	0.998	0.002	0.955	0.992
Average_90-91	1.053	1.051	0.048	0.952	0.992	1.116
Average_91-92	0.902	0.886**	1.000	0.000	0.841	0.932
Average_92-93	0.893	0.926**	0.998	0.002	0.881	0.974
Average_93-94	1.062	1.044**	0.000	1.000	1.015	1.078
Average_94-95	1.016	1.017	0.145	0.855	0.987	1.047
Average_95-96	1.028	1.025**	0.001	1.000	1.010	1.049
Average_96-97	0.928	0.938**	1.000	0.000	0.909	0.963
Average_97-98	1.078	1.122**	0.000	1.000	1.055	1.197
Average_98-99	0.921	0.927**	1.000	0.000	0.913	0.934
Average W_87-88	1.093	1.053	0.169	0.832	0.971	1.157
Average W_88-89	0.988	1.000	0.413	0.588	0.967	1.022
Average W_89-90	0.982	0.982	0.923	0.077	0.958	1.010
Average W_90-91	1.059	1.053	0.155	0.845	0.957	1.154
Average W_91-92	0.872	0.870**	0.997	0.003	0.796	0.956
Average W_92-93	0.782	0.799**	1.000	0.000	0.734	0.860
Average W_93-94	1.112	1.093**	0.000	1.000	1.040	1.159
Average W_94-95	1.033	1.042*	0.014	0.986	1.004	1.068
Average W_95-96	1.073	1.067**	0.000	1.000	1.039	1.099
Average W_96-97	0.859	0.866**	1.000	0.000	0.826	0.903
Average W_97-98	1.033	1.038	0.176	0.825	0.972	1.127
Average W_98-99	0.930	0.926**	1.000	0.000	0.905	0.952

XIV. Appendix XIV: Input-output-efficiency scatterplots

This appendix graphically summarises the results of our analysis. In particular for each bank we first present input output scatterplots, with the bank under examination identified by the markers with the red filling. We also map the bank under examination on the input-output space with respect to all banks in the sample which correspond to the other markers. The input and output variables are expressed in logs to help us identify clusters, though in all cases we observe that the observations are highly correlated and form one cluster. Also, given that we are using logs and movements on the input-output space can be thought of as proportional changes; in all cases we observe that the clusters lie on a straight line which has a slope close to one suggesting that a proportional increase in input leads to almost the same proportional increase in outputs, providing further support to the assumption of CRS.

After each set of “mapping scatterplots” we present the same input-output scatterplots, this time “zooming in” each bank and identifying its trajectory over time. It also provides information about our hypothesis tests and for this purpose we have used results from the moments bootstrap. In particular, the labels above each point indicate the mean bias-corrected efficiency score under the moments bootstrap and year identifier (where 1=1987 and 13=1999). If the one-sided tests of efficiency change at a 5% level of significance have indicated either a significant increase or decrease in efficiency we will denote this by linking the two consecutive markers with a solid black line. In the opposite case a light grey dotted line is used.

Figure XIV.1. Agricultural Bank

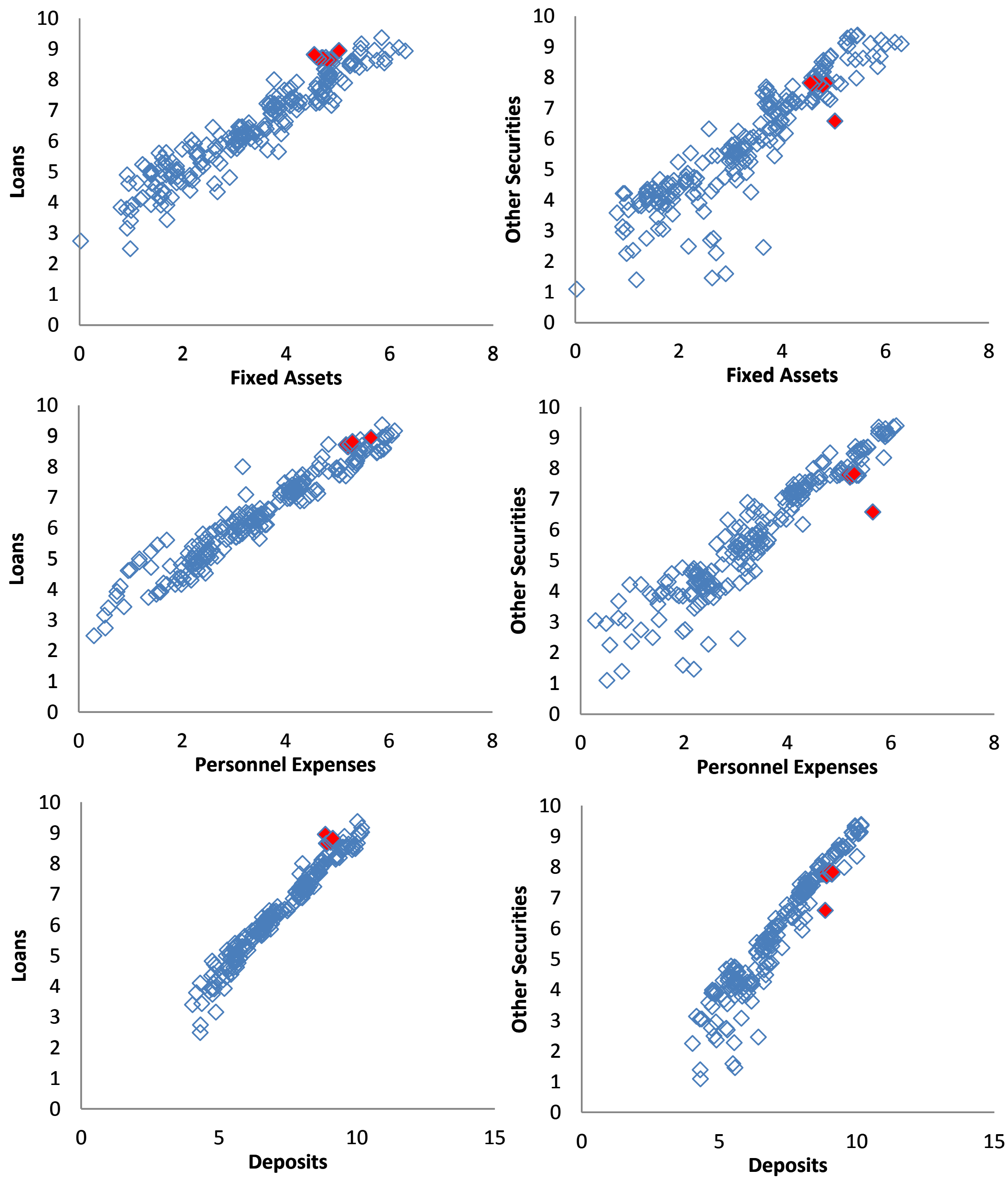


Figure XIV.2. Agricultural Bank

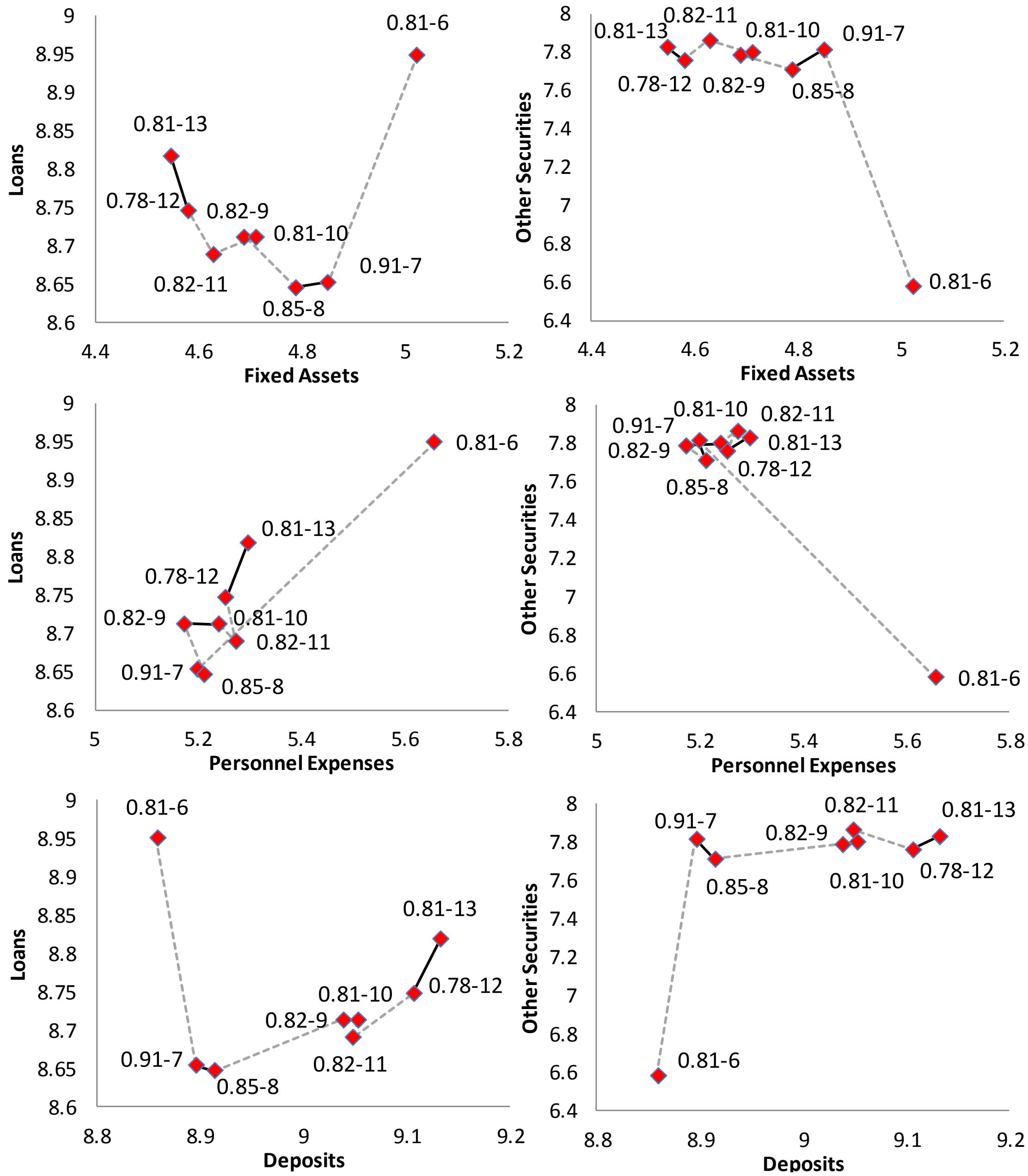


Figure XIV.3. Alpha Bank

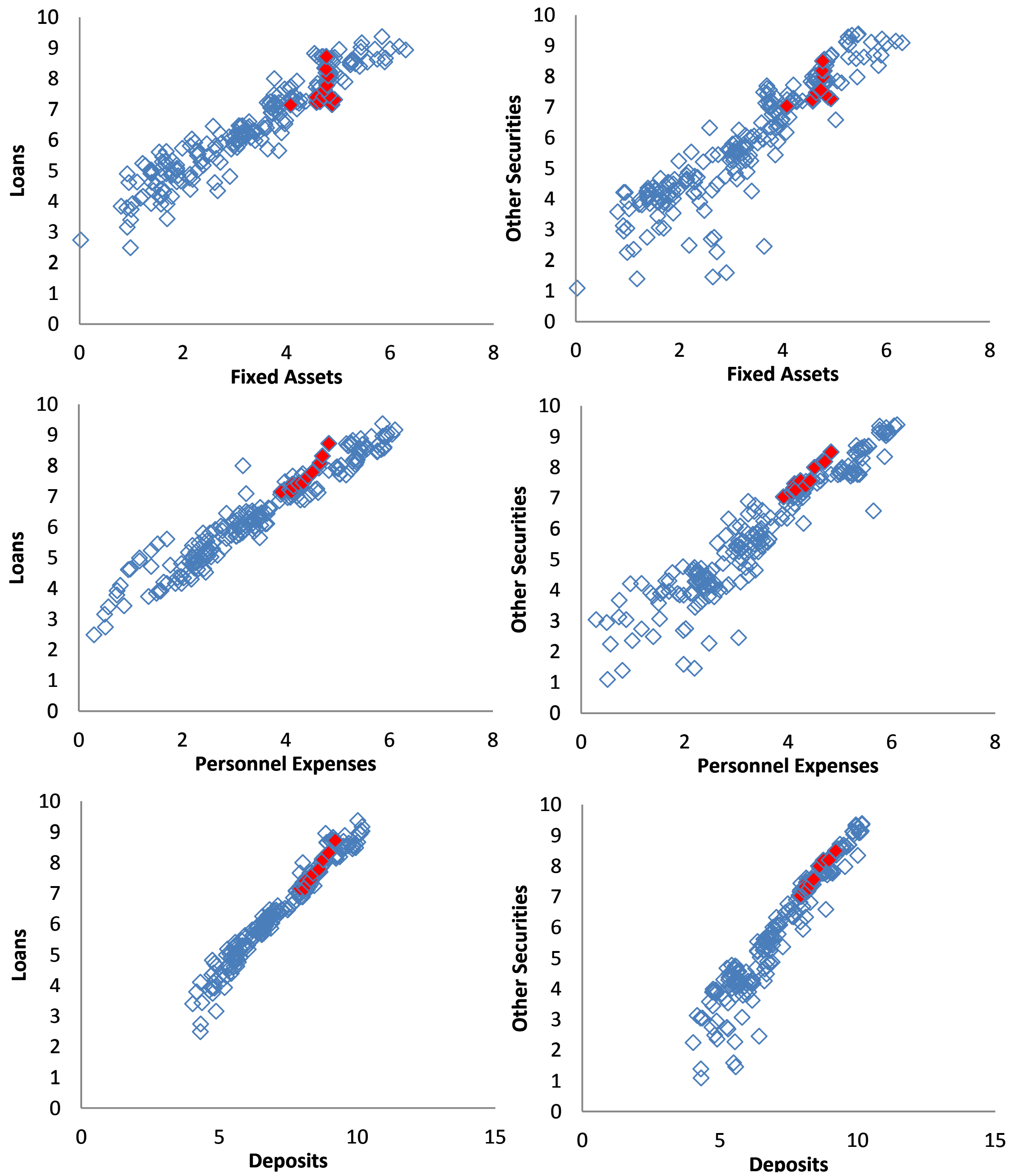


Figure XIV.4. Alpha Bank

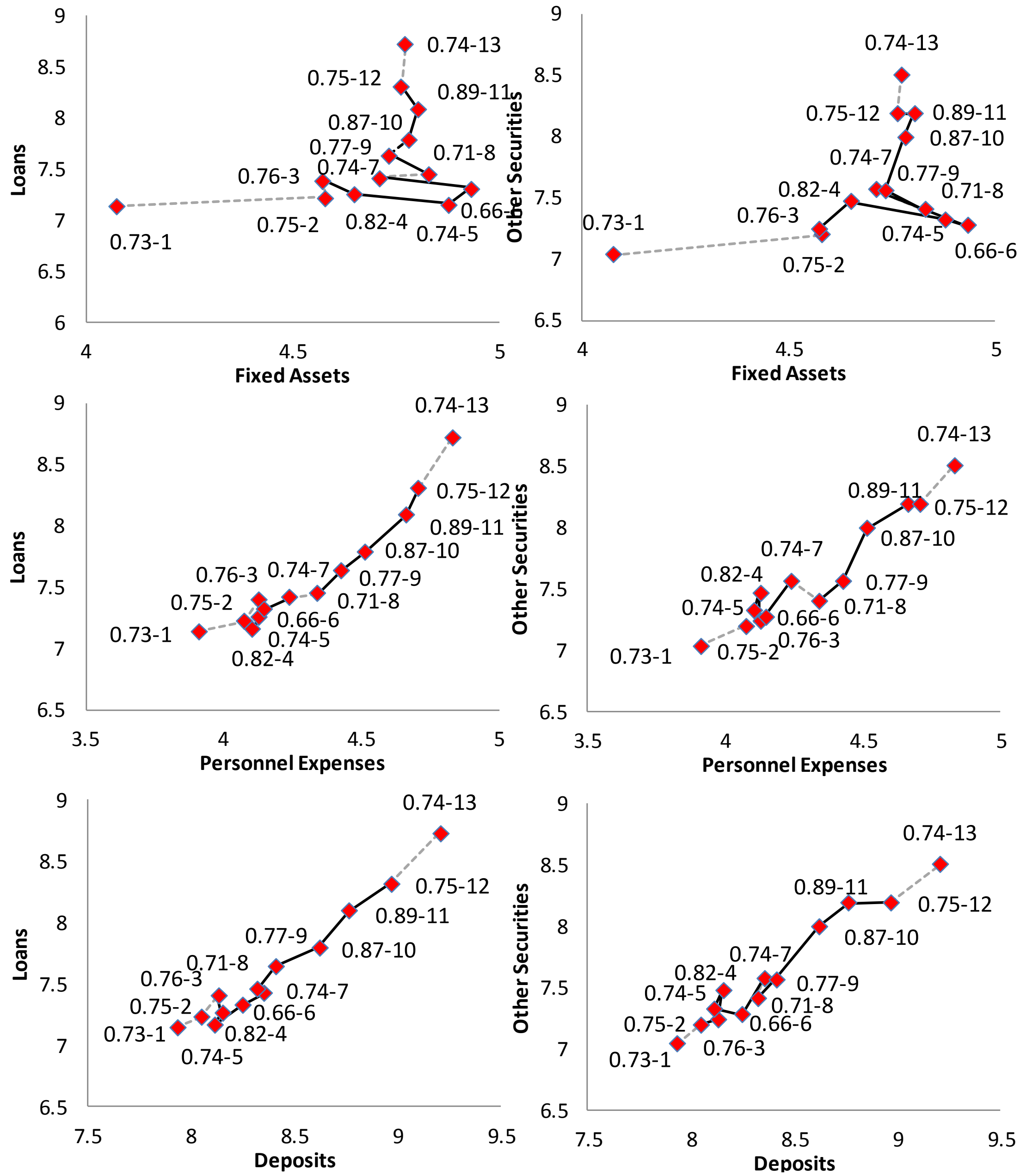


Figure XIV.5. Bank of Athens

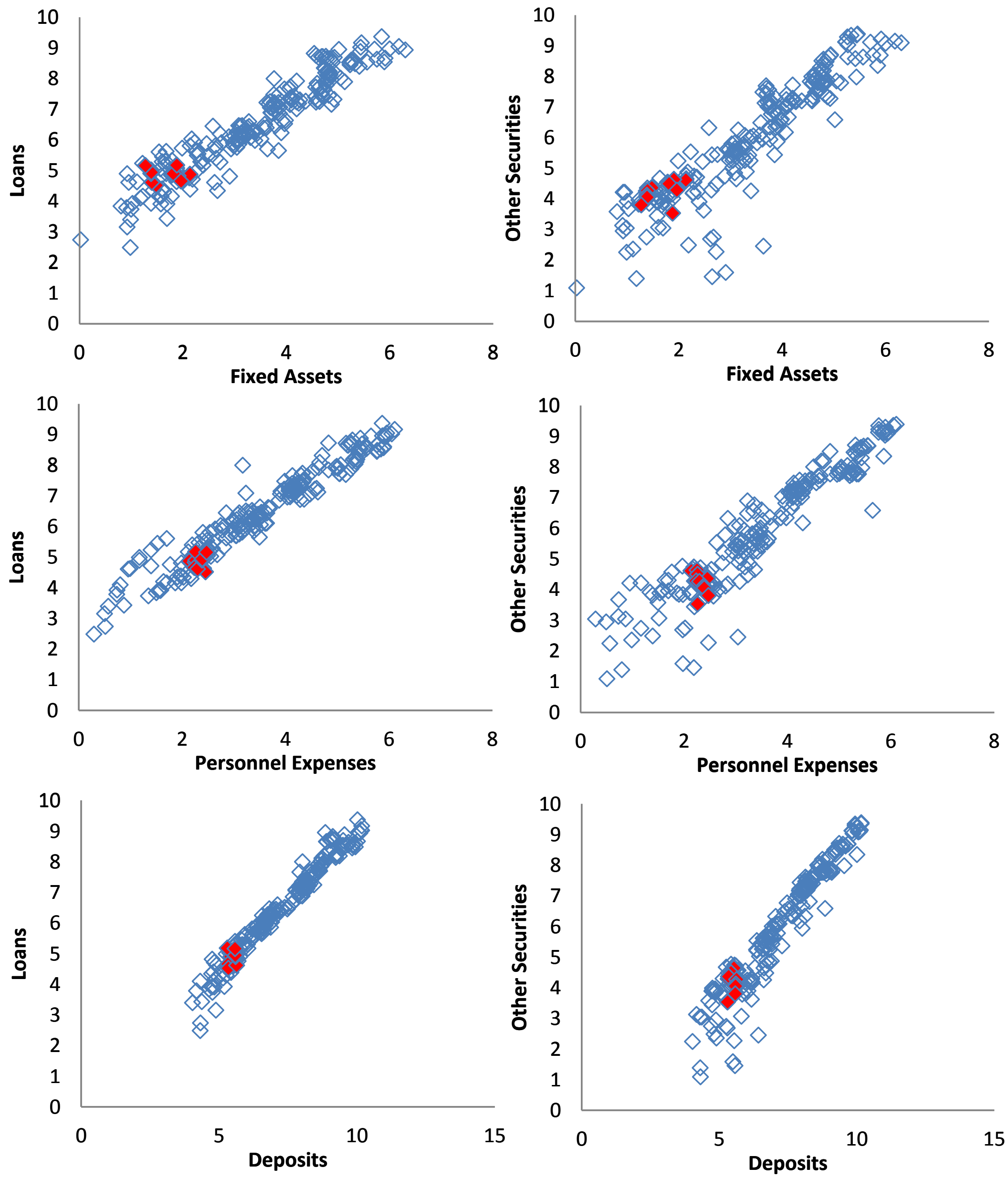


Figure XIV.6. Bank of Athens

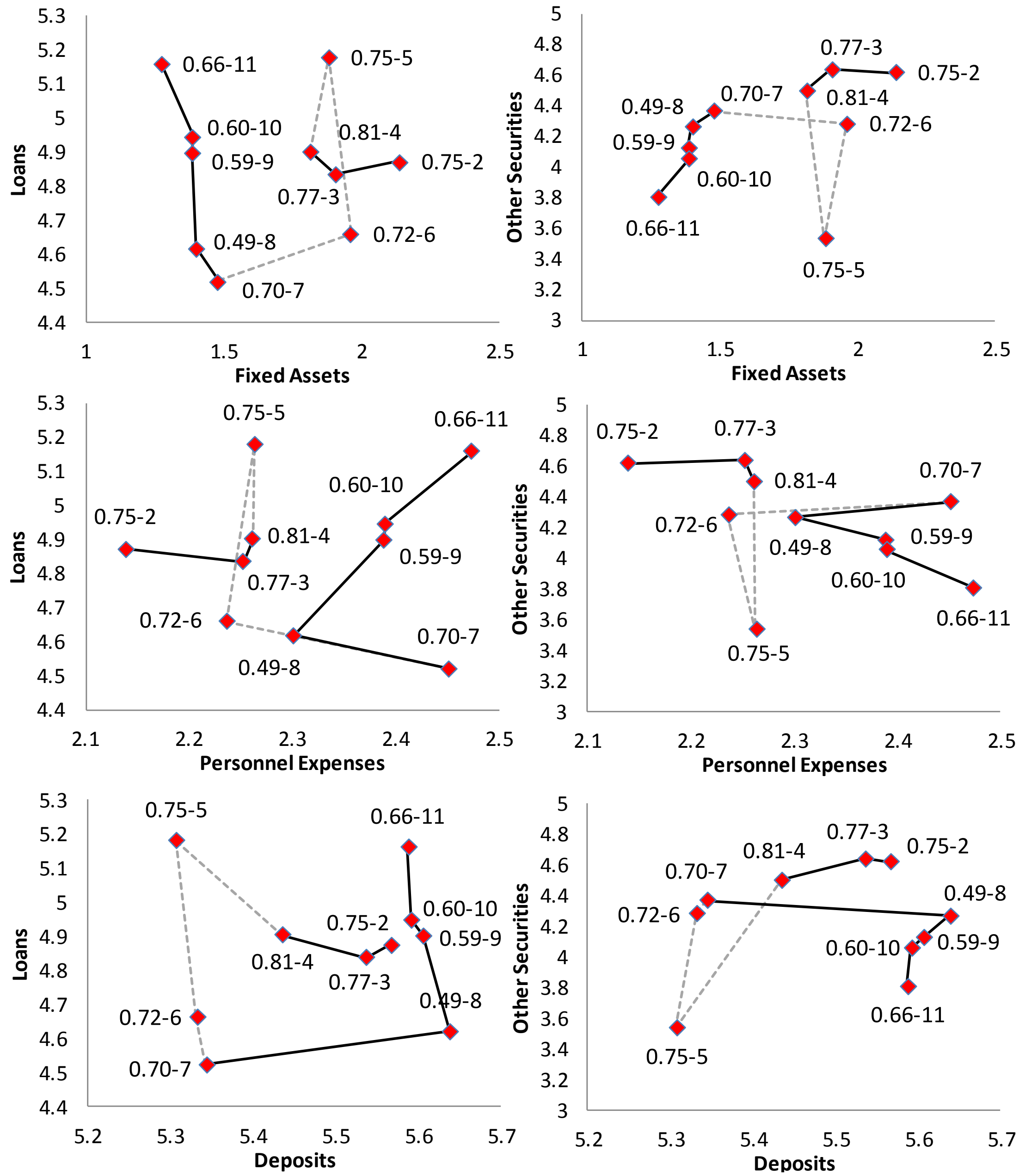


Figure XIV.7. Attica Bank

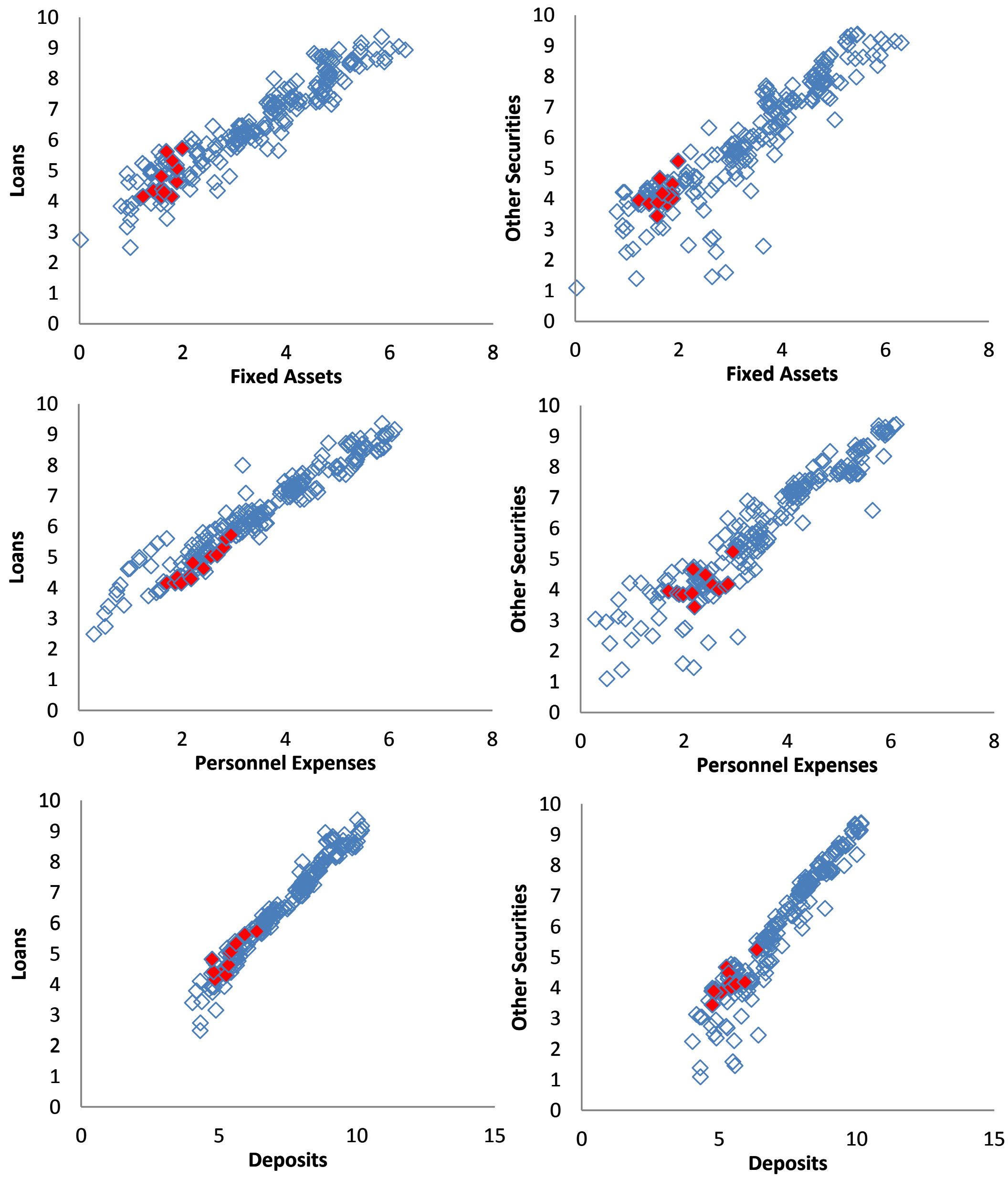


Figure XIV.8. Attica Bank

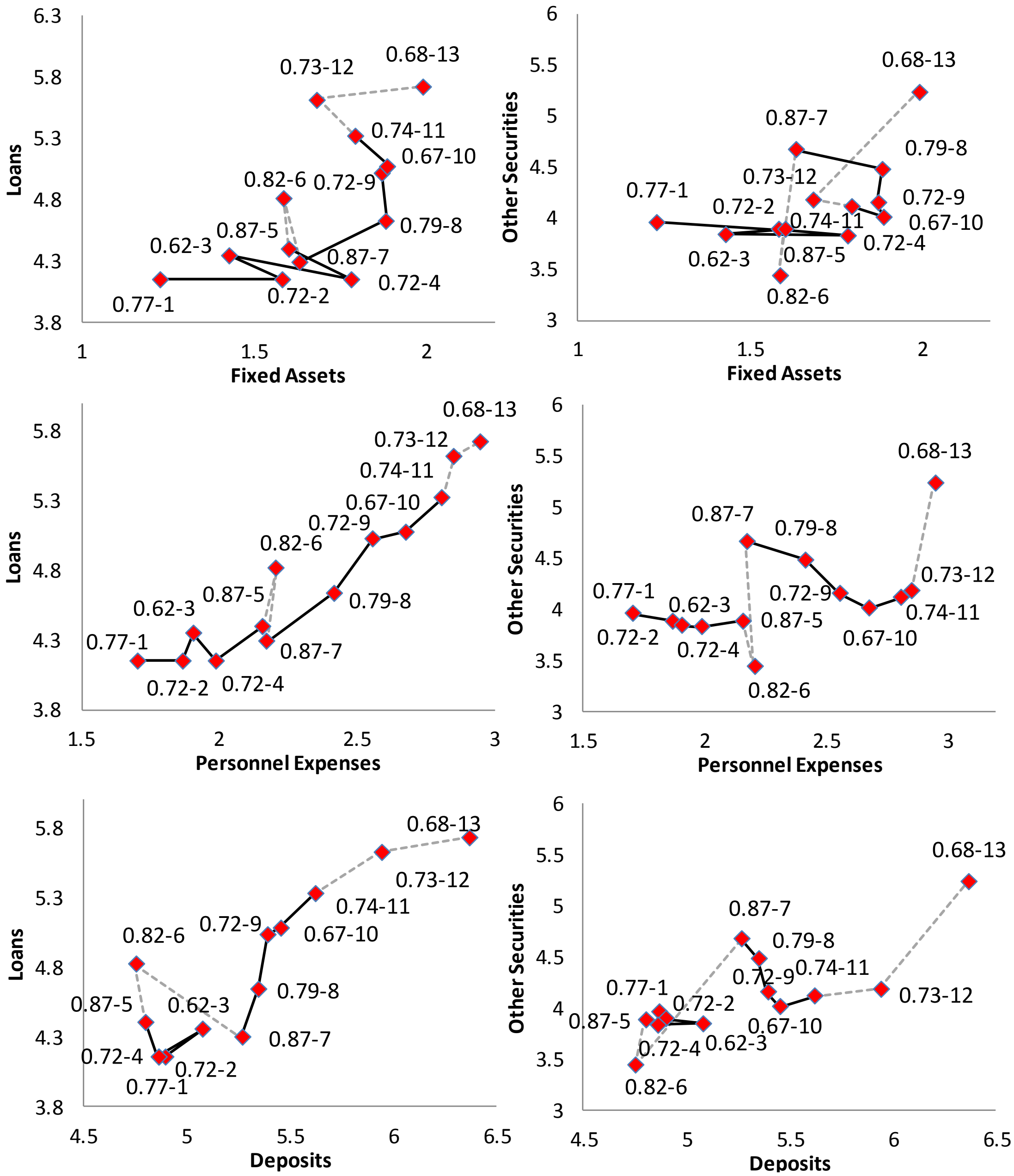


Figure XIV.9. Bank of Central Greece

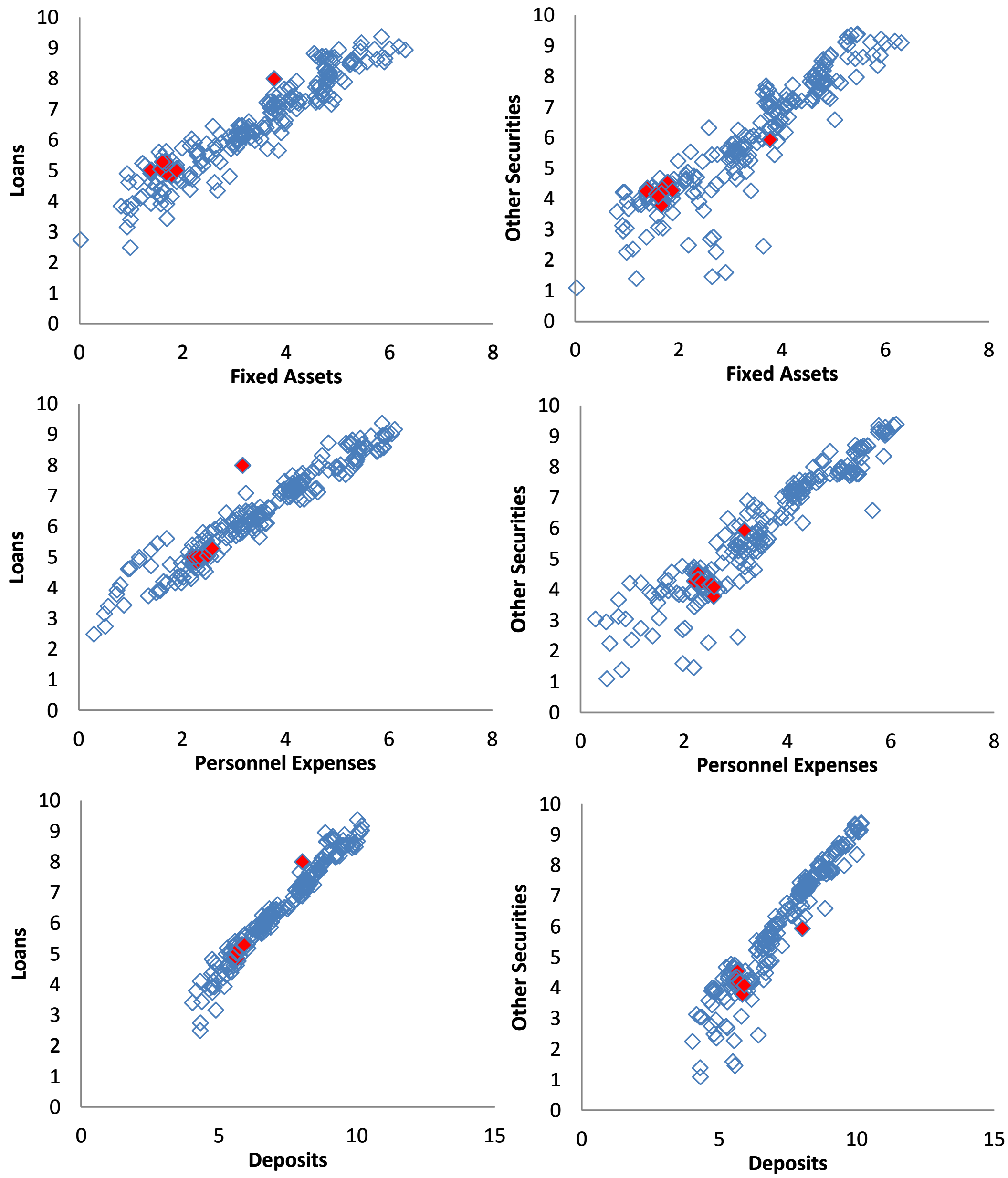
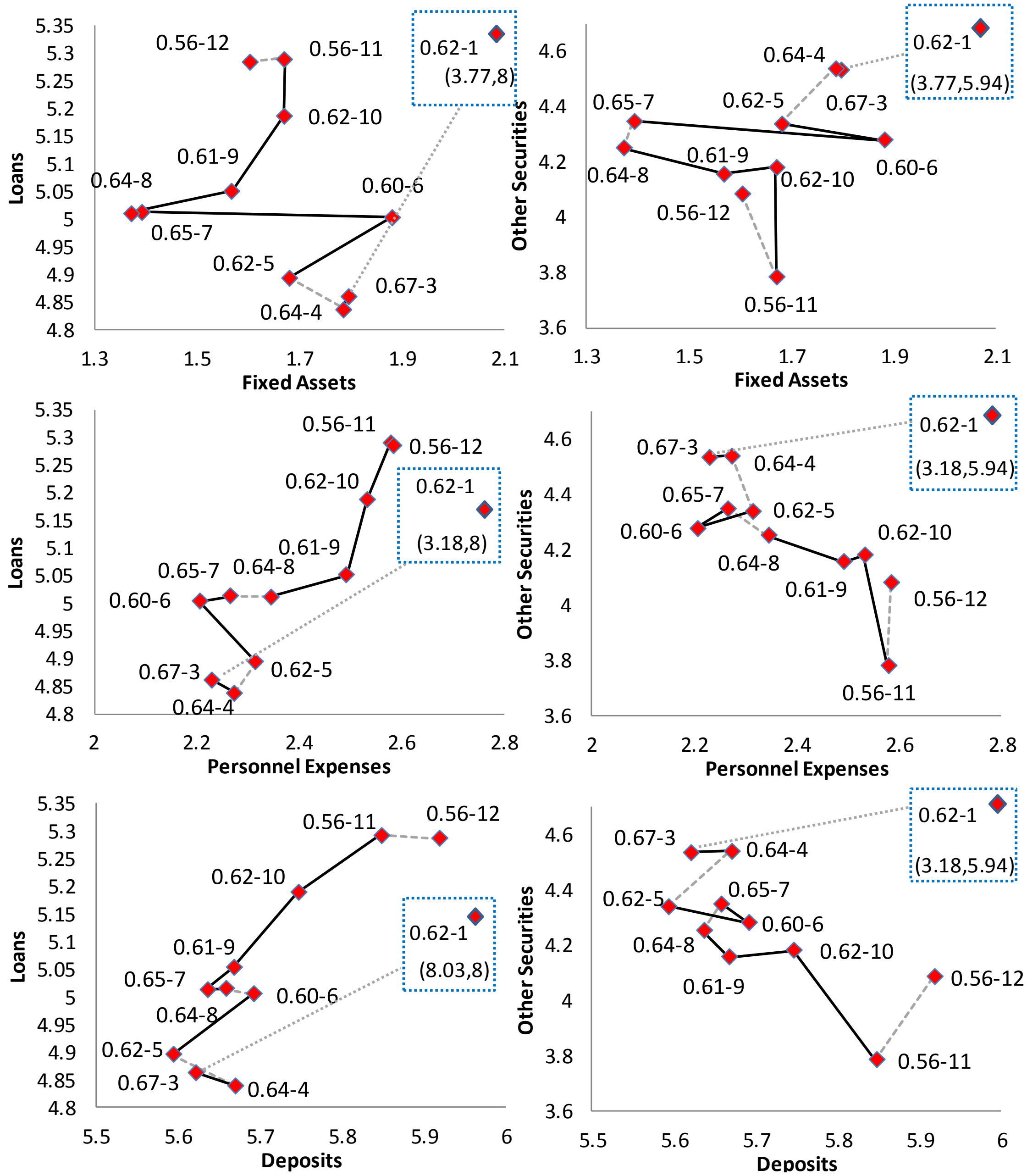


Figure XIV.10. Bank of Central Greece[153]



[153] The dotted boxes are the operations of Bank of Central Greece during 1987, which were substantially greater than in other years and hence it would affect the scaling of the axes to such an extent that it would be impossible to inspect the trajectory for this bank. The massive drop evidenced is due to a political scandal and resulted in public organisations switching their banking to other financial institutions. The jump from period 1 (1987) to 3 (1989) is due to lack of data for 1988.

Figure XIV.11. Bank of Crete - Cretabank

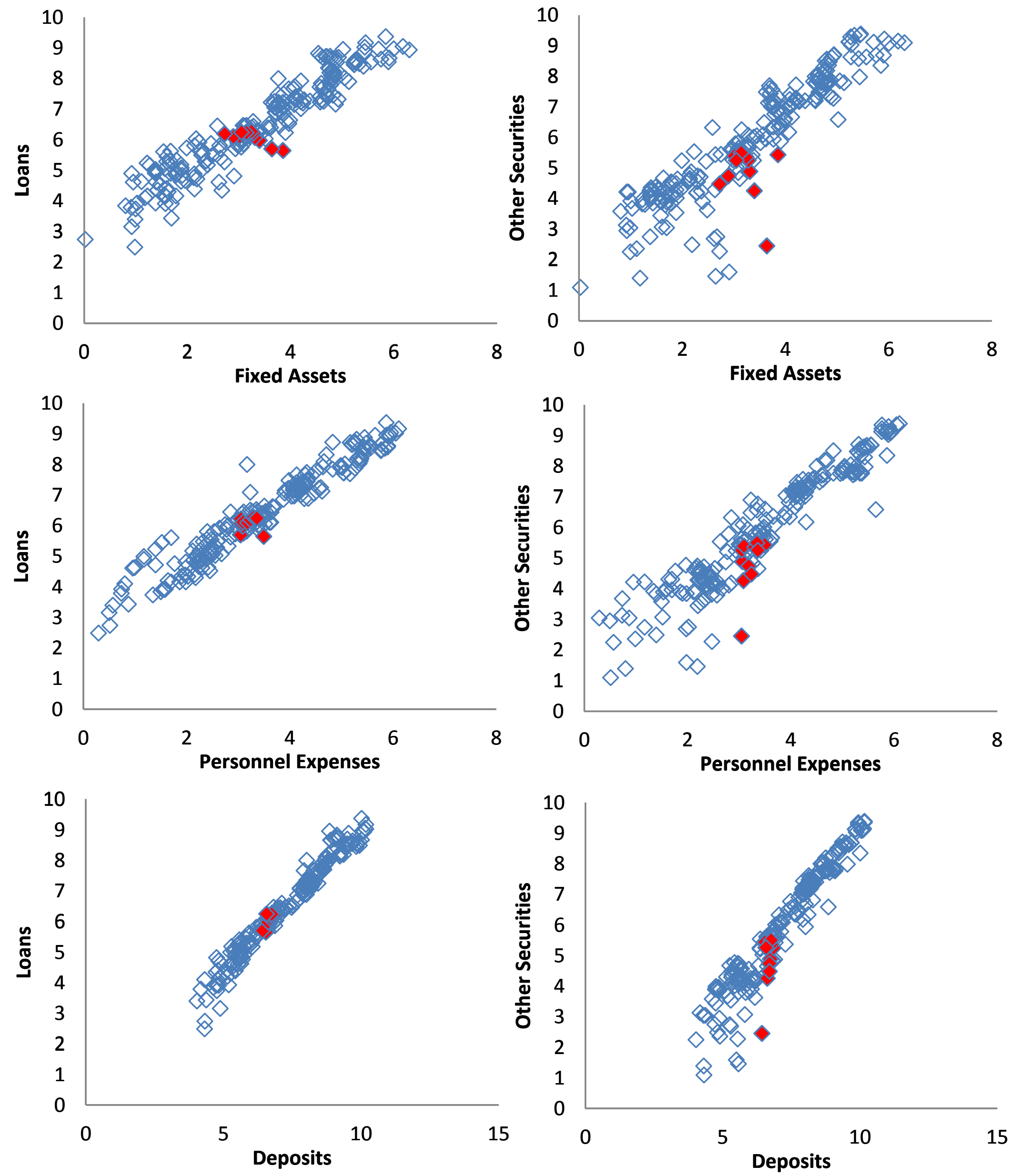


Figure XIV.12. Bank of Crete - Cretabank

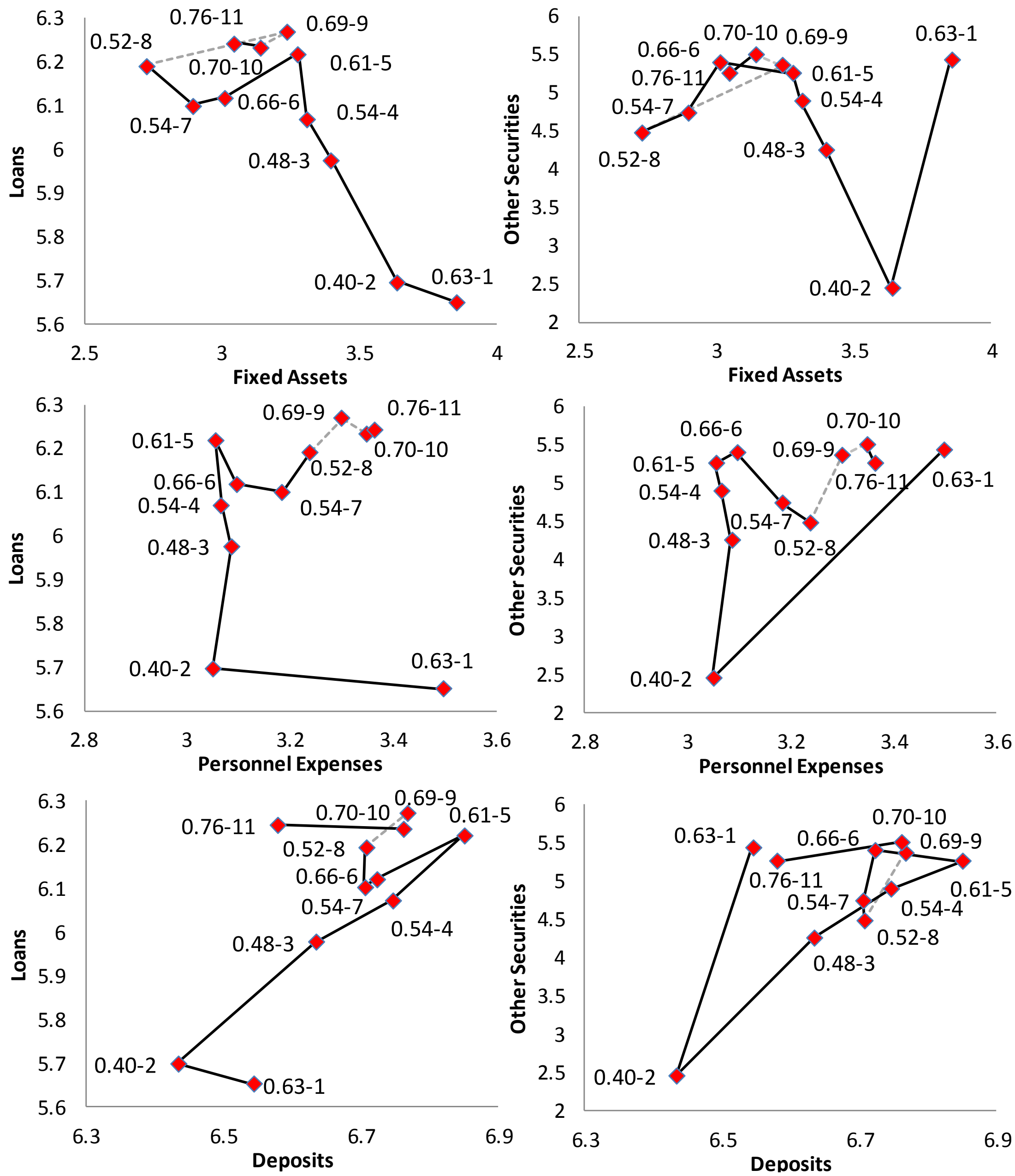


Figure XIV.13. Egnatia Bank

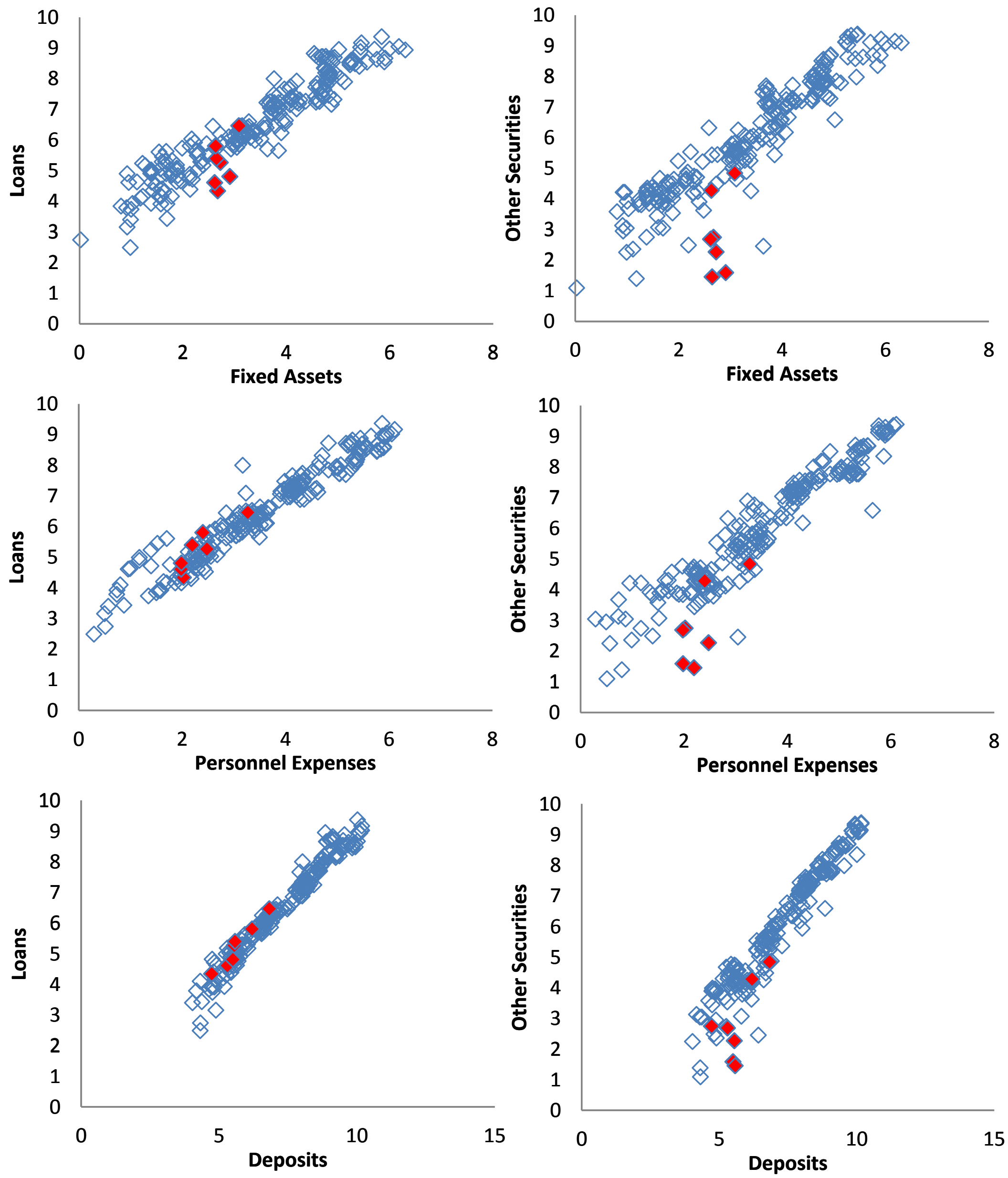


Figure XIV.14. Egnatia Bank

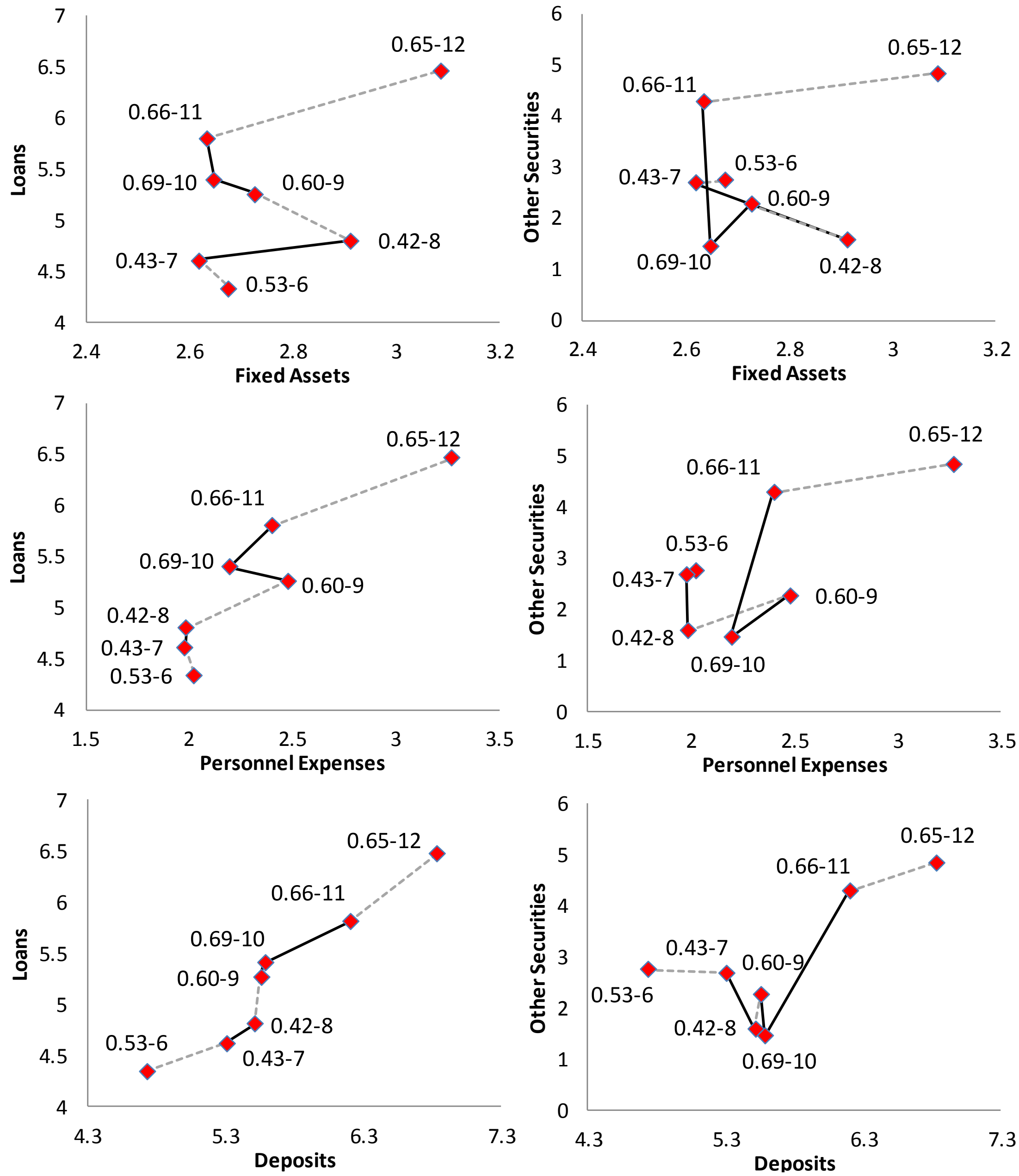


Figure XIV.15. Emporiki Bank

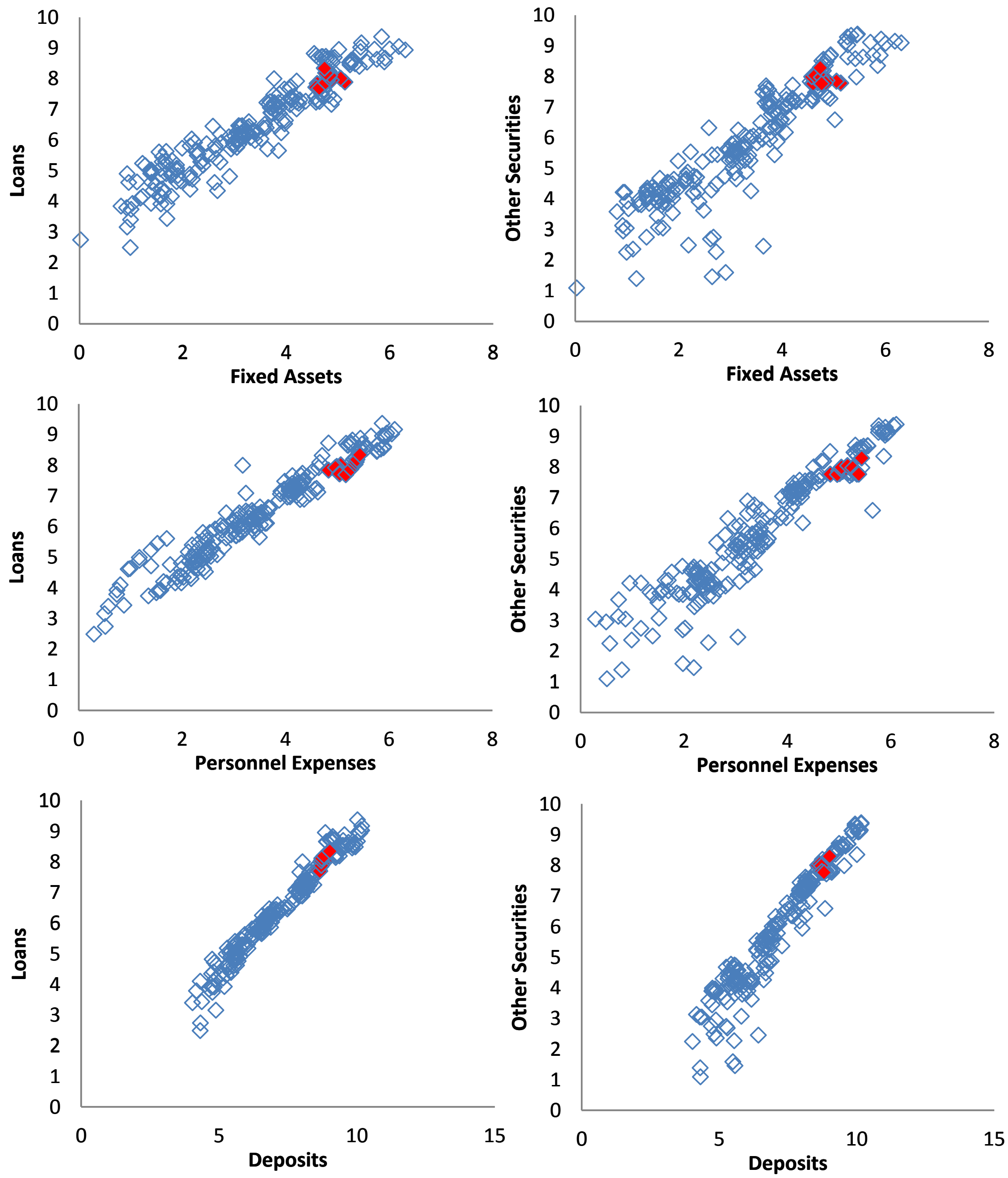


Figure XIV.16. Emporiki Bank

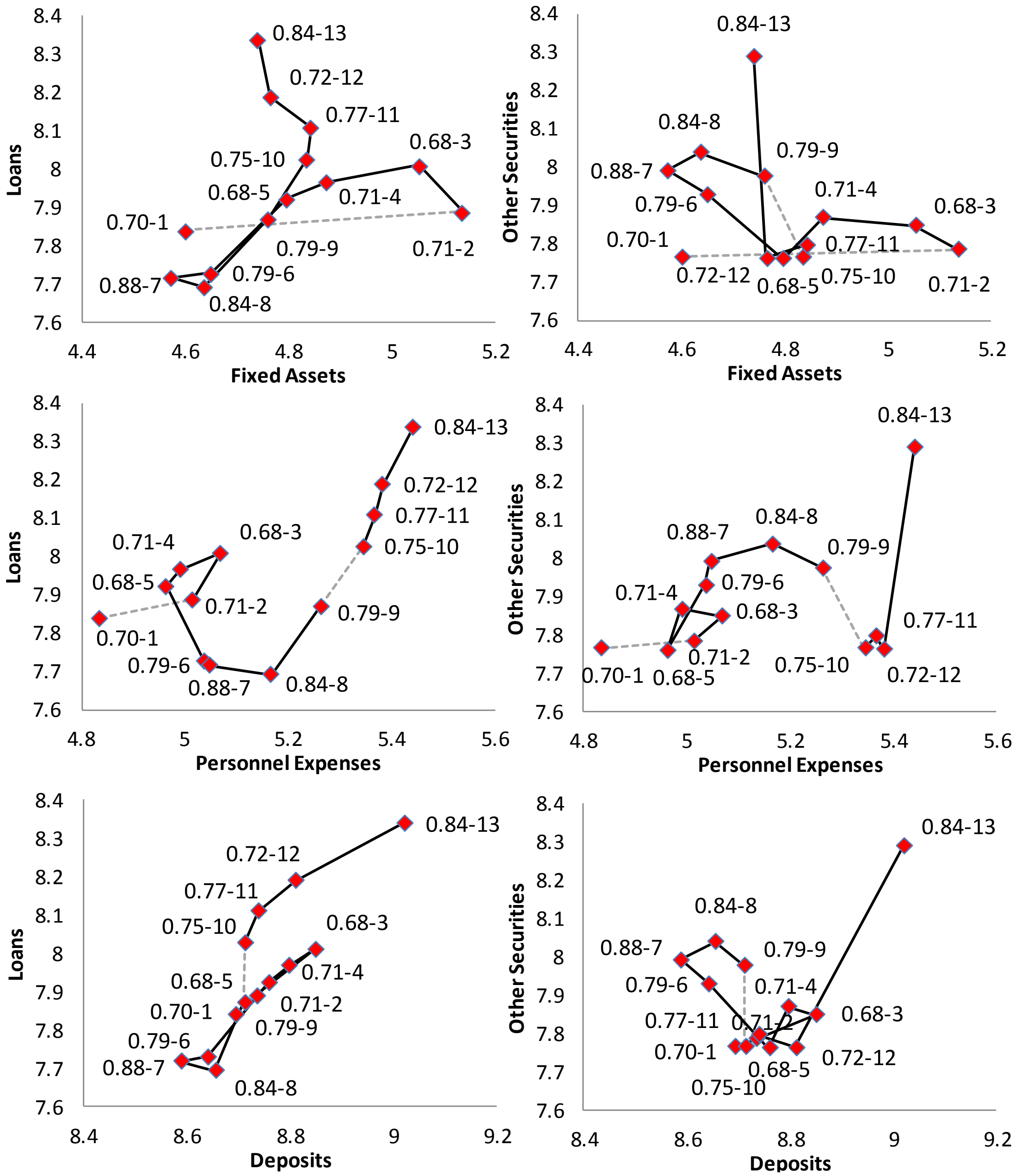


Figure XIV.17. Ergobank

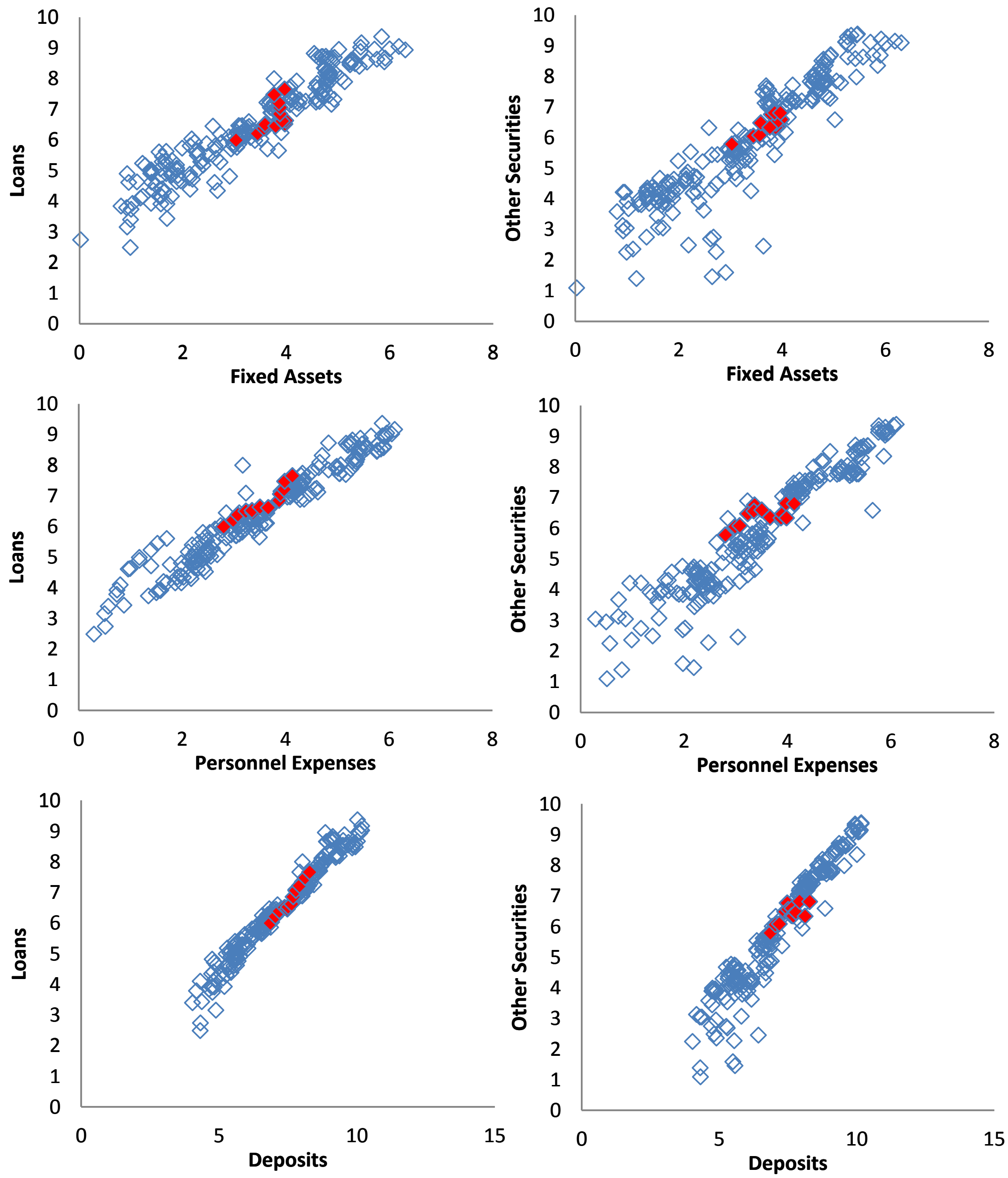


Figure XIV.18. Ergobank

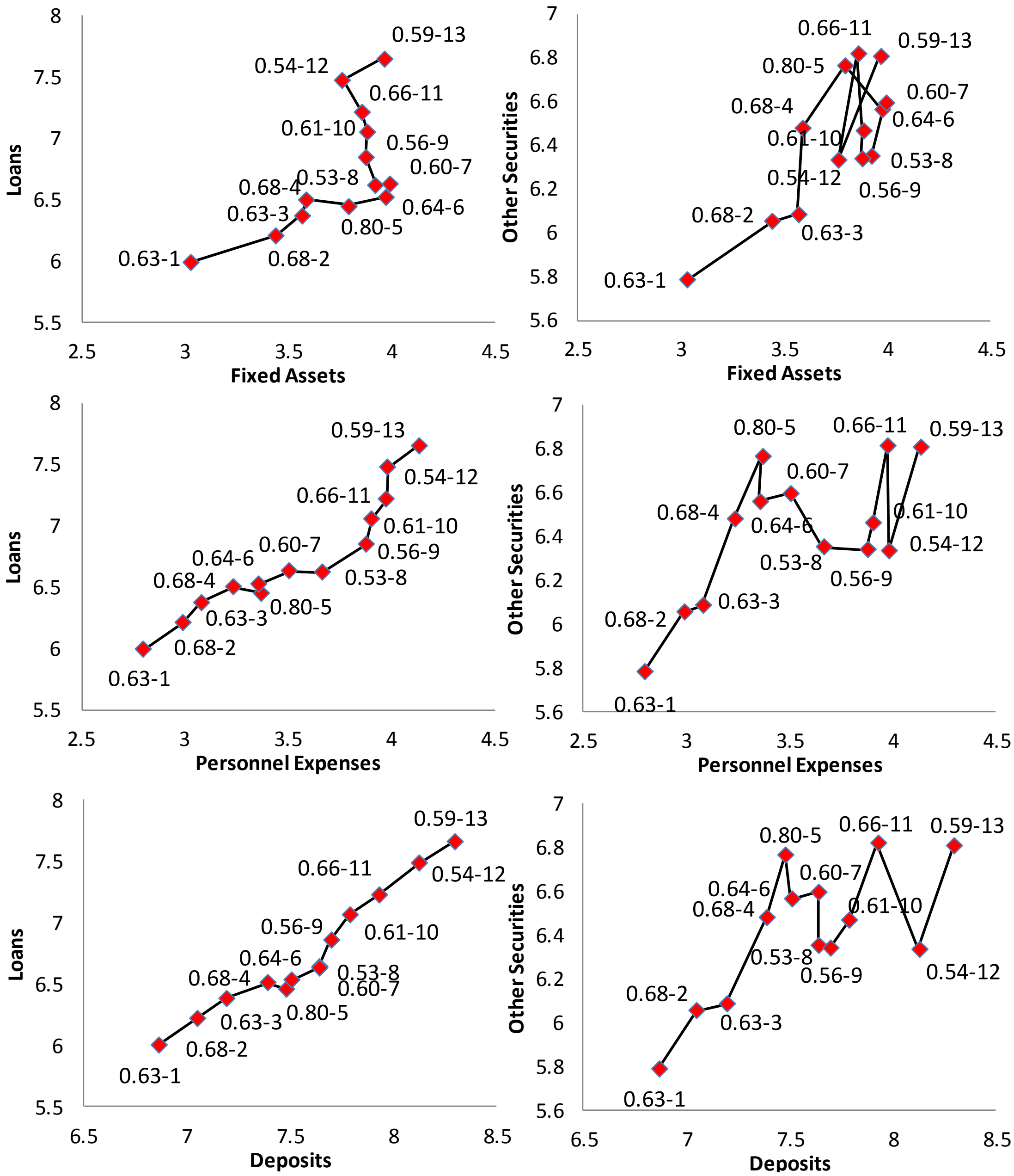


Figure XIV.19. EFG Eurobank

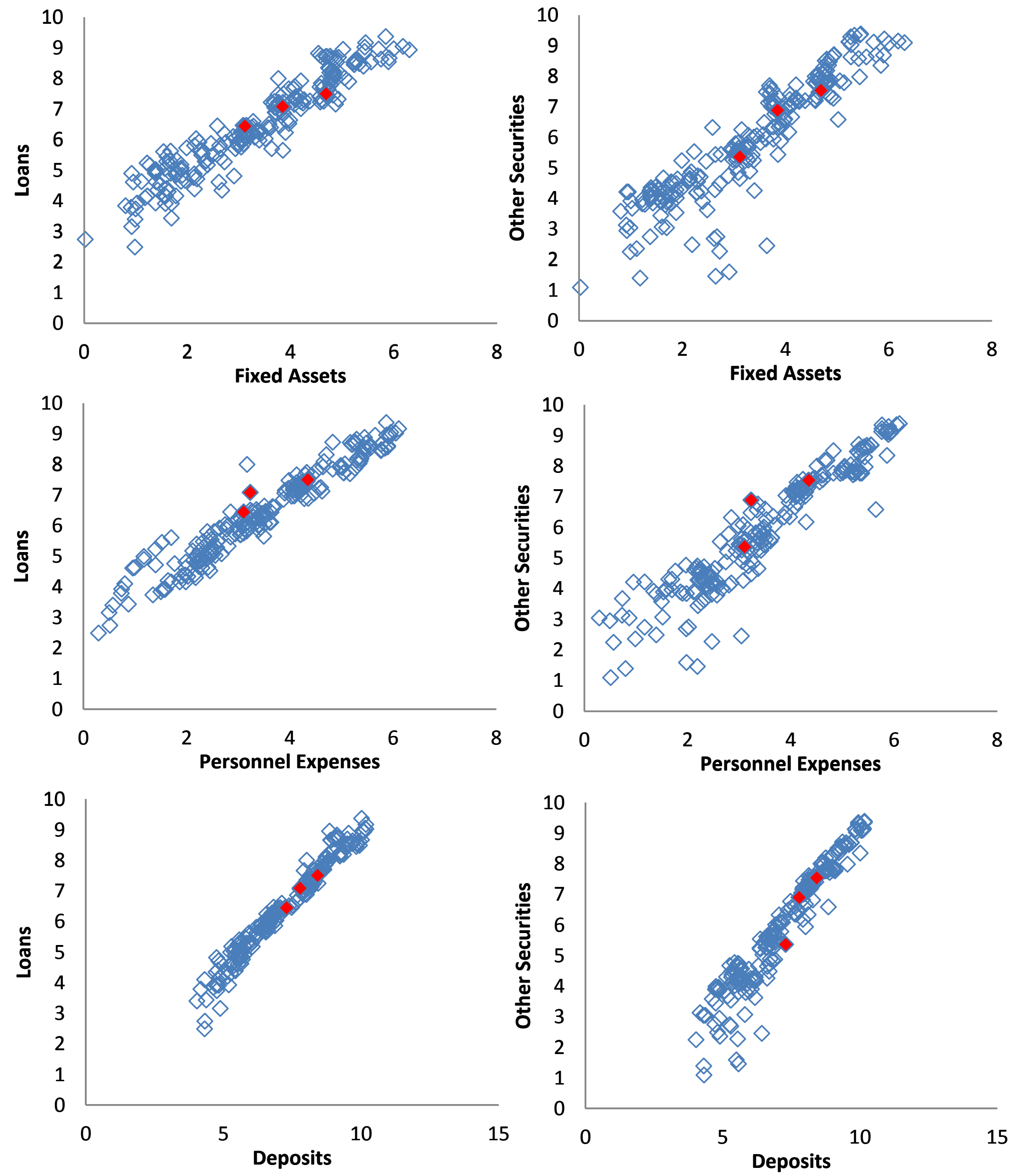


Figure XIV.20. EFG Eurobank

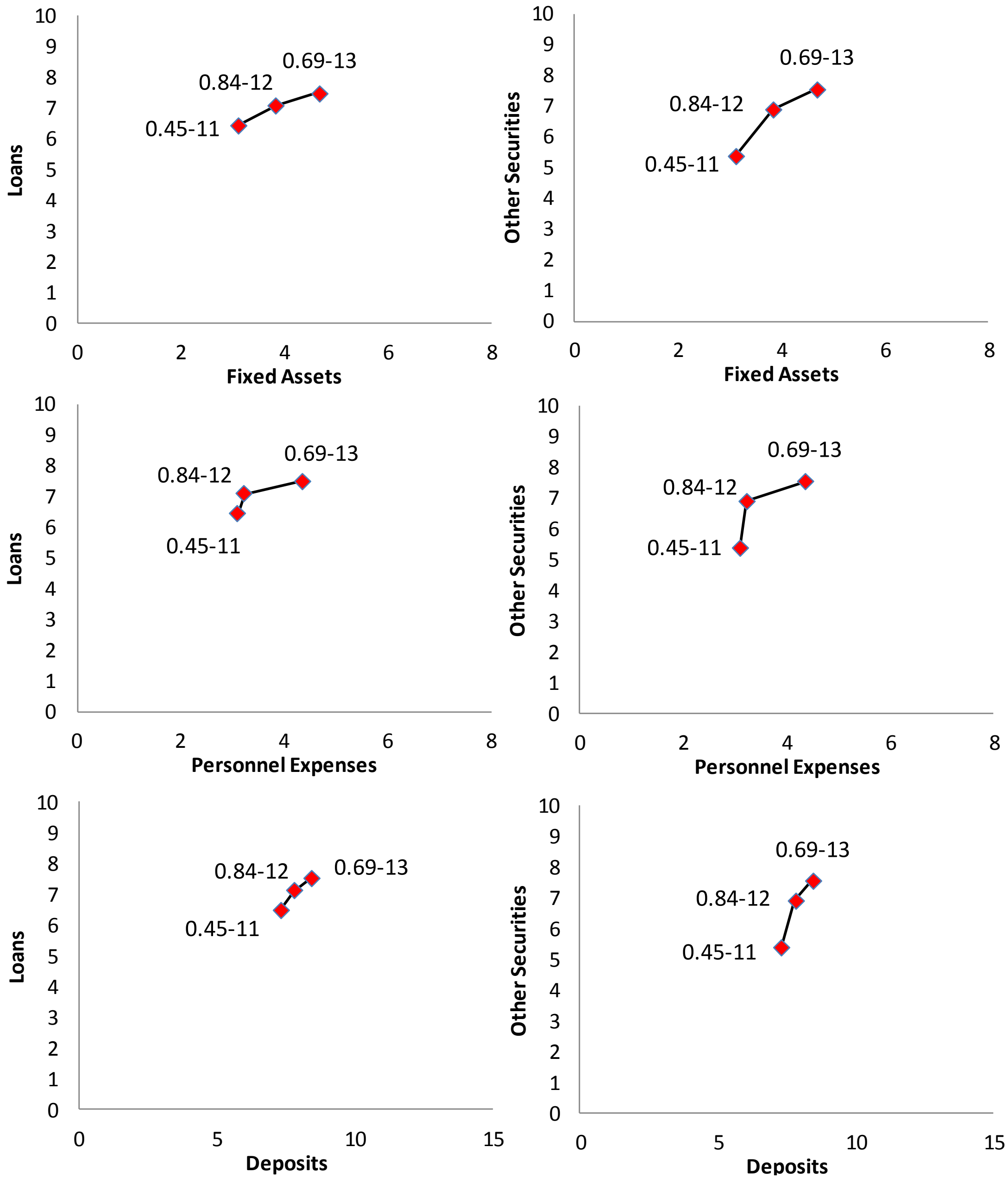


Figure XIV.21. General Bank

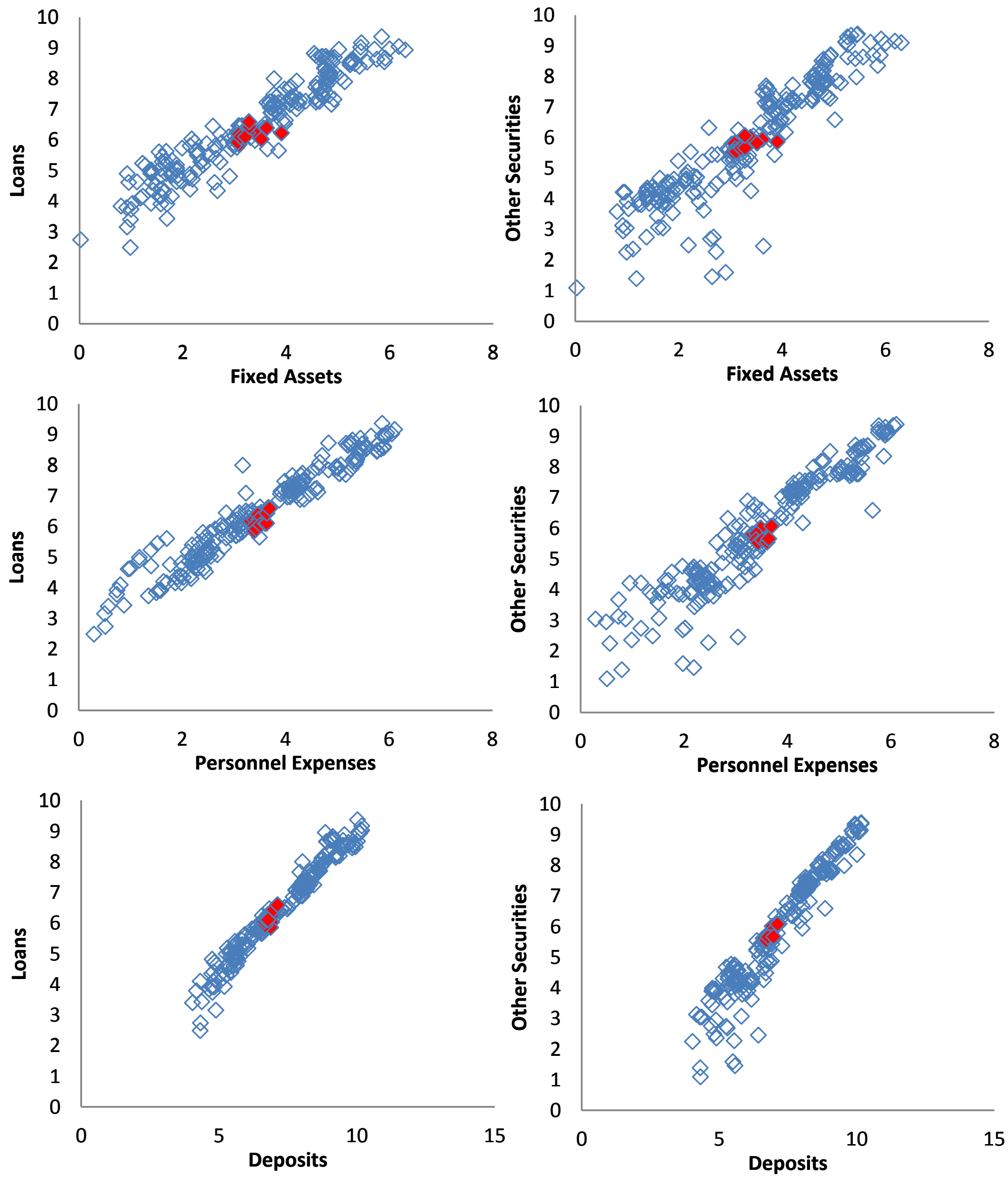


Figure XIV.22. General Bank

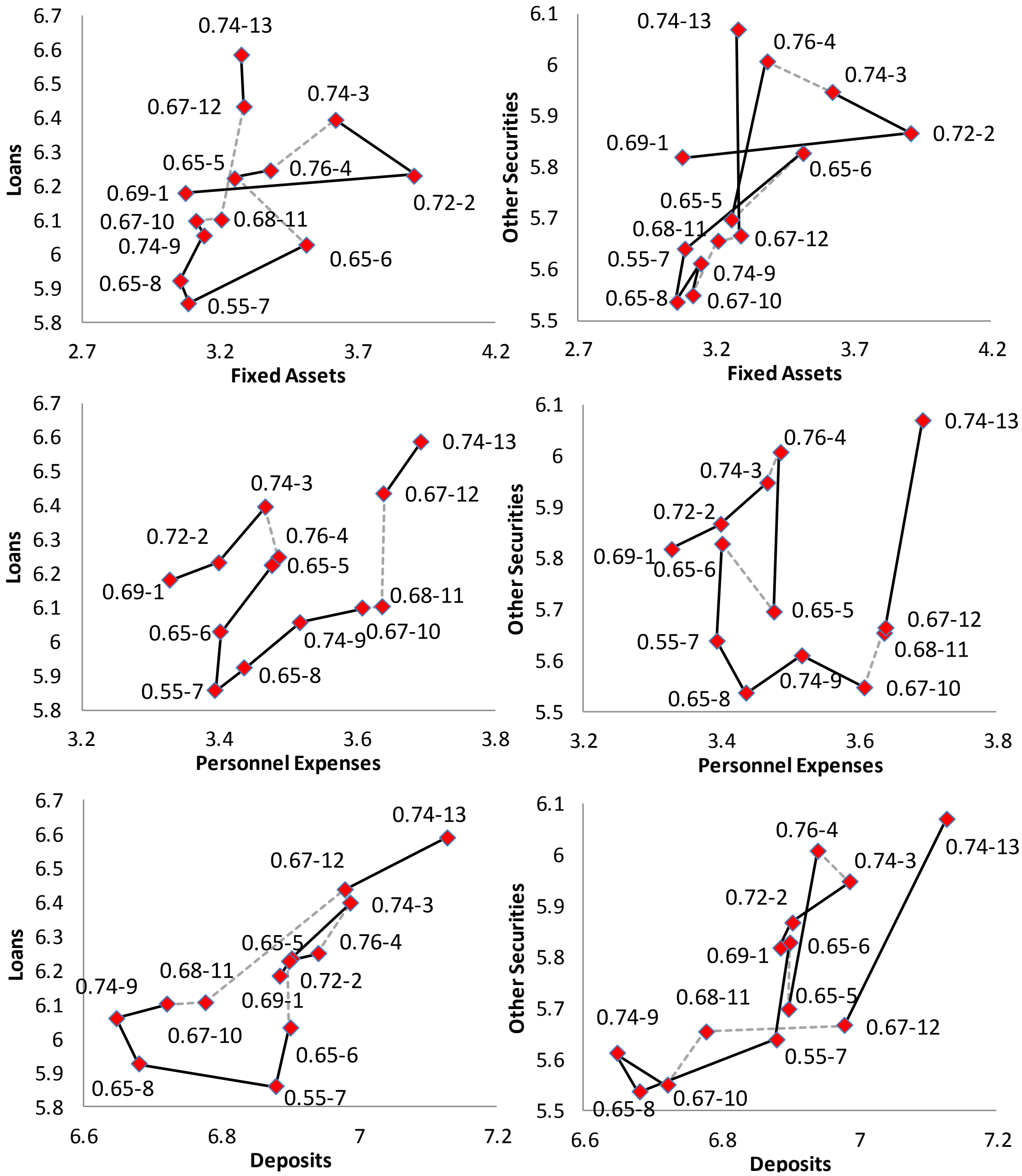


Figure XIV.23. Interbank

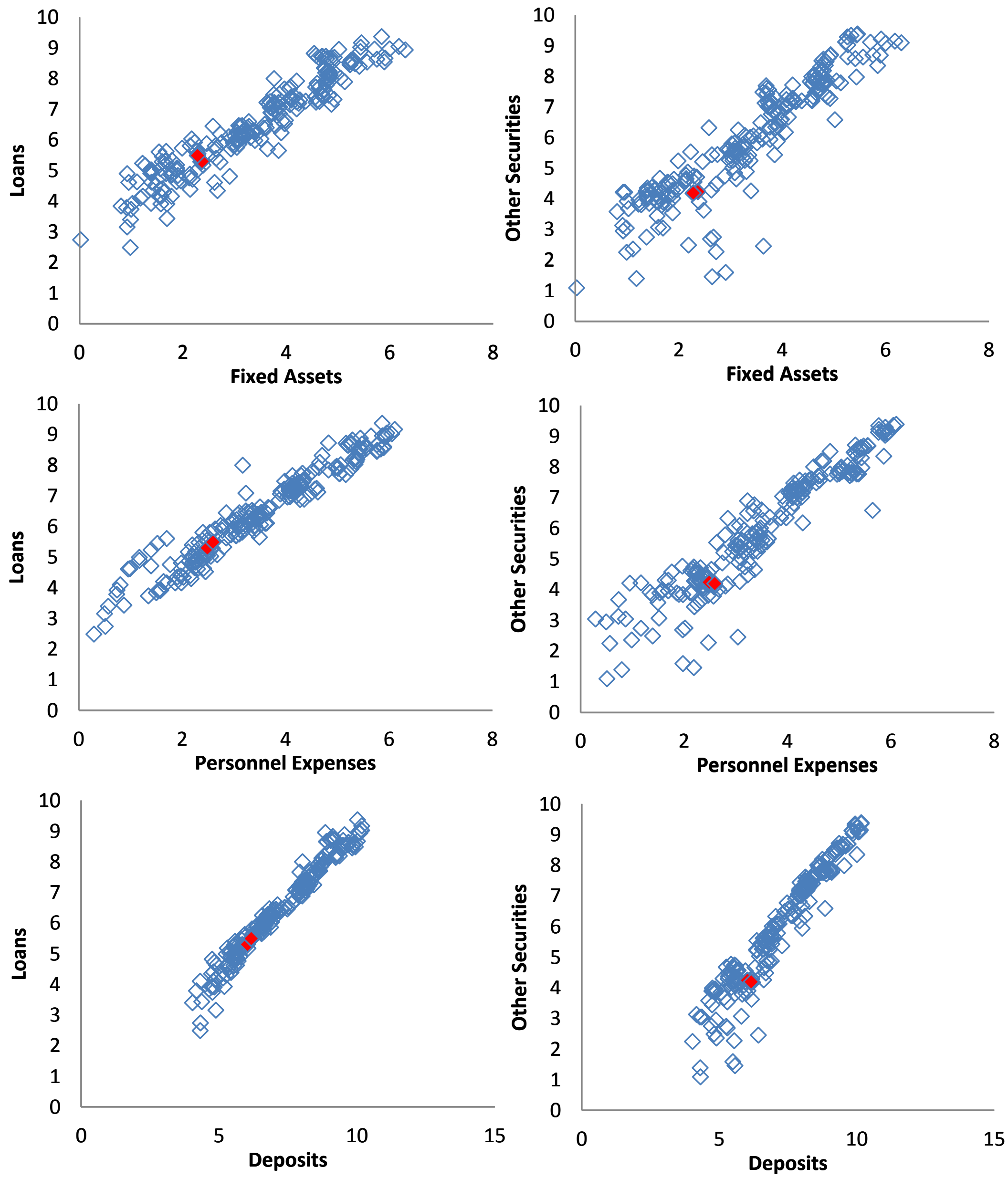


Figure XIV.24. Interbank

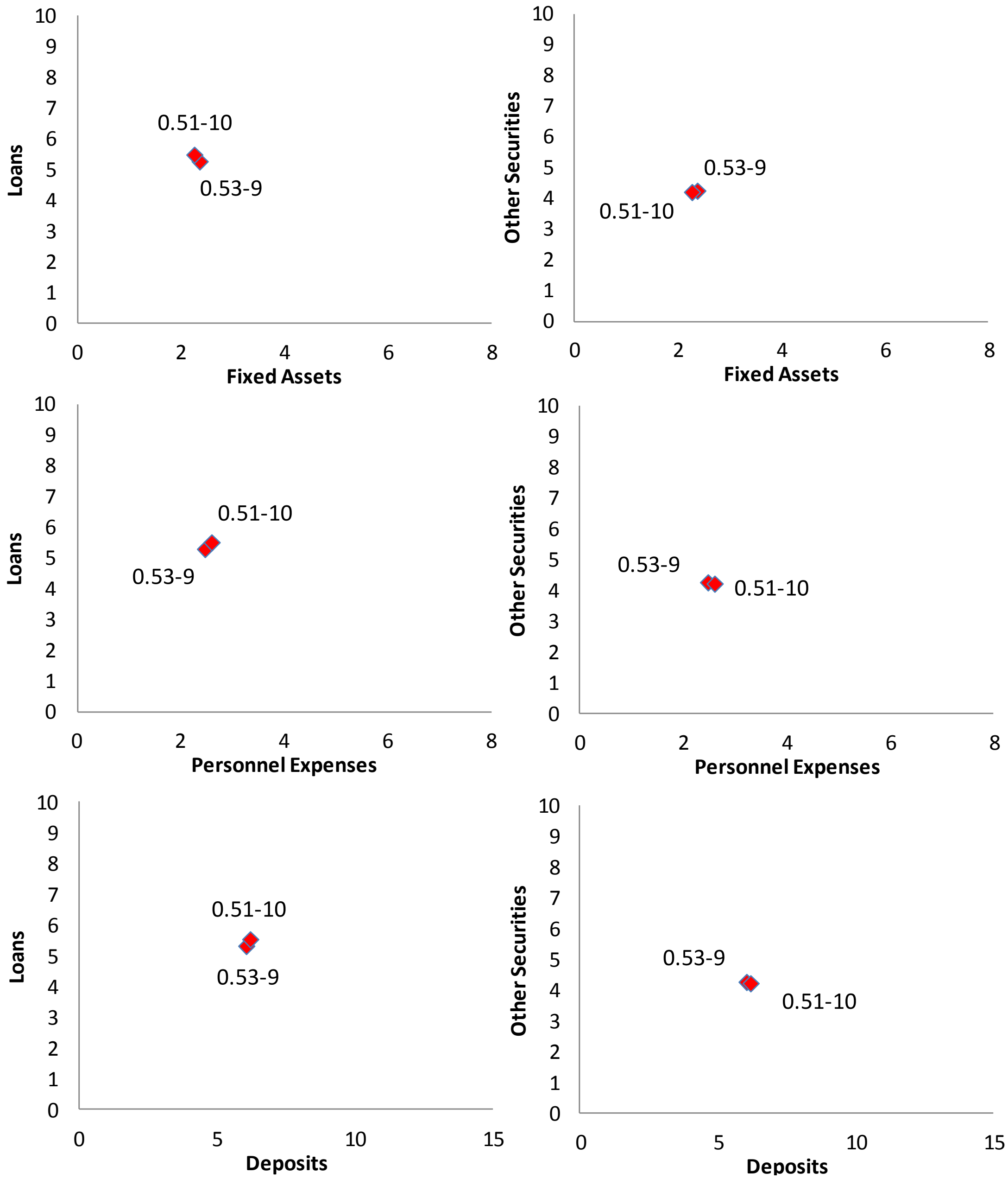


Figure XIV.25. Ionian and Popular Bank

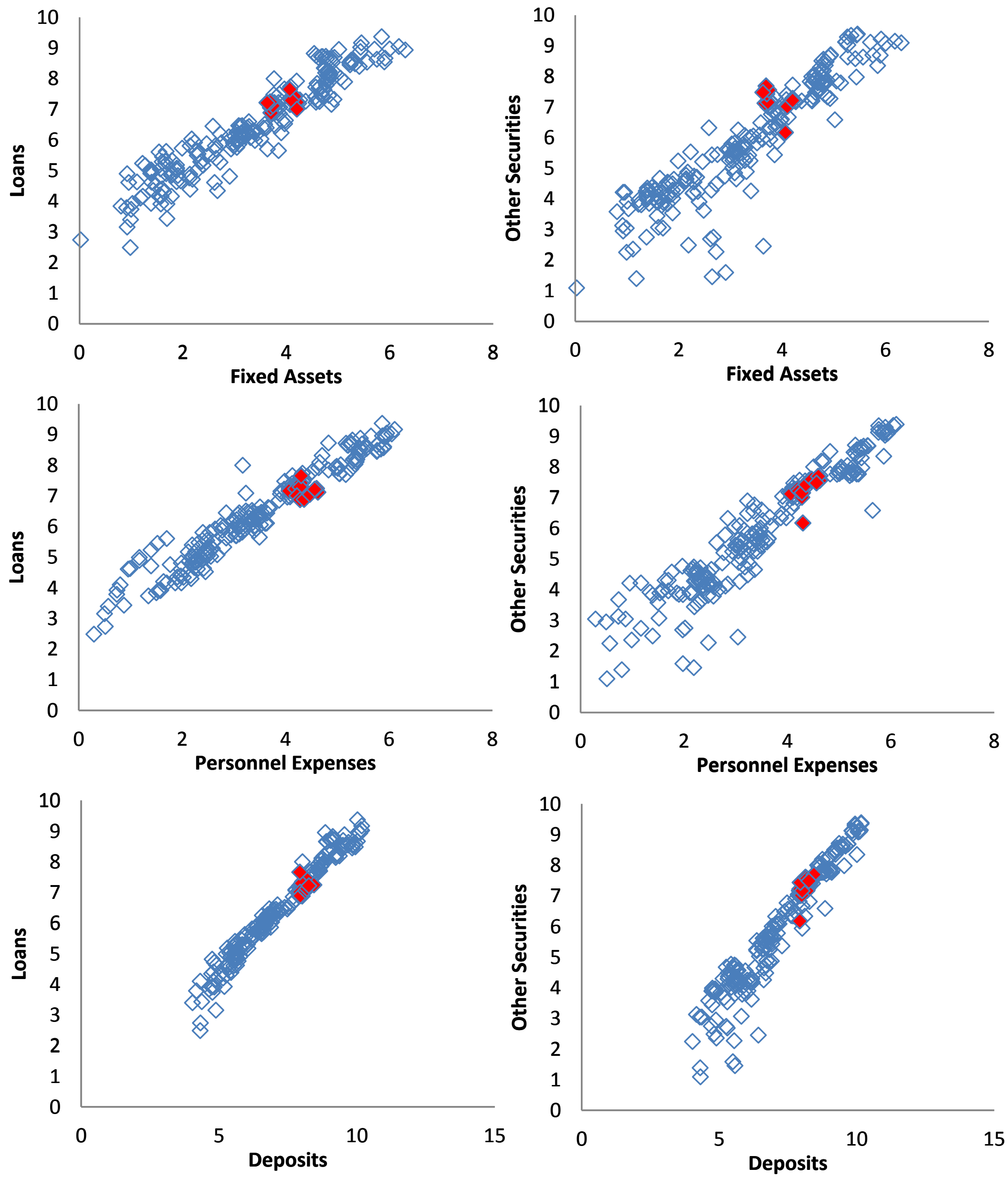


Figure XIV.26. Ionian and Popular Bank

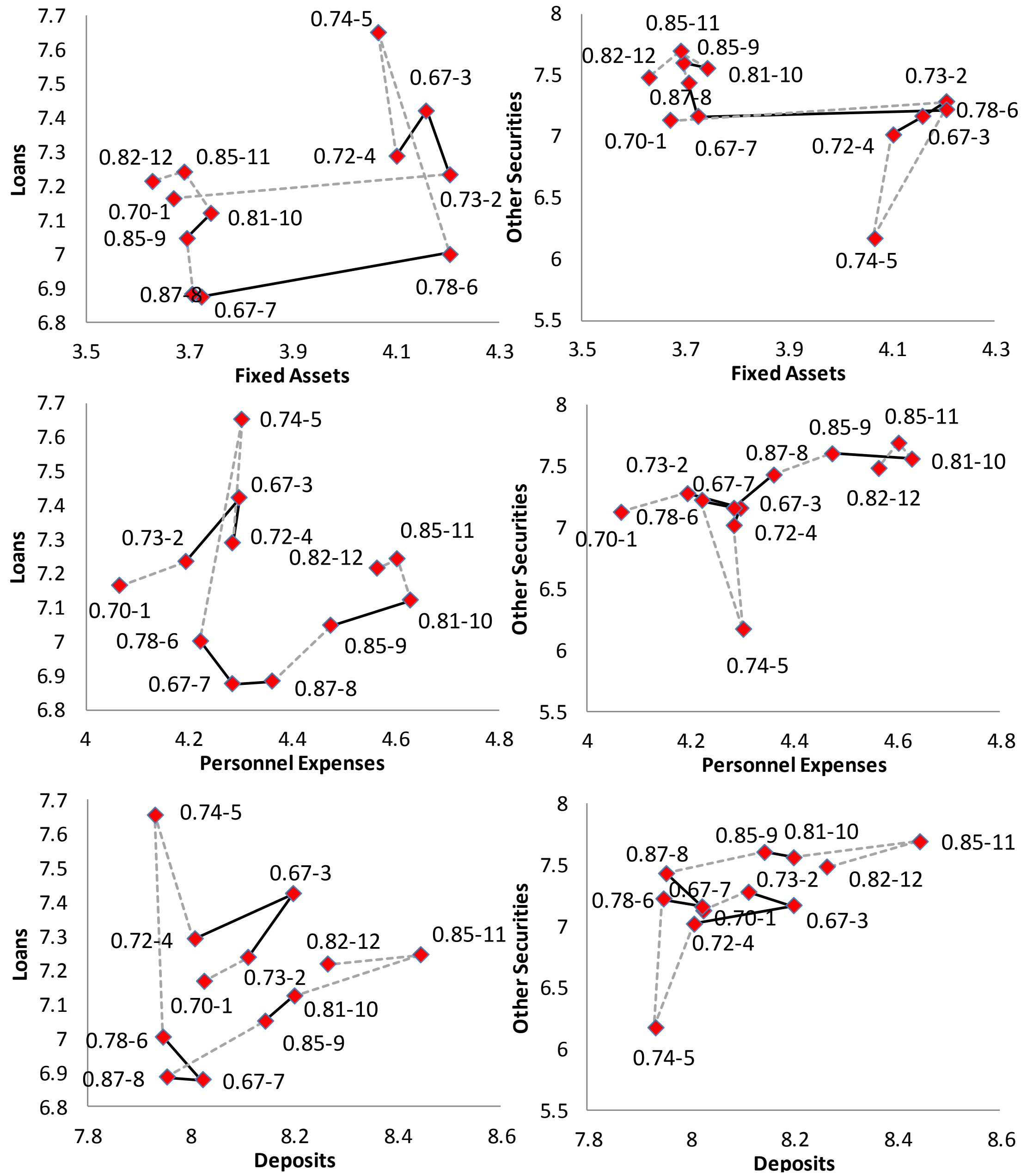


Figure XIV.27. Laiki Bank

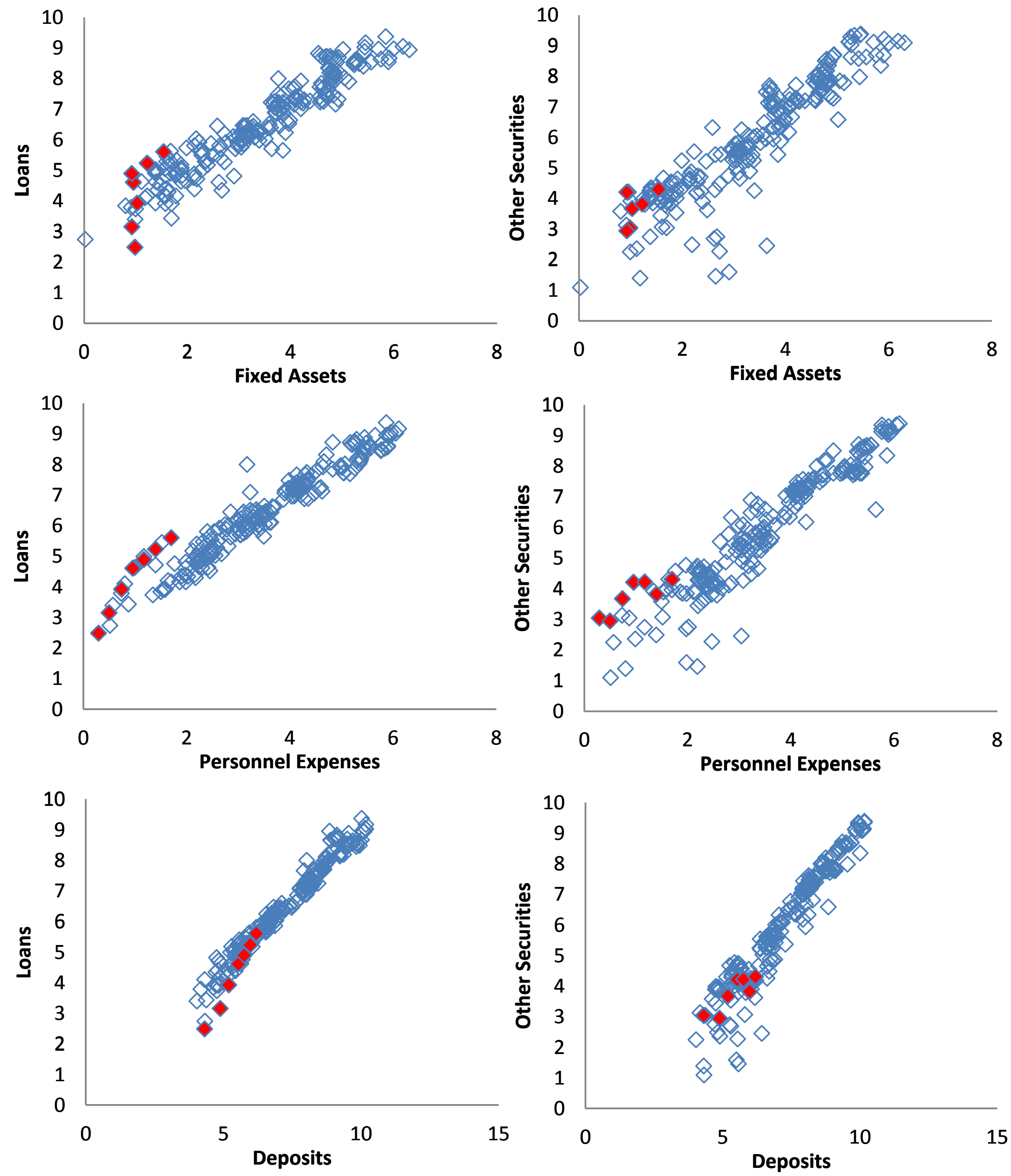


Figure XIV.28. Laiki Bank

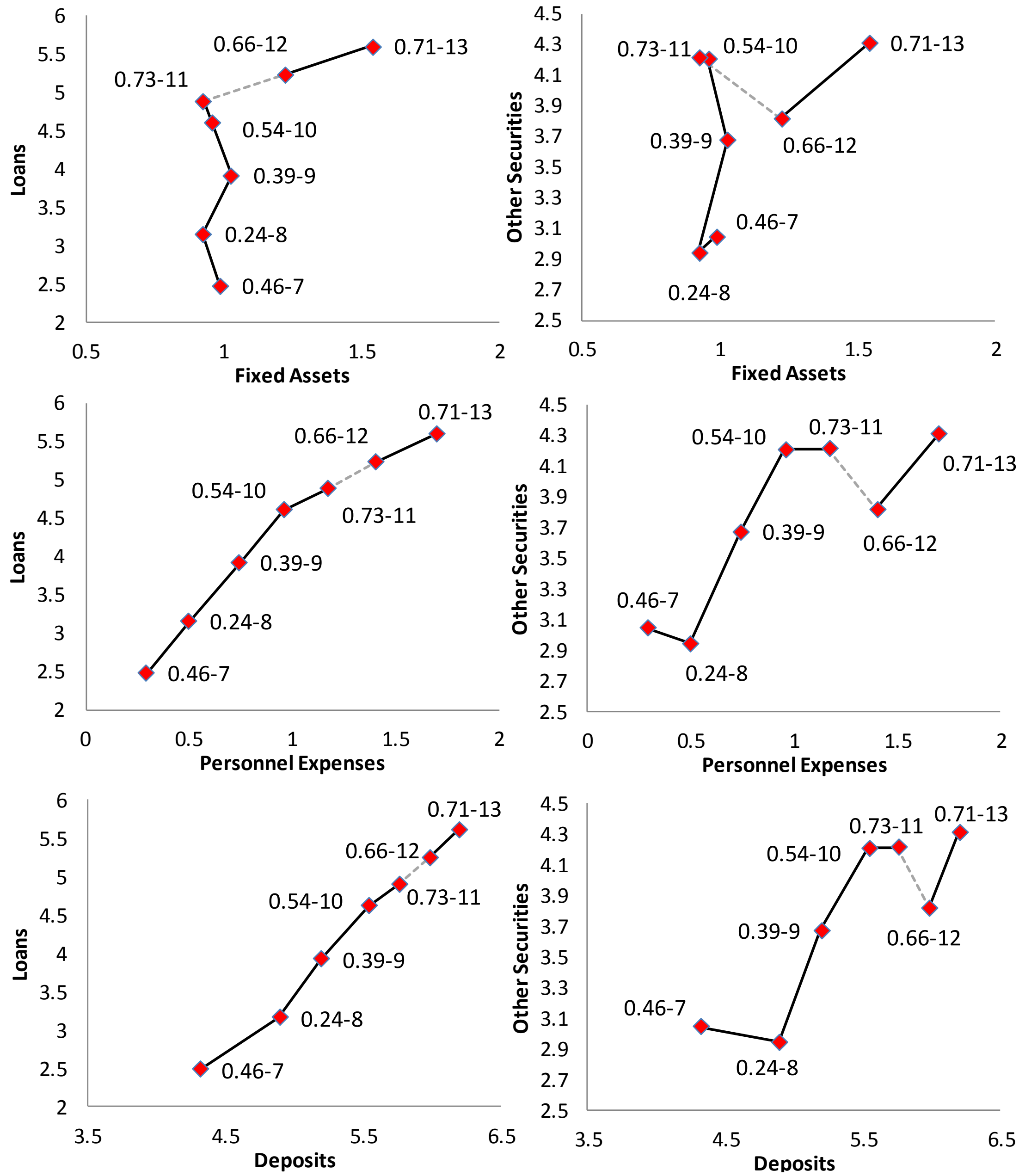


Figure XIV.29. Macedonia-Thrace Bank

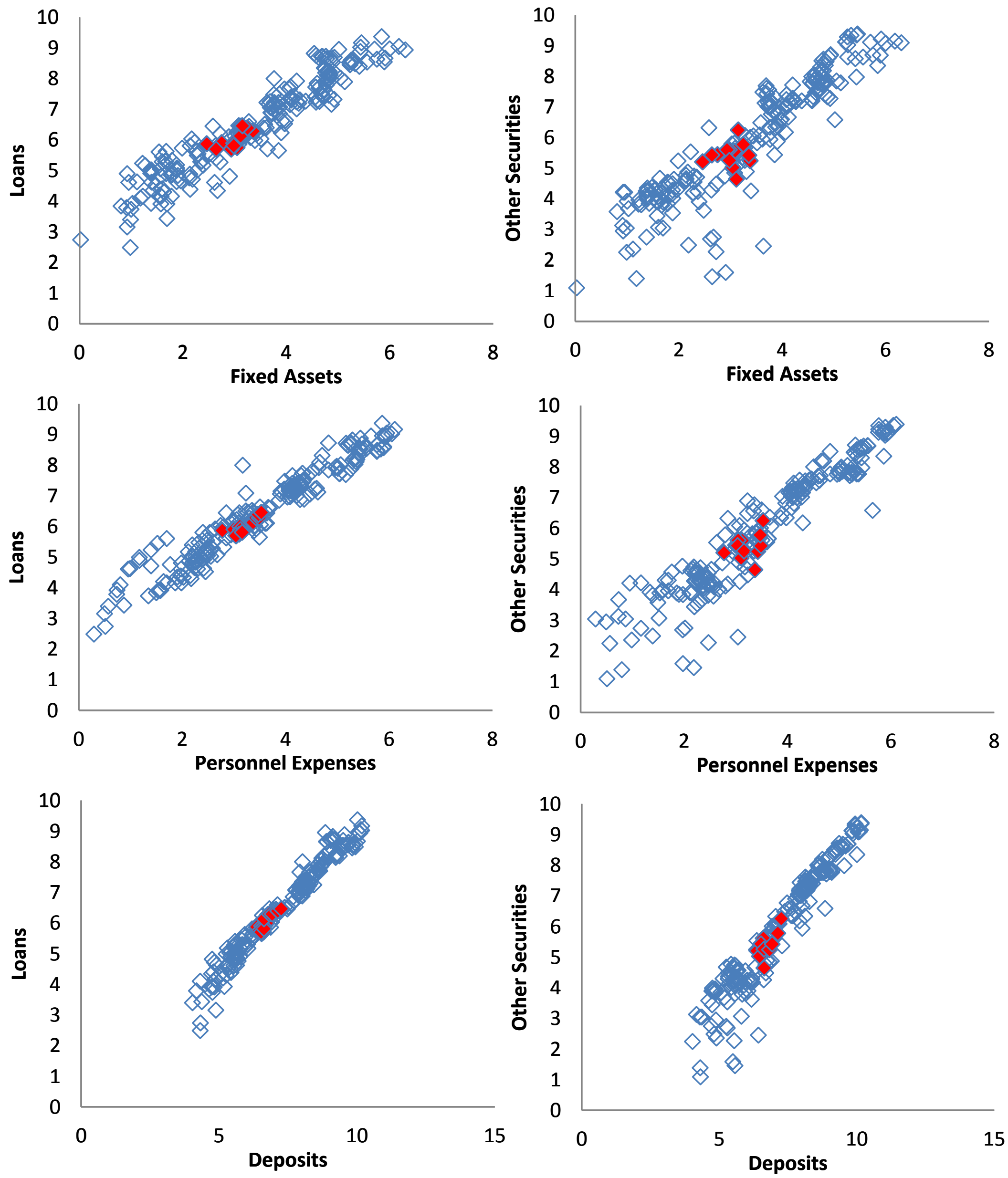


Figure XIV.30. Macedonia-Thrace Bank

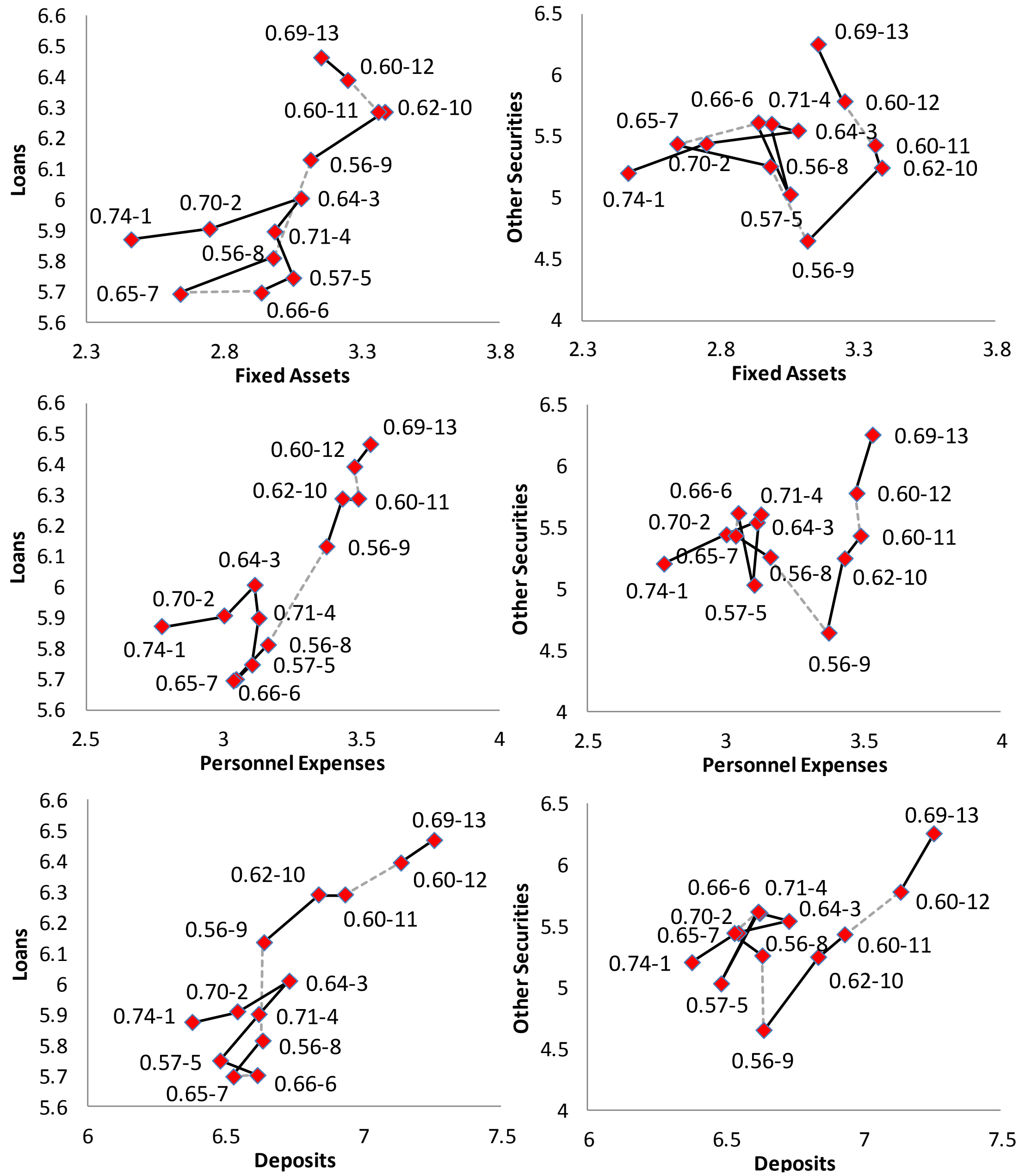


Figure XIV.31. National Bank

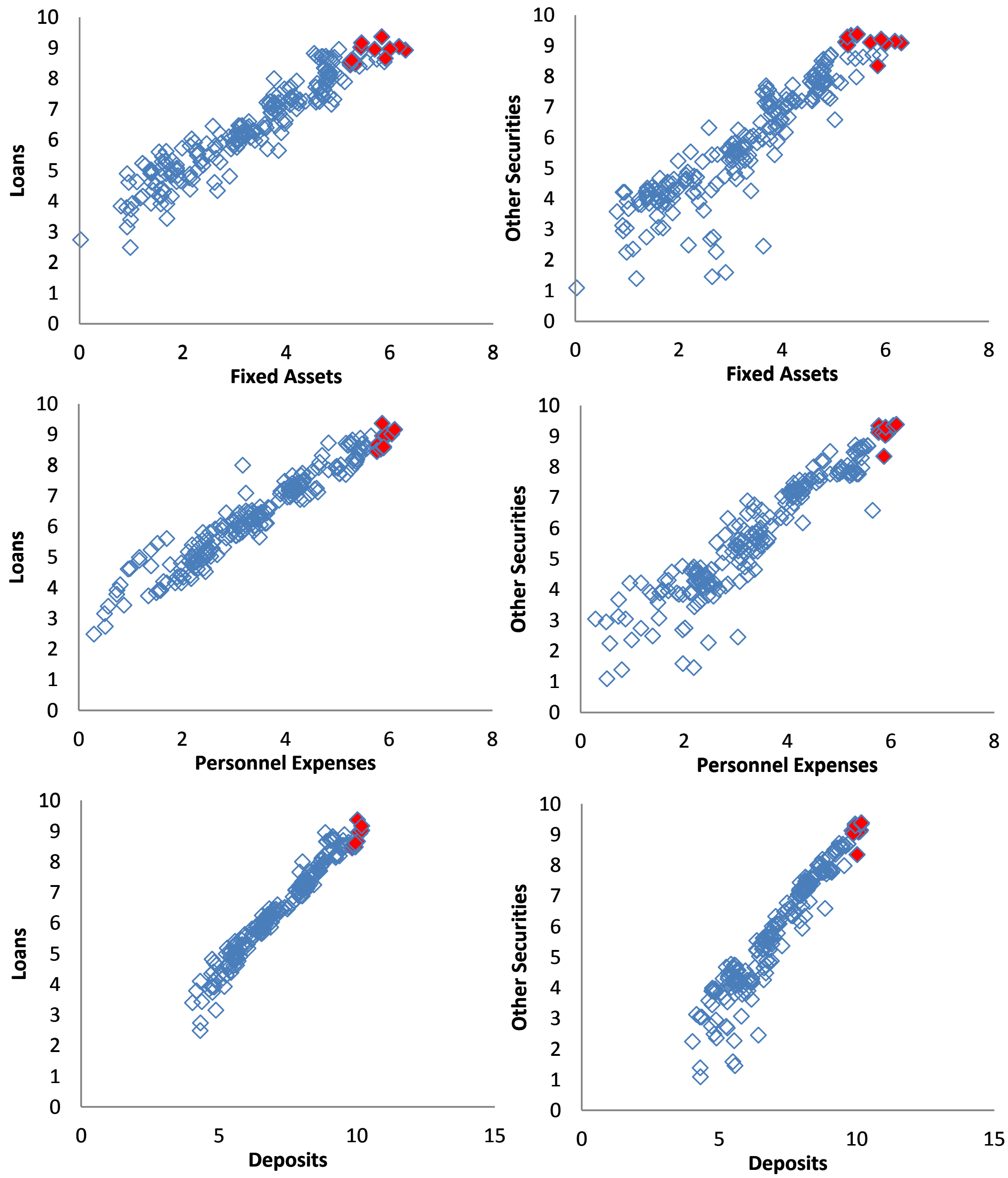


Figure XIV.32. National Bank

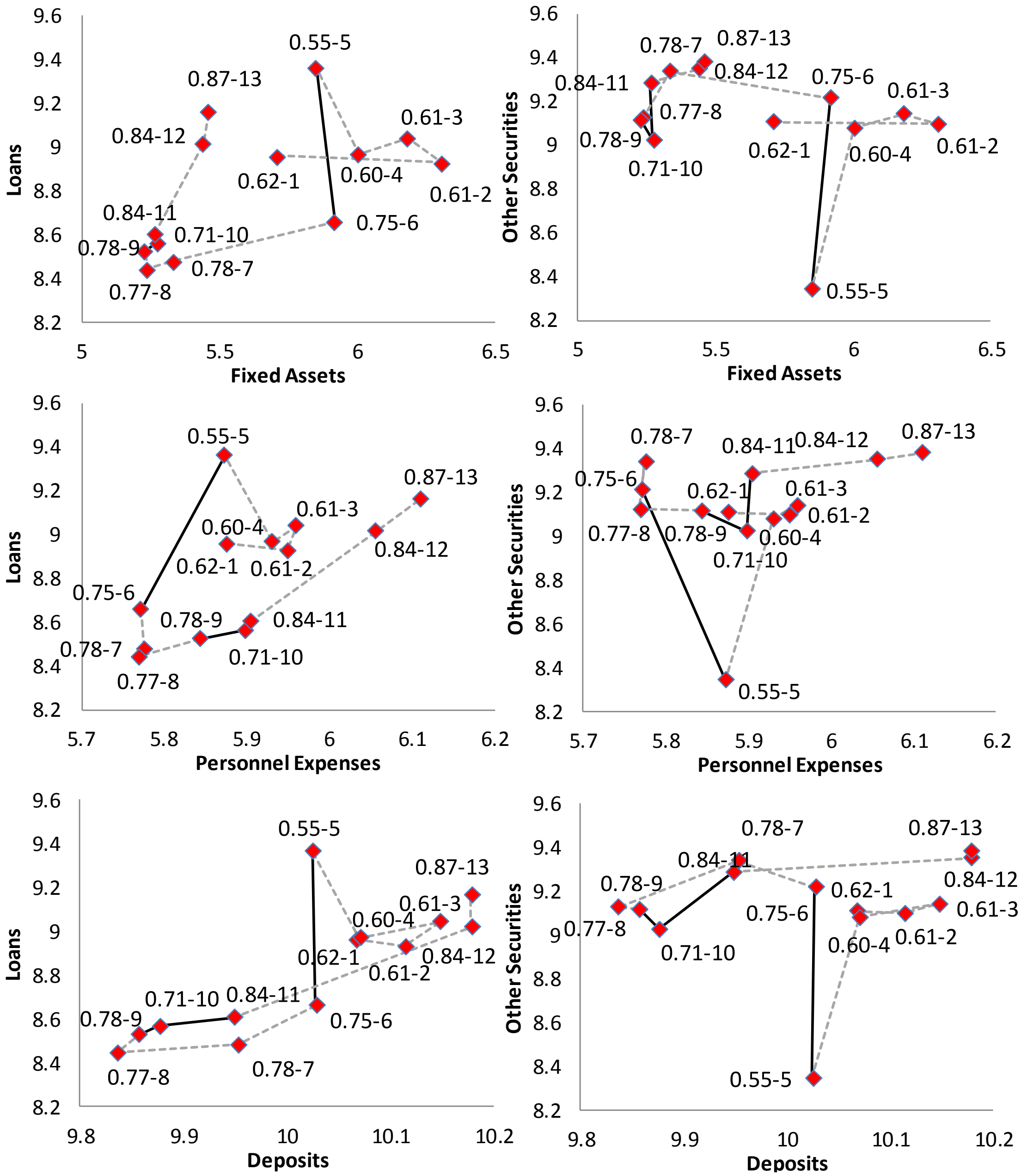


Figure XIV.33. Piraeus Bank

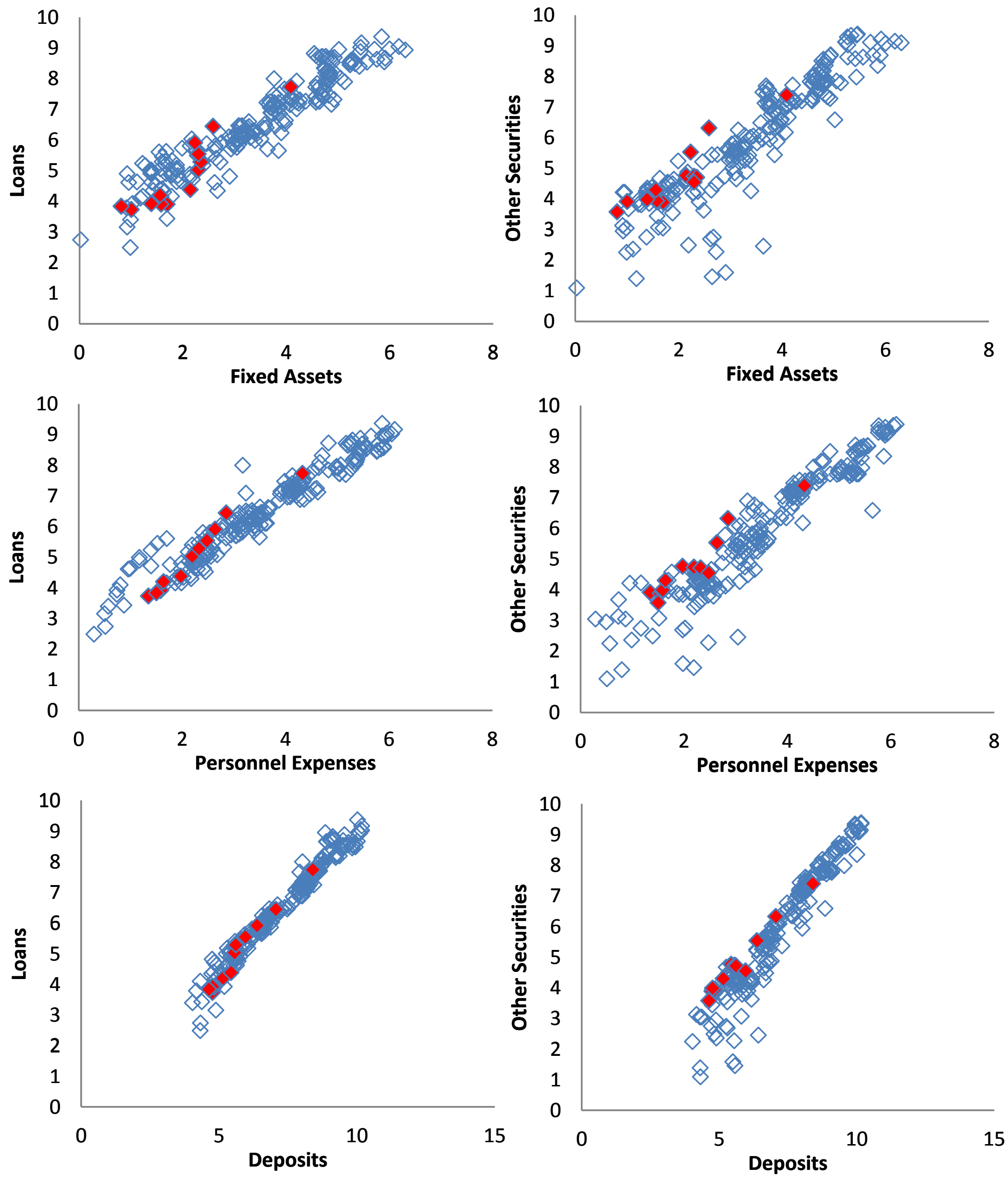


Figure XIV.34. Piraeus Bank

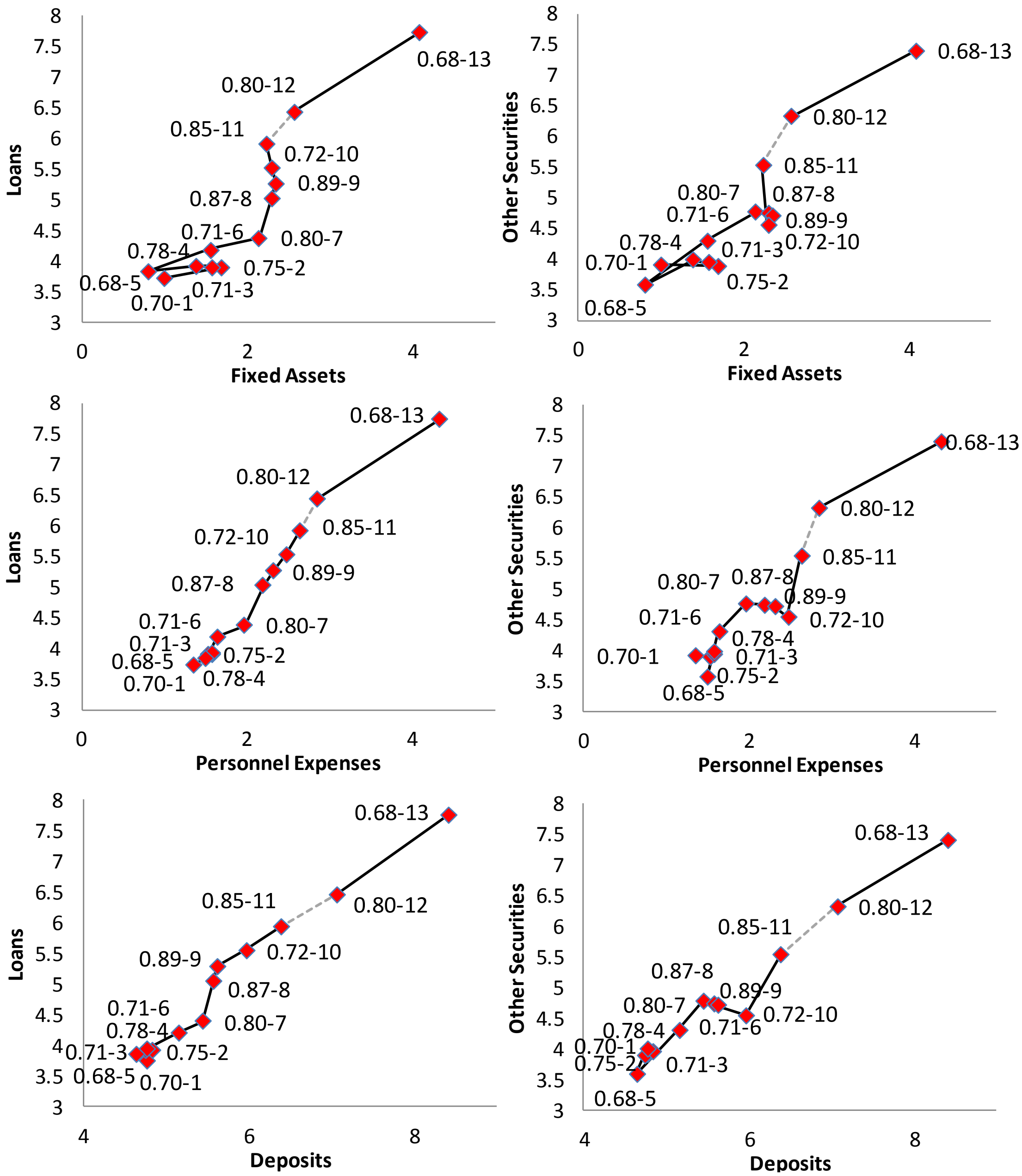


Figure XIV.35. TBank

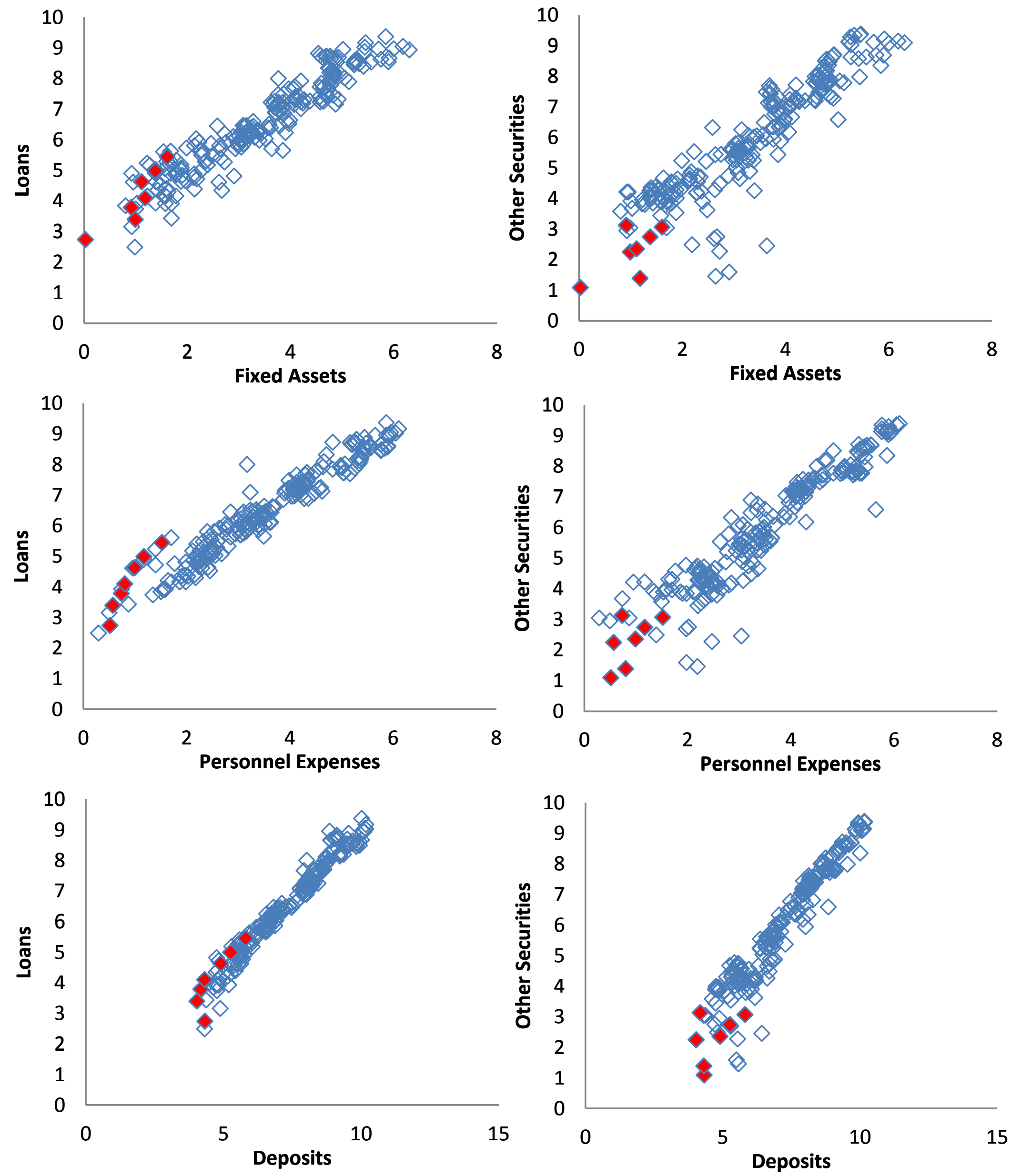


Figure XIV.36. TBank

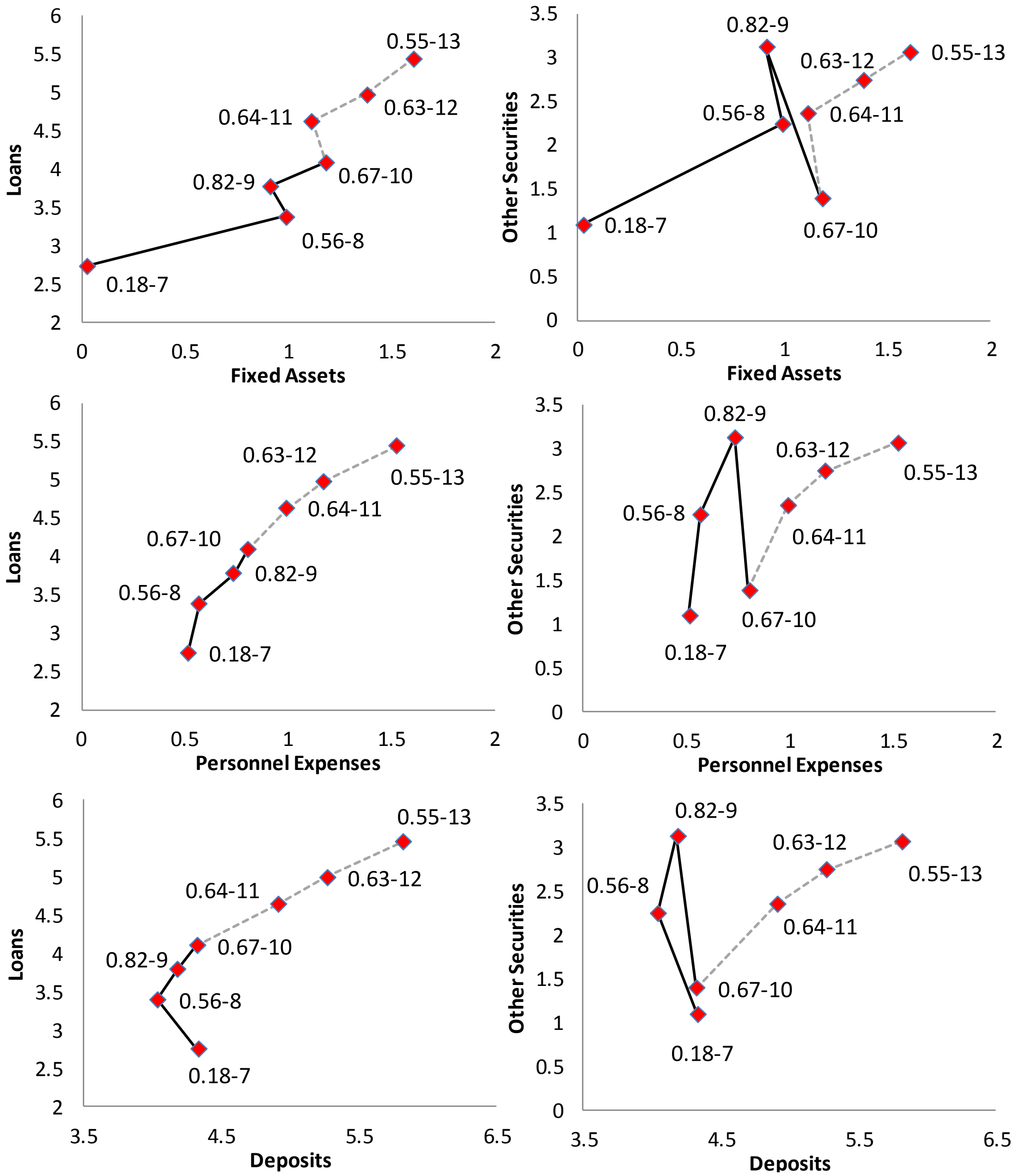


Figure XIV.37. Xiosbank

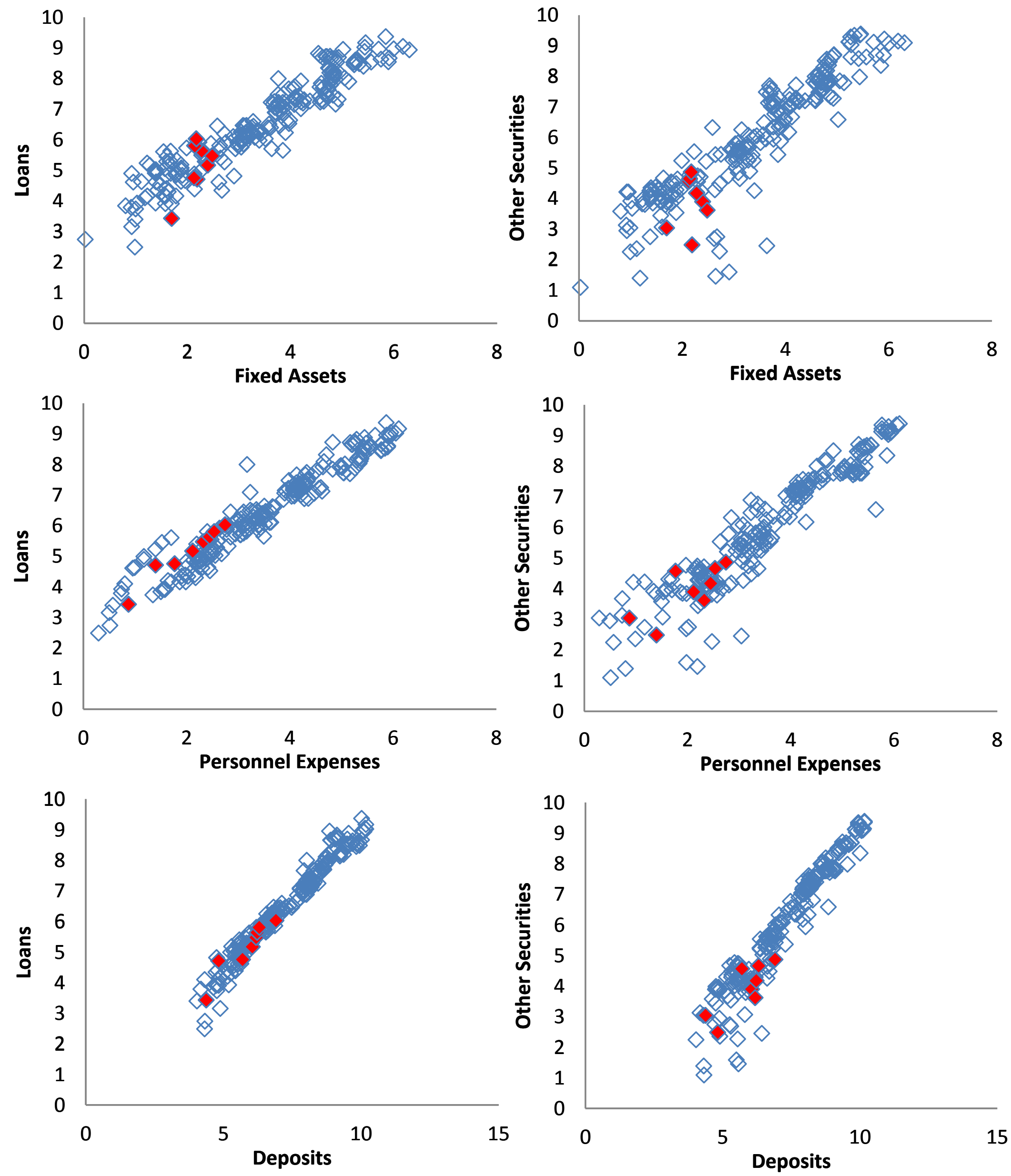


Figure XIV.38. Xiosbank

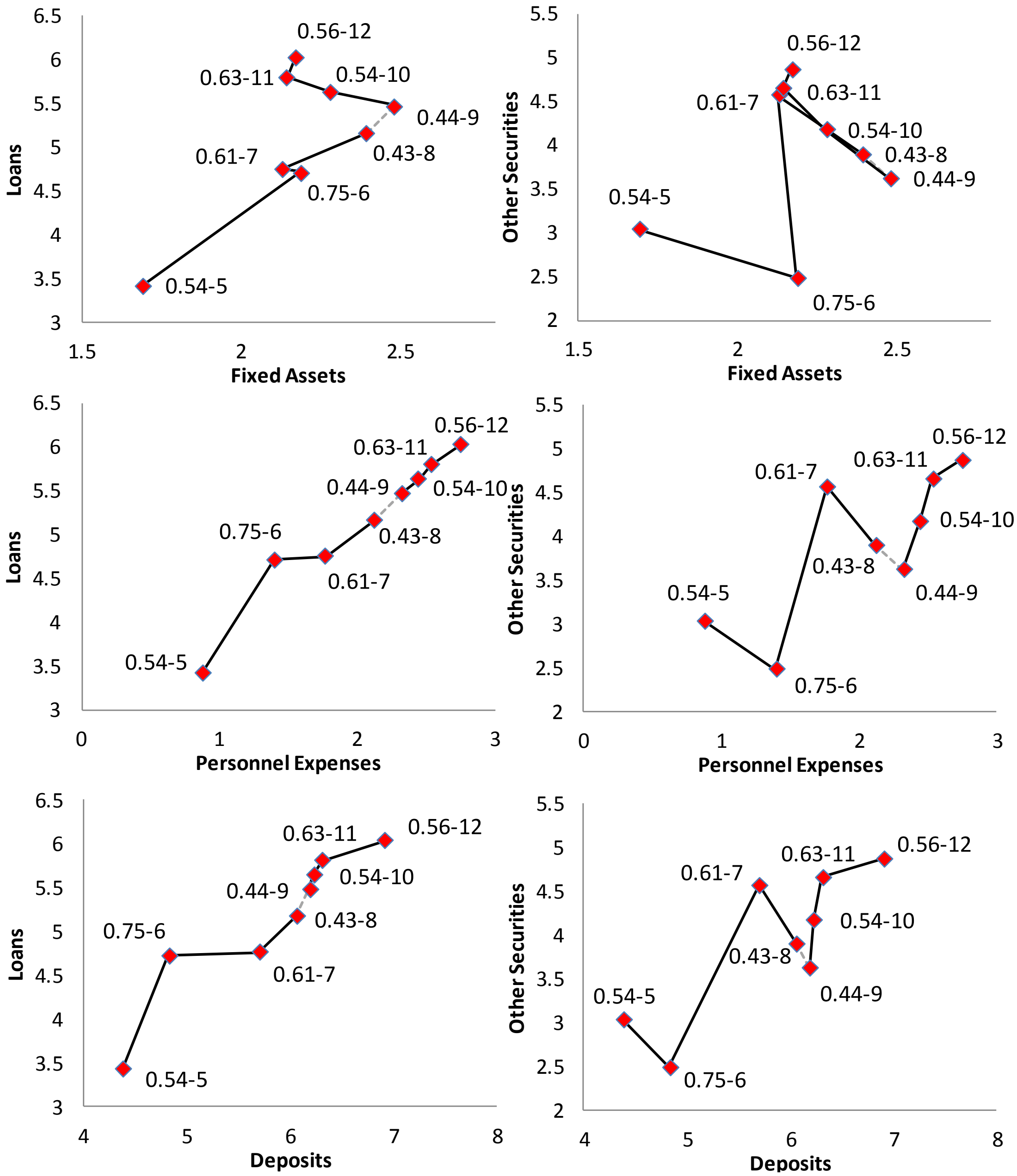


Figure XIV.39. Average Bank

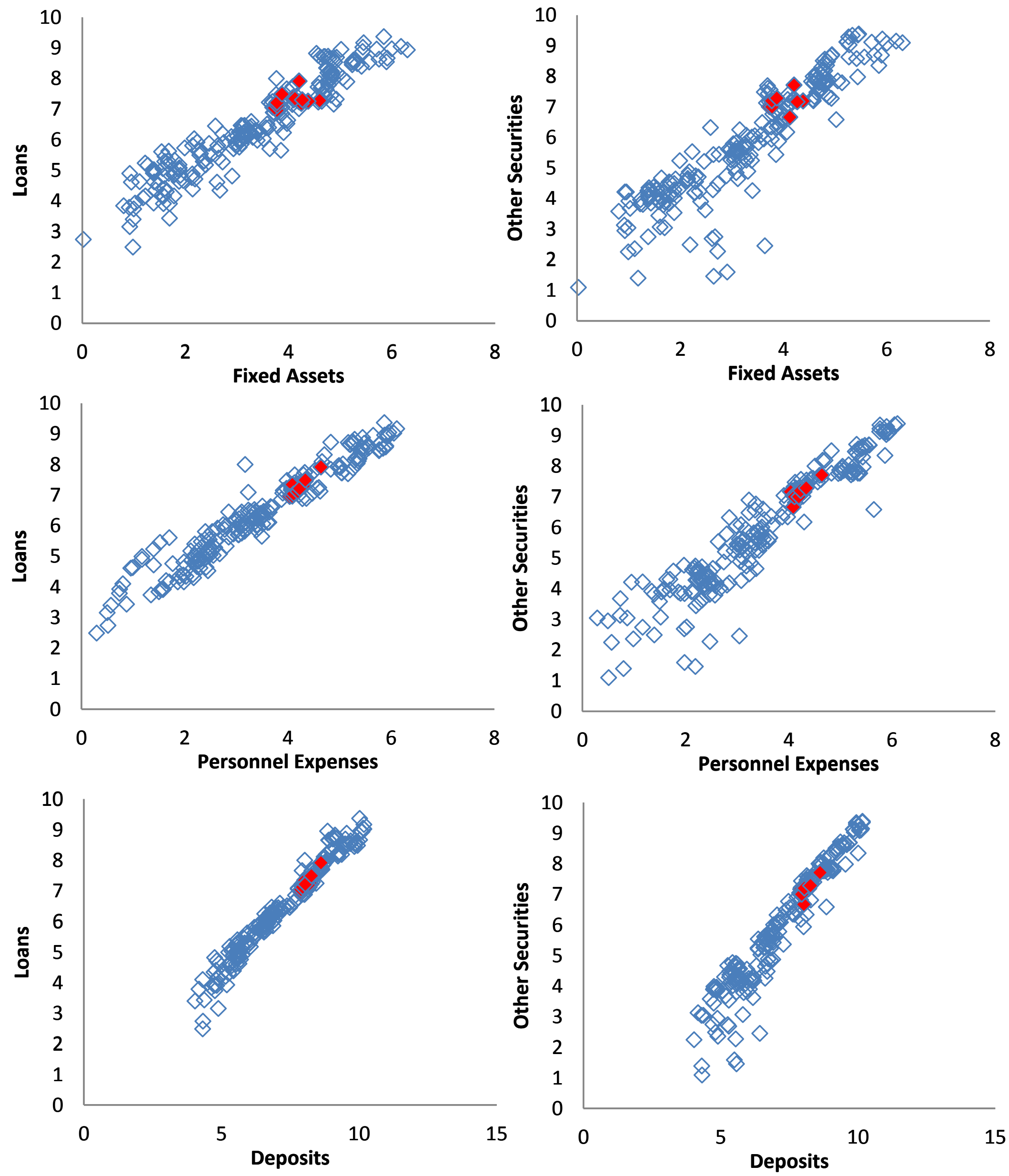


Figure XIV.40. Average Bank

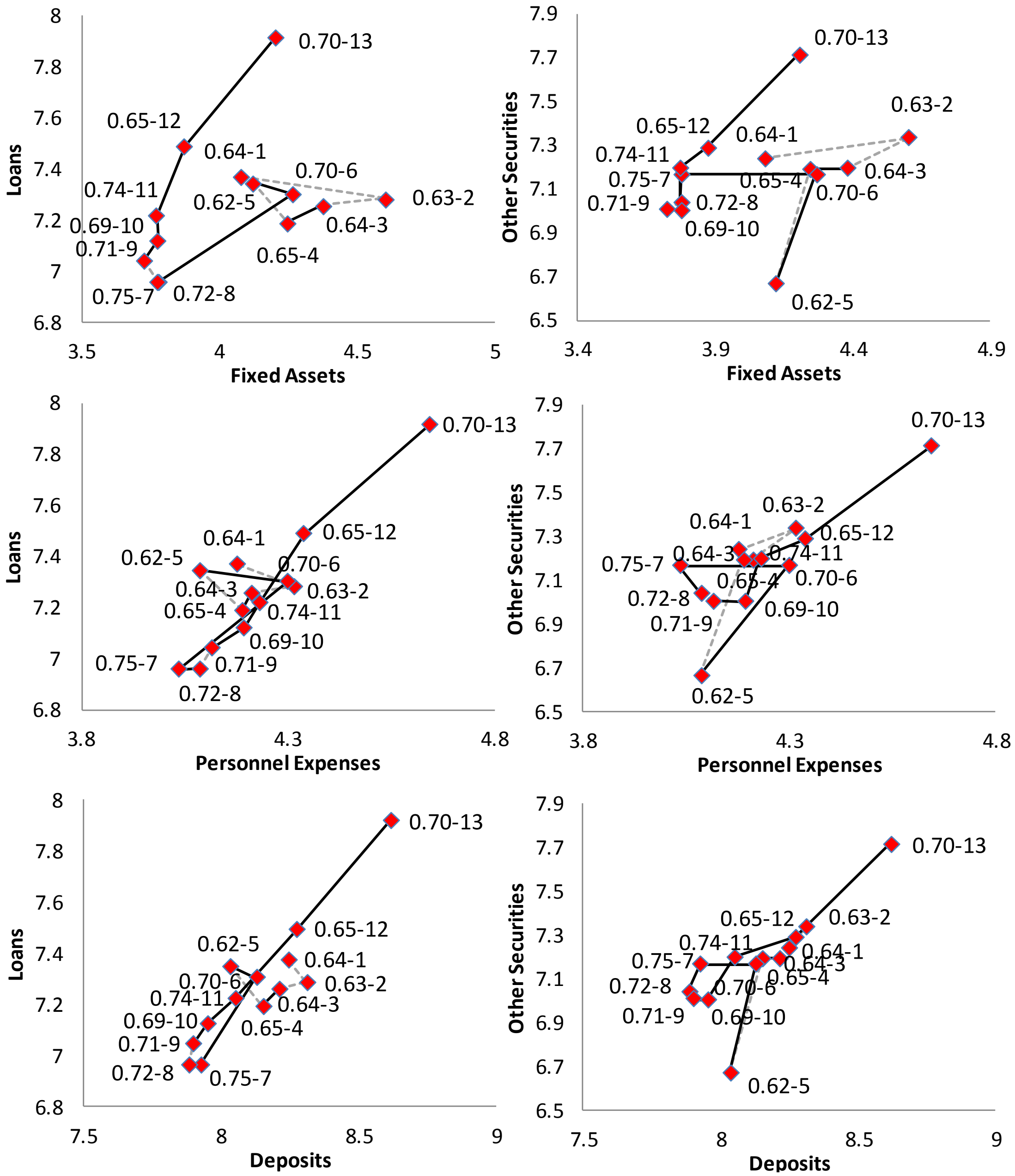


Figure XIV.41. Weighted Average Bank

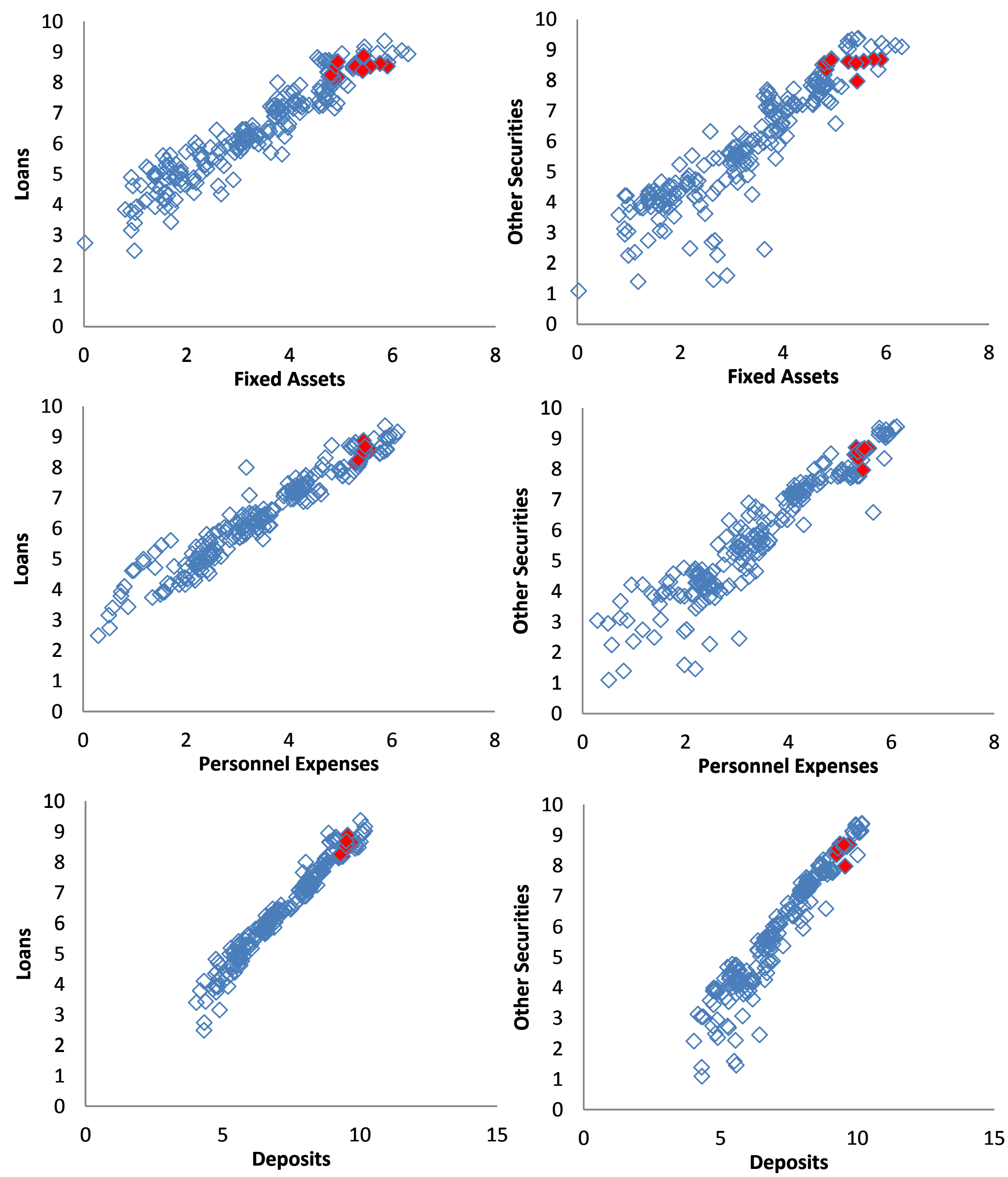
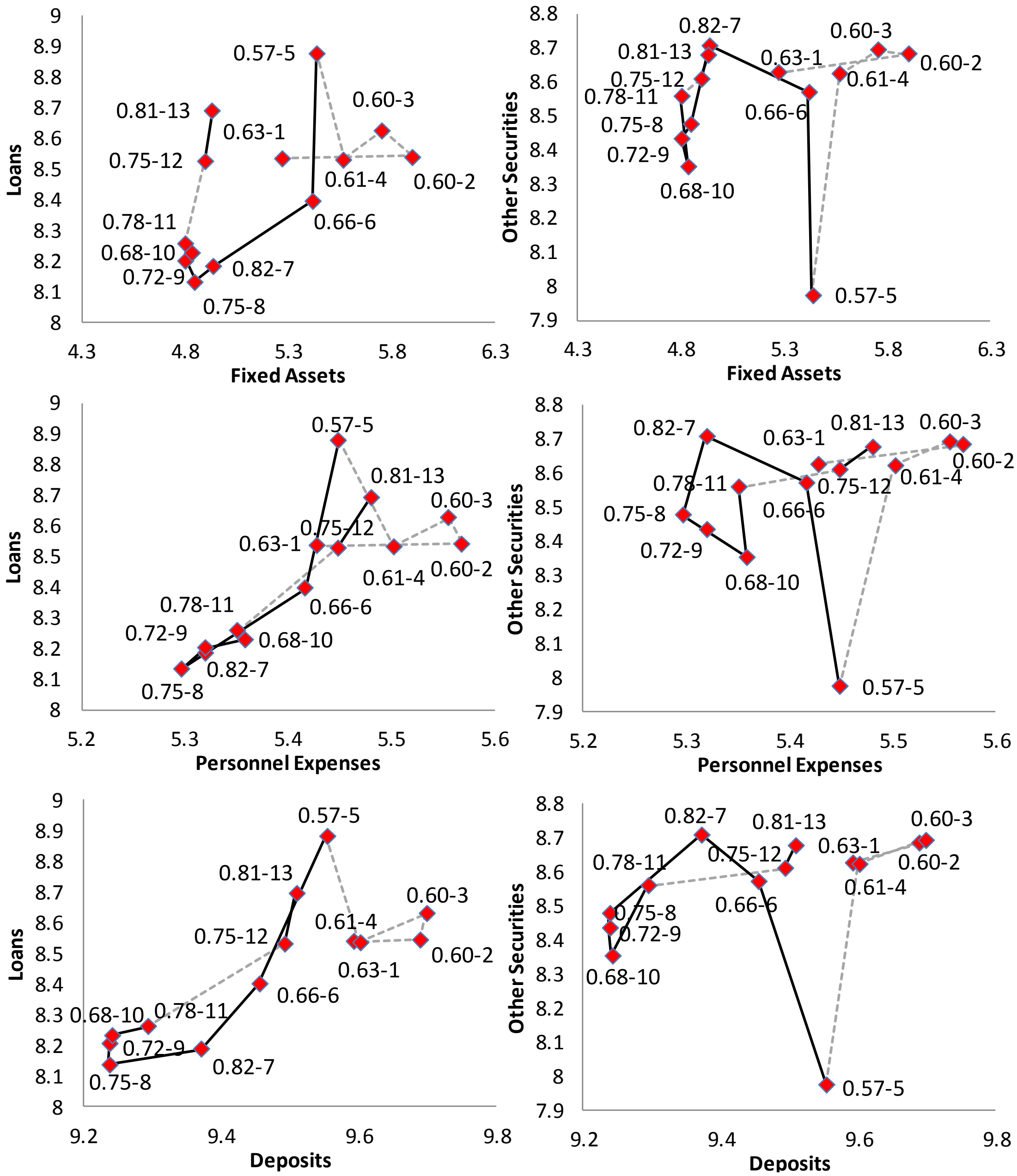


Figure XIV.42. Weighted Average Bank



References

Aigner, D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21–37.

Altunbas, Y., Evans, L., Molyneux, P., 2001. Bank Ownership and Efficiency. Journal of Money, Credit and Banking 33, 926–954.

Altunbas, Y., Goddard, J., Molyneux, P., 1999. Technical change in banking. Economics Letters 64, 215–221.

Amel, D., Barnes, C., Panetta, F., Salleo, C., 2004. Consolidation and efficiency in the financial sector: A review of the international evidence. Journal of Banking & Finance 28, 2493–2519.

Apergis, N., Rezitis, A., 2004. Cost structure, technological change, and productivity growth in the Greek banking sector. International Advances in Economic Research 10, 1–15.

Asmild, M., Tam, F., 2007. Estimating global frontier shifts and global Malmquist indices. Journal of Productivity Analysis 27, 137–148.

Ayadi, R., 2008. Banking mergers and acquisitions' performance in Europe. In: Molyneux, P., Vallelado, E. (Eds.), Frontiers of Banks in a Global Economy. Palgrave Macmillan, New York - Basingstoke (Hampshire), pp. 8–58.

Baltensperger, E., Dermine, J., 1987. Banking deregulation in Europe. Economic Policy 2, 64–109.

Banker, R.D., 1993. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science 39, 1265–1273.

Banker, R.D., 1996. Hypothesis tests using data envelopment analysis. Journal of Productivity Analysis 7, 139–159.

Banker, R.D., Chang, H., Cooper, W.W., 1996. Equivalence and implementation of alternative methods for determining returns to scale in data envelopment analysis. European Journal of Operational Research 89, 473–481.

Banker, R.D., Charnes, A., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30, 1078–1092.

Banker, R.D., Cooper, W.W., Seiford, L.M., Thrall, R.M., Zhu, J., 2004. Returns to scale in different DEA models. European Journal of Operational Research 154, 345–362.

Banker, R.D., Thrall, R.M., 1992. Estimation of returns to scale using data envelopment analysis. European Journal of Operational Research 62, 74–84.

Barth, J.R., Caprio, G., Levine, R., 2001. The regulation and supervision of banks around the world: A new database. Brookings-Wharton Papers on Financial Services 2001, 183–240.

Bauer, P.W., Berger, A.N., Ferrier, G.D., Humphrey, D.B., 1998. Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods. Journal of Economics and Business 50, 85–114.

Benston, G.J., 2000. Consumer protection as justification for regulating financial-services firms and products. Journal of Financial Services Research 17, 277–301.

Beran, R., Ducharme, G., 1991. Asymptotic theory for bootstrap methods in statistics. Centre de Reserches Mathematiques, University of Montreal, Montreal, Montreal.

Berg, S.A., Forsund, F.R., Jansen, E.S., 1992. Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. Scandinavian Journal of Economics 94, S211–28.

Berger, A.N., 1993. Distribution-free estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions. Journal of Productivity Analysis 4, 261–292.

Berger, A.N., Humphrey, D.B., 1992. Measurement and efficiency issues in commercial banking. In: Griliches, Z. (Ed.), Output Measurement in the Service Sectors. National Bureau of Economic Research, pp. 245–300.

Berger, A.N., Humphrey, D.B., 1997. Efficiency of financial institutions: International survey and directions for future research. European Journal of Operational Research 98, 175–212.

Berger, A.N., Hunter, W.C., Timme, G., 1993. The efficiency of financial institutions: a review and preview of reserach past, present and future. Journal of Banking and Finance 17, 221–249.

Berger, A.N., Mester, L.J., 1997. Inside the black box: What explains differences in the efficiencies of financial institutions? Journal of Banking & Finance 21, 895–947.

Bickel, P.J., Freedman, D.., 1981. Some assymptotic theory for the bootstrap. Annals of Statistics 9, 1196–1217.

Bickel, P.J., Gotze, F., R., Z.W., 1997. Resampling fewer than n observations: Gains, losses and remedies for losses. Statistica Sinica 7, 1–31.

Bowman, A.W., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71, 353–360.

Casu, B., Girardone, C., Molyneux, P., 2004. Productivity change in European banking: A comparison of parametric and non-parametric approaches. Journal of Banking & Finance 28, 2521–2540.

Casu, B., Girardone, C., Molyneux, P., 2006. Introduction to banking. Pearson Education, Harlow, Essex.

Caves, D.W., Christensen, L.R., Diewert, E.W., 1982. The economic theory of index numbers and the measurement of input, output and productivity. Econometrica 50, 1393–1414.

Charnes, a., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2, 429–444.

Charnes, A., Cooper, W.W., Lewin, A.Y., Seiford, L.M., 1994. Data Envelopment Analysis: Theory, Methodology and Applications, Journal of the Operational Research Society. Kluwer Academic Publishers, Boston.

Charnes, A., Cooper, W.W., Rhodes, E., 1981. Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. Management Science 27, 668–697.

Chen, X., Skully, M., Brown, K., 2005. Banking efficiency in China: Application of DEA to pre- and post-deregulation eras: 1993–2000. China Economic Review 16, 229–245.

Chortareas, G.E., Girardone, C., Ventouri, A., 2008. Efficiency and productivity change in Greek banking: Methods and recent evidence. In: Molyneux, P., Vallelado, E. (Eds.), Frontiers of Banks in a Global Economy. Palgrave Macmillan, pp. 211–233.

Chortareas, G.E., Girardone, C., Ventouri, A., 2012. Bank supervision, regulation, and efficiency: Evidence from the European Union. Journal of Financial Stability 8, 292–302.

Chortareas, G.E., Girardone, C., Ventouri, A., 2013. Financial freedom and bank efficiency: Evidence from the European Union. Journal of Banking and Finance 37, 1223–1231.

Christopoulos, D.K., Lolos, S.E.G., Tsionas, E.G., 2002. Efficiency of the Greek banking system in view of the EMU: A heteroscedastic stochastic frontier approach. Journal of Policy Modeling 24, 813–829.

Christopoulos, D.K., Tsionas, E.G., 2001. Banking economic efficiency in the deregulation period: Results from heteroscedastic stochastic frontier models. The Manchester School 69, 656–676.

Coelli, T., Rao, P.D.S., O'Donell, C.J., Battese, G.E., 2005. An introduction to efficiency and productivity analysis, 2nd ed. Springer, New York.

Cook, W.D., Tone, K., Zhu, J., 2014. Data envelopment analysis: Prior to choosing a model. Omega 44, 1–4.

Cooper, W.W., Seiford, L.M., Tone, K., 2006. Introduction to data envelopment analysis and its uses. Springer, New York.

Crafts, N., 2006. Regulation and productivity performance. Oxford Review of Economic Policy 22, 186–202.

Debreu, G., 1951. The coefficient of resource utilization. Econometrica 19, 273–292.

Delis, M.D., Molyneux, P., Pasiouras, F., 2011. Regulations and productivity growth in banking: Evidence from transition economies. Journal of Money, Credit and Banking 43, 735–764.

Deprins, D., Simar, L., Tulkens, H., 1984. Measuring labor inefficiency in post offices. In: Marchand, M., Pestieau, P., Tulkens, H. (Eds.), The Performance of Public Enterprises: Concepts and Measurements. North-Holland, Amsterdam, pp. 243–267.

Dermine, J., 2002. European Banking: Past, Present and Future. In: The Transformation of the European Financial System. Second Central Banking Conference, Frankfurt.

Dermine, J., 2006. European Banking Integration : Don ' t Put the Cart before the Horse. Financial Markets, Institutions & Instruments 15, 57–106.

Dewatripont, M., Tirole, J., 1994. The prudential regulation of banks. MIT Press, Campbridge, Massachusetts.

Diamond, D.W., Dybvig, P.H., 1983. Bank runs, deposit insurance, and liquidity. Journal of Political Economy 91, 401–419.

Diamond, D.W., Dybvig, P.H., 1986. Banking theory, deposit insurance, and bank regulation. Journal of Business 59, 55–68.

Dobratz, B.A., Whitfield, S., 1992. Does scandal influence voters’ party preference? The case of Greece during the Papandreou era. European Sociological Review 8, 167–180.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 9, 1–26.

Efron, B., 1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B., 1987. Better bootstrap confidence intervals. Journal of the American Statistical Association 82, 171–185.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Chapman and Hall, London.

Elrod, D.I.P., Tippett, D.D., 2002. The “death valley” of change. Journal of Organizational Change Management 15, 273–291.

Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115–146.

Färe, R., Grosskopf, S., 1985. A nonparametric cost approach to scale efficiency. Scandinavian Journal of Economics 87, 594–604.

Färe, R., Grosskopf, S., Kokkelenberg, E.C., 1989. Measuring plant capacity, utilization and technical change: A nonparametric approach. International Economic Review 30, 655–666.

Färe, R., Grosskopf, S., Norris, M., Zhang, Z., 1994. Productivity growth, technical progress, and efficiency change in industrialized countries. American Economic Review 84, 66–83.

Farrell, M.J., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society Series A (General) 120, 253–290.

Featherstone, K., 1990. The “party-state” in Greece and the fall of Papandreou. West European Politics 13, 101–115.

Featherstone, K., 1994. The Greek election of 1993: Backwards or forwards? West European Politics 17, 204–211.

Fried, H.O., Lovell, C.A.K., Schmidt, S.S., 2008. Efficiency and productivity. In: Fried, H.O., Lovell, C.A.K., Schmidt, S.S. (Eds.), The Measurement of Productive Efficiency and Productivity Growth. Oxford University Press, New York.

Garcia, J.A., Rixtel, A. Van, 2007. Inflation linked bonds from a central bank perspective (No. 62), Occasional Paper Series. Frankfurt.

Gilbert, R.A., Wilson, P.W., 1998. Effects of deregulation on the productivity of Korean banks. Journal of Economics and Business 50, 133–155.

Gocht, A., Balcombe, K., 2006. Ranking efficiency units in DEA using bootstrapping an applied analysis for Slovenian farm data. Agricultural Economics 35, 223–229.

Gonzales, X.M., Miles, D., 2002. Statistical precision of DEA : a bootstrap application to Spanish public services. Applied Economics Letters 9, 127–132.

Goodhart, C., 1988. The costs of regulation. Institute of Economic Affairs Readings 17–31.

Gortsos, C., 2002. Greece. In: Gardener, E.P.M., Molyneux, P., Moore, B. (Eds.), Banking in the New Europe. Palgrave-Macmillan, New York, pp. 128–159.

Greene, W.H., 2003. Econometric Analysis, 5th ed. Prentice Hall, New Jersey.

Halkos, G.E., Salamouris, D.S., 2004. Efficiency measurement of the Greek commercial banks with the use of financial ratios: A data envelopment analysis approach. Management Accounting Research 15, 201–224.

Hall, M.J.B., 1989. Handbook of banking regulation and supervision. Woodhead-Faulkner, Cambridge, England.

Hawdon, D., 2003. Efficiency , performance and regulation of the international gas industry — a bootstrap DEA approach. Energy Policy 31, 1167–1178.

Hondroyiannis, G., Lolos, S., Papapetrou, E., 1999. Assessing competitive conditions in the Greek banking system. Journal of International Financial Markets, Institutions and Money 9, 377–391.

Humphrey, D.B., 1990. Cost and technical change: Effects from bank deregulation. Federal Reserve Bank of Richmond Working Paper Series, 90 90-5.

Humphrey, D.B., Pulley, L.B., 1997. Banks' responses to deregulation: Profits, technology, and efficiency. Journal of Money, Credit and Banking 29, 73–93.

Isik, I., Hassan, K.M., 2003. Financial deregulation and total factor productivity change: An empirical study of Turkish commercial banks. Journal of Banking & Finance 27, 1455–1485.

Johnson, N.L., Kotz, S., BalakrishnanN., 1994. Continuous univariate distributions: Volume 1. John Wiley & Sons, New York.

Jones, M.C., Marron, J.S., Sheather, S.J., 1996. A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91, 401–407.

Kamberoglou, N.C., Liapis, E., Simigiannis, G.T., Tzamourani, P., 2004. Cost efficiency in Greek banking, Bank of Greece Working Paper Series.

Karafolas, S., Mantakas, G., 1994. A note on cost structure and economies of scale in Greek banking. Journal of Banking & Finance 20, 377–387.

Kneip, A., Park, B., Simar, L., 1998. A note on the convergence of nonparametric DEA estimators for production efficiency scores. Econometric Theory 14, 783–793.

Kneip, A., Simar, L., Wilson, P.W., 2008. Asymptotics and consistent bootstraps for Dea estimators in nonparametric frontier models. Econometric Theory 24, 1663–1697.

Kneip, A., Simar, L., Wilson, P.W., 2011. A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators. Computational Economics 38, 483–515.

Kneip, A., Simar, L., Wilson, P.W., 2012. Central limit theorems for DEA effciency scores: When bias can kill the variance. Universite Catholique de Louvain, Universite Catholique de Louvain.

Koopmans, T.C., 1951. Activity analysis of production and allocation. John Wiley & Sons, New York.

Korostelev, A., Simar, L., Tsybakov, A.B., 1995. Efficient estimation of monotone boundaries. Annals of Statistics 23, 476–489.

Krüger, J.J., 2012. A Monte Carlo study of old and new frontier methods for efficiency measurement. European Journal of Operational Research 222, 137–148.

Kumbhakar, S.C., Lozano-vivas, A., Lovell, C.A.K., Hasan, I., 2001. The effects of deregulation on the performance of financial institutions: The case of Spanish savings banks. Journal of Money, Credit and Banking 33, 101–120.

Kumbhakar, S.C., Sarkar, S., 2003. Deregulation, ownership, and productivity growth in the banking industry: Evidence from India. Journal of Money, Credit and Banking 35, 403–424.

Loader, C.R., 1999. Bandwidth selection: Classical or plug-in? Annals of Statistics 27, 415–438.

Löthgren, M., 1998. How to bootstrap DEA estimators: a Monte Carlo comparison. Stockholm School of Economics: Working Paper Series in Finance and Economics, Working Paper Series in Finance and Economics N.223.

Magnussen, J., Nyland, K., 2008. Measuring efficiency in clinical departments. Health Policy 87, 1–7.

Mas-Colell, A., Whinston, G.D., Green, J.R., 1995. Microeconomic theory. Oxford University Press, New York.

Matthews, K., Thompson, J., 2014. The economics of banking, 3rd ed. John Wiley & Sons, Chichester.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18, 435–444.

Molyneux, P., 2009. Do mergers improve bank productivity and performance. In: Balling, M., Gnan, E., Lierman, F., Schoder, J.-P. (Eds.), Productivity in the Financial Services Sector. SUERF Studies, Vienna.

Mouchart, M., Simar, L., 2002. Efficiency analysis of air controlers: first insights. Institut de Statistique Université Catholique de Louvain, Consulting Report, Consulting Report 0202.

Noulas, A.G., 2001. Deregulation and operating efficiency: The case of the Greek banks. Managerial Finance 27, 35–47.

Orea, L., 2002. Parametric decomposition of a generalized Malmquist productivity index. Journal of Productivity Analysis 18, 5–22.

Park, B., Marron, J.S., 1990. Comparison of data-driven bandwidth selectors. Journal of the American Statistical Association 85, 66–72.

Pasiouras, F., 2008. International evidence on the impact of regulations and supervision on banks' technical efficiency: An application of two-stage data envelopment analysis. Review of Quantitative Finance and Accounting 30, 187–223.

Pasiouras, F., Tanna, S., Zopounidis, C., 2009. The impact of banking regulations on banks' cost and profit efficiency: Cross-country evidence. International Review of Financial Analysis 18, 294–302.

Pastor, J.T., Lovell, C. a. K., 2005. A global Malmquist productivity index. Economics Letters 88, 266–271.

Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer-Verlag, New York.

Ray, S.C., Desli, E., 1997. Productivity growth , technical progress, and efficiency change in industrialized countries: A comment. American Economic Review 87, 1033–1039.

Rezitis, A.N., 2006. Productivity growth in the Greek banking industry: A non-parametric approach. Journal of Applied Economics IX, 119–138.

Rezitis, A.N., 2010. Evaluating the state of competition of the Greek banking industry. Journal of International Financial Markets, Institutions and Money 20, 68–90.

Rudemo, M., 1982. Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics 9, 65–78.

Sadjadi, S.J.Ã., Omrani, H., 2010. A bootstrapped robust data envelopment analysis model for efficiency estimating of telecommunication companies in Iran. Telecommunications Policy 34, 221–232.

Sanhueza, R., Rudnick, H., Lagunas, H., 2004. DEA efficiency for the determination of the electric power distribution added value. IEEE Transactions on Power Systems 19, 919–925.

Sealey, A.C.W., Lindley, J.T., 1977. Inputs , outputs , and a theory of production and cost at depository financial institutions. Journal of Finance 32, 1251–1266.

Shao, J., Tu, D., 1995. The jackknife and bootstrap. Springer-Verlag, New York.

Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society, Series B 53, 683–690.

Shepard, R.W., 1970. Theory of cost and production function. Princeton University Press, New Jersey.

Silverman, B.W., 1986. Density estimation for statistics and data analysisDensity estimation for statistics and data analysis. Chapman and Hall, London.

Silverman, B.W., Young, G. a., 1987. The bootstrap: To smooth or not to smooth? Biometrika 74, 469.

Simar, L., Wilson, P.W., 1998. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Management Science 44, 49–61.

Simar, L., Wilson, P.W., 2000a. A general methodology for bootstrapping in non-parametric frontier models. Journal of Applied Statistics 27, 779–802.

Simar, L., Wilson, P.W., 2000b. Statistical inference in nonparametric frontier models : The state of the art. Journal of Productivity Analysis 13, 49–78.

Simar, L., Wilson, P.W., 2002. Non-parametric tests of returns to scale. European Journal of Operational Research 139, 115–132.

Simar, L., Wilson, P.W., 2004. Performance of the bootstrap for DEA estimators and iterating the principle. In: Cooper, W.W., Seiford, M., Zhu, J. (Eds.), Handbook on Data Envelopment Analysis. Kluwer Academic Publishers, New York: London, pp. 265–298.

Simar, L., Wilson, P.W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 136, 31–64.

Simar, L., Wilson, P.W., 2008. Statistical inference in non-parametric frontier models. In: Fried, O.H., Lovell, C.A.K., Schmidt, S.S. (Eds.), The Measurement of Productive Efficiency and Productivity Growth. Oxford University Press, Oxford, New York, pp. 421–521.

Simar, L., Wilson, P.W., 2011. Inference by the m out of n bootstrap in nonparametric frontier models. Journal of Productivity Analysis 36, 33–53.

Simar, L., Zelenyuk, V., 2007. Statistical inference for aggregates of Farrell-type efficiencies. Journal of Applied Econometrics 22, 1367–1394.

Siriopoulos, C., Tziogkidis, P., 2010. How do Greek banking institutions react after significant events?—A DEA approach. Omega 38, 294–308.

Stine, R., 1989. An introduction to bootstrap methods: Examples and Ideas. Sociological Methods & Research 18, 243–291.

Swanepoel, J.W.H., 1986. An note on proving that the (modified) bootstrap works. Communications in Statistics – Theory and Methods 15, 3193–3203.

Tsionas, E.G., Lolos, S.E.., Christopoulos, D.K., 2003. The performance of the Greek banking system in view of the EMU: Results from a non-parametric approach. Economic Modelling 20, 571–592.

Tsolas, I.E., 2011. Performance assessment of mining operations using nonparametric production analysis: A bootstrapping approach in DEA. Resources Policy 36, 159–167.

Van Biesebroeck, J., 2007. Robustness of productivity estimates. Journal of Industrial Economics 55, 529–569.

Varian, H.R., 1992. Microeconomic analysis. W.W. Norton, New York.

Voridis, H., Angelopoulou, E., Skotida, I., 2003. Monetary policy in Greece 1990-2000 through the publications of the bank of Greece. Bank of Greece Economic Bulletin.

Wheelock, D.C., Wilson, P.W., 1999. Technical progress, inefficiency, and productivity change in U.S. banking, 1984-1993. Journal of Money, Credit and Banking 31, 212–234.